

计量经济学方法

JILIANG
JINGJI
XUE
FANGFA



中文版序言

英国曼彻斯特大学计量经济学教授 J. Johnston 所著《计量经济学方法》，是一部系统地论述计量经济学方法的教科书，它不仅简练地阐述了计量经济学的基本原理，同时也介绍了计量经济学方法的某些新发展。

随着我国社会主义现代化建设的发展，我国经济学界和经济管理部门越来越感到应用数量分析技术和方法解决经济及管理领域中问题的重要性，迫切要求学习和运用它来解决理论研究和实际工作中的问题。

计量经济学作为西方资产阶级经济学的一个分支，半个多世纪以来发展极为迅速。计量经济学的数量分析技术与方法，对于我国社会主义经济问题的研究有参考价值，对于改进经济计划和经济预测等方面，也有借鉴作用。

这本教科书的翻译出版，对于我国广大经济工作和管理工作者学习与研究计量经济学的理论与方法，发展我国的数量经济学将是有益的。

译者要我写几句话，限于本人水平，不妥之处请读者批评指正。

张 寿

1984年元月

第 二 版 序 言

本书第二版包括了自1963年第一版以来计量经济学迅速发展的一些主要课题。本版讨论的起始水平和第一版相同，并尽可能把计量经济学的新发展综合到第一版的结构和符号系统之中。

本版增加一章绪言，主要介绍计量经济学的性质。第二章和第三章与第一版一样，论述了两个变量的线性模型及其扩展。第四章阐述的是矩阵代数，并着重介绍了分块矩阵、秩和齐次方程的解、二次型及正定矩阵。第五章阐述的是一般线性模型，并增加了相关矩阵、预测、线性约束、多重共线性和设定误差等内容。第六章基本上是新的，包括虚变量，对方差分析作了特殊的介绍。第七章简单地介绍了广义最小二乘法，并新增加了纯粹估计和混合估计、观察值的分组以及方程的分组等内容。序列相关仍作为一章，并包括了最近的重要结果。第九章介绍了某些重要的渐近理论结果，也论述了工具变量及变量中的误差。第十章和第十一章是新增加的，第十章全部讨论的是滞后变量引起的问题；第十一章讨论的是其他重要的多变量方法。最后两章阐述的是联立方程问题，尤其包括了对各种估计式的小样本特性最新模拟研究的评价。本版并没有涉及贝叶斯(Bayesian)方法或谱分析，因为在计量经济学中这些方法发展的现阶段，最好在专门的论文中论述；同时限于本书假定的预备知识水平，介绍它们必将占去本书的大量篇幅，故在此从略这方面的内容。

本书的修改是我在Baruch College及Graduate Centre of the City University of New York做访问教授时开始的。在修改过程中充分考虑了来自世界各地的朋友、同事和学生们对第一版的批评和意见，我特别感谢University of Glasgow的数理统计学教授S.D.Silvey，他对本书各种底稿提出的非常宝贵的批评意见；也非常感谢我的同事R.W.Farebrother, S.A.Moore和J.Stewart，他们在编制注解、进行证明及整理索引等工作中给予了极大的帮助；在这里也感谢Paulin Millward和Dolores Whittaker，他们快速而又高质量地打印了我的原稿。

(其他谢语从略)

J. Johnston

第一版序言

本书的目的是为已经研究过一年左右统计理论和方法的读者提供计量经济学方法的完整论述。全书分两部分，第一部分是全面介绍标准线性回归模型，它是第二部分计量经济学理论的基础。第二部分阐述现在适用于估计计量经济学模型的主要统计方法。

对于已经学习过一年数理统计的读者可以跳过前两章，这两章的内容是作为连接统计方法中的普通课程插入的，对于社会科学方面的大多数读者已经掌握。第一章是对两个变量线性模型的详细介绍，包括与模型有关的估计、假设检验和预测等全部内容。大多数读者已经通晓本章论述的许多课题，本章利用了概率分布、期望值、估计和假设检验等基础知识。对于学习本章内容有困难的读者应当读一本好的统计学入门书^①。本章除了提供有帮助的复习材料之外，还介绍了基本推论问题，在本书其余部分的更加复杂的课文中还将进一步讨论这个问题。

第二章论述的是两个变量线性模型的扩展问题，包括非线性方面的扩充以及变量数目的增加，在引进标记法之前不准备多讲，因为其他复杂问题要求更强有力的技术。因此，第三章介绍矩阵代数基础，然后把它作为讲解本书其余内容的基本工具。第四章阐述 k 个变量的一般线性模型，它是第一部分的一章，也是第一部分最主要的一章，包括了一般线性模型的所有重要结果。第一、二、三章提供的内容是读者学习第四章的预备知识，第四章的内容是领会第二部分计量经济学方法的基本前提。

第二部分首先在第五章给出简短的绪言。第六章阐述变量中的误差，这个课题有时在计量经济学工作中不大引人注意，而实际上它往往是一个非常重要的问题。第七章论述的是序列相关扰动项存在时产生的一些问题，第八章介绍的是在单方程情况下可能产生一些其他问题，例如多重共线性、异方差性、滞后变量及虚变量等，并给出了统一处理这些问题的方法。最后两章论述的是联立方程问题，包括识别问题、间接最小二乘法、两阶段最小二乘法、有限信息方法、完全信息方法及三阶段最小二乘法等。

全书重点在于论述各种方法的基本原理，尽可能全面地说明各种方法的基本假设和给出各种结果的广泛研究情况。希望各种不同领域的读者能够通过本书提供的理论方法独立地开展研究工作，并在不同的实际应用中对各种方法的优点和限制做出实际评价。在课文中给出了一些数值例子，在大多数章的后面附上了理论分析和数值计算方面的练习题。对于皇家统计学会(Royal Statistical Society)、剑桥大学(University of Cambridge)、伦敦大学(University of London)、曼彻斯特大学(University of Manchester)和牛津大学(University of Oxford)等官方当局允许利用他们的考试

① 例如 P.G.Hoel, *Introduction to Mathematical Statistics*, 2d ed., Wiley, New York, 1954年。A.M.Mood, *Introduction to the Theory of Statistics*, McGraw-Hill, New York, 1950年。D.A.S. Fraser, *Statistics: An Introduction*, Wiley, New York, 1958年。

议题作为本书的例子表示感谢。由于越来越多的研究工作者使用具有各种应用程序的电子计算机，所以本书并没有给出计算问题的处理过程。

(其他谢语从略)

J. Johnston

目 录

中文版序言.....	1
第二版序言.....	2
第一版序言.....	3
第一章 计量经济学的性质.....	1
§1.1 变量之间的关系	1
§1.2 经济模型	1
§1.3 一个简单的宏观经济模型	1
§1.4 计量经济学的任务	4
第二章 两个变量的线性模型.....	6
§2.1 假设	6
§2.2 最小二乘估计式	10
§2.3 相关系数	26
§2.4 回归中的方差分析	28
§2.5 预测	31
第三章 两个变量线性模型的扩展.....	39
§3.1 两个变量之间的非线性关系	39
§3.2 三个变量之间的关系	45
§3.3 回归平面的拟合	47
§3.4 复相关系数	48
§3.5 偏相关系数	50
§3.6 三个变量情况的计算小结	52
第四章 矩阵代数基础.....	57
§4.1 矩阵	57
§4.2 行列式	65
§4.3 分块矩阵	74
§4.4 线性相关、秩与齐次方程组的解	79
§4.5 特征根和特征向量	84

§4.6 二次型和正定矩阵	86
§4.7 矩阵的微分计算	93
第五章 一般线性模型	100
§5.1 假 设	100
§5.2 最小二乘估计	101
§5.3 相关矩阵、偏相关系数和回归系数	109
§5.4 显著性检验与置信区间	112
§5.5 预测	128
§5.6 线性约束	130
§5.7 多重共线性	134
§5.8 设定误差	142
第六章 一般线性模型的扩展	150
§6.1 虚变量	150
§6.2 季节调整	157
§6.3 协方差分析	162
第七章 广义最小二乘法	176
§7.1 广义最小二乘 (AITKEN) 估计式	176
§7.2 预测	179
§7.3 异方差扰动	181
§7.4 纯粹估计和混合估计	187
§7.5 观察值的分组	193
§7.6 方程的分组	202
第八章 序列相关	207
§8.1 序列相关的性质	207
§8.2 序列相关扰动的后果	209
§8.3 序列相关的一般检验方法	212
§8.4 Theil的最优线性无偏标量协方差方法 (BLUS)	217
§8.5 估计方法	221
§8.6 预 测	226
第九章 随机解释变量、工具变量以及变量中的误差	228
§9.1 定义	228
§9.2 随机解释变量	233
§9.3 工具变量	238

§9.4 变量中的误差	240
第十章 滞后变量	250
§10.1 滞后解释变量	250
§10.2 滞后被解释变量	256
§10.3 估计方法	259
第十一章 其他多变量方法	277
§11.1 主分量方法	277
§11.2 典则相关分析方法	284
§11.3 判别分析方法	287
第十二章 联立方程方法——识别	294
§12.1 联立方程系统	294
§12.2 识别问题	302
§12.3 对结构参数的约束	306
§12.4 对方差和协方差的约束	313
第十三章 联立方程方法——估计	323
§13.1 递归系统	323
§13.2 两阶段最小二乘 (2SLS) 估计式	326
§13.3 有限信息估计式 (最小方差比估计式)	330
§13.4 k-级估计式	333
§13.5 两阶段最小二乘法 (2SLS) 和主分量	338
§13.6 三阶段最小二乘法 (3SLS) 和完全信息最大似然法 (FIML)	340
§13.7 预测和联合置信区间	343
§13.8 蒙特卡洛研究	351
附录	364
索引 (略)	

第一章 计量经济学的性质

§1.1 变量之间的关系

对于研究经济学的人们来说，首先需要建立的一个基本概念就是经济变量之间的关系。例如，可以把市场商品的需求量看作该商品价格的函数，可以把产品的生产成本假设为产量的函数，可以把消费支出作为收入的函数等等，上述这些都是两个变量关系式的例子。而更加切合实际情况的表述则要求在每个关系式中包括更多的变量。因而，可以把商品的需求量看作是价格、可支配收入、以及其他有关商品的价格的函数；生产成本依赖于生产率、生产要素价格和生产率的变化；可以把消费支出规定为收入、流动资产和以前消费水平的函数等。

§1.2 经济模型

研究经济理论的下一个步骤就是组合关系构造模型。经济模型所包括的经济关系的数目取决于构造模型的目的和探寻的解释程度。例如，传统的供给和需求模型研究的是在某一特定市场中，价格与商品量的组合；这个模型包括三个方程，即需求方程、供给方程和市场调节方程。这些方程除了包括商品量和价格之外，还包括一些其他的变量，例如需求方程中的可支配收入，供给方程中的生产要素价格等。因此，模型的解释能力也与其他变量有关，从这种意义上讲，模型是不完全的或是有条件的模型。目标更加广泛的模型将包含许多方程，并试图解释许多变量的性态。然而，大多数模型仍将是有条件的，因为模型中包含一些不能由它本身来确定或解释的变量。

所有的经济模型，不论是宏观的，还是微观的，不论是关于经济的、工业的、公司的还是市场的，都有一些共同的性质。第一，假设经济变量的性态是由若干个经济关系式联合同时运行决定的；第二，尽管模型是对实际的复杂情况的一种简化，但它必须把握所研究的经济部门或经济系统的关键特征；第三，希望通过对反映系统的模型的研究，预测系统的未来发展，并尽可能控制这些发展朝着改善经济状况的方向发展。

§1.3 一个简单的宏观经济模型

为了说明前述的观点以及便于解释计量经济学的特殊贡献，先来研究一个非常原始的模型。假设一个经济理论家提出了三个命题：第一，消费是可支配收入的递增函数，但消费的增长可能小于可支配收入的增长；第二，投资是国民收入的递增函数，同时还是政府调节因素（例如利率）的递减函数；第三，国民收入是消费、投资以及政府在商

品和服务方面支出的总和。

首要的任务就是将这些命题转换为数学形式，这里马上就碰到了一个问題，即在满足理论家的上述设想的情况下，可能有各种形式的模型。变量之间的关系式究竟是线性的，还是非线性的，如果是非线性的，那么是对数呢，还是多项式或其他别的函数形式呢？甚至在确定这些关系形式的时候，仍存在着各个方程中的时滞关系问题。例如，现期投资是与上一个时期的国民收入相对应，还是与以前几个时期的国民收入相对应呢？最初，解决这些困难的惯常作法是尽可能地简单地假定这些关系。于是，人们可以把上述命题转化为下列模型

$$C_t = \alpha_0 + \alpha_1 (Y_t - T_t) \quad (1-1)$$

$$I_t = \beta_1 Y_{t-1} + \beta_2 R_t \quad (1-2)$$

$$Y_t = C_t + I_t + G_t \quad (1-3)$$

这里可以把对模型的先验约束表示成

$$0 < \alpha_1 < 1, \quad \beta_1 > 0, \quad \beta_2 < 0$$

这三个关系式连同上述约束条件一起就构成了一个模型。其中 C 表示消费， I 表示投资， Y 表示国民收入， G 表示政府在商品和服务方面的支出， T 表示所得税， R 表示政府的调节因素。

这个模型是由两个解释消费和投资活动的行为关系式以及一个恒等式组成。所构造的这个模型称之为离散时间周期的模型，并在投资方程中规定了国民收入滞后一个周期。这里，把两个行为关系式暂时规定为精确的函数，但是正像在后面将要看到的，这是不符合实际情况的，计量经济学工作不可能不涉及某些附加的随机因素。

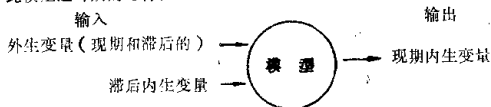
为了说明模型的功能，首先把模型中的变量作如下分类：

C_t, I_t, Y_t 是现期内生变量

T_t, R_t, G_t 是现期外生变量

Y_{t-1} 是滞后内生变量。

这种分类的根本理由在于，这一模型的意图是根据外生变量和滞后内生变量的值解释现期 (t) 内生变量的值。但是，模型并没有给出外生变量是如何确定的，当然，这意味着可以通过置这些外生变量的值来得到所希望的内生变量的值。然而，正如模型本身所表明的，这个过程要受到某种约束，因为一个内生变量的过去值也对现期内生变量产生部分影响。更为一般地说，滞后外生变量也将会在模型中出现，于是，两组外生变量（现期的和滞后的）以及滞后内生变量统一称为先决变量。因而，可以用下图概括描述此模型逐时期的运行状况



通常把关系式 (1-1) 到 (1-3) 称为模型的结构式。然而，如果把一个模型转换成

它的简化式，就可以更清楚地看出确定内生变量的过程。为了把每个现期内生变量表成仅仅是先决变量的函数，可以通过逐次替代的方法来实现。例如，把(1-2)式和(1-3)式代入(1-1)式便可得到

$$C_t = \alpha_0 + \alpha_1 (C_t + \beta_1 Y_{t-1} + \beta_2 R_t + G_t - T_t)$$

或者

$$C_t = \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_1 \beta_1}{1-\alpha_1} Y_{t-1} + \frac{\alpha_1 \beta_2}{1-\alpha_1} R_t + \frac{\alpha_1}{1-\alpha_1} (G_t - T_t) \quad (1-4)$$

投资方程已经是简化式形式了，因为它仅仅包含现期内生变量 I_t ，所以可以把(1-2)式简单地写作

$$I_t = \beta_1 Y_{t-1} + \beta_2 R_t \quad (1-5)$$

然后，利用(1-3)式、(1-4)式和(1-5)式得到

$$Y_t = \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_1}{1-\alpha_1} Y_{t-1} + \frac{\beta_2}{1-\alpha_1} R_t + \frac{1}{1-\alpha_1} G_t - \frac{\alpha_1}{1-\alpha_1} T_t \quad (1-6)$$

方程(1-4)、(1-5)和(1-6)就构成了模型的简化式。简化式系数均是原始的结构式系数的函数，每个外生变量的系数都有特定的意义，通常把这些系数称作影响乘数，因为它们显示了任一现期外生变量的变化对每个现期内生变量的影响。例如，当政府调节因素增加一个单位时， C_t 就变化 $\alpha_1 \beta_2 / (1-\alpha_1)$ ， I_t 变化 β_2 。因为模型是线性的，所以同一时期若干个外生变量变化的影响就是各个外生变量分别影响的简单和。于是，政府消费和税收同时增加一个单位，消费和投资不变时，国民收入增加一个单位。

在这个简单的模型中，对结构式参数施加的先验约束， $0 < \alpha_1 < 1$ ， $\beta_1 > 0$ ，及 $\beta_2 < 0$ ，是充分明确地确定方程(1-4)到(1-6)中简化式参数的符号所必须的。仅有结构式参数符号的定性信息是不够的，如果有可能，必须应用定量信息，之所以如此，至少有以下五个主要理由。第一，迄今为止有关该模型的定性信息纯粹是一种先验信息，人们希望看看这些先验假设是否符合实际数据。另外，模型的简单线性设定也需要对照数据进行检验，看看它是否对所观察到的行为给出了充分地表现，或者是否需要建立更加复杂的模型。第二，在更加复杂的模型中，仅有结构式参数符号的正确信息可能还不能给出影响乘数符号的明确指示。第三，由于某些外生变量可能是制定更加有效的政策的工具，为此，需要有影响乘数的某些定量信息。第四，如果希望分析模型包含经济行为类型（是衰减的还是蓬勃发展的，是振荡的还是单调的），最重要的是要掌握结构式系数的数值估计。例如，(1-6)式是国民收入的一阶非齐次差分方程，它的解依赖于系数 α 和 β 的数值，如果结构式方程包含一个时期以上的滞后，那么(1-6)式将是国民收入的二阶或更高阶的差分方程，解的性质就不可能根据差分方程系数的符号方面的信息来简单地确定。第五，外生变量的不同组合对内生变量数值预测的影响只有

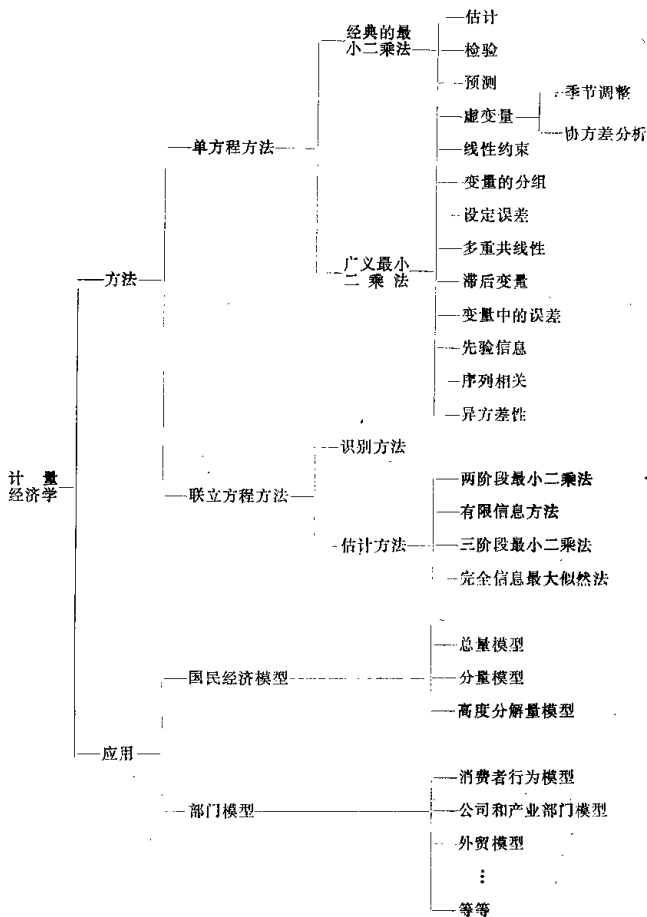
根据简化式参数的数值估计才能得到。^①

§1.4 计量经济学的任务

计量经济学的根本任务是估计经济模型和检验经济模型。进行计量经济学工作的第一步，就是设定模型的数学形式，正像我们已经看到的，从经济理论中引伸出来的先验约束并不足以给出精确的数学形式。第二步，是从组织机构或地区部门收集描述模型意图的相关数据。第三步，是利用所收集的数据估计模型的参数。第四步，是对所估计的模型进行检验，目的在于评定它是否反映了所研究的经济系统的实际状况，或者是否还需要估计不同形式的设定模型。也应当指出，如像方程(1—1)和(1—2)那种精确关系式的假设必须消除。在任何时候，经济数据总不能给出这类简单关系式的精确拟合，因为线性的或其他的简单形式仅是对可能复杂的而又未知的形式的一种近似，也因为任一模型的设定中通常可能包含的解释变量仅仅是所有可能解释变量的一小部分，所有这些因素都要求在每种关系式中规定某种随机误差或扰动项，而不是把每种关系式均指定为恒等式。的确，每种关系中误差方差的估计值是计量经济学加工的一个极为有用的结果，不仅在阐明计量经济学关系式的质量方面起着决定性的作用，而且对于估计系数的评价以及对于利用所估计的模型进行预测都是不可缺少的重要参数。

下图给出了计量经济学研究的领域，图中的大部分名称暂时将不是有意义的，但这个图示将帮助读者读完本书，为读者指出了一条研究路线。图的上半部分是本书涉及到的计量经济学方法，研究这些方法的目的在于学会如何对应用工作需要的计量经济学模型的参数做出尽可能好的估计。

① 读者若想透彻地了解经济模型数值估计的详细内容可参考D.B.Suit, "Forecasting and Analysis with an Econometric Model," *American Economic Review*, vol.LIII, PP.104—132, 1962年。具有矩阵知识的读者可以参考H.Theil, J.G.G.Boot, "The Final Form of Econometric Equation Systems", *Rev. Intern. Statist. Inst.*, vol.30, PP.136—152, 1962年。



第二章 两个变量的线性模型

§2.1 假 设

在第一章已经讲到, 计量经济学涉及到经济模型参数的估计问题。典型的经济模型是由若干个方程组成的, 每个方程又包含了若干个变量。考虑到经济模型可能具有的复杂性, 又由于不可能进行经济学的控制试验, 人们可以料想到模型参数的估计是一个非常困难的问题, 实际情况正是如此。但是, 复杂的方法是建立在比较简单的基本方法之上的, 对这些基本方法的透彻了解是处理更复杂情况的前提。特别要指出的是, 完全掌握最小二乘法原理, 是本书内容所要求的。因此, 这里将从最简单、最基础的情况开始, 即假设只有一个方程, 并且方程中只有两个变量。

现在, 用 Y 和 X 来表示这两个变量, 假设

$$Y = f(X) \quad (2-1)$$

这一步仅仅说明 X 是一个变量, 它的变化会影响到另一个变量 Y 。

下一步的工作是确定 Y 与 X 之间关系的形式。基于研究(2-1)式的理论, 可以提出一种可用的精确函数形式, 或者仅仅提出函数的截距、斜率及曲率等某方面的条件, 而可以满足这些条件的函数又可能有若干个。因此, 就需要借助于统计分析的方法对这些函数进行选择。

两个变量之间最简单的一种函数关系是线性关系。即:

$$Y = \alpha + \beta X \quad (2-2)$$

在此式中, α 与 β 是未知的参数, 表示函数的截距和斜率, 两个变量之间的其它关系有

$$Y = \alpha e^{\beta X} \quad Y = \alpha X^{\beta} \quad Y = \alpha + \beta \frac{1}{X}$$

上述的第三个关系式表示的是变量 Y 与 $1/X$ 之间的线性关系, 第一、二个关系都可以用等式两边取对数的方法分别化成:

$$\log_e Y = \log_e \alpha + \beta X$$

和

$$\log_e Y = \log_e \alpha + \beta \log_e X$$

前一个方程表示的是 $\log_e Y$ 与 X 之间的线性关系, 后一个方程表示的是 $\log_e Y$ 与 $\log_e X$ 之间的线性关系。

大多数的传统经济理论, 或是以图形表示, 或是以代数公式表示, 都假定在经济变量之间存在着精确的函数关系。但用最常见的经济数据摘出来的那些点并不是精确地落在某一条直线或其它的光滑曲线上。因而, 对于量度和检验经济关系来说, 仅有象

(2-1)式或与其有联系的其它函数形式是不够的。为此，就需要在经济关系中引入随机扰动项，以扩展它的应用。

例如，假设研究的是用某一时期横剖面数据给出的家庭消费支出和可支配收入之间的关系。令 Y 表示消费支出， X 表示可支配收入，这样，从10,000户家庭的收支资料中就可以取得10,000对数据 $X_i, Y_i (i=1, 2, \dots, 10,000)$ 。又假设根据家庭规模和构成，这些家庭划分为不同的组，然后，在某个组里观察 X 与 Y 之间的关系。这里，不可能期望在具有某一给定收入 X' 的组内，所有家庭的消费支出 Y 均相等。实际上，这些家庭的消费支出总会有多与少的。但的确可以期望这些消费支出数字会聚集在一个与问题中的收入值相适合的值周围。这种想法可以用下面新的线性假设来表示。

$$Y = \alpha + \beta X + u \quad (2-3)$$

式中的 u 表示一个可为正值也可为负值的变量，于是，如果所研究的那些家庭子组具有一定的收入 X' ，其消费支出的中心值将是 $\alpha + \beta X'$ ，而这个子组中各个家庭的实际消费数字可用 $\alpha + \beta X' + u_1, \alpha + \beta X' + u_2, \dots$ 等来表示，其中 $u_1, u_2, \dots$ 表示各个家庭的消费支出超过或少于 $\alpha + \beta X'$ 的数量。

在(2-3)式中加入 u 的合理性可用以下三点理由来说明，这三点之间也并不是互不相干的。第一，如果知道了每个家庭消费支出的全部作用因素，并且掌握了所有必需的数据，那就可以对每一个家庭户的消费支出进行详尽的解释，即使是家庭的规模和结构完全相同，在很多方面仍会存在着区别。例如，父母和孩子的准确年龄，结婚的年限，丈夫是否喜欢打高尔夫球、喝酒、打扑克或养鸟，妻子是否喜欢高耸的帽子和巴黎时装，是否喜欢游泳、或是外国赛车，家庭收入是在增加还是在减少，这个家庭中的父母本人是来自勤俭、谨慎的家庭，还是来自挥霍浪费的家庭，等等。在解释人们行为的时候，涉及的相关因素可能是无穷无尽的。但是，有许多因素是无法定量描述的，即使是可以定量描述的，通常在实践中也不可能获得它们的完整资料，即使人们可以得到若干因素的完整资料，这些因素的数目几乎可以肯定会超过可用的观察值数目，而估计这类因素影响的统计方法并不存在。况且，许多变量的影响都非常小，即便是有了实际的定量资料，对这些变量的影响进行统计估计也是非常困难和不确定的。如果用函数形式来表示这种情况，则有 $Y = f(X_1, X_2, \dots, X_n)$ ，这里的 n 是一个非常大的数字。因此，在这种情况下，就把 Y 表示成少数几个比较重要的 X 的显函数，而把其它变量的影响统统用 u 来表示。在单一显变量的极限情况下，就有

$$Y = f(X_1, u) \quad (2-4)$$

在某个家庭中，可能有很多因素都对这个家庭的收入与消费支出发生影响，其中不少的因素可能是在相反的方向上互相牵扯。因此，可以料想在大多数的情况下， u 的值都是较小的。由此，可以把 u 想象成是一个以零为中心，有有限方差 σ_u^2 的某种概率分布的变量，这就是把 u 看成是一种随机扰动项（或误差）的理由。从所包含的这些因素影响来看，它们之间的大多数都是独立的，根据中心极限定理，可以把 u 看成是服从正态分布的一个变量。^①

① 见W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed., Wiley, New York, 1957, vol.1, pp.238-241.

在经济关系中引进扰动项的第二个理由是假设除了所有的相关因素之外，还存在一个最基本但又不可预测的随机性因素，这就是人对问题的反映，它只能用一个随机变量来描述。对于实用统计学来说，并不需要把上面关于 u 的合理性的两个理由加以区别，由于理论和数据两个方面的原因，在任何一个经济关系中，都不可能把所有可分离和相关的因素都包括进来。因此，嵌入一个随机项主要是考虑到第一个理由，如果第二种现象出现的话，只不过是增加了随机项的方差。这些随机项有时都被称为方程中的扰动项或误差项。

误差的第三个来源是观察误差或测量误差。例如：变量 Z 与 X 以 $Z = \alpha + \beta X$ 确定了它们之间的精确的线性关系，但由于有测量中的误差，使得人们难以看准 Z 的真值，这样，实际观察到的就不是 Z ，而是 $Y = Z + u$ ，这里的 u 就表示测量误差。因此有

$$Y = Z + u$$

也就是

$$Y = \alpha + \beta X + u$$

当然，也可能有测量误差和方程误差叠加在一起的情况。如下例所示，

$$Z = \alpha + \beta X + u \quad \text{方程误差}$$

$$Y = Z + v \quad \text{测量误差}$$

因此有

$$Y = \alpha + \beta X + w$$

式中的 $w = u + v$ 。这里若仅用一个变量来研究测量误差，就可能是不符合实际情况的。对于测量误差的详细讨论将在第九章中进行，前面的一些章节将着重研究方程中的误差。

最初拟定经济关系时，必须建立一些关于扰动项概率分布的假设，而这些假设必定与均值、方差、协方差有关。最简单的假设是设均值为零，方差为常数， X 是独立的，各个 u 值也是相互独立的。现在，为继续讨论上面提到过的例子，假设把可支配收入分成各个不同的水平组。为了方便起见，按收入水平，从小到大地把这些组排列起来，即 $X_1, X_2, \dots, X_n$ 。如果线性假设是真实的，以及上面关于扰动项的假设也都成立，就可以用图2—1来说明这种情况。以直线 $\alpha + \beta X$ 为中心的分布就被假设为 u 的分布。在这个例子中，如果假设扰动项的方差随着 X 的增加而增大，可能是更符合实际情况的。对这个问题的详细讨论将留到第七章中去进行。

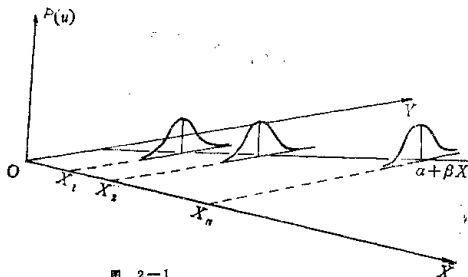


图 2—1

现在, 如果选一个包含 n 个家庭的样本, 每个家庭都是选自不同的收入水平组, 就可以用图2-2的散布图来表示这些样本点。如果 u 值是独立抽取的, 那么样本点将随机地聚集在一条直线的周围。如果一个正的 (或负的) u_1 值并不以任何方式影响 $u_2, u_3, \dots$ 的值, 就是这种情况。另一方面, 如果收入偏低的家庭多数都是由那些结婚不久的年青人组成, 而收入偏高的家庭多数都是由那些即将退休的老年人组成, 那么, 就可以料到对较小的 X 值有正的扰动项, 而对较大的 X 值则有负的扰动项。扰动项的这种非随机性指出了一个被省略的重要解释变量, 即年龄。在扰动项中包含年龄这个因素就会阻碍

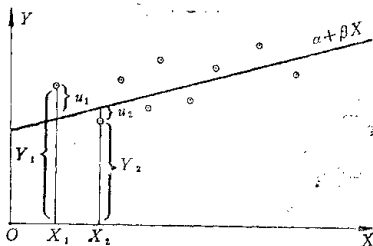


图 2-2

随机性的显露。但实际情况与图2-2有一个本质上的区别, 就是图中的直线 $\alpha + \beta X$ 是未知的。为此, 给出一组假设:

$$\begin{aligned}
 Y_i &= \alpha + \beta X_i + u_i & i &= 1, 2, \dots, n, \\
 E(u_i) &= 0 & \text{对所有的 } i \\
 E(u_i u_j) &= \begin{cases} 0 & i \neq j; i, j = 1, 2, \dots, n \\ \sigma_u^2 & i = j; i, j = 1, 2, \dots, n \end{cases}
 \end{aligned}$$

(2-5)

在 (2-5) 式中, α 、 β 和 σ_u^2 都是未知的参数。根据 X 和 Y 的样本观察值, 在统计上可以估计出这些参数, 还可以检验有关这些参数的假设。例如, 是否可以假设消费同收入成正比关系 ($\alpha = 0$)? 消费的边际倾向是否大于 0.5? 线性函数是否能较好地反映出这些数据之间的关系? 关于扰动项方差是常数的假设对于这些数据来看是否能成立? 这些都是典型的统计推断问题。下面的任务就是要讨论适合于处理这些问题的标准方法。正如读者将要在后面看到的, 许多计量经济学理论都涉及到研究经济模型中处理统计推断问题的新方法, 这是因为标准方法并不是在所有情况中都适用的。但是, 对标准方法的透彻了解是掌握那些更复杂技术的基本前提。