

学校编码: 10384

分类号_____ 密级_____

学号: 19020141152626

UDC_____

厦 门 大 学

硕 士 学 位 论 文

面板数据模型分析及其应用

Analysis and Application of Panel Data Model

苟博一

指导教师姓名: 谭 忠

专业名称: 应 用 数 学

论文提交日期: 2017 年 4 月

论文答辩日期: 2017 年 5 月

学位授予日期: 2017 年 月

答辩委员会主席: _____

评 阅 人: _____

2017 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

中文摘要

随着人们生活和经济活动的日益复杂，以及信息技术的不断发展，相比较传统横截面数据和时间序列，面板数据更能反映人们复杂的行为和数据背后的深层含义，对面板数据模型的分析研究成为科研工作者关注的新热点。面板数据是包含多个时间序列截面的数据，因此它也比横截面数据和时间序列包含了更多的现实意义和内涵。

本文对面板数据静态模型和动态模型进行了研究，着重讨论了各个模型的参数估计，在不同的模型下利用合适的估计方法得到对参数的一致无偏估计。在静态模型中加入滞后项，得到了面板数据的动态模型，对两类模型的对比分析发现，由于滞后项和变量之间存在相关性，静态模型中的估计方法并不完全适用于动态模型。然后本文对非平稳面板数据的单位根检验进行讨论，分析了不同检验方法的特点与局限，发现IPS检验比LL检验更具普适性，也研究了在截面相关时的单位根检验方法。接着引入了面板数据协整理论和协整检验，为研究非平稳面板数据提供了方法。最后，本文选取了我国9个省市的农业生产数据，分析了影响粮食生产的各个因素在不同地区的显著程度，得到了符合我国的农业生产现状的结果。又对我国9个省市的居民消费水平和固定资产投资两个非平稳的面板数据进行了单位根检验，得到他们都是一阶单整的面板数据，然后对他们进行协整检验，分析结果显示两者存在协整，即它们之间存在长期的平衡关系。

关键词：面板数据，面板数据单位根检验，面板数据协整分析。

Abstract

With the increasing complexity of people's lives and economic activities, and the continuous development of information technology. Compared to traditional cross-sectional data and time series, panel data better reflect the deep meaning of people behind complex behavior and data. The analysis and research on the panel data model has become a new hot spot for scientific research workers. Panel data is data that contains multiple time series sections, so it also contains more real meaning than cross-section data and time series.

In this paper, the static model and dynamic model of panel data are studied, and the parameters estimation of each model is discussed emphatically. The unbiased estimation of the parameters is obtained by using the appropriate estimation method under different models. The hysteresis is added to the static model, and the dynamic model of the panel data is obtained. The comparison between the two models shows that the estimation method in the static model is not completely applicable to the dynamic model because of the correlation between the lag items and the variables. Then, the unit root test of non - stationary panel data is discussed, and the characteristics and limitations of different test methods are analyzed. We found that the IPS test was more general than the LL test, and the unit root test method at the time of cross-sectional correlation was also studied. And then introduced the panel data cointegration theory and cointegration test, to study the non-stationary panel data provides a method. In the end, this paper chooses the data of agricultural production in nine provinces and cities, analyzes the significant factors influencing the factors of grain production in different regions, and obtains the results of agricultural production in line with our country. The unit consumption of two non-stationary panel data of the residents' consumption level and fixed assets

investment in nine provinces and cities in our country was tested. We find them are first-order single-set panel data, and then they co-integration test, the results show that there is cointegration of the two, that is, there is a long-term balance between them.

Key words: Panel data, Unit root test of panel data, Cointegration test of panel data.

厦门大学博硕士论文摘要库

目 录

中文摘要	I
英文摘要	II
中文目录	IV
英文目录	VI
第一章 绪论	1
1.1 面板数据的特点	1
1.2 面板数据的应用	2
1.3 研究现状	3
1.4 论文结构安排和主要创新	3
第二章 面板数据静态模型	5
2.1 协方差分析	5
2.2 固定效应模型	9
2.3 随机效应模型	11
2.4 变系数模型	13
第三章 面板数据动态模型	17
3.1 协方差估计量	18
3.2 随机效应模型	19
3.3 固定效应模型	27
3.4 固定效应向量自回归模型	29

第四章 单位根检验和协整检验	32
4.1 时间序列平稳性和单位根检验	32
4.2 非平稳大维面板数据线性回归的极限理论	33
4.3 面板数据的单位根检验	35
4.4 面板数据的协整检验	45
第五章 实证研究	49
5.1 农业面板数据模型	49
5.2 面板数据单位根检验和协整检验	52
第六章 总结与展望	55
参考文献	56
致谢	60

Contents

Chinese Abstract	I
English Abstract	II
Chinese Contents	IV
English Contents	VI
1 Introduction	1
1.1 Characteristics of the panel data	1
1.2 Application of panel data	2
1.3 Research status	3
1.4 Paper structure and innovation	3
2 Static model of panel data	5
2.1 Analysis of covariance	5
2.2 Fixed effects model	9
2.3 Random effects model	11
2.4 Variable coefficient model	13
3 Dynamic model of panel data	17
3.1 Covariance estimator	18
3.2 Random effects model	19
3.3 Partial quadratic eigenvalue assignment	27
3.4 Fixed effects of vector autoregressive model	29

4	Unit root test and cointegration test	32
4.1	Time series stationarity and the unit root test	32
4.2	Theory of limit linear regression in panel data	33
4.3	Unit root test of panel data	35
4.4	Cointegration test of panel data	45
5	Empirical research	49
5.1	Panel data model of agriculture data	49
5.2	Unit root test and cointegration test of panel data	52
6	Summary and prospect	55
	References	56
	Acknowledgements	60

厦门大学博硕士学位论文摘要库

第一章 绪论

1.1 面板数据的特点

面板数据，又称为“纵列数据”，是含有多个个体的时序观测数据构成的数据集，是一类样本容量较大的数据类型。它同时包含了时间序列数据和横截面数据的信息，面板数据模型也兼具了横截面数据模型和时间序列模型的优点。

与传统的横截面数据或时间序列相比，面板数据有如下的优点^[1][参见Hisao(2003)]:

1、控制个体的异质性

传统的时间序列分析和横截面数据分析都没有考虑个体的异质性，这样有可能导致对模型的估计不是无偏估计。

2、包含更多的数据点，增加了数据的自由度，降低了解释变量之间的线性相关，使估计模型更为高效。

面板数据包含了更多的个体和时间上的数据，使得面板数据模型可以分析每个个体与其影响因素之间的关系，也可以分析不同个体之间的异同关系。面板数据既反映了个体之间的差异情况，又包含了个体内部的动态特征。

3、面板数据可以处理一些单纯横截面数据和单纯时间序列无法确定的关系。

由于面板数据同时具有横截面数据和时间序列，在研究实际问题是往往具有更广泛的适用性。例如一个经典案例^[2]：假设我们从已婚女士的横截面数据样本中发现，她们平均每年有50%的劳动参与率。对这个数据一种可能的解释是：样本中的女性来自于不同的总体，其中50%一直在工作。另外50%一直未就业。而另一种可能的解释是：样本中的女士来自于同一个总体，在给定时间段的就业几率为50%。第一种

情况说明女士在频繁的换工作，而第二种情况说明女士们的工作状态没有变化。只有面板数据才能对这两种情形进行区分。

1.2 面板数据的应用

随着人们生活生产中产生的数据变得更加的多种多样，纯粹的横截面数据和时间序列数据已经无法满足人们的分析研究需求。面板数据不仅拥有横截面数据的优点，而且也具有时间序列数据的优点，能更加真实的反映数据包含的信息。因此，面板数据的相关研究在近几十年成为了热点议题，在社会科学的研究中有着重要的地位。

如赵春华等^[3]利用面板数据，对影响我国居民消费水平的三大因素：收入、物价和利率进行研究，分别建立静态和动态面板数据模型，得出：收入、前期消费、价格指数对于城镇居民的本期消费有着正面影响，并且居民以前的消费影响着他们现在的消费。消费对前期消费的弹性系数与消费对收入的弹性系数大小相当，利率对当前消费虽然存在负面影响，但影响程度不大，收入对消费的影响有限。

又如王丽^[4]利用省级面板数据分析，发现可支配收入、家庭平均人口、受教育水平以及汽油成本是影响我国私人轿车消费者购车需求的主要因素。进一步通过分析得到我国消费者购车意向与我国日益增长的私有汽车保有量和日益严重的空气污染之间密切相关，并利用相关数据从以上角度来讨论汽车消费和空气污染之间的关系。

又如梁怀学等^[5]在分析我国行业能源消费特征中，对面板数据进行分析，建立了能源消费的面板数据模型。具体分析了包含消费总量、汽油、柴油、天然气、电力消费量等5种变量的变系数面板数据模型，以及煤炭、原油、煤油、燃油消费量等4种变量之间的变截距面板数据模型，利用面板数据模型详细分析了工业、农业、建筑业、交通业和批发业等6个行业的能源消费倾向和能源消费平均水平的特点和联

系。充分说明了在分析能源消费特征时面板数据模型的重要性。

1.3 研究现状

最初，在假设解释变量与被解释变量之间无关的前提下，建立了简单的面板数据回归模型。随后，1961年Mundlak将个体效应通过虚拟变量的形式引入到模型中，建立了固定效应面板模型。1963年Kuh将个体效应认为是随机变量分离出来，建立了一维误差分析模型。

1966年，Balestra、Nerlove等认为个体影响应该假设为随机变量，建立新的模型。1978年，Hausman提出了Hausman检验，2002年Wooldridge等提出密度比率准则、AIC准则、Schwarz准则等方法来解决模型选择的问题。

近年来，对面板数据单位根检验相关研究发展迅速，成为面板数据研究的新热点和新方向。面板数据的单位根理论和协整理论也渐渐完善，Maddala和Wu^[6]、Change^[7]等一些列学者提出了各种检验方法，并且在多个领域得到应用。如Chio^[8]利用单位根检验证明购买力平价理论。

对面板数据的分析和研究在国外已有几十年的历史，如Hsiao^[1]等对面板数据分析的著作都较为全面的介绍和分析了面板数据的模型建立以及参数估计等知识理论，有着比较完善的理论体系研究和广泛的应用案例。相对来说，国内对面板数据的研究起步较晚，大部分还是以应用现有的模型和方法进行实证研究为主，对面板数据的理论研究涉足较少，但也有一些较好的研究结果。如白仲林基于变截距模型构造了一种联合LM检验；齐晓丽等人提出了建模数据的分位数回归思想。

1.4 论文结构安排和主要创新

本文主要介绍和分析了面板数据模型以及其单位根检验和协整理论，并将其应用到我国实际数据中，进行相关分析研究，论文章节安排为：第一章绪论，介绍了面板数据的特点和应用，以及国内外的研究现状，让我们初步了解了研究面板数据

的意义；第二章静态面板数据模型，主要研究了静态面板模型的建立以及参数估计，在不同的情况下建立合理的模型进行分析，得到对参数的非偏一致估计；第三章动态面板数据模型，讨论了动态面板数据模型的建立，以及动态面板数据与静态面板数据模型的异同点，使用合适的估计方法得到参数估计；第四章面板数据的单位根检验和协整检验，主要研究了现有的一些检验方法，并对各种方法进行比较，分析了它们的优点和局限性；第五章实证研究，结合我国实际数据进行面板数据模型分析，得到了符合我国实际情况的相关结论，又对非平稳的面板数据进行单位根检验，进而分析了它们的协整关系；第六章总结展望，总结了本文的主要工作，对未来研究进行展望，期待对面板数据的相关研究可以取得新的突破。

本文创新主要在：将面板数据模型应用到中国农业数据分析中来，结合国内实际情况，对所分析省份的粮食生产情况做了研究讨论；对比分析了面板数据单位根检验和协整检验的不同方法，对它们各自的特点和局限做了研究讨论。

第二章 面板数据静态模型

本章主要研究静态的面板数据模型，着重考察固定效应模型和随机效应模型，以及各种常用的估计方法，静态模型的最大特点是在解释变量中没有出现因变量的任何滞后项。

2.1 协方差分析

与单纯的横截面数据和时间序列数据相比，我们可以用面板数据建立更为复杂有效的模型。我们发现某些效应产生的真实原因是存在与解释变量相关的遗漏变量，利用面板数据，我们可以很好的处理这个问题。

设我们对 N 个个体进行观测，得到了在 T 个不同时期的 K 个属性的样本观测值，用 $y_{it}, x_{kit}, i = 1, \dots, N, t = 1, \dots, T, k = 1, \dots, K$ 表示。我们建立如下的线性模型进行因素分析：

$$y_{it} = \alpha_{it}^* + \beta'_{it}x_{it} + u_{it}, \quad i = 1, \dots, N, t = 1, \dots, T \quad (2.1.1)$$

其中 α_{it}^* （截距）和 $\beta'_{it} = (\beta_{1it}, \beta_{2it}, \dots, \beta_{Kit})$ （斜率）分别是随着 i 与 t 变化的 1×1 和 $1 \times K$ 的常数向量， $x'_{it} = (x_{1it}, x_{2it}, \dots, x_{Kit})$ 是 $1 \times K$ 的外生变量， u_{it} 是均值为0，方差为 σ_u^2 的误差项。

模型(2.1.1)仅用来进行理论描述，它既无法估计也不能做出预测，因为该模型的可用自由度 NT 比模型参数的个数 $NT(K + 1) + u_{it}$ 分布的参数小。故在进行研究时，给模型(2.1.1)添加限制。假定参数不随时间的变化而变化，只根据个体的不同而变化，得到一般面板数据模型：

H_1 ：回归斜率不同，截距也不同，即：

$$y_{it} = \alpha_i^* + \beta'_i x_{it} + u_{it}, \quad i = 1, \dots, N, t = 1, \dots, T \quad (2.1.2)$$

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库