

离散选择模型 理论与应用研究

ISANXUANZEMOXING
LILUNYUYINGYONGYANJIU

段鹏 著



中国出版集团



世界图书出版公司

离散选择模型理论与应用研究

段 鹏 著



中国出版集团



世界图书出版公司

广州·上海·西安·北京

图书在版编目(CIP)数据

离散选择模型理论与应用研究/段鹏著.

—广州:世界图书出版广东有限公司,2014.11

ISBN 978-7-5100-8844-5

I. ①离… II. ①段… III. ①计量经济学-经济模型-研究 IV. ①F224.0

中国版本图书馆 CIP 数据核字(2014)第 260104 号

离散选择模型理论与应用研究

责任编辑 李汉保

封面设计 杨 冰

版式设计 白 杨

出版发行 世界图书出版广东有限公司

地 址 广州市新港西路大江冲 25 号

电 话 020-84459702

印 刷 虎彩印艺股份有限公司

规 格 710mm×1000mm 1/16

印 张 11.75

字 数 236 千字

版 次 2014 年 11 月第 1 版 2015 年 3 月第 2 次印刷

I S B N 978-7-5100-8844-5/F·0162

定 价 49.00 元

版权所有 翻印必究

内 容 提 要

作为微观、宏观经济问题研究中常用的建模方法,本书在梳理离散选择模型相关理论方法体系的基础上,对有关理论问题进行了探索性研究。同时,也运用离散选择模型对现实问题进行了分析。

本书内容分为七章。第1章为绪论,在交代本书写作依据和研究意义之后,对离散选择模型的相关文献进行回顾,并在此基础上提出本书研究内容和创新点;第2章根据样本数据的类型对离散选择模型进行分类,同时也对 Bayes 计量分析的理论方法进行回顾;第3章在梳理截面数据离散选择模型理论方法的基础上,分析不同 Bayes 抽样算法及参数先验分布的设定对抽样估计量性质的影响,并对频率学派与 Bayes 学派思路下参数估计的有效性和假设检验的准确性进行比较;第4章整理了非平稳时间序列数据离散选择模型的既有理论,在提出待研究问题的基础上,对“带漂移项随机游走过程”解释变量条件下二元选择模型 MLE 的收敛性及收敛速度进行了探索性研究;第5章归纳了常用固定效应、随机效应面板数据离散选择模型的估计方法;第6章运用 Bayes 计量和离散选择模型的理论方法对现实问题进行研究;第7章总结全文研究内容,同时对以后研究方向进行展望。

本书以方法论研究为主,在理论研究上做了以下工作:①以二元选择 Probit 模型为例,研究了 Bayes 抽样算法和参数先验分布的设定对不同样本容量条件下参数估计量统计性质的影响。结论显示:在小样本容量条件下,“偏倚”的先验分布设定使 Gibbs 抽样估计量“有偏”,而 M-H 抽样估计量和拒绝抽样估计量则具有无偏性特点。在大样本容量条件下,“偏倚”先验分布的设定不影响各种抽样估计量的无偏性;②以 MNL 模型和有序选择 Probit 模型为研究对象,以 ML 估计方法和 M-H 及 Gibbs 抽样算法为例,对不同样本容量、不同“选项”数量条件下模型频率学派和 Bayes 学派参数估计的有效性和假设检验的准确性进行比较。结论显示:就 MNL 模型和有序选择 Probit 模型而言,在给定参数无信息先验条件下,综合参数估计的有效性和假设检验的准确性两方面考虑,在小样本容量条件下, Bayes 学派方法优于频率学派方法;③针对宏观经济变量表现出明显时间趋势的特点,运用“响应面函数”分析方法,对“带漂移项随机游走过程”解释变量条件下二元选

择模型 MLE 的收敛性及其收敛速度进行研究。结论显示:参数 MLE 为其真值的无偏估计量,且参数 MLE 以 $T^{4/3}$ 的速度收敛于其渐近分布。

在实证研究部分,本书进行以下分析:①在充分考虑各国甲流感死亡率可能存在个体混合效应、独立效应、相关效应及空间相关效应的基础上,运用 Bayes 计量分析方法对各国甲流感死亡率建模;②为探究村民的行为选择对农村贫困的影响,在合理界定我国农村贫困程度的基础上,利用微观统计数据,通过建立有序选择 Probit 模型来分析村民的行为选择在“农村脱贫”中的作用;③在我国银行间同业拆借市场交易量剧增、利率波动变大的背景下,为使商业银行有效地规避利率风险,运用时间序列数据二元选择模型对同业拆借市场实际利率调整的决定因素进行研究。

Abstract

As a usually used method in microeconomics and macroeconomics, in the paper we firstly listed systematically the theoretical methods of different types of discrete choice models, and then did some research works on the theoretical problems, at last, we gave some applications of discrete choice model.

The contents of the paper can be divided into seven chapters. Chapter1, which is the Foreward, we introduced the writing background of the paper, the reviews of the discrete choice model, the research problems and the innovations of the paper; Chapter 2 makes a reviews of the discrete model and Bayes Econometrics. In Chapter3, after listing systematically the estimation methods of the discrete choice models using the cross-section data, we analysed the impacts of different algorithms and prior specifications to the properties of the estimator. At the same time, we compared the effectiveness of the estimators and the accurateness of hypothesis tests of different methods belong to Frequency School and Bayes School; In Chapter4, we reviewed the theoretical methods of discrete choice models using time series data, proposed some problems, and then analysed the convergence and its rate of the MLE when the explanatory variable is a random walk process with a drift in the binary choice model; In Chapter5, we generalized the commonly used methods for the fixed and random effects discrete choice models of panel data; In Chapter6, using the methods of Bayes econometrics and the discrete choice models, we analyzed some problems in the reality; In Chapter7, we summarized the paper and brought forward some problems that can be considered in the future.

We focused on the theoretical analysis in the paper. we mainly did three works in the paper: Firstly, given the binary probit model, we analysed the effects of different algorithms and prior specification on the properties on the estimator. The results revealed that: When the sample size is small, "biased" prior specification can lead the Gibbs sampler estimator be biased, but the phenomenons can't be happened for the M - H and Rejection sampler estimators. When the sample size is large, the "bias" of the prior specification has no affect on the unbias of the estimators; Secondly, given the MNL

and ordered Probit model, when the prior distributions of the parameters are flat, we found that from the estimation and hypothesis test point of view in small sample size, the methods of Bayes School are more priorior than that of Frequency School; Thirdly, in order to discribe the obvious trends in macroeconomic time series, we put the random walk procee with a drift into the binary model, using the “response surface function”, we analysed the the convergence and its rate of the MLE of the model, the Monto Carlo similations revealed that the MLEs are consistly and unbaised, the rate of convergence is $T^{4/3}$.

Empirical analysis included the following four aspects: Firstly, In the context of the population of Pandemic Influenza H1N1 in the world, we think of the possible mixed effects, independent effects, correlated effects and spatial correlated effects of the mortality rates of Pandemic Influenza H1N1 in different countries. Using the model choice methods in Bayesian econometrics, we decide the most appropriate model to discribe the mortality rates of different countris. Base on the model, we estimate the mortality rates of different countris. The result reveals that: the individual independent and spatial correlated effects model can best fit the sample data, and the estimated average mortality rate of Pandemic Influenza H1N1 in the world is 0.577%; Secondly, in order to ananalysis the impact of households' behavior selections on rural poverty in China, after defining the degree of rural poverty reasonably, using the suevey data, we employ an ordered probit model to find whether the households' behavior selections can influence the degree of rural poverty. the results reveal that: the numbers of enterprise, the efficiency of rural management, the employment structure, the degree of education of the households, the public investement and the state of popularization of agricultural technology all determinate the degree of rural poverty in China; ThirdlyIn the context of increasing volume in the inter-bank loan market, the offered real interest rate fluctuated violently, so the commercial banks facing the interest rate risk. In order to reduce the risks they faced, we analysed the determinants of the adjustments of the offered real interest rates in the inter-bank loan market. The results revealed that: all the changes of the government's expenditure, the investment of the fixed assets and the money supply can influence the adjustments of the real interest rate significantly. however, both the changes of people's consumer expenditure and the import and export commercial



trading can't. The findings reflect the effectiveness of our macroeconomic policy in the currency market.

目 录

第 1 章 绪论	1
1.1 写作依据与研究意义	3
1.2 文献综述	4
1.3 研究内容	9
1.4 论文创新	10
第 2 章 离散选择模型与 Bayes 计量理论回顾	13
2.1 离散选择模型分类	15
2.2 Bayes 计量相关理论	16
第 3 章 截面数据离散选择模型	31
3.1 二元选择模型	33
3.2 多项选择模型	40
3.3 有序选择模型	72
第 4 章 非平稳时间序列数据离散选择模型	81
4.1 非平稳时间序列数据二元选择模型	83
4.2 非平稳时间序列数据有序选择模型	87
4.3 存在的问题及探索性研究	90
第 5 章 面板数据离散选择模型	101
5.1 固定效应模型	103
5.2 随机效应模型	109
第 6 章 离散选择模型的应用研究	113
6.1 甲型 H1N1 流感的死亡率高吗 ——Bayes 计量分析框架下的建模与估算	115
6.2 我国农村贫困地区贫困的决定因素 ——基于村民行为选择视觉的实证分析	127
6.3 我国银行间拆借市场利率调整的决定因素 ——基于二元选择模型的实证分析	140

第7章 研究总结与展望	149
7.1 研究总结	151
7.2 研究展望	152
附录	155
参考文献	166
后记	177

第 1 章

绪 论

1.1 写作依据与研究意义

1.1.1 写作依据

计量经济学按研究内容可以划分为微观计量经济学、宏观计量经济学及金融计量经济学。在国外,各类计量经济理论方法蓬勃发展。作为微观计量经济学的重要内容,离散选择模型理论方法发展迅速。由于具有微观统计数据的支持,其应用研究也涉及到社会科学的方方面面。与此同时,国内计量经济学的研究现状可以概括为:宏观、金融计量经济学研究的快速发展与微观计量经济学研究的徘徊不前。统计数据的缺乏是造成国内微观计量经济理论发展缓慢的主要原因。没有微观统计数据,就不可能运用离散选择模型进行实际问题的分析。再进一步,就会使研究者失去对离散选择模型理论方法研究的兴趣。但随着社会的进步、我国统计制度的日益完善,微观统计数据的获取不再是离散选择模型实证应用的“瓶颈”。因此,该背景下对国外离散选择模型的理论方法进行追踪、对相关理论问题研究,同时运用离散选择模型对我国现实经济问题进行分析,这对于推动我国离散选择模型理论方法与应用研究的发展具有现实意义。

在确定研究主题后,本书内容的组织基于这样一个事实:同一模型,不同的数据类型对应不同的计量方法理论。在计量经济学中,给定相同的模型设定形式,当样本数据的类型不同时(截面数据、时间序列数据、面板数据),模型对应不同的理论方法。那么对于离散选择模型,本书分别就样本数据是截面数据、时间序列数据和面板数据时,对其相关的理论方法进行梳理,并就有关理论问题进行研究。

在论文技术分析方法的选择上,本书同时运用频率学派方法和 Bayes 学派方法对模型进行分析。作为计量经济理论分析中的两种不同思路,频率学派方法和 Bayes 学派方法各有优劣:在参数估计过程中,频率学派方法基于样本信息对参数进行估计,比较客观。而 Bayes 学派方法则在样本数据的基础上引入参数的先验信息(反映研究者的主观判断),因而其估计结果带有一定的主观性(特别是在样本容量较小时);就假设检验而言,由于频率学派估计量具有“渐近有效性”的特点,故小样本容量条件下基于估计量的“渐近结论”进行假设检验其相关结论值得怀疑。而 Bayes 学派方法恰好能弥补频率学派方法在假设检验中的不足,Bayes 学派估计方法的一大优势是能够实现小样本容量条件下参数后验分布的准确模拟,而基于参数后验分布的高概率密度(hpd)区间能实现相关假设检验的准确推断。为使两类分析方法“取长补短”,本书运用 Bayes 学派方法和频率学派方法同时对离散选择模型进行分析。并在截面数据离散选择模型部分,对于特定模型,本书从参数估计的有效性和假设检验的准确性角度对两类分析方法进行了比较。

1.1.2 研究意义

本书对离散选择模型的理论方法和实证分析进行了探索性研究,研究内容无论在理论上还是在实际工作中都具有较强的现实意义。

在理论研究部分,本书做了三方面工作,分别具有以下理论意义:首先,以MNL模型、有序选择Probit模型为例,在不同样本容量条件下,通过比较不同估计方法条件下参数估计量的有效性和假设检验的准确性,为特定模型在实证分析中选择合适分析方法提供了理论依据;其次,以二元选择Probit模型为例,运用蒙特卡洛模拟方法分析了参数先验分布的设定、不同Bayes抽样算法的选择对参数抽样估计量统计性质的影响。这对于Bayes计量分析中设定合理的参数先验、选择合适的抽样算法具有重要的理论指导意义;最后,在运用时间序列数据离散选择模型分析宏观经济问题中,针对宏观经济变量表现出明显时间趋势的特点,将带漂移项随机游走过程解释变量引入二元选择模型,并创新性的运用“响应面函数”的分析方法对参数MLE的渐近性质(收敛性、收敛速度)进行研究。这对于完善时间序列数据离散选择模型的理论方法体系、丰富估计量渐近性质的研究方法意义重大。

在应用研究部分,围绕三个问题进行了分析:首先,运用Bayes计量分析中的模型选择标准、空间相关性的建模方法对全球甲型H1N1流感死亡率建模,这对于消除甲流感受感染群体的恐慌情绪、加强甲流感病毒的基础性研究有重大现实意义;其次,从村民自身行为选择的视角,运用有序选择Probit模型对我国农村贫困地区贫困的决定因素建模;作为农村扶贫活动中的另一个主体、受益者,村民的行为选择能否影响农村的贫困状况?村民可以通过哪些行为方式改善农村的贫困状况?在传统扶贫模式成效递减的背景下,以上问题的研究对于我国实现“农村脱贫”战略意义重大。最后,运用时间序列数据二元选择模型对我国同业拆借市场实际利率调整的决定因素进行研究。在我国银行间同业拆借市场交易量剧增、利率波动变大的背景下,该研究内容对于商业银行有效规避利率风险具有现实意义。

1.2 文献综述

1.2.1 截面数据离散选择模型综述

最初,截面数据条件下个体的行为选择是离散选择模型的建模对象。根据“选择项”的数量,离散选择模型分为二元选择模型和多元选择模型。就二元选择模型而言,Amemiya(1981)和Maddala(1983)根据频率学派的方法对Logit模型和Probit模型的估计和选择等问题进行了研究。Albert和Chib(1993)则从Bayes学派的角度提出运用“潜变量数据扩展分析方法”对二元选择模型进行Gibbs抽样估

计。此外,Zellner 和 Rossi(1984)研究了 Logit 模型的 Bayes 推断问题,Poirier (1994)也对 Logit 模型的先验信息选择等问题进行了研究;就多元选择模型而言,根据“选项间”的关系有不同分类:根据选项间是否存在特定顺序,可以分为“有序选择模型”和“无序选择模型”。就“有序选择模型”而言,既可以运用传统估计方法(ML)对其进行估计,也可以运用 Albert 和 Chib(1993)提出的“潜变量数据扩展分析”思路对模型的相关参数进行 Gibbs 抽样估计(具体步骤见第3章)。“有序选择模型”的估计理论相对简单^①,而“无序选择模型”的估计方法则比较复杂。

当“无序选择模型”的“选项”间无相关关系时,可以建立 MNL(Multinomial Logit)模型。频率学派运用 ML 模型对 MNL 模型进行估计。从 Bayes 学派角度,Koop 和 Poirier(1993)研究了 MNL 模型中参数的抽样估计方法及其先验分布的选择等问题。此外,Allenby,McCulloch 和 Rossi(2005)提出运用“独立”MH 抽样算法和“随机游走”MH 抽样算法对 MNL 模型进行抽样估计。而当 MNL 模型的解释变量表现出不同的属性时,MNL 模型有不同的分类。各类 MNL 模型的建立依赖于较强的假设条件:IIA(independent of irrelevant alternatives)假设^②,IIA 假设的成立是 MNL 模型正确设定的充要条件。Hausman,McFadden(1984)等从频率学派角度提出 IIA 假设的各种检验方法(Hausman 检验、Wald 检验、LR 检验及 LM 检验),并对不同检验方法的检验功效和检验水平进行了比较。值得说明的是,IIA 假设是很强的约束条件,现实中更为常见的是不同选项间存在相关关系或具有一定程度的替代关系(substitutes)。当 IIA 假设不成立时,可以建立 GEV(Generalized extreme value)模型、嵌套(Nested)MNL 模型、随机系数 MNL 模型及 MNP 模型。

McFadden(1978)提出了 GEV 模型的一般形式,该模型通过设定随机扰动项联合分布的形式来描述各选项间的相关关系。而当各选项间存在特定“嵌套”^③结构时,可以对个体行为建立“嵌套”MNL 模型,除运用 ML 方法估计“嵌套”MNL 模型外,Poirier(1996)也讨论了“嵌套”MNL 模型的 Bayes 估计、推断等问题。在实证分析中,为反映选项间的相关性,用得最多的是 MNP 模型,MNP 模型通过随机扰动项的协方差矩阵来描述选项间的相关结构。在很长一段时间内,由于计算高维积分的难度极大^④,MNP 模型的分析、应用一直处于停滞状态。直到 Keane

① 实际上是施加较强的约束条件而简化了相关分析。

② 对于特定个体,其选择一组选项(alternatives)中任意两项的概率比(the ratio between their choice probabilities)不随其他选项的增加或减少而变化。

③ 具体结构见第3章。

④ 运用 ML 方法估计 MNP 模型时,计算“选择概率”涉及高维积分的运算,直到 GHK 算法的提出,该问题才得以解决。

(1994)、Hajivassiliou, McFadden and Ruud(1996)提出了 GHK 算法、MNP 模型的 ML 估计不再是难题之后, MNP 模型才有了广泛的应用。与此同时, 众多学者也运用 Bayes 学派的相关理论对 MNP 模型的估计问题进行了研究: McCulloch 和 Rossi(1994)提出了 MNP 模型的不完全识别估计方法。而针对 McCulloch 和 Rossi(1994)不完全识别估计方法收敛速度较慢、对参数初始值选择较敏感的特点, Nobile(1998)提出了 MNP 模型的混合 MCMC 估计方法。该算法弱化了参数抽样值 Markov 链对其初始值的依赖性, 同时加速了抽样值 Markov 链收敛的速度, 因而是 MNP 模型较为有效的估计方法。虽然 McCulloch 和 Rossi(1994)、Nobile(1998)估计方法“计算简单”^①, 但他们提出的估计方法的局限性同样明显: 均不能对模型所有参数进行识别。为了对 MNP 模型的所有参数进行有效识别, McCulloch、Polson 和 Rossi(2000)提出了 MNP 模型的完全识别估计方法。该方法的缺陷是参数抽样值构成的 Markov 链具有较高的自相关系数, 参数抽样值的收敛速度较慢。在比较以上各 Bayes 抽样估计方法的基础上, Imai 和 Dyk(2005)在“数据扩展”分析方法的基础上, 提出 MNP 模型的一种新估计方法, 该估计方法具有简单的参数先验分布形式, 同时参数抽样值具有较快的收敛速度。最后, 随机系数 MNL 模型也是描述选项相关性的常用建模方法。对于频率学派的学者, 他们提出运用 SML 方法估计随机系数模型^②。除此之外, 根据“等似然函数”和“成比例似然函数”的相关结论, Chen 和 Kuo(2001)提出通过估计“泊松非线性模型”和“泊松对数线性模型”来实现随机系数 MNL 模型的估计。从 Bayes 学派角度, Train(2001, 2003)提出了随机系数 MNL 模型的 Bayes 抽样估计方法。

1.2.2 非平稳时间序列数据离散选择模型综述

换一个角度, 许多宏观经济问题可以转化成离散选择模型的建模问题, 如居民贫困问题、经济增长问题、汇率制度决定问题以及宏观经济政策调整等问题。但对以上宏观经济问题建立离散选择模型时, 一个显著的特点是离散选择模型的数据基础为非平稳时间序列数据(宏观经济变量大多表现出明显的时间趋势, 为非平稳解释变量)。此时, 非平稳时间序列解释变量的出现是否对离散选择模型的估计产生影响? Park 和 Phillips(2000)最早对这一问题进行了研究。他们给出了非平稳时间序列数据的二元选择模型, 并假设模型的解释变量均为随机游走过程, 且模型的参数向量为非零向量。对于这种模型, Park 和 Phillips(2000)研究了模型参数 MLE 的渐近性质(包括参数 MLE 的一致性、收敛速度及其渐近分布形式)。同时,

① 它们都是对相关标准分布进行随机抽样。

② 运用 ML 方法对随机系数 MNL 模型估计时, 特定选项的“选择概率”不能解析, 此时运用特定模拟方法估计“选择概率”, 此思路下的 ML 方法称为 SML 方法。

他们也对该模型相关统计量的渐近性质进行了分析。在其基础上,Guerra 和 Moon(2002)讨论了单变量非平稳时间序列数据 Logit 模型,当解释变量为随机游走过程、且对应系数为 0 时,模型参数极大似然估计量的渐近分布及对应 t 统计量的渐近性质。可以看出,以上理论研究均具有较强的假设条件:解释变量是并且都是随机游走过程。但在运用离散选择模型对宏观经济问题的研究中,更为常见的情形是:时间序列解释变量中既可能含有平稳解释变量,同时也可能含有非平稳解释变量;而在非平稳解释变量中,既可能含有 $I(1)$ 过程,也可能同时含有 $I(2)$ 过程。Wojciech Grabowski(2006)研究了这种更一般情形下二元选择模型参数 MLE 的渐近性质。同时,对不同可能情形下二元选择模型参数 MLE 的小样本性质进行了分析。相关结论表明:在小样本容量条件下,参数 MLE 经常偏离其真值较远,其小样本性质较差(poor)。

Hu 和 Phillips(2004)、Phillips, Jin 和 Hu(2007)对非平稳时间序列数据有序选择模型的建模、估计及推断等相关问题进行了研究。值得说明的是,与截面数据有序选择模型的设定形式不同,非平稳时间序列数据有序选择模型的门限值为“规模化(scaled)”的门限值,该门限值随样本容量而变化。模型“规模化”门限值的设定,其基本原理是保证门限值与“潜变量”具有相同的数量级,只有这样,才不至于弱化样本数据的样本信息。^① 在他们的研究中,分别对不同情形下(情形 1:仅含有随机游走解释变量、系数向量不为零的有序选择模型;情形 2:既含有随机游走解释变量、又含有平稳解释变量,且系数向量不为零的有序选择模型)有序选择模型参数 MLE 的渐近性质和相关统计量的渐近分布进行分析,不同情形下相关结论见 Hu 和 Phillips(2004)、Phillips, Jin 和 Hu(2007)。

1.2.3 面板数据离散选择模型综述

当研究对象为“同质”个体时,可以对面板数据建立“混合效应”离散选择模型;当研究对象为“异质”个体时,可以对面板数据建立“固定效应”离散选择模型和“随机效应”离散选择模型。

对于固定效应二元选择模型,在给定 T 条件下(N 可以自由取值),由于模型的非线性设定形式,基于模型对数似然函数的极大似然估计会产生“偶发性参数问题”,其后果是模型参数(无论是固定效应参数还是结构参数)的 MLE 是参数真值的非一致估计量(个体效应参数随个体数目 N 变化,个体效应参数估计量的非一致性传递给结构参数,导致参数估计量不具一致性),有关该问题的论述及证明详见 Andersen(1973),Heckman(1981),Hsiao(2005)。Neyman 和 Scott(1948)最早

^① 详见第 4 章“时间序列有序选择模型”的理论回顾。