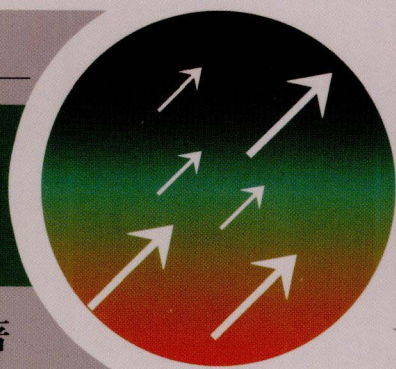


面板数据模型的 检验方法 ■■■■

Testing Methods for
Panel Data Model

陈海燕 著

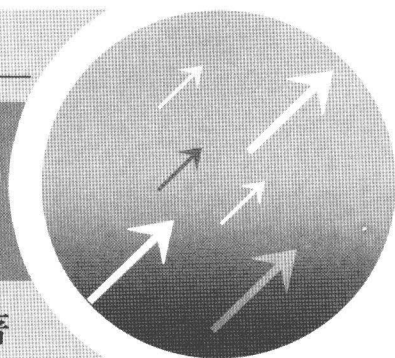


经济科学出版社
Economic Science Press

面板数据模型的 检验方法 ■■■■

**Testing Methods for
Panel Data Model**

陈海燕 著



经济科学出版社
Economic Science Press

图书在版编目 (CIP) 数据

面板数据模型的检验方法 / 陈海燕著. —北京: 经济科学出版社, 2012. 7

ISBN 978 - 7 - 5141 - 2038 - 7

I. ①面… II. ①陈… III. ①经济统计 - 统计数据 - 数据模型 - 检验方法 IV. ①F222 - 34

中国版本图书馆 CIP 数据核字 (2012) 第 132442 号

责任编辑: 周国强

责任校对: 苏小昭

责任印制: 邱 天

面板数据模型的检验方法

陈海燕 著

经济科学出版社出版、发行 新华书店经销

社址: 北京市海淀区阜成路甲 28 号 邮编: 100142

编辑部电话: 88191350 发行部电话: 88191537

网址: [www. esp. com. cn](http://www.esp.com.cn)

电子邮件: [esp@ esp. com. cn](mailto:esp@esp.com.cn)

北京密兴印刷有限公司印装

710 × 1000 16 开 13.75 印张 230000 字

2012 年 8 月第 1 版 2012 年 8 月第 1 次印刷

ISBN 978 - 7 - 5141 - 2038 - 7 定价: 38.00 元

(图书出现印装问题, 本社负责调换。电话: 88191502)

(版权所有 翻印必究)

目 录

第一章 引言	1
第一节 面板数据概况	1
第二节 面板数据的特点	4
第三节 面板数据模型的分类	6
第四节 面板数据模型的研究概况	11
第五节 本书主要内容	18
第二章 面板数据模型的基本估计方法	20
第一节 最小二乘估计	22
第二节 广义矩估计	30
第三节 最大似然估计	34
第四节 面板分位数回归	36
第五节 其他估计方法	41
第三章 面板数据模型的基本检验	46
第一节 面板数据模型的协方差分析	46
第二节 基于最大似然估计的检验	49

第三节	面板随机/固定效应检验	52
第四节	面板结构突变点的检验	56
第五节	面板 Granger 因果关系检验	57
第六节	其他检验方法	60
第四章	面板异方差和序列相关	62
第一节	面板异方差	62
第二节	面板异方差的确定性检验	68
第三节	面板 GARCH 效应检验	75
第四节	面板序列自相关检验	80
第五节	面板截面相关检验	86
第六节	联合检验	91
第五章	面板单位根检验	94
第一节	截面独立的面板单位根检验	94
第二节	截面相关的面板单位根检验	100
第三节	结构变化的面板单位根检验	106
第四节	面板单位根检验的其他问题	126
第六章	面板协整检验	131
第一节	面板数据伪回归	131
第二节	基于残差的单一协整关系检验	133
第三节	多个协整关系的检验	138
第四节	截面相关的面板协整检验	141
第五节	面板变结构协整检验	146
第六节	面板误差修正模型	151
第七章	面板协整模型的估计与推断	153
第一节	单一方程估计	153
第二节	系统方法估计	156

第八章 基于 LSTR 的面板变结构协整检验	159
附录 A 基本数学工具	178
附录 B 概率和统计	183
附录 C 极限理论	194
参考文献	199
后 记	215

第一章

引言

随着全球化程度的加深，国际经济贸易市场的不断发展和信息技术的不断进步，经济因素间的相互作用日益增强，不同市场上进行的区域交易越来越多地被人们所关注。人们能通过互联网了解到更多的信息，这些信息影响着他们的行为与决策，仅仅应用横截面数据或者时间序列数据进行经济理论的检验、经济变量变化进程的预测、经济结构的解析等已不能完全刻画经济数据的演变，不能满足人们的研究需求，而需要同时分析和比较横截面观察值和时间序列观察值结合起来的数据形式。

面板数据模型（Panel Data Models）是现代计量经济学理论方法的重要发展之一，它能更好地识别和度量单纯时间序列模型和单纯横截面数据模型所不能发现的影响因素，能够构造和检验更复杂的经济行为，具有更好的理论价值和应用价值。

第一节 面板数据概况

在计量经济学中，用于理论研究和实证研究的经济数据，其数据结构形

式主要包括横截面数据 (Cross-Sectional Data)、时间序列数据 (Time Series Data)、混合横截面数据 (Pooled Cross-Sectional Data) 和面板数据 (Panel Data)。在经典计量经济学模型中, 主要是以时间序列数据或横截面数据形式作为研究对象。

- 横截面数据, 就是在给定时点对不同的个体进行抽样所构成的数据集, 即由多个被观测对象在单一时期的观测值组成。有时候, 个体数据并非完全对应绝对的同一段时间。比如, 几个企业在同一年中的不同星期被调查。在横截面数据分析中, 时间的细微差别可以忽略。横截面数据广泛地被应用于经济学和其他社会科学研究领域中, 表现着不同的个体在同一时期的样本值。

- 时间序列数据, 是由一个变量在不同时期的观察值组成的数据集, 即由单一的被观测对象在多个时期的观测值组成。比如, 不同年份的国民生产总值, 不同月份的消费者价格指数, 企业销售量等。按着时间先后顺序排列的时间序列数据能表现过去事件对于现在或未来的影响。由于数据采集频率的不同, 时间序列数据可以以分钟、小时、天、周、月、季度等进行记录, 表现出不同的周期性和循环性。随着时间的延续, 还会有时间上的趋势性特征。

- 混合横截面数据, 是结合时间序列数据和横截面数据的数据集。比如, 对中国的家庭进行三次横截面数据的抽样调查, 时间分别是 2000 年、2005 年和 2010 年, 在 2000 年时假设抽到了家庭 A, 但是到 2005 年和 2010 年进行抽样时, 再次抽到家庭 A 的概率较小, 即同一个体在不同时期样本中都出现纯属偶然。这是混合横截面数据有别于面板数据的最大特征。

- 面板数据, 又名综列数据, 是由横截面数据中每个个体的时间序列所组成, 其每个横截面个体在不同的时期是不变的。比如前面提到的家庭抽样, 在 2000 年抽到了 A、B、C 三个家庭进行调研, 那么在 2005 年和 2010 年, A、B、C 家庭必然还会被跟踪调查。

因此, 混合横截面数据和面板数据都是结合时间序列数据和横截面数据的数据形式。所不同的是, 面板数据是固定的横截面对象在时间上的重复观测样本数据, 而混合横截面数据所观测的对象不固定, 每一段时间上的观测都是总体上进行随机样本抽样所得。因此在数据收集和模型假设上, 面板数据比混合横截面数据更实用更方便, 成为现代计量经济学研究的热点之一。

表 1-1 给出了不同收入层次家庭的平均每人全年消费性支出情况数据, 第七行 2005 年不同收入层次家庭的每人消费数据为一个横截面数据样本, 样本容量为 7, 代表的是在 2005 年不同收入层次家庭平均消费支出的差异。第五列为中等收入家庭从 2000~2010 年的每人消费性支出数据为一个时间序列数据样本, 样本容量为 11, 代表的是中等收入家庭在不同年份消费性支出的变化。整个表格表示的就是不同收入层次家庭从 2000~2010 年的每人消费性支出的面板数据, 样本容量为 $7 \times 11 = 77$ 。

表 1-1 不同收入层次家庭的平均每人全年消费性支出情况 单位: 元

年份	最低收入家庭	低收入家庭	中等偏下收入家庭	中等收入家庭	中等偏上收入家庭	高收入家庭	最高收入家庭
2000	2540.13	3274.93	3947.91	4794.56	5894.92	7102.33	9250.63
2001	2690.98	3452.27	4197.57	5131.55	6241.5	7495.09	9834.2
2002	2387.91	3259.59	4205.97	5452.94	6939.95	8919.94	13040.69
2003	2562.36	3549.28	4557.82	5848.02	7547.31	9627.58	14515.68
2004	2855.15	3942.23	5096.15	6498.36	8345.70	10749.35	16841.82
2005	3111.47	4295.35	5574.32	7308.06	9410.77	12102.51	19153.73
2006	3422.98	4765.55	6108.33	7905.41	10218.25	13169.82	21061.68
2007	4036.32	5634.15	7123.69	9097.35	11570.39	15297.73	23337.33
2008	4532.88	6195.32	7993.67	10344.7	13316.63	17888.18	26982.13
2009	4900.56	6743.09	8738.79	11309.73	14964.37	19263.88	29004.41
2010	5471.84	7360.17	9649.21	12609.43	16140.36	21000.42	31761.63

一般情况下, 面板数据的记数方式如表 1-2, 为七个收入层次在 2000 年、2005 年和 2010 年的数据, 一共有 21 个。首先, 不同收入层次即面板数据的不同截面, 都会有一个编号, 从 1 到 7, 哪个层次标记为 1 或 2 并不重要, 即不同截面之间的排序位置无关紧要。其次, 数据排序时, 将不同年份的个体数据列在一起, 接下来才是下一个个体的不同年份数据序列。

表 1-2 面板数据的一般展现形式

序号	年份	不同个体序号	收入层次	消费性支出
1	2000	1	低收入家庭	2540.13
2	2005	1	低收入家庭	3111.47
3	2010	1	低收入家庭	5471.84
4	2000	2	中等偏下收入家庭	3274.93
5	2005	2	中等偏下收入家庭	4295.35
6	2010	2	中等偏下收入家庭	7360.17
⋮	⋮		⋮	⋮
19	2000	7	最高收入家庭	9250.63
20	2005	7	最高收入家庭	19153.73
21	2010	7	最高收入家庭	31761.63

在经济全球化的背景下，各经济区域间的联系日趋紧密，对研究数据形式提出了更高的要求，对模型的解释范畴也有了更多的需要。面板数据是集截面数据和时间序列数据优点的数据形式，能多方位的反映数据信息，提供更好的拟合模型。相对只利用截面数据或只利用时间序列数据进行经济分析而言，面板数据模型具有开拓性意义。

比如分析我国的结构性失业问题，失业率既受到各地区产业结构的影响，也受到国家在各个时期的宏观政策的影响。只利用截面数据，即选择同一时间上不同省份的数据作为样本观测值，可以分析各省份不同的产业结构对结构性失业的影响，但是不能分析国家的宏观政策对各省份结构性失业的影响；只利用时间序列数据，即选择同一省份或者全国整体在不同时间上的数据作为样本观测值，可以分析国家的宏观政策对某地区的结构性失业的影响，但是不能分析不同的产业结构对结构性失业的影响。如果采用面板数据模型，即在不同的时间上选择不同省份的数据作为样本观测值，既可以分析不同的产业结构对结构性失业的影响，也可以分析国家的宏观政策对结构性失业的影响。

第二节 面板数据的特点

面板数据的优点主要表现在可以克服多重共线性的困扰，能够提供更多

的信息、更多的变化、更高的自由度和更高的估计效率，减少共线性。

(1) 面板数据是指 N 个不同对象在 T 个不同时期被观测的数据，是可以控制某些类型的遗失变量。多元回归是控制变量的一种有力的工具，但它所要求的变量必须有全面的数据，如果得不到一些变量的数据，则某些遗失变量就不能包括在回归中，参数估计的最小二乘估计量就会存在估计偏差。比如研究因变量随时间的变化情况，面板数据就有可能消除那些在对象之间有差异但随时间不变的遗失变量的影响。

例如，考虑我国粮食生产的地区差异，用我国 26 个省份的 1997 ~ 2003 年这 7 年间每一年的粮食产量、播种面积、劳动力投入量、化肥投入量以及有关变量处理此问题。这个面板数据集允许某些遗失变量被控制，比如各地区气候状况的不同和土壤的质量不同，这些难以观测的变量在地区之间不同但不随时间变化。面板数据也允许那些随时间变化但在各地区不变的变量被控制，比如度量政府对于农业生产的扶持态度的变量等。

(2) 面板数据集含有 $N \times T$ 个数据，能提供更多的数据信息和更大的自由度。既可以分析个体之间的差异情况，又可以描述个体的动态变化特征；既可以从不同角度反映已有数据的信息，又可以反映被遗失变量的信息。比如对上一例子，面板数据模型可以分析各省份粮食产量与其影响因素的动态发展关系，也可以分析各省份间粮食生产的异同关系，了解我国的粮食生产结构。

(3) 面板数据模型提供了具有挑战性的模型方法，把研究对象划分为更多的类别，能更准确地估计模型参数，更准确地捕捉社会的复杂行为。由于面板数据具有双下标的形式，一方面能更好地代表来自不同个体和不同时段的数据信息；同时，其模型检验与估计也因此变得更加复杂。面板数据有时间和个体上的两个维度，根据维度上的不同效应，可以分为个体固定效应、个体随机效应、时间固定效应、时间随机效应和混合效应等模型形式，为了得到模型参数的最优估计量，每一种形式都有不同的假设和估计方法；面板数据除了每个截面序列存在序列自相关性外，截面与截面之间还存在空间相关性。

比如，异方差的类型，对于横截面数据而言，异方差表现为不同个体的随机干扰项估计方差的不确定性，而对于面板数据，异方差不仅体现在个体方差的不确定上，时点上的因素也存在异方差。相比时间序列分析中的异方

差性和相关性检验方法，违背经典假设条件下的面板数据模型中的模型检验、估计和修正，显得更加复杂。

(4) 由于计算机计算速度和能力提高，现有时间序列分析理论的完善，使得面板数据的计算和统计推论更方便。

第三节 面板数据模型的分类

巴尔塔吉 (Baltagi)、肖政 (Hsiao) 和阿里拉诺 (Arellano) 都出版了关于面板数据理论全面系统的书籍，但在介绍面板数据模型时，出发点是不相同的。巴尔塔吉 (Baltagi, 2005) 从面板数据模型受个体截面和时间影响的差异出发，首先把模型分为单向 (One-Way) 和双向 (Two-Way) 两种情况进行说明，再分别在这两种情形下讨论了模型的随机效应与固定效应。萧政 (Hsiao, 2003) 是把模型分为变截距 (Variable Intercepts) 和变系数 (Variable Coefficient) 模型，分别在这两种情形下讨论了模型的随机效应与固定效应，并先从单向情况入手进行参数估计，最后考虑了双向情况下的参数估计。阿里拉诺 (Arellano, 2003) 从静态和动态模型出发，考虑了面板数据的异质性和动态性，强调了微观面板的计量经济分析。本书主要从面板数据模型的各种检验方法出发，探讨不同模型形式下与面板数据模型有关的检验问题。

面板数据模型的一般回归式：

$$y_{it} = \mu + X'_{it}\beta + v_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T \quad (1.1)$$

记 $y_t = (y_{1t}, \dots, y_{Nt})'$ ^① 为因变量， $X_t = (X_{1t}, \dots, X_{Nt})'$ 为自变量，其中 $X_{it} = (x_{it1}, \dots, x_{itK})'$ 表示有 K 个解释变量； $\beta = (\beta_1, \dots, \beta_K)'$ 表示待估参数； μ 表示模型在时空上的共同均值项； v_{it} 为误差项； i 为截面个体下标， t 为时间下标， N 表示面板数据中含有的个体数， T 表示时间序列的长度。

误差项 v_{it} 可分解为三个变量： $v_{it} = \alpha_i + \lambda_t + u_{it}$ ，一般假设 $u_{it} \sim IID(0, \sigma^2)$ 。

^① 注：若无特别说明，“'”在本书中表示对向量或矩阵求转置，“⁻¹”在本书中表示对矩阵求逆。

σ_v^2)^①。一是随截面个体变化而变化,但不随时间推移而变化的变量 α_i ,如性别、籍贯、家庭背景变量等;二是随时间变化而变化,但不随个体变化而变化的变量 λ_t ,如某产品价格、宏观经济变量等;三是随时间和截面个体变化而变化的变量 u_{it} ,如企业销售额、利润等。

为了得到模型 (1.1) 的最优估计量,对模型中变量给出假设如下:

$$Eu_{it} = E\alpha_i = E\lambda_t = 0, \quad Ev_{it}\alpha_i = E\alpha_i\lambda_t = E\lambda_tv_{it} = 0$$

$$E\alpha_i\alpha_j = \begin{cases} \sigma_\alpha^2, & i=j \\ 0, & i \neq j \end{cases}, \quad E\lambda_t\lambda_s = \begin{cases} \sigma_\lambda^2, & t=s \\ 0, & t \neq s \end{cases}, \quad Eu_{it}u_{js} = \begin{cases} \sigma_u^2, & i=j \text{ 和 } t=s \\ 0, & \text{其他} \end{cases}$$

且与解释变量 x_{it} 是相互不相关的,即 $Ev_{it}x_{it} = E\alpha_ix_{it} = E\lambda_t x_{it} = 0$ 。

式 (1.1) 的矩阵表示为:

$$\mathbf{Y} = \mu\mathbf{l}_{NT} + \mathbf{X}\boldsymbol{\beta} + \mathbf{V} \quad (1.2)$$

$$\text{其中, } \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_{N \times 1}, \quad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}_{N \times 1}, \quad \mathbf{V} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix}_{N \times 1}, \quad y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}_{T \times 1}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}_{K \times 1},$$

$$l_{NT} \text{ 为 } NT \text{ 维单位向量, } l_{NT} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{NT \times 1}, \quad v_i = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{pmatrix}_{T \times 1}, \quad X_i = \begin{pmatrix} x_{i11} & x_{i21} & \cdots & x_{iK1} \\ x_{i12} & x_{i22} & \cdots & x_{iK2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1T} & x_{i2T} & \cdots & x_{iKT} \end{pmatrix}_{T \times K}。$$

面板数据的双下标形式决定了其模型类型比传统意义上的时间序列模型或横截面模型更具特殊性。下面根据不同的情形对面板数据模型进行分类归纳。

1. 维数不同

当 $\lambda_t = 0$ 时,即 $v_{it} = \alpha_i + u_{it}$,模型中没有仅随时间改变的因素影响 λ_t ,而只有仅随个体改变的因素时,式 (1.1) 称为单向面板数据模型 (One-Way Panel Data Models),此时只考虑了截面个体的异质性,忽略了个体在时间上的不同;

① IID 表示独立同分布,英文全称为: independently and identically distribute。

当 $\alpha_i = 0$, 即 $v_{it} = \lambda_t + u_{it}$, 模型中没有仅随个体改变的因素影响 α_i , 而只有仅随时点改变的因素时, 式 (1.1) 也为单向面板数据模型;

当 $\lambda_t \neq 0$, $\alpha_i \neq 0$ 时, 式 (1.1) 为双向面板数据模型 (Two-Way Panel Data Models), 即时点和个体上都存在相应的变量。

除了有单向、双向模型外, 面板数据模型可以发展到多向。关于多向模型的详细阐述请参见另外的文献 (Antweiler, 2001; Baltagi, Song, Jung, 2001; Davis, 1999)。

本书若无特别说明, 全部指的是单向面板数据模型。

2. 影响因素不同

记 $\alpha = (\alpha_1, \dots, \alpha_N)'$, 当个体效应 α_i 为随机变量时, 式 (1.1) 为个体随机效应模型 (Entity Random-Effects Models), 此时 $Var(\alpha_i) = \sigma_\alpha^2 ee'$, $e = (1, \dots, 1)'$ 为 T 维列向量, 一般采用广义最小二乘法 (Generalized Least Squares, GLS) 进行估计; 当个体效应 α_i 为固定常量时, 式 (1.1) 为个体固定效应模型 (Entity Fixed-Effects Models), 有 $Var(\alpha_i) = 0$, 此时一般采用最小二乘虚拟变量法 (Least Squares Dummy Variable, LSDV) 进行估计。

另外, 还可以根据 $E(v_{it} | X_{it}) = 0$ 是否成立来区分固定效应和随机效应。当 $E(v_{it} | X_{it}) = 0$ 成立时, 说明模型中不可观测因素是随机变化的, 与自变量没有关系, 模型可以确定为随机效应模型; 而当 $E(v_{it} | X_{it}) = 0$ 不成立时, 说明模型中不可观测因素与可观测因素即自变量之间具有相关性, 对模型的影响具有可测性, 可考虑为固定效应模型。

同理, 根据 λ_t 的情况, 模型可以分为时间固定效应模型 (Time Fixed-Effects Models) 和时间随机效应模型 (Time Random-Effects Models)。实质上, 固定效应模型是把个体特征之间的差异用不同的截距项 $\alpha_1, \dots, \alpha_N$ 来反映, 而随机效应模型是把各组的不同的截距项看成是不同的随机扰动项 α_i 相加而成。即是说如果模型中的系数为确定性变量, 即模型中省略因素对个体差异的影响是固定不变的, 就是固定效应模型, 反之就是随机效应模型。

3. 参数不同

根据模型中参数的不同要求, 可以把面板数据模型分为变截距模型和变系数模型, 具体的检验方法详见第三章的协方差分析。对于变系数模型, 又

可以根据系数是否具有随机性分为固定系数模型和随机系数模型。

对于变系数模型：

$$y_{it} = \beta_i x_{it} + u_{it}, i = 1, \dots, N, t = 1, \dots, T$$

若参数 $\beta_i = \beta + \varepsilon_i, \varepsilon_i \sim IIDN(0, \delta^2), u_{it} \sim IIDN(0, \delta_i^2), E(\varepsilon_i) = 0, \beta$ 为均值项，则为随机系数模型。有

$$y_{it} = \beta x_{it} + u_{it} + \varepsilon_i x_{it}$$

且满足 $\text{cov}(u_{it} + \varepsilon_i x_{it}, u_{is} + \varepsilon_i x_{is}) = \begin{cases} \sigma_i^2 + \delta^2 x_{it}^2, & t = s \\ \delta^2 x_{it} x_{is}, & t \neq s \end{cases}, \text{cov}(u_{it} + \varepsilon_i x_{it}, u_{jt} + \varepsilon_j x_{jt}) = 0, i \neq j。$

4. 数据类型不同

根据所研究的对象、数据来源不同，面板数据模型可以分为宏观面板和微观面板两个方面。如果把一组不同国家、地区的时间序列数据混合在一起，就得到了一个宏观面板数据，可以用来研究跨国增长和经济收敛模型等；如果抽样数据是来自家庭、企业等，就是微观面板数据，具有较大的横截面维度和较小的时间段。

5. 矩不同

应用面板数据的矩来进行建模，可以分为均值模型、方差模型、高阶矩模型等，它们分别从不同角度考察了面板数据的性质。格里尔等（Grier et al, 2002）对面板数据自回归条件异方差模型进行了研究，考虑了固定效应条件下的均值方程和方差方程的变截距模型，给出了基于时间序列分析传统方法的假设检验和参数估计。北泽（Kitazawa, 2003）对面板数据的波动持续性进行了研究，提出了面板数据随机波动（stochastic volatility, SV）模型。陈海燕（2007）对随机效应的变截距 GARCH（1, 1）模型进行了研究，给出了相应的拉格朗日乘数（Langrange Multiple, LM）检验和最大似然估计（Maximum Likelihood Estimation, MLE）。

6. 结构不同

根据数据结构是否发生突变，可以分为结构变化模型和结构稳定模型。

结构变化模型又可以根据发生变化的参照变量进行划分，比如面板门限模型、面板结构变化模型、面板时变模型等。根据结构变化发生的时滞，可以分为突变式和渐变式面板数据模型。在非平稳面板数据建模上，除了传统的检验方法外，还有针对结构变化数据的面板变结构单位根检验、面板变结构协整检验等。

另外，根据模型中是否含有时间滞后效应，可以分为静态模型和动态模型。具有滞后作用的变量称为滞后变量，含有滞后变量的模型统称为滞后变量模型，而动态模型为含有滞后被解释变量的模型。由此，可以将面板数据模型根据不同情况进行简约分类，见图 1-1。

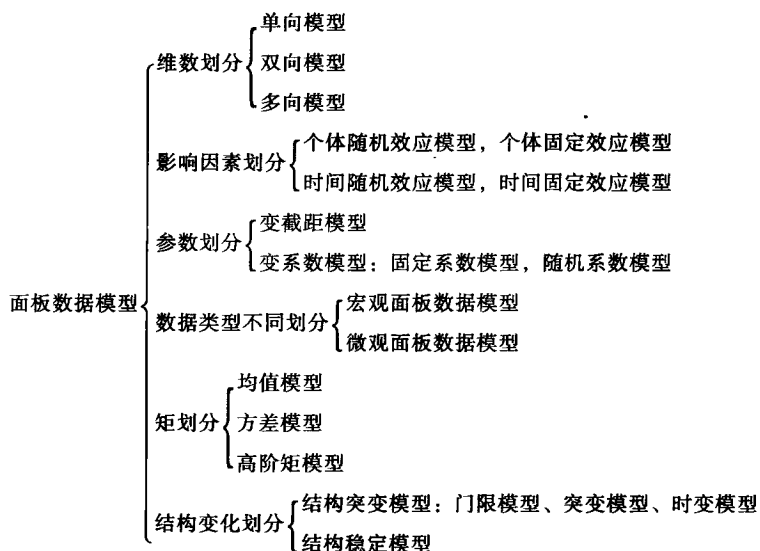


图 1-1 面板数据模型的分类

当然，面板数据模型并不仅仅局限于以上几种分类。随着现代化计算机技术的进步和经济发展的多元化，在对实际问题进行模拟和实验时，衍生出许多适应实际发展需求的模型形式。比如面板离散选择模型、面板受限因变量模型、非线性面板数据模型等。面板数据模型形式并不固定和唯一，是相互渗透和相互影响的，比如可以考虑单向随机效应影响下的变截距模型、双向门限自回归模型等。

第四节 面板数据模型的研究概况

自 20 世纪 60 年代以来, 计量经济学界开始关注面板数据。现在, 面板计量经济学已成为现代计量经济学研究的重要分支之一。

早在 1986 年, 萧政就写了关于面板数据模型方法研究的书籍, 到 2003 年有了第二版。伍德里奇 (2002)、李 (Lee, 2002)、阿里拉诺 (2003)、巴尔塔吉 (2005) 都对面板数据模型各方面给出了比较全面完整的说明。哈拉比 (Halaby, 2004) 对社会学研究中涉及的面板数据模型进行了深入而细致的探讨。近年来对于面板数据模型的研究已在国内外掀起一股热潮, 相关面板数据主题主要包括: 动态面板数据模型、非线性面板数据模型、非平稳面板数据数据、受限因变量面板数据模型、面板数据变结构模型等。

在回归模型研究中, 为了得到系数的最优估计量, 总是对模型中的自变量和误差项赋予诸多的假设限制。经典假设包括线性、同方差性、序列无关性、无多重共线性等, 但是这些完美的假设并不符合实际研究情况, 学者们通过不同的检验方法去发现数据的实际特性, 并通过不同的修正方法对模型进行调整, 以得到最优估计量。怀特异方差检验 (White Heteroscedasticity Test)、DW 序列相关检验等都是对违背经典假设的研究发现, 广义 OLS、差分法、逐步回归法等都是针对不同假设情形而发展起来的修正估计方法。在面板数据回归模型中, 模型假设条件的设置更加复杂, 处理方式也更加多变。

探索经济数据的动态结构, 研究它们的统计性质, 理解数据生成过程的特点和性质, 从而建立更有效的计量经济模型, 用于经济预测和各种经济理论的可靠性分析, 是现代计量经济学的一个重要研究课题。在进行这些研究的时候, 通常都假设变量为平稳过程, 在此基础上对模型进行参数估计和假设检验。但是, 现实中的许多经济数据的生成过程并不是稳定的过程, 非平稳数据使得传统的平稳序列研究方法失效。单位根过程是最常见的非平稳过程之一, 其一阶差分过程为平稳过程。在进行建模之前, 对时间序列进行单位根检验, 可以避免伪回归的发生, 已成为现代经济计量分析的首要步骤。时间序列分析中常用单位根检验有 DF 检验、ADF 检验、PP 检验、KPSS 检