第 1 章　面 板 数 据

随着人们对经济活动现象认识的深化，单纯应用截面数据或时间序列数据来检验经济理论、寻找经济规律和开展经济预测均受到样本数据异质性与序列相关性的严重影响。为了进一步提高计量经济学模型的可靠性，自 20 世纪 70 年代以来，计量经济学家利用面板数据不仅可以揭示经济行为的动态变化，而且可以保证样本信息充分和全面。目前，面板数据计量经济分析已经成为计量经济学研究的重要分支之一。

1.1　面板数据及其分类

1. 面板数据的定义

“面板数据”(panel data)一词指的是**一部分**个体(个人、家庭、企业或国家等)的某个经济属性在一段时期内的观测值所构成的二维数据集合。这样的数据可以通过在一段时期内对一些家庭或个体进行跟踪调查来获得，也可以从相应的统计汇总资料，如统计年鉴采集得到。

例如，1984—2014 年我国 31 个省、自治区、直辖市(不含港澳台)的“居民消费价格指数(CPI，上年＝100)”数据，如表 1-1 所示。

表 1-1　31 个省、自治区、直辖市居民消费价格指数的面板数据

地　　区	1984 年	1985 年	1986 年	…	2014 年
北京	102.20	117.60	106.80	…	101.62
天津	101.80	113.10	106.80	…	101.85
河北	102.50	106.80	105.70	…	101.72
…	…	…	…	…	…
新疆	102.40	107.80	107.30	…	102.11

这样的二维表格数据就是变量 CPI 的面板数据。

从每个时期的截面看，面板数据是由若干期截面数据组成，每个截面是所有个体在某一时点的观测值；并且，从每个个体来看，面板数据是由各个体的时间序列数据组成。

显然，与截面数据(差异的静态性)比较，面板数据能够反映个体差异性的动态行为(差异的动态性)。例如，研究各省、自治区、直辖市物价水平差异随时间的变化性。与时间序列(动态行为的同质性)比较，面板数据可以揭示个体动态行为的差异性(动态行为的异质性)。例如，研究各省、自治区、直辖市 CPI 动态演化机制的差异性。

所以，**面板数据是对截面数据静态性和时间序列数据同质性的推广，不仅能够揭示个体差异性的动态行为，也可发现个体动态行为的差异性**。

2. 微观面板数据和宏观面板数据

在经济学的应用中，通常将面板数据分为微观面板(micro panels)数据和宏观面板(macro panels)数据两类。微观面板数据是由微观个体的调查数据组成的面板数据，其特点是个体数 N 较大(通常是数百个个体)，而时期数 T 较小(最短是 2 期，最长不超过 10 期或 20 期)。例如，近 5 年上海证券交易市场上市公司的年末资产负债率数据；或者，对一年内某一地区城镇居民家庭每月家庭食品支出调查获得的数据。宏观面板数据是由一段时期内不同国家的宏观经济数据得到的面板数据，这类数据一般具有适度规模的个体数 $N(>10)$，时期数 T 一般较大(在 20 期到 60 期之间)。例如，1952 年至 2017 年中国 31 个省、自治区、直辖市(不含港澳台)的 GDP(国内生产总值)增长率数据、改革开放以来中国国民经济 30 个行业大类[①]的年末就业人数，等等。

数据结构上的差异，使得微观面板和宏观面板的计量分析方法有所区别，它们主要表现在如下三个方面：①渐近性：对于微观面板数据模型的估计量与检验统计量必须研究 T 固定、N 较大趋于无穷的渐近性质；而研究宏观面板数据模型的渐近性质时，应讨论在联合极限、对角极限和序贯极限下的统计性质。②平稳性：对于宏观面板数据，当时间序列较长时，需要考虑数据的非平稳问题，如单位根和协整等问题；而微观面板数据通常不需要处理非平稳问题，特别是每个家庭或个体的时期数 T 较小时。③空间相关性：在建立宏观面板数据模型时必须考虑国家之间的相关性(包括空间面板数据的空间相关性)；而在微观面板数据中，如果个体是随机抽样产生，则个体之间不大可能存在相关性。

3. 常用面板数据库及其应用

早在 1968 年，为了研究美国的贫困特征及其原因，密歇根大学社会科学研究所就建立了研究收入动态行为的面板数据(panel study of income dynamics，PSID)，俄亥俄州立大学人力资源研究中心开发了国家劳动力市场长期调查面板数据(national longitudinal surveys of labor market experience，NLS)。之后，美国又相继建立了面板数据 LRHS (longitudinal retirement history study)、CPS(current population survey)和 HRS(health and retirement study)。1989 年德国建立了德国社会经济面板数据集(German socio-economic panel，GSOEP)；1993 年加拿大建立了加拿大劳动力收入动态调查面板数据(Canadian survey of labor income dynamics，CSLID)；2002 年，欧共体统计办公室建立了欧共体家庭面板数据(European Community household panel，ECHP)。Borus(1982)、Wagner(1993)和 Peracchi(2002)等西方经济学家应用这些微观面板数据对微观经济学、发展经济学和劳动经济学等众多经济学的热点问题进行了广泛研究。

在国内，学术界常用的微观面板数据的原始资料来源于下面三个数据库。

1) 中国健康与养老追踪调查数据库

2011 年北京大学国家发展研究院开发建立了中国健康与养老追踪调查(China health and retirement longitudinal study，CHARLS)数据库，覆盖 150 个县级单位、450 个

① 资料来源：中国国家统计局发布的《2017 年国民经济行业分类(GB/T 4754—2017)》，http://www.stats.gov.cn/tjsj/tjbz/hyflbz/201710/t20171012_1541679.html.

村级单位、约 1 万户家庭中的 1.7 万人。这些样本以后每两年追踪一次，调查结束一年后，数据将对学术界公开。调查收集了代表中国 45 岁及以上中老年人家庭和个人的高质量微观数据，用以分析我国人口老龄化问题，推动了老龄化问题的跨学科研究。截至 2017 年，CHARLS 数据用户量达到两万余人，基于 CHARLS 数据在国内外发表的学术论文已达 708 篇。例如，鲁万波等（2018）和李超等（2018）分别利用 CHARLS 数据研究了农村老年人健康问题和老龄化社会家庭储蓄率的决定问题。

2）中国工业统计数据库

始建于 1998 年的中国工业统计数据库（Chinese industry statistical database）涵盖了中国销售额 500 万元人民币以上的 43 万多家大中型制造企业（不含港澳台），即包括国有企业、集体企业、股份合作企业、联营企业、有限责任公司、股份有限公司、私营企业、其他内资企业、港澳台商投资企业和外商投资企业等。工业统计指标包括工业增加值、工业总产值、工业销售产值等主要技术经济指标以及主要财务成本指标和从业人员、工资总额等数据。

3）中国海关进出口贸易数据库

2000 年中国海关总署建立的中国海关进出口贸易数据库按照中国所有进出口代码（4 位国际码、8 位中国海关指定商品代码）记录了每年 2 000 多万条 12 000 多种商品的进出口资料；每月完整的进出口数据信息涉及产销国、关区、经营单位、使用地区、数量、金额、进出口公司等。

例如，蒋银娟（2016）采用中国工业企业数据库和海关企业数据库的匹配数据分别研究了企业进口中间品多样化与企业产出波动之间的关系和价值链参与对企业出口产品质量的影响。

近年来，应用宏观面板数据研究宏观经济问题的文献也层出不穷。例如，在国际金融学领域，Chinn 和 Johnston（1996）与 MacDonald 和 Nagayasu（2000）等使用一些国家宏观面板数据检验购买力平价理论（PPP），研究实际汇率决定问题；在世界经济学领域，Michael（2003）和 Jansen（2000）等应用宏观面板数据研究国际资本流动问题、东欧转型经济国家的出口变化和经济增长问题以及欧美国家的失业问题；在发展经济学中，Strauss（2000）、Nerlove（2002）和 Migue（2002）分别应用面板数据的计量经济学方法研究经济体经济增长的决定因素和经济增长收敛理论；Baicker（2005）建立空间面板模型检验了美国州政府财政支出的空间相关性，研究发现一个州政府的财政支出与其邻近州政府的财政支出显著地正相关。再如，邵红伟和靳涛（2016）利用世界银行 WDI（世界发展指标）数据库中 149 个国家和地区 1981—2013 年的跨国面板数据再次证实了收入分配的库兹涅茨倒 U 曲线，特别地，发现中国大致在 2011 年以后已经进入库兹涅茨拐点区，收入差距会在一定时期内维持稳定。邓慧慧和赵家羚（2018）基于 2006—2014 年 249 个地级市的面板数据，应用空间面板计量模型从社会互动视角检验地级市政府设立开发区的动机和激励机制，研究发现，地方政府设立开发区的“同群效应”更多是源于对经济发展水平较高城市的模仿行为，且自身经验并不能抑制模仿冲动。

1.2 面板数据的优势与局限

1. 面板数据的优势

与传统的截面数据和时间序列数据结构比较，面板数据具有如下优势。

1）控制异质性、降低参数估计的偏倚

面板数据能反映个体、企业、州或国家之间存在的异质性，即时间上和空间上的异质效应。而时间序列数据和截面数据分析没有控制这种异质性，因而其结果很可能是有偏的。

例如，Baltagi 和 Levin(1992)研究 1963—1988 年美国 46 个州的香烟需求问题时，设定香烟需求模型

$$D_{it} = g(D_{i,t-1}, p_{it}, I_{it}, \text{reli}_i, \text{edu}_i, \text{adv}_t, \cdots)$$

则模型中解释变量包括可观测的和不可观测的两大类。

可观测的解释变量可分为如下三种。

(1) 可观测的时变异质因素，即随个体(州)和时间的变化而变化的可观测变量，如香烟消费量的滞后项、价格和收入等可观测的变量。

(2) 随个体(州)变化而不随时间变化的可观测变量，如宗教和教育等变量，即可观测的非时变异质因素。

(3) 不随个体(州)变化而随时间变化的可观测变量，如电视和广播中的广告等变量，即可观测的时变同质因素。

不可观测的解释变量包括下述三种。

(1) 个体效应，非时变的异质性不可观测因素(ξ_i)。

(2) 时间效应，时变的同质性不可观测因素(λ_t)。

(3) 剩余效应，时变的异质性不可观测因素(u_{it})。

这样，模型可设定为

$$D_{it} = f(D_{i,t-1}, p_{it}, I_{it}, \text{reli}_i, \text{edu}_i, \text{adv}_t) + \xi_i + \lambda_t + u_{it}$$

事实上，对于宗教变量，人们不可能得到每年每个州某一宗教人数占总人口的百分比，所以一般认为不同年份的百分比也不会有太大变化。同样，完成高中或大学学业的人数占总人口的百分比也是如此。电视和广播中的广告是全国性的，它对需求的影响不会随着州的不同而变化。

显然，面板数据模型能够基于这六类中的所有变量建模，遗漏六类中任何一种都可能导致估计结果的偏倚。

2）降低多重共线性、增强有效性

面板数据具有更多的信息、更大的变异性，降低了解释变量间的多重共线性；使面板数据模型自由度提高，从而增强了参数估计的有效性。

时间序列研究中令人烦恼的问题之一是多重共线性；如在上述香烟需求的研究中，如果从总量的角度看价格和收入就具有很强的共线性，而使用一国各省市(州)的面板数

据，存在共线性的可能就很小了，因为增加截面个体维度的同时也增加了数据的变异，也增加了有关价格和收入的信息。

事实上，数据中的变异可以分为两个部分：一是各省市（州）之间由于规模和经济特征的不同所表现出的变异性；二是各省市（州）内部不同时间上表现出的变异性，前者的变异程度往往更大。使用更多、更有信息的数据就可以得到更可靠的参数估计值。当然，这要求不同省市（州）的变量间应具有相同的关系式，换句话说，这要求数据是可混合的（poolable）。

3）更适合于研究经济状态的持续性

面板数据可以将个体在某个时点的经历和行为与另一个时点的其他经历和行为联系起来，包含了个体差异性的动态行为，所以，面板数据适用于研究经济状态的持续性问题。并且，如果这些面板数据的时期足够长，它们就能够更清楚地反映经济政策对个体差异性的动态影响。

例如，在测度失业率时，截面数据只能估计出人口中多大一部分比例在给定的时间处于失业状态，截面时间序列数据（统计调查数据）仅仅表明这一比例如何随时间而变化。然而，只有面板数据才能确切地估计出在某个时期的失业者中有多少在另一个时期仍处于失业状态，即反映了失业状态的持续性。并且，能够揭示政策对解决失业、贫困问题的效果。因此，面板数据在研究失业、贫困等经济状态的持续性方面具有特殊的作用。

4）简化了统计推断

在研究面板数据模型的参数估计量或检验统计量的渐近分布时，尽管理论上具有三种收敛方式，但是，最常用的渐近方式是序贯极限的收敛方式。所以，面板数据的统计量往往具有标准的渐近分布。例如，与时间序列分析中进行单位根检验遇到的非标准分布问题不同，面板单位根检验统计量通常具有标准的渐近分布。因此，面板数据简化了统计推断。

5）模型更具有普遍性

面板数据可以构建并检验更复杂的行为模型。例如，对技术效率问题使用面板数据建模研究效果更好（Baltagi 和 Griffin，1988；Baltagi，Griffin 和 Rich，1995；Koop 和 Steel，2001）。另外，在分布滞后模型中使用面板数据比使用纯时间序列数据需要的约束条件更少，因为通常使用 GMM（广义矩方法）估计。因此，基于面板数据建立的模型更具有普遍性。

由此可见，因面板数据及其分析方法的诸多优点，面板数据的计量分析方法及其应用研究主导了近年来社会科学界的实证研究。

2. 面板数据的局限性

半个世纪以来，基于自然实验数据进行因果关系推断是社会科学界统计分析的基本问题之一。尤其，根据面板数据进行反事实因果推断也已成为学术界的共识。然而，在分析面板数据的方法上，并未取得一致。因此，在运用面板数据研究社会科学问题时，广大应用者也应该清楚地认识面板数据分析方法的局限性。面板数据的局限性包括以下几方面。

1）微观调查面板数据极少

在针对微观个体的调查中，因追踪调查设计中对象选择和采访频率确定的复杂性、数据采集时样本自选择、无应答、个体的非随机流失和威望偏倚（prestige bias）等问题，以及数据管理的多维性，使得真正意义的微观面板数据很少（Kasprzyk 等，1989）。人们逐渐使用伪面板（pseudo-panel）和轮换面板（rotating panel）开展微观经济分析。

2）制约了经济稳态性质研究

微观面板各个体的时期较短，主要依赖个体数趋于无穷进行渐近统计分析，在时间维度不能进行渐近性分析，制约了对经济稳态性质的研究。

3）宏观面板数据的截面相关性

在宏观经济问题的研究中，由于国际贸易、外商直接投资和地缘经济特征等因素的作用，经济的外溢性和内联性非常突出。所以，对于国家或地区间的宏观面板数据，建立计量经济模型时通常需要考虑截面相关性。并且，对于时期较长的面板数据，如果未考虑国家之间的相关性，将会导致错误的统计推断结论。事实上，早期的面板数据分析方法通常假设个体间是相互独立的。例如，LLC、IPS 和组合 p 值的面板单位根检验方法。近年来，学者们也提出了考虑这种相关性的面板单位根检验方法。并且，空间面板数据计量模型已成为目前的重要计量分析工具。因此，在应用面板数据计量方法时，必须慎重研究各种方法的前提假设及其适用性。

总之，面板数据不是灵丹妙药，它并不能解决传统计量经济分析（时间序列或截面研究）中解决不了的所有问题。例如，面板单位根检验比单一时间序列的单位根检验功效更高，这应该能更好地推断购买力平价和增长收敛问题。事实上，面板数据在导致大量经验应用研究的同时，也引来了一些批评，Maddala（1999）和 Banerjee 等（2004，2005）认为面板数据也不能解决 PPP 以及增长收敛的问题。

1.3 扩展的面板数据

近年来，关于面板数据建模的理论与应用研究发展迅速，并取得了一些重要的研究成果，研究论文涵盖了经济学、管理学、统计学、人口学、环境科学、医学、遗传学、社会学等领域。然而，发展并不均衡，有关平衡面板数据的线性回归模型的应用研究比较丰富，而有关面板数据模型的新近理论研究及其应用尚待深入，下面将对面板数据模型的新进展进行简单梳理。

1. 非平衡面板数据

绝大多数研究都是基于完整的面板数据或者说平衡面板数据进行的，即涉及的个体都是在整个样本范围内进行观测的。然而，在经济学研究中，非平衡面板数据可能是更普遍的样本数据。例如，在研究某行业上市公司的动态行为时，在观测的样本区间内，一些公司可能退市，而同时又有新的公司上市。相似地，当利用城镇居民家庭消费支出调查面板数据时，一些家庭迁出，另一些家庭迁入。另外，在收集国家跨期数据时，可能会发现一些国家的历史数据比其他国家的更久远。于是，引发了对非平衡或者不完整面板数据的

研究。关于非平衡面板数据模型的估计及其检验理论已比较成熟。例如，单误差分量模型、双误差分量模型和非平衡嵌套误差模型的分析方法。可是，计量分析软件中相应功能的制约，影响了非平衡面板数据模型的应用。

2. 空间面板数据

在微观个体水平的随机抽样样本中，很少关注截面个体之间的相关性。然而，当考虑国家、地区、州、县等相关截面数据时，这些总量个体可能表现出必须处理的截面相关性。现在有大量运用空间数据的文献处理这种相关性。这种空间相依模型在区域科学和城市经济学中比较普遍。具体来说，这些模型使用经济距离测度设定了面板数据的空间自相关性和空间结构(空间异质性)，这方面文献的详尽介绍可以参见 Anselin(1988，2001)的研究。近年来，在经济学的实证研究中，空间面板模型变得越来越有吸引力。在包含空间误差自相关和空间滞后被解释变量的情形下，Elhorst(2003)讨论了固定效应和随机效应面板数据模型的 ML(最大似然法)估计，他们也对随机系数模型做了相应的扩展。在包含空间误差自相关，或者空间滞后被解释变量的情形下，Elhorst(2005)研究了固定效应动态面板数据模型的估计。

3. 轮换面板数据

因为同一个家庭可能不愿一次又一次地被回访，为了保持调查中家庭数目相同，在第二期调查中退出的部分家庭，被相同数目的新的家庭所替代，这在获得调查面板数据时是必要的。Biorn 和 Jansen(1983)认为轮换面板允许研究者检验抽样时间(time-in-sample)偏倚[①]效应的存在性。对于轮换面板数据，每批新增于面板数据的个体组提供了检验抽样时间偏倚效应的方法。事实上，在面板数据调查中普遍存在着轮换组偏倚效应，而实践中调查条件并没有保持不变，因而很难把抽样时间偏倚效应同其他效应区分开。例如，Solon(1986)等研究发现第一次轮换所报告的失业率比基于全样本的失业率高出 10 个百分点。

4. 伪面板数据

1985 年，Deaton 指出，“由于统计调查的样本轮换和样本非随机流失问题，绝大多数国家并不存在较长时间跨度的真正面板数据，或者这样的真正面板数据是难以获得的，对于发展中国家的微观经济变量尤其如此。”并且，Deaton 发现“虽然某变量的统计抽样不能连续调查到各个体的观测数据，但是，如果按照某种属性(例如，年龄、职业和身份等)将各期调查对象分成不同的**群**(cohort)；对于各个观测期，选择各群内观测数据的均值(中位数或分位数)，即可构造以群为‘个体’单位的面板数据”。于是，对于截面时间序列的统计调查数据，基于某种属性分群，称以群为个体而构造的人工面板数据为**伪面板数据**(pseudo panel data)。

众所周知，面板数据的本质是在观测期内的每期都能观测到相同个体的相关数据。显然，伪面板并非如此。在观测期内，它允许每期观测的个体不同，并且重点关注的是个体群的统计特征，即通过群均值和群方差的统计特征揭示相应变量的总体分布特征。

---

① 抽样时间偏倚是指初次采访和随后的采访之间的回答有显著的改变。

例如，Inoue(2008)还研究了包含群效应的(动态)伪面板数据模型的 GMM 估计。白仲林(2012)基于天津市住户抽样调查数据，根据同龄群(户主具有相同年龄段的家庭的群体)和同生群(户主具有相同出生年度段的家庭的群体)建立动态伪面板数据模型研究了我国城镇居民跨期消费选择行为。

应用群体分析方法得到的伪面板数据还具有以下优点：①伪面板数据是由各群群内个体属性的总体统计量组成，与一般面板中的个体数据相比，前者消除了个体的测量误差，且避免了样本流失。②由于不需要在每期中追踪固定的个体，这样可得到更长时间跨度的数据。但是，也产生了新的问题，如无意义的个体效应、滞后数据的不可观测。③McKenzie(2004)的研究发现，当个体数 $N\to\infty$ 时动态伪面板数据的 FE(固定效应)估计还是一致的，即避免了动态面板数据 FE 估计的 Nickell 偏倚(Nickell，1981)。

例如，为了基于城市住户抽样调查数据研究城市居民消费选择行为，常见的分群标准是户主出生年的区间、户主年龄段和户主职业类别。

按户主出生年的区间分群，在各观测期，同群中的不同家庭都是户主在同一出生年区间的家庭，不同群的家庭是户主在不同出生年区间的家庭。于是，对于不同的群就可构造一个关于家庭消费的面板数据，称之为按出生年分群的家庭消费伪面板数据。

类似地，也可以构造按年龄段分群的家庭消费伪面板数据。

应用群体分析方法得到的伪面板数据还具有如下两个优点。

(1) 伪面板数据是由各群群内个体属性的总体统计量组成，降低了个体的测量误差。

(2) 由于不需要在每期追踪固定的个体，避免了样本非随机流失问题，可得到更长时间跨度的数据。

但是，伪面板数据的计量分析也产生了以下两个新问题。

(1) 无意义的个体效应。

(2) 滞后数据的不可观测性，使动态模型复杂化。

第 2 章　面板数据线性回归模型

2.1　面板数据线性回归模型

多元线性回归模型是最基本的计量经济学模型，被广泛应用于经济学的各个领域。鉴于面板数据的二维特征，面板数据线性回归模型能够较好地控制不可观测的非时变异质性与时变同质性对回归参数估计量无偏性和有效性的影响，使得面板数据线性回归模型不仅能更好地识别和度量时间序列模型或者截面数据模型所不能发现的影响因素，而且它能够构造和检验更复杂的行为模型。

事实上，利用面板数据不仅可以建立静态线性回归模型，而且可以建立动态线性回归模型。本章主要讨论面板数据静态线性回归模型，通常可划分为七大类模型，分别是混合回归模型(pooled regression models)、单因素效应模型、双因素效应模型、确定变系数回归模型、随机系数回归模型(random coefficient regression model)、组间回归模型、平均个体回归模型。

2.1.1　面板数据线性回归模型的一般形式

面板数据线性回归模型的一般形式如下：

$$y_{it} = \beta_{0i} + \sum_{k=1}^{K} \beta_{ki} x_{kit} + u_{it}, \quad u_{it} \sim \text{i.i.d.}(0, \sigma_i^2) \tag{2-1}$$

其中，$i = 1,2,\cdots,N$ 表示 N 个个体；$t = 1,2,\cdots,T$ 表示已知的 T 个时期。y_{it} 是个体 i 在 t 时期被解释变量的观测值；x_{kit} 是第 k 个非随机解释变量对于个体 i 在 t 时期的观测值；β_{ki} 是待估计的回归系数；u_{it} 是具有方差 σ_i^2 的随机误差项，其矩阵形式表示是

$$\boldsymbol{Y}_i = \beta_{0i}\boldsymbol{\iota}_T + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{U}_i, \quad i = 1,2,\cdots,N \tag{2-1'}$$

其中，$\boldsymbol{Y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}_{T\times 1}$，$\boldsymbol{X}_i = \begin{pmatrix} x_{1i1} & x_{2i1} & \cdots & x_{Ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{Ki2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{KiT} \end{pmatrix}_{T\times K}$，$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \\ \vdots \\ \beta_{Ki} \end{pmatrix}_{k\times 1}$，$\boldsymbol{U}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}_{T\times 1}$，

$\boldsymbol{\iota}_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1}$。

2.1.2 面板数据线性回归模型的分类

为了利用频率统计学方法建立模型(2-1),必须设定一些限制性假设,使估计模型时具有足够的自由度。一般来说,常用的面板数据回归模型有如下七种模型,下面分别进行介绍。

1. 混合回归模型

如果模型(2-1)中,设定了假设:

$$H_0^1:\beta_{k1}=\beta_{k2}=\cdots=\beta_{kN}=\beta_k,\quad k=0,1,\cdots,K$$

则称模型

$$y_{it}=\beta_0+\sum_{k=1}^{K}\beta_k x_{kit}+u_{it},\quad u_{it}\sim \text{i. i. d.}(0,\sigma_u^2) \tag{2-2}$$

为面板数据的混合回归模型,其矩阵形式如下:

$$\boldsymbol{Y}=\boldsymbol{\iota}_{NT}\beta_0+\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{U} \tag{2-2'}$$

其中,$\boldsymbol{Y}=\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\\ \vdots\\ \boldsymbol{Y}_N\end{pmatrix}_{NT\times 1}$,$\boldsymbol{X}=\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\\ \vdots\\ \boldsymbol{X}_N\end{pmatrix}_{NT\times K}$,$\boldsymbol{\beta}=\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_k\end{pmatrix}_{K\times 1}$,$\boldsymbol{U}=\begin{pmatrix}\boldsymbol{U}_1\\ \boldsymbol{U}_2\\ \vdots\\ \boldsymbol{U}_N\end{pmatrix}_{NT\times 1}$。

显然,对于面板数据的混合回归模型,可以直接把面板数据混合在一起视为截面数据。并且,在满足回归模型的基本假设条件下,使用普通最小二乘法(OLS)估计模型参数。

例如,如果劳动力市场、资本市场和商品市场是有效的,那么,在技术进步非时变和各地区(个体)技术效率相同的假设下,根据各地区的面板数据利用 Cobb-Dauglas 生产函数估计劳动和资本对产出的贡献时,可以将模型设定为混合回归模型。

实际上,混合回归模型假设了解释变量对被解释变量的影响与个体无关。模型中的回归方程部分表示了被解释变量中可观测的时变同质性效应、截距项反映不可观测的非时变同质性效应的平均水平,而且独立同分布的误差项包含了被解释变量中所有不可观测的时变异质性,它未包含不可观测的非时变异质性和时变同质性效应。因此,混合模型不仅不能体现面板数据模型的信息优势,而且对实际问题进行了严格的约束。如 Cobb-Dauglas 生产函数中施加了技术水平不变的条件。

尽管混合回归模型在早期被广泛应用,但是,Mairesse 和 Griliches(1990)指出在许多问题的研究中面板数据的混合回归模型并不适用。

2. 单因素效应模型

鉴于面板数据混合回归模型并未揭示被解释变量中的不可观测非时变异质性,如各地区的地缘经济特征;以及未考虑被解释变量中的不可观测时变同质性,如全国性经济政策对各地区经济的影响。更广泛使用的面板数据线性回归模型是单因素效应模型。

所谓单因素效应模型,就是在模型中考虑了不可观测非时变(个体)异质性效应,或者,考虑了不可观测时变(个体)同质性效应的模型。因此,单因素效应模型分为个体单因