

清 华 大 学

综 合 论 文 训 练

题目：基于选择模型和机器学习的
消费者需求研究

系 别：工业工程系

专 业：工业工程

姓 名：竭卓妮

指导教师：邓天虎 副教授

2018 年 6 月 12 日

清华大学工业工程系本科综合论文评阅意见反馈

论文题目	基于选择模型和机器学习的消费者需求研究	姓 名	竭卓妮
		学 号	2014010882
		系别/学校	工业工程/清华大学
<p>论文修改说明/针对评阅意见提出的问题反馈</p> <p>本论文评阅意见如下：本文对从理论和实证方面，研究消费者需求预测和分析问题，为今后基于选择模型和机器学习的消费者需求预测模型集成提供新的思路。论文选题和工业工程在服务行业中的应用密切相关，具有较好的理论意义和实际应用的价值，文献调研充分，工作内容充实，方法丰富并且正确，结论明确。符合本科毕设要求，同意学生参加答辩。</p> <p>在论文改善方面，文章在集成学习中选择 stacking 方法，其原因需要进一步阐述。此外，建议增加一些离散选择模型和机器学习模型的互补讨论。最后，参考文献中的中文文献数量过少。</p> <p>本文针对评阅意见、指导老师意见及自我检查进行了以下几方面的修改。首先，在综述中阐述现有研究的模型估计方法时，在模拟矩估计方法的介绍中加入了下文提到的广义矩估计方法的叙述，这两种估计方法为包含关系。第二，完善了“聚合型消费者需求数据集类别定义及相关研究”表格部分的内容引用。第三，将数值案例中第二种选择模型名称由“改进 logit 模型”改为更明确的“考虑消费者购买量的 logit 模型”。第四，在模型集成介绍部分更明确地阐述了本研究选择 stacking 方法的原因。第五，在数值案例结果讨论中增加了离散选择模型和机器学习模型互补特点的讨论。第六，完善了中英文摘要和外文文献调研报告中的语言表述。第七，检查统一了模型推导部分的符号表示。最后，由于国内相关研究还较少，很多中文文献也大多直接参考引用国外研究，国内该领域的优秀前沿研究也往往选择在国际刊物上发表研究成果，所以我在根据论述需要综合考虑之后增加了 2 篇所参考的中文文献。</p>			
修改日期：		签字：	

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签 名：_____导师签名：_____日 期：_____

中文摘要

在市场营销领域中，进行消费者需求预测与分析有很重要的实际意义，而选择模型和机器学习模型是进行需求预测的两类重要工具。

理论方面，本论文梳理了离散选择模型的发展历程，对比了常用几类模型各自特点，并在经典选择模型的基础上进行了推导改进和编程实现。同时，本文归纳了现有少量使用机器学习方法进行消费者需求预测的研究，并研究了三种机器学习模型和它们的模型集成方法，发现模型集成可以改进预测效果。本文提出可以运用集成学习的方法，来结合选择模型与机器学习模型进行预测，发挥这两大模型各自的优势。

实证方面，本论文根据公开数据集做了数值案例分析，进行了模型结果比较，探究了影响需求的因素，进行了消费者行为解释，提出了拓展数据集、考虑供给侧信息、考虑多层次模型等未来改进方向。

选择模型和机器学习模型结合领域的相关研究目前还非常少，本文的理论研究和实例分析可以为今后基于选择模型和机器学习的消费者需求预测模型集成提供新的思路。本文研究成果可以运用到捆绑销售、推荐系统设计、电商和大零售商系统设计等实证领域。

关键词：需求预测；离散选择模型；机器学习；集成学习；模拟矩估计

ABSTRACT

In the field of marketing, it is of great practical significance to carry out consumer demand forecasts and analysis. Choice models and machine learning models are two important tools for demand forecasting.

On the theoretical side, this dissertation summarizes the development of discrete choice models, compares the characteristics of several models, and derives, improves and programmatically implements the classical choice model. Besides, this paper summarizes the existing research on the use of machine learning methods for consumer demand forecasting. This paper then studies three machine learning models and their integration methods, and found that model integration can improve forecasting results. This dissertation proposes that we can use ensemble learning methods to combine choice model with machine learning models for demand forecasting.

On the empirical side, a numerical case study was conducted based on a public dataset. This dissertation compared the model results, explored the factors affecting the demand, and interpreted consumer behaviors. For future work, we can expand the datasets, add the supply side information, and maybe consider the multi-level models.

There is yet very little research on the combination of the choice model and the machine learning models. The theoretical research and empirical analysis of this paper can provide new ideas for the future integration of consumer demand forecasting models based on the choice model and machine learning models. The research results of this paper can be applied to empirical fields such as bundling sales, recommending system design, e-commerce and large retailer system design.

Keywords: Demand forecasting; discrete choice model; machine learning; ensemble learning; simulated method of moments

目 录

第 1 章 引言	1
1.1 研究背景及意义	1
1.2 研究现状	4
1.3 研究方案及论文结构	8
第 2 章 离散选择模型	10
2.1 发展历程	10
2.2 理论基础	11
2.2.1 模型假设	11
2.2.2 随机效用理论	12
2.2.3 效用最大化假说	13
2.2.4 常用离散选择模型	13
2.3 模型推导	16
2.4 模型估计	21
第 3 章 机器学习模型与模型集成	23
3.1 岭回归	23
3.2 随机森林	24
3.3 支持向量机回归	25
3.4 机器学习模型实例	26
3.5 模型集成	30
第 4 章 数值案例	33
4.1 数据介绍	33
4.2 模型结果	36
第 5 章 总结与展望	40
5.1 研究总结	40
5.2 创新点	40
5.3 研究展望	41

插图索引.....	42
表格索引.....	43
参考文献.....	44
致 谢	51
声 明	52
附 录 外文资料调研阅读报告	53

第1章 引言

本章首先从消费者需求分析的实际应用领域引入课题背景，系统回顾了应用离散选择模型与机器学习模型做消费者需求预测的理论研究成果，点明了需求预测模型面临的待估计参数维度大、消费者品味异质性、商品价格内生性三大问题，并对以 logit 模型为代表的离散选择模型家族技术发展与所应用的数据类型发展进行了综述，介绍了目前少量的机器学习预测研究与机器学习应用于离散选择模型的研究，在此基础上归纳总结出了本课题的研究意义。接着在研究现状方面，本章归纳介绍了现有研究中离散选择模型的种类、数据集种类、模型估计方法种类和机器学习方法。本章最后阐明了研究方案与本文结构，为进一步展开叙述奠定基础。

1.1 研究背景及意义

在市场营销领域的社会学分析中，消费者需求分析一直占有非常重要的地位。消费者需求分析可以用于分析市场干预和政策实施给市场带来的影响，还可以用于预测销售额，描述和检验消费者行为，并分析公司具体的营销策略来实现公司利益和消费者福利的最大化^[1]。近年来，消费者需求预测的运用也得到了飞速发展，从传统的零售业、汽车业等销售数据的分析与需求预测，到现在广告投放、视频推荐等用户推荐智能算法，消费者需求预测都能发挥巨大作用。

对于不同的市场营销策略，消费者可能会作出更换品牌、改变购买数量、既考虑调整品牌选择又改变购买数量、不采取措施等各种反应。研究消费者需求而发展出的各种需求模型，主要用于定量分析消费者和市场在不同营销策略下的效益反应。随着大数据时代的到来，相关消费数据来源越来越丰富，数据处理方法不断创新突破，运算设备硬件性能、存储能力更新换代，相关需求预测模型的种类、性能和它们各自的应用特点也在不断发展。这些模型通过求解顾客最大化效用的条件，来得到相应的最优条件，这些条件也可以反应消费者不同选择之间的相互影响和权衡^[2]。

通常情况下，消费者在发生购买时，会作出以下三种决策：买还是不买商品，选择哪一种或多种商品购买，每种的商品购买数量。这些消费者行为数据通常作为每次购买的行为特点被单独记录^[3]。研究对差异化产品的顾客需求，通常有以下几

个关注点。首先，大量的差异化产品给模型带来了大量的待估计参数，即使通过一些经济学理论对模型进行限制也不能解决这个问题。第二，消费者对产品具有异质性的偏好，怎样解释这种偏好异质性会对经济与政策结论产生重大影响。第三，产品的价格可能和某些销售者能观测到但研究者观测不到的因素有关，这会带来产品价格的内生性问题^[4]。消费者个人品味异质性和市场营销因素可能带来的价格内生性，若不被正确地处理，会给模型参数估计带来偏差。

McFadden 在 1973 年提出了 logit 需求模型，通过把产品映射到一个由产品特征来表示的空间上来解决变量维度问题，但这个模型对商品和消费者偏好异质性的解释效果很有限^[5]。正如 Allenby 和 Rossi（1999 年）所研究的，许多传统计量经济学研究中，消费者偏好的异质性常常被作为干扰因素从目标函数中积分抵消^[6]。然而在需求预测实证领域，使用正确合理的模型来拟合消费者品味的异质性对提高选择预测精确度很重要。后来的研究在 McFadden 的模型基础上引入了人口异质性，同时保留了这个需求模型的优点。Train（2009 年）^[7]，Allenby 和 Rossi（1999 年）^[8]运用了混合 logit 模型来描述消费者偏好异质性，并通过仿真来估计选择概率和相关系数。另外一种用来处理消费者品味异质性的方法是采用机器学习之中经常运用的正则化手段。Evgeniou 等人在 2007 年采用凸优化建立相应方程^[9]。2014 年，Mishra 等人研究了边际分布模型（the marginal distribution model, 即 MDM）的特性，发现广义极值分布模型家族（包括 multinomial logit, 即 MNL 和 nested logit, 即 NestL 等）均可以由这个模型推导得到。不像极值分布模型家族要求错误项必须独立，边际分布模型可以用于推断更灵活、更有代表性的选择模型。另外，将边际分布模型和机器学习结合起来，可以处理消费者和产品间的成分效用和尺度异质情况^[10]。Berry 等人在 1994 年^[11]、1995 年^[12]提出了 BLP 模型估计方法，可以运用到个体水平和聚合水平的数据中，首次解决了价格内生性问题。后来 BLP 模型估计方法在 Nevo（2001 年）^[13]、Petrin（2002 年）^[14]、Goolsbee and Petrin（2004 年）^[15]，Chintagunta, Dubé 和 Goh（2005 年）^[16]等人的研究中得到广泛应用。

最早的消费者选择模型中，研究者大多分别仅研究消费者是否发生购买或者消费者购买产品种类及数量。比如 Wales 和 Woodland^[17]，Lee 和 Pitt^[18]先后在 1983、1986 年研究了消费者在什么时候不会产生购买行为，King^[19]，Dubin 和 McFadden^[20]等人则在 1980 年、1984 年研究了消费者在产生购买行为的条件下的离散和连续的选择行为。Gupta 在 1988 年建立了模型，分别独立研究了消费者商品品牌，购买数量和购买时间这三种决策^[21]。后来，Chiang 等人将模型发展，同

时考虑了是否发生购买及具体选择的品牌以及相应购买数量这三种决策过程，并论述了考虑不购买行为对最终购买概率及数量的影响关系^[22]。

销售数据有个体水平和聚合水平之分。Krishnamurthi 和 Raj (1988 年)^[23], Chiang (1991 年)^[24], Chintagunta (1993 年)^[25]等学者先后提出了一些使用个体水平的销售数据的离散选择模型。然而，在实际情况中，这种非聚合型的细分数据很难得到，我们更容易得到的是商店水平、连锁店水平或市场水平的各种聚合数据。因此，很有必要建立针对相应聚合数据的需求预测模型。Besanko (1998 年)^[26]和 Sudhir (2001 年)^[27]提出了针对聚合数据的 Logit 需求系统分析方法。Logit 选择模型方法和传统离散选择模型是非常简洁的，适用于有很多细分商品的产品大类，但同时这种模型也有缺点，它建立在消费者发生的都是单件商品购买的假设前提下。在这种模型假设下，所计算出的各种商品市场份额就会和真实值有偏差，所估计的市场销售额和商品的需求价格弹性系数也会有偏差，可能会误导定价策略和营销利润措施。随后 Nair 提出了一个基于聚合数据的离散/连续需求模型，这个模型可以在商店水平的数据中获得比以前 logit 模型更好的估计结果，同时可以解决消费者个人品味异质性和市场营销因素内生性的问题^[28]。但是他们的模型目前没有区分消费者主动购买大批货物的行为和预测市场情况所采取的囤货行为。Erdem 等人在 2003 年利用家庭层级的数据，对消费者囤货行为已经有了初步研究成果^[29]。

除离散选择模型之外，也有一些研究探究了机器学习模型在需求预测领域的作用。以机器学习模型为代表的结果预测型模型，将在以自动建模、智能软件助手为代表的实际市场运用中扮演越来越重要的角色。支持向量机在营销领域的应用还相对较少。Cui 和 Curry 在 2005 年发表的文章中系统性研究了市场营销领域中不同预测环境下支持向量机模型的预测情况，并指出机器学习方法和传统消费者选择模型之间可以优势互补^[30]。他们用实验证明了选择模型较支持向量机模型更容易受到选择集大小和顾客特征维度的影响。与传统选择模型相反，更大的选择集和更高的特征维度反而会提升机器学习模型的预测正确率，机器学习模型更适合复杂关系中的预测工作。Cui 和 Curry 指出，支持向量机模型依然有很多不足之处，在投入实际运用之前，它的发展还需要进一步的理论推断支持。Viaene 等人研究发现，一种最简单的 OLS 类支持向量机都已经表现出了比人工神经网络、k 近邻算法、决策树、贝叶斯多层感知器等算法更好的预测效果^[31]。Platt (2000 年)^[32], Vapnik 和 Chapelle (2000 年)^[33]针对如何得到购买概率估计进行了研

究。他们提出，未来如何从支持向量机模型中得到除单纯结果预测外的其他模型结构性预测，也是一个研究方向^[34]。

还有一些研究探索了机器学习方法在离散选择模型的建立与估计中的应用。Norets 等人在 2012 年运用人工神经网络来近似动态规划解，用于处理动态离散选择模型中的高维积分，推断随机系数^[35]。Benson 等人在 2018 年提出一种子集选择模型的启发式算法，运用机器学习模型学习子集效用，然后得出选择概率^[36]。Chiong 等人（2016 年）^[37]在具有高维选择集的离散选择模型中用机器学习中的随机映射来降维，然后利用离散选择模型的凸分析性质，通过周期单调矩不等式来估计模型。根据调研，目前还没有研究直接结合离散选择模型和机器学习模型的预测结果来进行需求预测，本论文将在总结与展望部分提出今后可能的模型结合方向。

综上所述，本课题的研究意义主要有以下两方面。第一方面，梳理了目前离散选择模型的发展和特点，在现有选择模型基础上进行实现和改进，同时归纳了现有使用机器学习方法进行消费者需求预测的研究，使用了三种机器学习模型实现预测，并为今后基于选择模型和机器学习的消费者需求预测模型集成提供思路。第二方面，根据公开数据集做了数值案例分析，探究了影响需求的因素，进行了消费者行为解释，提出了拓展数据集、考虑供应侧信息、考虑多层次模型等未来改进方向。我们的研究成果可以运用到捆绑销售、推荐系统设计、电商和大零售商系统设计等实证领域。

1.2 研究现状

自从 1970 年 McFadden 在旧金山市铁路交通系统方式需求预测中提出基于消费者效用最大化理论的多项 logit 模型以来，离散选择模型研究得到了愈发广泛的运用与发展^[38]。离散选择模型由效用函数随机项所服从的分布可分为两大类：第一类模型，即 logit 模型家族，假设效用函数中包含的随机项服从极值分布；第二类模型，即 probit 模型家族，假设效用函数中包含的随机项服从正态分布。其中某些 logit 模型的求解不能只通过解析方法，而需要随机模拟完成模型估计。由选择项之间相关性强度的不同假设条件，logit 模型又可分为多元 logit（multinomial logit）模型、嵌套 logit（nested logit）模型和随机系数 logit（random coefficient logit）模型等^[39]。Probit 模型家族和嵌套 logit 模型、随机系数 logit 模型，可以解决传统 logit 模型所面临的 Luce（1959 年）提出的“不相关选项间的独立性”（Independence

from Irrelevant Alternatives，即 IIA）问题^[40]。该条件会在 2.2.4 部分进行更详细的阐述。传统 logit 模型的 IIA 特性，会使得相近选择支被选择的概率被夸大。

各类离散选择模型特征概览总结如下表：

表 1.1 各类离散选择模型概览^[41]

基本假设	模型名称	主要研究人员	年代	模型特点
极值分布	多项 logit 模型	McFadden	1970	随机项独立同分布，具有 IIA 特性，具有数值可解性。
	嵌套 logit 模型	Ben-Akiva Lerman	1973	选择集具有嵌套结构，能部分解决 IIA 问题。
	混合 logit 模型	McFadden Train	2000 2003	参数具有随机性，选择集中各项都可以都相关，无封闭解。
正态分布	probit 模型	Daganzo	1979	选择集中各项可以相关，求解较 logit 模型复杂很多，应用意义不大。

数据方面，在需求预测中，个体水平的数据常常很难获得，即使获得，也很难完全保证数据的完整性，并且缺少很强的市场代表性。在实际操作中，我们更容易得到的市场数据是聚合类的数据，而使用这种数据即面临着解决商品和消费者偏好异质性、商家操作和商品价格存在内生性的问题。本课题主要关注聚合数据下的需求预测。在估计过程中，我们经常会用到聚合需求数量、价格、商品特性、市场总量、人口统计学数据等变量。现有研究中已有一些公开发表的数据集，经典聚合型消费者需求数据的三大类别、类别定义及部分相关研究见下表：

表 1.2 聚合型消费者需求数据集类别定义及相关研究^[42]

类别	定义	相关研究
纵向数据	一个市场/商店，不同时期	BLP Ecta 1995; Sudhir Mkt Sci 2001; Chintagunta Mkt Sci 2001
横截面数据	多个市场/商店，一个时期	Datta and Sudhir 2011
面板数据	多个市场/商店，多个时期	Nevo Ecta 2001; Chintagunta, Singh and Dube QME 2003

离散选择模型估计方面，现有研究主要通过仿真模拟来得到购买概率的估计，主要的参数估计方法有以下三种^[43]：

1. 最大模拟似然（Maximum Simulated Likelihood）：和最大似然法类似，但使用的不是确切的购买概率，而是模拟概率。其特性由 *Gourieroux* 和 *Monfort*（1993 年）^[44]，*Lee*（1995 年）^[45]，与 *Hajivassiliou* 和 *Ruud*（1994 年）^[46]等人的研究得到。
2. 模拟矩估计法（Method of Simulated Moments）：模拟矩估计法和广义矩估计法（Generated Method of Moments）实质相同，只是模拟矩估计法的矩条件是由模拟得到的。该方法最早由 *McFadden*（1989 年）^[47]提出，适用于传统矩函数很难直接得到的情况。
3. 模拟评分法（Method of Simulated Scores）：一个观测的对数值的梯度被定义为此观测的分数。此方法寻找使平均分数为 0 的参数值。当确切的概率值被使用时，这种方法和最大似然法等效，*Hajivassiliou* 和 *McFadden*（1998 年）^[48]提出较使用准确分数而言，使用模拟的分数可以得到更松弛条件下的一致性和效率。

Berry，*Levinsohn* 和 *Pakes* 在 1995 年提出了一种可以利用聚合型数据估计有差别的商品的需求的离散选择模型^[49]。他们在模型中引入了随机系数解决随机性问题，尝试解释了研究者观测不到的对消费者选择有关的公司策略等内生性问题。在他们的方法中，每个市场中的消费者面临相同的商品集合和各商品相同的特性，但商品特性可能在不同市场间发生变化。*Sudhir*（2001 年）^[50]和 *Chintagunta*（2001 年）^[51]很快意识到了这种估计方法的重要性并进行了相关研究，在此之后，BLP 方法很快得到广泛应用。下表是运用过 BLP 方法的一些著名期刊上的文章：

表 1.3 使用 BLP 估计聚合数据的部分文章总结

期刊	文章
American Economic Review	Berry, Levinsohn, and Pakes (1999)
Econometrica	Berry et al. (1995); Goeree (2008); Nevo (2001)
Journal of Political Economy	Berry, Levinsohn, and Pakes (2004); Petrin (2002)
Review of Economics and Statistics	Nevo (2003); Rekkas (2007)
Review of Economic Studies	Berry, Linton, and Pakes (2004); Nakamura and Zerom (2010); Villas-Boas (2007)
Rand Journal of Economics	Akerberg and Rysman (2005); Armantier and Richard (2008); Berry (1994); Bonnet and Dubois (2010); Chu (2010); Copeland, Dunn, and Hall (2011); Davis (2006); Iizuka (2007); Nevo (2000); Song (2007); Villas-Boas (2009)

另一方面，在消费者需求预测领域，机器学习模型具有一些选择模型没有的优点，它们可以摆脱对许多模型结构类型假设的依赖。机器学习模型还非显性地自动化了模型预测过程，在参数预测阶段可以考虑一种可能性的结构组合，而不只是单个可能性。以支持向量机模型为例，支持向量机模型会通过核变换，把非线性问题映射到高维空间从而转换成线性问题求解，较极大似然等方法具有更好的稳定性和鲁棒性。即使使用数据量相对较小的数据，也会得到相对稳定的预测结果。机器学习模型也因其特定优越性，将在聚合型数据的需求预测下发挥越来越大的作用，但目前相关研究还相对较少。机器学习也可能被运用到分析消费者的非互补型选择策略中，分析消费者会根据哪些特性采取哪一些特定的策略^[52]。Bajari 在 2015 年的研究中^[53]详细探讨了包括逐步回归、前向分段回归、LASSO 和支持向量机在内的多种机器学习模型在消费者需求预测的应用，并根据 Bates 和 Granger 在 1969 年提出的模型线性结合方法对这些机器学习模型进行了集成，成功改进了预测效果^[54]。

1.3 研究方案及论文结构

本课题的主要工作分为三部分。一是对现有消费者选择模型进行综述梳理、理论推导与优化，同时介绍了三种机器学习模型及其集成方法，运用机器学习方法进行了消费者需求预测，并对机器学习模型进行了线性和 **stacking** 两种集成结合，比较了不同模型的特点与表现。二是利用建立并实现的需求预测模型进行基于仿真和实际数据集的模型检验，对消费者行为特点进行了分析解释。三是提出模型改进的方向和未来离散选择模型和机器学习模型的可能结合方法。

我的研究方案如下：

首先，我阅读了大量文献，深入学习和研习推导各种离散选择模型，总结目前不同模型所使用的方法、解决的问题及侧重点。在此同时我也进行了一些基于开放数据集或仿真数据的案例应用研究，以加深学习。在此基础上，运用已有的公开数据集，对改进的离散选择模型进行编程实现。

第二步，在选取和处理数据得到有关特征后，运用机器学习方法进行需求预测，比较不同方法的模型表现并调整选择最佳参数。在此基础上，通过模型集成结合各模型特点获得更好表现。

第三步，综合以上结果，进行数值案例分析与评价，分析消费者行为特点，提出机器学习和离散选择模型的结合方向。并在最后结合可能的实际应用领域情况，从模型延伸、数据拓展等层面对模型可能的改进之处进行分析和展望。

本课题所采用的研究路线可由下图来概括：

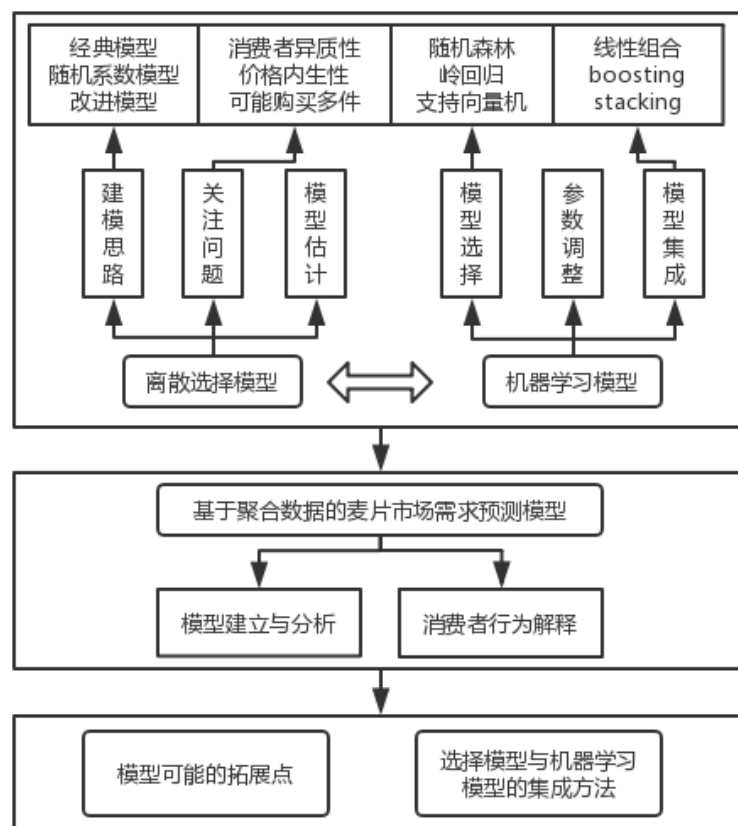


图 1.1 本课题研究路线框图

本论文结构如下：

第一章为引言，系统综述了与本课题研究有关的消费者需求预测理论及相关应用，介绍了研究背景及意义、相关研究现状，梳理了研究方案与论文结构。

第二章总结与推导了离散选择模型，探讨了不同模型的特点，解释了模型的变量选择及意义，指出了目前离散选择模型可以解决的价格内生性和消费者异质性问题，在此基础上对经典模型进行了优化改进。

第三章介绍了机器学习模型及模型集成方法，包括随机森林、岭回归和支持向量机，并在此基础上介绍了目前现有的几种模型集成方法。

第四章结合了一个具体的即食谷物市场数值案例来估计与实现预测离散选择模型和机器学习模型，并进行模型比较与消费者行为分析。

第五章对本论文的研究进行了总结，提出了未来的模型拓展与集成方向，归纳了研究的创新点，并对之后可能的模型改进等工作进行了展望。

第 2 章 离散选择模型

本章首先简要介绍了离散选择模型的发展历程及模型特点演变，接着在阐述决策模型四方面假设和随机效用理论、消费者效用最大化理论的基础上，介绍了各种离散选择模型的理论推导过程及其各自特点，并在此基础上建立并推导了一个较传统 logit 模型有所改进的选择模型，最后分析了估计这种模型的方法。

2.1 发展历程

离散选择模型属于微观计量经济学领域，描述决策者在各可选项之间的决策过程（Train，1985 年）^[55]，可以对决策者的行为进行经验性的统计分析。离散选择模型最早起源于对交通方式选择的分析和预测。1975 年，著名经济学家 McFadden 等人在旧金山交通系统项目研究中提出了多元 logit 模型^[56]，这种关于交通方式选择的统计模型后来经常被有关部门用于道路运载能力估计、道路城市规划等工作中。后来，离散选择模型的应用从交通需求领域拓展到教育及职业选择、消费者商品需求、居住地的选择等领域（McFadden，2001 年）^[57]。结合经济学研究，离散选择模型在劳动经济学领域也有大量应用。研究者利用选择模型对劳动者职业的选择、是否参加工会等活动进行建模，评估群体的职业选择路径并对经济回报制定给出参考^[58]。

在 80 年代离散选择模型发展初期，该领域的相关研究受到多元 logit 模型个人效用随机项独立同分布特性和选择项间独立不相关特性的限制。这些限制条件使模型求解更加容易，但也大大限制了 logit 模型在需求预测中的实际应用。90 年代以来，离散选择模型的研究取得了持续突破，研究者通过在传统 logit 模型中引入模型嵌套、各种形式的随机项等方式，获得了嵌套 logit 模型、混合 logit 模型等新模型，可以使得选择集中各选择具有不同的相关性^[59]。模型中引入的新元素也能解释一些由于市场和消费者特性、产品本身特性等因素带来的需求震荡。运用离散选择模型，可以由模型系数的估计值正负号、显著性分析得到人群的行为规律。还可以由其他要素与价格要素的系数之比，得到该要素的货币化价值，有利于估计不同消费者的支付意愿。同时，选择模型中大量使用的模拟分析，也使得研究者可以通过模拟仿真方式前瞻性地探究政策方案实施、选项和条件的变化对系统表现的影响^[60]。

2.2 理论基础

2.2.1 模型假设

离散选择模型不是为了测量消费者的偏好，而是研究消费者在面临不同选择集的情况下会产生怎样的行为。为了使模型能描述人们做出选择的过程，对于每次决策，我们从以下四个方面进行假设：

1. 决策者 (the decision-maker)：即做出选择的主体。需要定义谁是决策者，以及其自身特点。
2. 选择方案 (the alternatives)：确定决策者的可能选择，为决策变量，通常有多种方案供决策者选择。
3. 属性 (the attributes)：确定各种决策者在做出选择时可能考虑的各选择属性。
4. 决策规则 (the decision rules)：各决策者做出方案选择时的行为准则不尽相同，需要描述确定决策者做出选择的这个过程的依据。

需要注意的是，在通常情况下，为了缩小采用的模型的范围，建模时会固定某些假设^[61]。

决策者是选择行为的主体。通常，决策者可以以个人为单位做出决策，此时的模型被称为非聚合型 (disaggregate) 模型。同时，决策者也可以以群体（如家庭、政府等）为单位，决策群体做出的选择将作为整体被考虑，此时的模型被称为聚合型 (aggregate) 模型。实际应用的具体性质和可得到的数据粒度将决定这个模型的种类。

对于一种商品，消费者会产生“买”或“不买”两种行为。在分析消费者选择时，不仅要关注消费者选择的选项，也要关注消费者没有选择的选项。选择集 (choice set) 是所有选择的集合。选择集具备互斥性、完备性和有限性等三个基本性质。选择集同时有离散、连续之分，在本课题中我们主要考虑离散选择集，其中的选择可以被有限、可数地列出来。选择集按层次可以分为通用方案集 (universal choice set)、减少后的选择集 (reduced choice set)，在减少后的选择集中，现实情况下不可选择的方案都被除去，实际建模中我们通常考虑这种减少后的选择集。

各选项的属性和消费者的属性类似，都是有利于将该选项与其他选项区别开并可能影响消费者选择的因素。这里的单个属性不一定是直接观测到的量，它也可以是观测数据的函数变化形式（如对数等）。

不同的决策准则可能会使决策者做出不同结果的决策。常见的理性决策方式包括优势准则、下限准则、多重排序准则和效用最大化准则等。离散选择模型的基础原理是随机效用理论（random utility theory）^[62]，其构建通常建立在决策者采用效用最大化决策准则的假设之上。效用（utility）是选择支带给决策者的好处，根据 Marschak（1960 年）^[63]的定义，这里的效用可以理解为消费者做出某项决策行为后在个人偏好、所花费用、所用时间及感受到的舒适度等多方面获得的满足感，决策者会依据效用最大化原则进行选择。在这种情况下信息是不完全的，因此效用的不确定性必须纳入考虑。Manski（1977 年）^[64]提出了四大类不确定性因素：未观察到的选择属性，未观察到的决策者属性（被称为 unobserved taste variations），测量误差，工具变量。这些随机性催生了随机效用理论的研究。

2.2.2 随机效用理论

随机效用理论^[65]认为消费者为每个备选项所赋予的效用是一个随机变量，由固定和随机两部分加和构成。此效用 U 包括能被一些可观测的特性解释的效用部分 V ，和受不可观测的因素以及误差影响的部分 ε 。用数学表达式表示如下：

$$U = V + \varepsilon \quad (1)$$

对决策者 i 而言，选择 j 的效用表述为：

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (2)$$

事实上，作为研究者或销售者，我们均不能完全掌握决策者为每个选项赋予实际效用值，此处的 V_{ij} 中也可能会包含一部分随机变量。随机效用理论描述了决策者在选择行为中的随机偶然性。此外，这里的模型要求 ε_{ij} 均值为 0，且与 V_{ij} 不相关。可以证明，当消费者效用有一部分是和选项有关的常数项时， ε 均值为 0 的假设是有效的。若 ε_{ij} 为 Gumbel 分布，则得到的模型属于 logit 模型类别；若为正态分布，则可以得到 probit 模型类别。

确定效用项 V 必为选项本身和消费者特性的函数，令 x 是包含所有商品和消费者特征的向量，考虑有 n 种属性，令 β 为待估计参数，则决策者 i 对选择 j 的可观测效用可表示为：

$$V_{ij} = V_{ij}(x_{ij}) = \sum_{k=1}^n \beta_k x_{ij}(k) \quad (3)$$

2.2.3 效用最大化假说

效用最大化理论^[66]认为，在特定选择条件下，决策者总是会从选择集中理智地选择其所认知到的对自己效用最大的方案。于是决策者 i 从选择集 J 中选择 j 的概率可以表示为：

$$\begin{aligned} P_{ij} &= \text{Prob}(U_{ij} \geq U_{im}, \forall m \in J) \\ &= \text{Prob}(V_{ij} + \varepsilon_{ij} \geq V_{im} + \varepsilon_{im}, \forall m \in J) \\ &= \text{Prob}(V_{ij} - V_{im} \geq \varepsilon_{im} - \varepsilon_{ij}, \forall m \in J) \end{aligned} \quad (4)$$

在选择行为之中，任何效用的绝对值是无意义的，对选择行为起作用的是选择集中各选择支之间的相对效用值。效用函数具有线性性。

2.2.4 常用离散选择模型

如第 1 章所提到的，常用的离散选择模型有多元 logit、GEV（即 generalized extreme value model，推广极值模型）、混合 logit、probit 模型等几类。

第一大类离散选择模型，logit 模型，又称 logistic 回归模型。假设随机效用 ε 服从极值分布，该模型可以由 Luce（1959 年）对于选择概率提出的非相关选项独立性假设公理（即 IIA）得到^[68]。Luce 公理如下：

条件 1（IIA 条件）

假设选择集 B 中有选项 x 和 y ，给定影响决策的特征 $s \in S$ ，有

$$\frac{\text{Pr}(x|s, \{x, y\})}{\text{Pr}(y|s, \{x, y\})} = \frac{\text{Pr}(x|s, B)}{\text{Pr}(y|s, B)} \quad (5)$$

即对决策者而言，选择 B 中任意两个选项 x 和 y 的概率之比不受该选择集中其他任何选项的影响，即使其他选项可能和这两个选项特别类似。这意味着选择集中的其他任何选项均对选项 x 、 y 有相同的影响效果。

条件 2（概率为正）

假设决策者选择任何一个选项的概率均严格大于零。

$$\Pr(y|s, B) > 0 \quad \forall y \in B \quad (6)$$

由 Luce 公理，我们可以推得 logit 模型的基本形式。当决策者 i 选择选项 j 时，概率为：

$$P_{ij} = \frac{e^{v_{ij}}}{\sum_m e^{v_{im}}} \quad (7)$$

1974 年，McFadden 研究指出通过 Luce 公理和一个关于 ε 的限制条件，可以得到 I 型极值分布^[69]。然而 logit 模型也有这个公理带来的局限性，我们假设了不同 ε 之间的独立性，但实际上它们也有可能相关。由于 IIA 特性，相近选择支的被选择概率容易被夸大。可以采用以嵌套 logit 模型为代表的广义极值模型解决 IIA 特性带来的问题。

广义极值模型的得名来源于其随机项的性质，假设 ε 服从极值分布，但允许各选择支之间存在不同程度的相关性。若我们总共有 J 个选择，把选择集分为 S 个嵌套，每个嵌套包含的选择集为 B_s ， λ_s 为各个嵌套对应的权重系数， $s \in S$ 。我们可以得到如下的联合累积分布函数：

$$F(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J) = \exp \left(- \sum_{s=1}^S \left(\sum_{j \in B_s} e^{\left(\frac{\varepsilon_j}{\lambda_s} \right)} \right)^{\lambda_s} \right) \quad (8)$$

所对应的选择选项 i 的概率为：

$$i \in B_s: P_i = \frac{e^{\frac{V_i}{\lambda_s}} \left(\sum_{j \in B_s} e^{\frac{V_j}{\lambda_s}} \right)^{\lambda_s - 1}}{\sum_{t=1}^S \left(\sum_{j \in B_t} e^{\frac{V_j}{\lambda_t}} \right)^{\lambda_t}} \quad (9)$$

广义极值分布可由不同条件得到不同的推广。当所有随机项的相关性为零时，广义极值分布就是标准的极值分布，即为标准的 logit 模型^[70]。一般的广义极值分布即为嵌套 logit 模型，一个嵌套内的选择项的不可观测因素之间存在相同的相关性，满足 IIA 条件，而各嵌套还可以互相嵌套，且不同嵌套之间各选项没有相关性，因此可通过合理建立嵌套解决 IIA 特性带来的问题。更复杂的广义极值模型还包括嵌套包含嵌套（nests within nests）和嵌套重叠（overlapping nests）等模型。

混合 logit 模型（随机系数 logit 模型）在衡量产品特性时允许通过各种随机系数表示消费者偏好异质性，相对于简单 logit 和嵌套 logit 来说更灵活。而且这个模型还能包括各市场中各产品特异的需求冲击，可以描述能影响消费者选择但是研究者无法控制的产品特性。Berry 等人在 1995 年即提出了考虑到异质性的采用广义矩估计（GMM）和固定点迭代方法的估计方法，可以得到在每个市场和特定产品有关的需求冲击效应^[71]。

第二大类离散选择模型，probit 模型^[67]，假设 ε 服从正态分布，其概率计算为非封闭型，需要通过数值计算和计算机模拟来进行参数估计。Probit 模型能较好克服传统 logit 模型 IIA 特性的缺点。如果假设 v_i 是一个有关选择特征的简单线性函数（我们可以推广这个线性函数来包括个人特点和其他的一些交互项），我们可以得到以下的效用表达式：

$$U_i = Z_i \beta + \eta_i \quad (10)$$

其中 $\eta_i \sim N(0, \sigma_i^2)$, $\eta_i \perp Z_i, \beta, \eta_i, \forall i, j$ 。 β 也是一个随机变量，有 $\beta \sim (\bar{\beta}, \Sigma_\beta)$ 。针对两个选项的情况，若我们选择选项 1，可推算得到：

$$\begin{aligned} & \Pr(U_1 - U_2 \geq 0 \text{ and } U_1 - U_3 \geq 0) \\ &= \Pr [\sqrt{\text{Var}(U_1 - U_2)}t_1 + (Z_1 - Z_2)\bar{\beta} \\ &\geq 0 \text{ and } \sqrt{\text{Var}(U_1 - U_3)}t_2 + (Z_1 - Z_3)\bar{\beta} \geq 0] \end{aligned} \quad (11)$$

其中 t_1 和 t_2 是可能有相关性的标准正态分布。类似地，可以利用下式计算得到多种选择情况下 probit 模型的选择概率：

$$\Pr(\text{选择 } 1) = \int_{-\infty}^a \left(\int_{-\infty}^b \left(\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{\frac{1}{2}(\varepsilon_1^2 - 2\rho t_1 t_2 + t_2^2)}{1-\rho^2}} \right) dt_2 \right) dt_1 \quad (12)$$

$$a = \frac{(Z_1 - Z_2)\beta}{\sqrt{\text{Var}(U_1 - U_2)}}, \quad b = \frac{(Z_1 - Z_3)\beta}{\sqrt{\text{Var}(U_1 - U_3)}}$$

2.3 模型推导

根据 Hanemann 在 1984 年提出的框架^[72]，本文接下来简要介绍基于 Chiang（1991 年）^[73]、Chingtagunta（1993 年）^[74]、Nevo（2000 年）^[75]和 Nair（2005 年）^[76]的消费者水平模型。其中涉及的模型变量、符号及相关假设总结如下：

表 2.1 模型涉及变量及相应描述

变量	描述
J	商品总数，每个商品为 1,2,...,J
x	商品购买量
p	商品价格
y	消费者总预算
z	消费者潜在购买预算
ψ_j	消费者心中对商品赋予权重
ψ_z	消费者心中对不购买选项赋予权重
u	消费者直接效用
ω_j	商品 j 特性
w	消费者个人偏好特性
ϵ	能影响消费者但不能被观察到的随机项
γ_j	顾客对商品 j 天生的品味
β	消费者对促销的敏感度

d_j	商品 j 是否促销宣传
ξ_j	研究者未观察到的变量引起的错误项
α_s	和具体商店有关的系数
$D'_s\tau$	和人口统计学因素有关的变量
δ	所有消费者之间相同的效用值
μ	因消费者而异的部分
I	消费者若选择购买这类物品, 则 $I=1$
C	消费者若选择购买这个物品, 则 $C=1$
\tilde{Q}_j	消费者对商品 j 的平均购买量

假设这里有一个研究者感兴趣的物品集合, 集合中每种商品的购买量为 $\mathbf{x} = (x_1, \dots, x_J)$, 它们对应的价格为 $\mathbf{p} = (p_1, \dots, p_J)$ 。同时我们有一个外部产品对应项 z , 表示该顾客总预算 y 中有一部分用来购买除这 J 种商品以外的产品, 在这里根据实际情况我们假设 z 是严格为正的, 因为消费者总是不会把自己所有积蓄花光。在计量经济学中, 在模型中引入 z 对于允许该大类商品销售总额变化非常重要, 而 z 也相当于离散选择模型中的不发生购买的选项。我们同时假设 ψ_j 为顾客心目中为商品 j 赋予的品质权重, ψ_z 为顾客心目中对外部产品赋予的品质权重。这样我们可以得到关于需求的直接效用 $u(\sum_{j=1}^J \psi_j x_j, \psi_z z)$, 且受到预算的限制 $\mathbf{x} \cdot \mathbf{p} + z = y$ 。于是根据效用最大化假说, 我们可以通过求解以下方程得到消费者需求:

$$\begin{aligned} \max_{x_1, \dots, x_J, z} u &= u^* \left(\sum_{j=1}^J \psi_j x_j, \psi_z z \right) \\ \text{s.t. } \sum_{j=1}^J p_j x_j + z &= y; \quad x_j \geq 0; \quad z > 0. \end{aligned} \quad (13)$$

由于 u 是递增的, 顾客总会把预算花光来至少购买一件商品, 又由于线性性, 在优化条件中仅有一种商品会最终被选择。根据以往的研究, 为了保持权重的非负性和能够解释一些随机影响因素, 我们假设顾客心目中的商品品质权重为如下形式:

$$\psi_j = \exp(\psi_j(\omega_j, \epsilon_j; w)) \quad (14)$$

$$\psi_z = \exp(\psi_z(\omega_z, \epsilon_z; w)) \quad (15)$$

其中 ω_j 表示商品 j 特性, w 表示消费者个人的偏好特性, $\epsilon = (\epsilon_1, \dots, \epsilon_j, \epsilon_z)$ 是能反应一些和商店、营销模式等相关的能影响消费者偏好权重的未被观测到的不确定项, 我们假设 ϵ_j 呈极值分布。我们认为顾客在发生购买之前能够观测到会影响其不确定项构成的商品特性、商店营销、品牌效应等因素。

具体地, 类似于 Nair 的模型, 我们在此将这个权重细化为:

$$\psi_j = \exp\left[\frac{1}{\alpha_s}(\gamma_j + \beta d_j + \xi_j + \epsilon_j)\right], \epsilon_j \sim EV(0, \mu) \quad (16)$$

$$\psi_z = \exp\left[\frac{\epsilon_z}{\alpha_s}\right], \epsilon_z \sim EV(D'_s \tau, \mu) \quad (17)$$

其中 γ_j 代表顾客天生的品味, β 代表消费者对促销的敏感度, d_j 代表是否促销宣传, ξ_j 表示研究者未观察到的变量引起的错误项, α_s 是一个和具体商店有关的系数, 而极值分布中的参数 $D'_s \tau$ 是和人口统计学因素有关的变量。

过去的研究已有许多对引入随机错误项的讨论。Akerberg 等人在 2002 年的研究中发现, 消费者选择模型中这些随机错误项包含了一些隐含的前提假设, 而这些假设可能会带来预测的偏差。这些随机错误项的维度是由产品数量决定的, 市场中每增加一个商品, 则会增加一个随机错误项的维度, 这会使商品未被观测到的特征维度有很少的冲突, 进而会夸大新产品进入市场时带来的福利效应, 也会带来商品之间的替代效应估计偏差。商品的价格弹性会强行与选择集大小相关^[77]。Berry、Levinsohn 和 Pakes (1995 年) 尝试在这个错误项中包含一些能得到的消费者异质性和产品特性数据, 减小错误项本身对模型的影响^[78]。Berry 和 Pakes (1999 年)^[79]以及 Bajari 和 Benkard (2001 年)^[80]在他们的研究中去除了这个随机项, 而仅采用单维未观测到的特性来表示这个不同产品针对不同顾客的空间。Akerberg 总结出现这种问题是因为随机项包含了一些过度的灵活性, 他在原本模型中仍包含了这个随机项, 但是把市场内的总商品数、每个嵌套内的产品数也纳入了模型考虑。

而 Nair（2005 年）在这个模型的品质权重中包含了和具体商店特性、顾客对商品自有品味、商品展示情况、顾客对商品营销敏感度、商家能观测到并可能会做出反应的影响因子、其他未被观测到的项目相关的变量。这些项能很好地解决顾客品味的异质性问题，引入了研究者未知但商家能观测到并可能会做出反应的影响因子，比如商品在货架上的摆放位置等，也能从一定程度上解决商品价格的内生性问题^[81]。

我们可以将消费者的直接效用重写为消费者群体的平均效用和因消费者而异的偏差以及效用随机项之和：

$$u = \delta + \mu + \varepsilon \quad (18)$$

其中 δ 表示所有消费者之间相同的效用值， μ 表示因消费者而异的部分。在此模型的权重假设下，为简化表示，我们令：

$$\delta_j = \gamma_j + \beta d_j + \xi_j - \alpha_s \ln p_j, \forall j \quad (19)$$

由于效用函数的直接表达式复杂且难得到，我们用间接效用函数来解决这个问题。间接效用函数是关于商品价格和预算的函数， $v(p, p_z, y) = u(\cdot)$ 。根据 Chiang（1991 年）的研究^[82]，我们使用同质 Translog 间接效用函数（Homothetic Translog indirect utility function）：

$$v(p, p_z, y) = \ln y - a \ln p - (1 - a) \ln p_z + \frac{b}{2} (\ln p)^2 + \frac{b}{2 (\ln p_z)^2} - b (\ln p) (\ln p_z) \quad (20)$$

其中 a 和 b 是两个正参数。结合 Roy 不等式，我们可以得到消费者对商品 k 的最优需求：

$$x_k^*(p, p_z, y) = - \frac{\partial v(p, p_z, y) / \partial p_k}{\partial v(p, p_z, y) / \partial y} \quad (21)$$

我们可以由拉格朗日算子和 KKT（Karush-Kuhn-Tucker）条件推得顾客会在以下这种情况选择购买非零数量的商品 k ：

$$\frac{p_k}{\psi_k} = \min_{j=1,\dots,J} \left(\frac{p_j}{\psi_j} \right) \text{ 且 } \frac{p_k}{\psi_k} \leq \frac{1}{\psi_k} e^{a/b} \quad (22)$$

这意味着当商品性价比超过某个临界值之后，消费者会选择购买价格品质之比最小的商品。由此我们可以推得消费者对于商品 k 的需求函数为：

$$\begin{aligned} x_k^*(p_k, \psi_k, \psi_z, y | c_k = 1, I = 1) &= x_k^*\left(\frac{p_k}{\psi_k}, \frac{1}{\psi_z}, y | c_k = 1, I = 1\right) \\ &= \frac{y}{p_k} \left[\alpha_1 - \alpha_3 \ln\left(\frac{p_k}{\psi_k}\right) + \alpha_3 \ln\left(\frac{1}{\psi_z}\right) \right] \\ &= \frac{y}{p_k} \left[\alpha_1 + \frac{\alpha_3}{\alpha_s} (\delta_k + \varepsilon_k - \varepsilon_z) \right] \end{aligned} \quad (23)$$

进一步地，可推得消费者在此类商品中发生购买行为以及选择购买商品 k 的概率：

$$\Pr(I = 0) = \Pr\left(\frac{1}{\psi_z} e^{a_1/a_3} < \frac{p_j}{\psi_j}, \forall j\right) = \frac{1}{1 + \sum_{j=1}^J e^{V_j}} \quad (24)$$

$$\Pr(I = 1) = 1 - \Pr(I = 0) = \frac{\sum_{j=1}^J e^{V_j}}{1 + \sum_{j=1}^J e^{V_j}} \quad (25)$$

$$\Pr(C_k = 1, I = 1) = \Pr(C_k = 1 | I = 1) \Pr(I = 1) = \frac{e^{V_k}}{1 + \sum_{j=1}^J e^{V_j}} \quad (26)$$

这里的 $V = \delta + \mu$ 。由此，积分得出其期望需求数量，并计算其市场份额，即消费者对商品 k 的平均购买需求 \tilde{Q}_k ：

$$\tilde{Q}_k = \int [\Pr(C_k = 1, I = 1) E_\epsilon(x)] \phi(\Lambda) \partial \Lambda \quad (27)$$

传统的 logit 需求系统并未考虑每次选择购买的数量，在传统 logit 需求模型中，消费者平均购买数量可以被写为：

$$\tilde{Q}_k = \int [\Pr(C_k = 1, I = 1)] \phi(\Lambda) \partial \Lambda \quad (28)$$

2.4 模型估计

根据 Nevo 在 2001 年^[83]与 Nair 在 2005 年^[84]的方法，我们采用模拟矩估计方法（method of simulated moments）进行模型系数估计。

上述商品效用中的随机错误项 ξ 可能包含了研究者不可见，但零售商可见的会影响消费者需求的因素。如果直接采用极大似然（maximum likelihood）方法进行估计，可能会产生内生性偏差。Besanko（1998 年）^[85]，Sudhir（2001 年）^[86]和 Chintagunta（2003 年）^[87]均在零售数据研究中讨论了这种偏差。

代入相关变量并进行计算后，对于商品 k ，我们可将 \tilde{Q}_k 表达式重写为：

$$\tilde{Q}_k = \int \frac{y}{\alpha_s p_k} \frac{e^{\delta_k + \mu_k}}{\sum_{j=1}^J e^{\delta_j + \mu_j}} \phi(\Lambda) \partial \Lambda \quad (29)$$

我们的模型估计过程分为以下几个主要步骤：

首先，对人口统计学变量等变量进行仿真随机抽取，给定搜索系数初始值，通过同质 logit 模型算出与消费者无关的效用 δ_k 的初始值。

第二，通过模型和自行给定的非线性部分系数初始值，进行蒙特卡洛仿真求得计算顾客期望购买数量的积分 \tilde{Q}_k 。

第三，通过给定非线性部分的系数给定值与已知数据，采用压缩映射（contraction mapping）方法求得 δ_k ，Nair 在他的估计方法中证明了压缩映射方法的有效性。

我们设在聚合数据集中消费者平均购买数量为 q_k ，则压缩映射函数 $g(\cdot)$ 可定义为：

$$g(\delta) = \delta + \ln(q) - \ln[\tilde{Q}(\delta)] \quad (30)$$

通过迭代求解，我们可以得到 δ 的一个理想收敛值。

第四，预测线性部分系数，生成有关非线性部分系数的 GMM (generalized method of moments) 函数。对于每个给定的 δ_k ，可以根据 $\xi = \delta(\theta) - xb + \alpha_s \ln(p)$ 建立矩条件，在这里有假设 $E[\xi X|X] = 0$ 。价格内生性问题产生意味着有条件 $E[\xi \ln(p) | \ln(p)] \neq 0$ 。在这里，我们引入工具变量解决内生性问题，使得 $E[\xi Z|Z] = 0$ 。

常用的工具变量包括产品特性（除去价格等可能有内生因素）、该公司提供的其他产品的相同特性之和、其他公司提供的产品的相同特性之和等。在假设城市或商店之间的 ξ 独立但可能随时间相关的情况下，该商品在其他城市或商店的价格也是有效的工具变量。

第五，对于每个给定的非线性系数序列，循环 2~4 步，最小化 GMM 目标函数，得到最优系数估计结果。

第3章 机器学习模型与模型集成

本章首先介绍了岭回归、随机森林和支持向量回归三种机器学习模型的理论背景，介绍了一个机器学习方法用于需求预测的具体数学模型，接着提出了模型集成的概念并分析了集成模型的优势，归纳了用于结合多种模型的集成学习方法，并在最后重点解释了本课题所使用的 stacking 集成方法。

3.1 岭回归

在处理复杂数据回归问题时，普通的线性回归会遇到预测精度和模型解释能力等方面的问题。岭回归 (ridge regression) ^[94] 实际上是一种改良的最小二乘估计法，它是一种有偏的回归估计方法，尤其适合共线性数据分析。岭回归在模型原来的目标函数上加上一个惩罚项，进行正则化，以解决数据具有多重共线性、对输入变量中的噪声敏感的问题。

岭回归对于一个线性模型，在原来的损失函数中加入惩罚项，其损失函数为如下形式：

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \{ \|y - X\beta\|^2 + \alpha \|\beta\|^2 \} \quad (31)$$

这里的 α 是平衡损失与正则项之间的系数， $\alpha \geq 0$ 。 α 的数值越大，则惩罚项的作用越明显； α 数值越小，惩罚项的作用就越弱。我们有：

$$\|y - X\beta\|^2 + \alpha \|\beta\|^2 = (y - X\beta)^T (y - X\beta) + \alpha \beta^T \beta \quad (32)$$

关于参数 β 求导之后，可得其解为：

$$\hat{\beta}^{ridge} = (X^T X + \alpha I)^{-1} X^T y \quad (33)$$

$\hat{\beta}^{ridge}$ 是有偏的。即使 X 不是满秩的, $(X^T X + \alpha I)$ 也是可逆的。 $\hat{\beta}^{ridge}$ 的估计较普通最小二乘估计法得到的结果总有更小的偏差, 因此岭回归可能得到更小的均方误差。总体来说, 岭回归倾向于给强关联的特征赋予相似的权值。

3.2 随机森林

Breiman 在 2001 年提出了随机森林 (random forest) 模型^[95], 随机森林是一种基于分类树 (classification tree) 或回归树 (regression tree) 的算法。随机森林扩展了使用预测变量集合进行预测的想法, 改变了分类或回归树的构建方式, 将随机性引入在节点级别上考虑的用于分割的变量集合, 随机化各分类或回归树, 再汇总分类或回归树的结果。在标准树中, 每个节点使用所有变量之间的最佳分割进行分割。而在随机森林中, 每个节点使用在该节点随机选择的预测变量的子集中的最佳分割。

随机森林的训练采用 bootstrap aggregating (缩写为 bagging)^[96]的方法。给定训练集 $X = x_1, \dots, x_n$ 和响应集 $Y = y_1, \dots, y_n$ 。Bagging 方法随机选取 B 次有放回的抽样并选取一个随机的特征子集来拟合相应的回归树。对于 $b = 1, \dots, B$, 从 X 和 Y 中选取的样本为 X_b 和 Y_b , 基于此训练的回归树为 f_b 。则对于样本 x' , 预测值可以由平均这 B 个回归树的预测值来得到:

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x') \quad (34)$$

这种 bagging 的方法可以减少模型的方差而不增加其偏差, 虽然单个回归树可能会对噪声数据高度敏感, 但如果这些回归树不是相关的, 使用它们的平均值进行估计就可以降低模型对数据的敏感性。此处的有放回抽样可以给每个树提供不同的训练样本, 从而减少它们的相关度。对预测的不确定性估计可以由每个回归树对于 x' 的预测标准偏差得到, 为:

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(x') - \hat{f})^2}{B - 1}} \quad (35)$$

随机森林只含有两个参数：每个随机子集中每个节点的变量数目、森林中树的数目。恰当选取参数能获得更好的预测或回归效果，也能获得更优的收敛速度。在调参时，要取得良好性能，所需的树木数量随着预测因子数量的增加而增加。确定需要多少树木的最好方法是将森林的预测与一个森林子集的预测进行比较，当子集有和全森林一样的效果时，就意味着树木数量足够。足够数量的树是无偏的，但是可能有较大的方差，最好取平均值。在实际代码应用中，我通过网格搜索方式运用交叉验证评价寻找到这两个最佳参数。

3.3 支持向量机回归

支持向量机回归 (support vector regression, SVR) ^[98] 是一种带有惩罚的回归，它利用有限的样本信息，权衡了回归模型的复杂度与学习能力。支持向量机希望把各种问题转化成线性问题，由于需要分类或回归的数据常常在原空间非线性可分，与降维思想相反，支持向量机利用非线性映射函数将样本映射到高维线性空间，并在此空间寻找满足分类要求的最优分类器或满足回归要求的最优回归函数。对分类器而言，直观上讲，求得的最优超平面离最近的各类训练数据点有最大距离，因为越大的距离表示着分类器具有越低的错误率。支持向量机可以选择合适的核函数，低维输入空间的核函数的类型和参数可以间接决定样本在高维空间的分布，从而影响支持向量机的性能^[97]。样本在高维空间的点积运算也由核函数实现。

构造 SVR 时，需要选择合适的损失函数。SVR 的参数包括核参数、损失函数中的 ε 以及误差惩罚参数 C 。常用的参数确定方法包括经验确定和网格搜索，实际运用中常将这两种方法结合，从准确率、学习时间、均方差 (mean squared error, MSE) 来判断。SVR 的目标函数可表述如下：

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & \text{subject to } \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad \text{for } i = 1, \dots, N \end{aligned} \quad (36)$$

其中 $\varepsilon \geq 0$ ，表示 SVR 预测值与实际值最大的差距，在加入惩罚项之后，允许部分样本落在 ε 之外，而每个训练样本各自的 ξ_i, ξ_i^* 即用以决定该样本是否可以落在 ε 之外。参数 C 用以调整训练模型的拟合度。

由于 SVR 问题是一个凸优化问题，我们可以得到其 Lagrange 对偶函数及相应优化条件：

$$\begin{aligned} \max L_D = & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j < x_i, x_j > \\ & \text{subject to } 0 \leq \alpha_i \leq C \\ & \text{and } \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \quad (37)$$

若我们对 SVR 引入核函数 $K(x, x') = < h(x), h(x') >$ ，则此时 Lagrange 对偶函数及对应解为：

$$L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j < h(x_i), h(x_j) > \quad (38)$$

$$\hat{y}(x) = \sum_{i \in S} \hat{\alpha}_i y_i < h(x_i), h(x_j) > = \sum_{i \in S} \hat{\alpha}_i y_i K(x_i, x_j) \quad (39)$$

3.4 机器学习模型实例

对于消费者品味异质性，除了在第 2 章提到的通过带有各种随机参数的选择模型来描述消费者品味的异质性以外，以往也有多个采用机器学习领域常用的正则化方法来处理消费者异质性的研究。在介绍了三种有代表性的机器学习方法后，本文接下来介绍 Evgeniou 的岭回归机器学习模型^[88]和 Cui 和 Curry^[89]、Levis 和 Papageorgiou^[90]的支持向量机机器学习模型研究实例。

我们假设现在有 I 个消费者，消费者序数编号用 $i \in \{1, 2, \dots, I\}$ 表示。每个消费者对 J 个商品进行评价（ J 对于不同消费者可能不同），用行向量 \mathbf{x}_{ij} 表示， $j \in \{1, 2, \dots, J\}$ 。我们假设成分效用值的个数为 p ，即每个 \mathbf{x}_{ij} 有 p 列。由此我们可以用 $J \times p$ 维度的矩阵 X_i 表示消费者 i 的设计矩阵（每行表示对一个商品的评价）；

用 $p \times 1$ 维度的列向量 \mathbf{w}_i 表示消费者 i 的成分效用值；用 $J \times 1$ 维度的列向量 \mathbf{Y}_i 表示消费者 i 的评价偏好（ y_{ij} 表示该消费者对商品 j 的评价情况）。为了简洁，我们引入随机项 ϵ_{ij} ，采用效用函数的标准假设：

$$U_i(\mathbf{x}_{ij}) = \mathbf{x}_{ij}\mathbf{w}_i + \epsilon_{ij} \quad (40)$$

在这里，我们建立岭回归模型（ridge regression, RR）。关于这种模型及其相关的拓展已经有了许多详尽的研究，如 Tikhonov 和 Arsenin 在 1977 年的研究^[91]，Wahba 在 1990 年^[92]，Hastie 在 2003 年的研究^[93]。岭回归模型通过最小化一个关于 \mathbf{w}_i 的凸损失函数来估计出个人水平的成分效用值，这个损失函数如下：

$$\min_{\mathbf{w}_i} \frac{1}{\gamma} \sum_{j=1}^J (y_{ij} - \mathbf{x}_{ij}\mathbf{w}_i)^2 + \|\mathbf{w}_i\|^2 \quad (41)$$

如前面介绍的，交叉检验经常被用于确定机器学习模型中的参数，这个损失函数通常使用一个由交叉检验确定的正的权重 γ 来参数化。这个损失函数的第一部分衡量了估计的效用和观察到的消费者评级之间的相符度。对于一个固定的 γ 而言，这个可以被解释为对应一个均值为 0 方差为 γ 的正态错误项似然值的对数（如 Hastie 等人，2003 年）。第二个部分，控制了成分效用值 \mathbf{w}_i 的复杂度（如 Vapnik，1998 年；Hastie 等人，2003 年）。正参数 γ 可以设定控制拟合度和收缩度的权衡，它的值可由交叉检验得到。注意到当 $\gamma \rightarrow 0$ 时我们可以得到 OLS 解 $\mathbf{w}_i = (X_i^T X_i)^{-1} X_i^T Y_i$ ，其中当 $X_i^T X_i$ 不可逆时，采用它的伪逆矩阵（Hastie et al., 2003 年）。

岭回归的损失函数还可以通过用其他种类的错误函数替换第一部分的错误项平方来得到。只要这第一部分误差函数是凸的，整个损失函数就会是凸的。比如对于有关消费者选择的数据来说，如果我们采用 logistic 错误项，就可以得到 SVM 方法（Vapnik，1998 年）。Cui 和 Curry（2005 年），Evgeniou（2005 年）就将这种方法引入到了市场营销领域。模型接下来对损失函数进行拓展，来解释消费者异质性：

$$\min_{\mathbf{w}_i} \frac{1}{\gamma} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mathbf{x}_{ij} \mathbf{w}_i)^2 + \sum_{i=1}^I (\mathbf{w}_i - \mathbf{w}_0)^T D^{-1} (\mathbf{w}_i - \mathbf{w}_0) \quad (42)$$

使得 D 是一个半正定矩阵，矩阵的迹为 1。和之前的损失函数不同的是，这个损失函数涉及到凸优化过程，联合估计所有的成分效用值。和之前使成分效用缩小到 0 不同，这里通过最小化 $(\mathbf{w}_i - \mathbf{w}_0)^T D^{-1} (\mathbf{w}_i - \mathbf{w}_0)$ 使成分效用缩小到 \mathbf{w}_0 ，可证得其为样本平均数。矩阵 D 和成分效用的协方差矩阵有关。和之前同样地，正参数 γ 可以设定控制拟合度和收缩度的权衡，它的值越大，消费者的偏好之间越同质。

于是，基于决策者做出选择的数据，可以得到以下模型：

$$\begin{aligned} \gamma^* &= \arg \min_{\gamma} \text{cross-validation}(\gamma), (\{\mathbf{w}_i^*\}, \mathbf{w}_0^*, D^*) \\ &= \arg \min_{\{\mathbf{w}_i\}, \mathbf{w}_0, D} - \frac{1}{\gamma^*} \sum_{i=1}^I \sum_{j=1}^J \frac{\log(e^{\mathbf{x}_{ij}(q^*) \mathbf{w}_i})}{\sum_{q=1}^Q e^{\mathbf{x}_{ijq} \mathbf{w}_i}} \\ &\quad + \sum_{i=1}^I (\mathbf{w}_i - \mathbf{w}_0)^T D^{-1} (\mathbf{w}_i - \mathbf{w}_0) \end{aligned} \quad (43)$$

使得 D 是一个半正定矩阵，矩阵的迹为 1。其中 \mathbf{x}_{ijq} 对应第 q 个选择， $\mathbf{x}_{ij}(q^*)$ 对应决策者的最终选择。 J 代表需要选择的问题， Q 代表每个问题的选项数目（不同问题和回答者之间的选项个数可以不同）。这种方法理论上和分层贝叶斯理论具有一定相似性。

除了岭回归模型以外，Cui 和 Curry（2005 年），Levis 和 Papageorgiou（2005 年）还运用了支持向量机模型来预测分析消费者需求。假设有以下训练数据：

$$\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\} \subset R^Z \times R \quad (44)$$

其中 \mathbf{x}_t 是 t 时刻的输入向量， y_t 是相应的顾客需求。我们的目标是找到这个回归函数 $F(\mathbf{x}_t)$ ：

$$F(\mathbf{x}_t) = \mathbf{w}^T \mathbf{x}_t + \beta \quad \mathbf{w}, \mathbf{x}_t \in R^Z, \beta \in R \quad (45)$$

根据统计学习理论，为了获得有推广意义的模型，我们需要控制模型复杂度和训练的误差容忍（Chalimourda 等人，2004 年）。在这个模型里的一种方法是最小化欧几里德项 $\|w\|$ 。同时，这个回归函数也需要保持能够和拟合实际情况相符合的复杂度，因此我们引入对 ε 不敏感的损失函数 $|\xi|_\varepsilon$ ：

$$|\xi|_\varepsilon = \max(0, |F(x_t) - y_t| - \varepsilon) \quad (46)$$

这样可以保证比 ε 小的错误项不被纳入考虑，同时对比 ε 大的偏差进行处罚。于是我们得到以下优化条件：

$$\min_{w, \beta, \xi_t, \xi_t^*} \frac{1}{2} \|w\|^2 + C \cdot \sum_{t=1}^N (\xi_t + \xi_t^*) \quad (47)$$

subject to:

$$\begin{aligned} y_t - w^T x_t - \beta &\leq \varepsilon + \xi_t \quad \forall t = 1, \dots, N \\ w^T x_t + \beta - y_t &\leq \varepsilon + \xi_t^* \quad \forall t = 1, \dots, N \\ \xi_t &\geq 0 \quad \forall t = 1, \dots, N \\ \xi_t^* &\geq 0 \quad \forall t = 1, \dots, N \end{aligned}$$

根据拉格朗日函数和 KKT 条件，令 $K(x_i, x_j) = \Phi(x_i^T) \cdot \Phi(x_j)$ ，我们可以根据以下两个优化方程组求解得到非线性支持向量回归：

$$\begin{aligned} \max_{\lambda_t, \lambda_t^*} & -\frac{1}{2} \sum_{t=1}^N \sum_{t'=1}^N (\lambda_t - \lambda_t^*)(\lambda_{t'} - \lambda_{t'}^*) K(x_i, x_j) - \varepsilon \sum_{t=1}^N (\lambda_t + \lambda_t^*) \\ & + \sum_{t=1}^N y_t (\lambda_{t'} - \lambda_{t'}^*) \end{aligned} \quad (48)$$

subject to:

$$\begin{aligned} \sum_{t=1}^N (\lambda_t - \lambda_t^*) &= 0 \\ 0 &\leq \lambda_t \leq C \quad \forall t = 1, \dots, N \\ 0 &\leq \lambda_t^* \leq C \quad \forall t = 1, \dots, N \end{aligned}$$

$$\min_{\beta, \xi_t, \xi_t^*} \sum_{t=1}^N (\xi_t + \xi_t^*) \quad (49)$$

subject to:

$$y_t - \sum_{t'=1}^N (\lambda_{t'} - \lambda_{t'}^*) K(x_t, x_{t'}) - \beta \leq \varepsilon + \xi_t \quad \forall t = 1, \dots, N$$

$$\sum_{t'=1}^N (\lambda_{t'} - \lambda_{t'}^*) K(x_t, x_{t'}) + \beta - y_t \leq \varepsilon + \xi_t \quad \forall t = 1, \dots, N$$

$$\xi_t \geq 0 \quad \forall t = 1, \dots, N$$

$$\xi_t^* \geq 0 \quad \forall t = 1, \dots, N$$

我们还可以在这个基础模型上进行延伸拓展，采用不同的核方法或加入各种成分效用之间的交互项，取得更好的模型表现。

3.5 模型集成

和单一预测模型不同，组合预测往往可以取得更优的预测效果。首先，有时很难识别出哪种模型能带来最优预测效果。此外，预测模型为数据生成过程常常做了简化，而实际情况可能比模型假设的要复杂得多。因此，单个模型预测不太可能涵盖所有模型。每种方法利用的数据信息均有差别，即使某种模型总是产生比所有其他预测产生更低的预期损失，只要该模型的预测误差与其他竞争模型的预测误差不完全相关，总可以通过将部分权重分配给其他预测模型来达到模型多样化。这种多样化很可能会取得比最优单一模型更好的表现。模型集成能根据不同的情景调整组合方式，取得各情景下的最优表现。

集成学习 (ensemble learning)^[99]指结合多个学习方法来获得更好的预测结果。与统计力学中统计集成的集合通常无限不同，集成学习仅包含具体的有限的备选模型集合，但通常允许在这些备选模型中存在灵活的组合方式。Bates 和 Granger 在 1969 年提出了一种基于模型预测误差的预测组合方式，其主要思想是将许多单一预测模型进行加权组合^[100]。他们证明，多种无偏单项预测组合起来可以获得模型预测信息。此后组合预测领域得到了持续发展。Granger 和 Newbold (1973 年)^[101]与 Diebold 和 Pauly (1987 年)^[102]分别提出了基于方差协方差分析和回归架构的随时间动态变化的模型结合框架，实证经验表明这种动态集成方式往往比静态架构获得更好的表现。

组合预测的关键在于各模型权重的确定，而采用不同的结合原则就会有不同的组合预测模型。通常情况下，预测效果更准确的模型可以获得更大的权重，预测误差项对各单一模型的权重影响很大，而结合权重也会影响每次预测结果之间的关联度。模型集成可以用到国内生产总值、股票回报、城市人口、政策风险等各领域的预测评估中。加权组合预测模型的权重选择方法有算术平均法、方差倒数法等直接计算方法，此外还有基于模型本身的集中最优组合预测权系数确定方法。Bajari 在 2015 年运用了线性回归的方式确定各子模型在结合模型中所占的权重^[103]。

常用的集成学习方法包括 Bagging、Boosting 与 Stacking，它们之间的主要特点比较见下表：

表 3.1 常用集成学习方法比较^[104]

	Bagging	Boosting	Stacking
子集划分方式	随机	给被错误分类的样本更高权重	多样的
集成目标	最小化方差	增强解释力	前两者结合
使用方法	随机子空间	梯度下降	混合
结合函数	加权平均	加权主要投票	logistic 回归

本文 3.2 节所介绍的随机森林其实就是一种 bagging 方法，而 boosting 算法侧重将弱学习器转化为强学习器。较其他两种方法来讲，stacking 方法的特点是可以改进模型预测效果，由于我们这里研究消费者需求预测模型的预测效果，于是采用 stacking 方法进行集成。

接下来介绍本研究用到的 stacking 方法^[105]。Stacking 法，有时也称为 stacked generalization，其核心思想是训练一种学习算法来集成其他学习算法的预测结果，适用于集成不同类型的学习算法。它引入了元学习（meta learning）的概念。首先，使用可用数据训练所有单一算法，然后训练一种组合器算法以使用其他算法的所有预测作为附加输入来进行最终预测。基本算法所得的结果在这里构成一个元数据集（meta dataset）。逻辑回归和线性回归模型通常用作组合器。为了防止组合器再次回归带来的过拟合问题，常常使用交叉验证或留一法，采用初级学习器模型未使用的样本来产生组合器学习的训练样本。

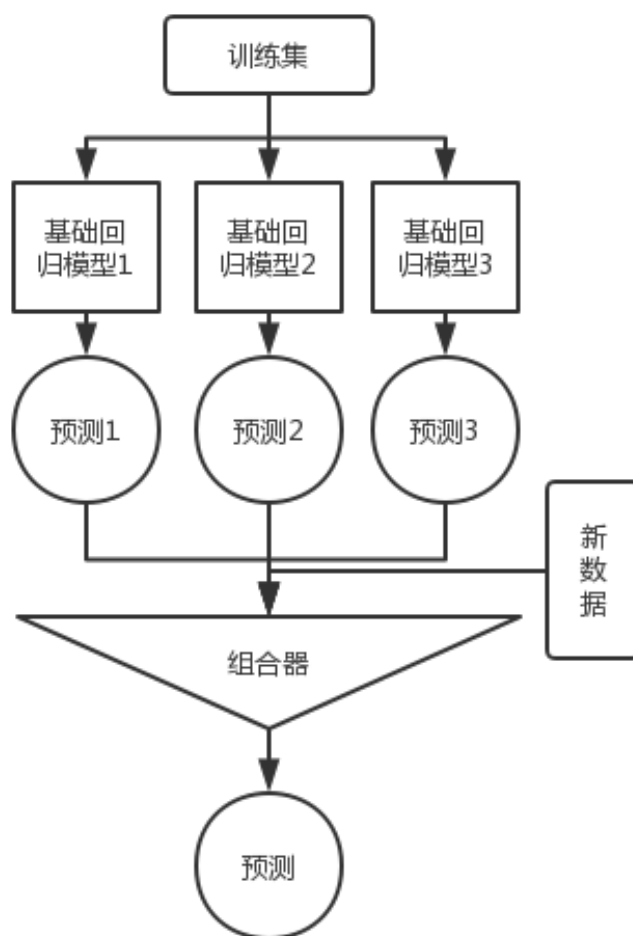


图 3.1 Stacking 方法应用原理图

Stacking 方法的应用原理如上图所示。它得到的模型通常可以比任何单一训练模型拥有更好的性能，目前已被成功用于监督学习任务（回归、分类和远程学习等）和无监督学习任务（密度估计等）。

第 4 章 数值案例

本章将基于离散选择模型与机器学习模型，介绍一个公开的即食谷物市场数据集，在考虑消费者偏好异质性和产品价格内生性的情况下建立并估计模型，对模型参数所反映的消费者行为进行解释，并比较与分析了各种机器学习模型和集成模型的预测表现差异。

4.1 数据介绍

我们在这里关注即食谷物行业。即食谷物行业的主要特点包括：品牌高度集中、具有高利润率、广告销售比率大，以及经常有大量新产品推出。以前的研究人员发现即食谷物行业是一个接近共谋定价和具有强烈的非价格竞争的行业，但此结论仍待验证。在即食谷物行业中，有主导地位的少数公司旗下有一系列不同的产品，并且可以影响大家对商品品质的印象。对产品进行需求分析，有利于之后结合更多经济学理论来细分影响利润率的因素。

我们使用 Nevo 在 2000 年发表的文章中使用的数据集进行案例分析^[106]。该数据集包括 24 种不同麦片商品的商品特征及其在 47 个城市、2 个季度的价格与市场份额数据，是由一个需求供给模型生成的。这里我们把每个城市每个季度看作一个市场。这个数据集经过了 Nevo 的简化，侧重于用于展示方法实现原理和一些基础分析。生成数据集的需求模型采用 Nevo 描述的 logit 模型，其商品边际价格和相关参数由 Nevo 在 2000 年另一个研究^[107]估计的系数确定；供给模型采用标准多产品公司的差异化产品 Bertrand 模型。其中市场份额数据已经由 Nevo 将谷物商品的重量（pounds）处理成为了份数（servings）。

对于市场总份额，Nevo 假定每人每天均消耗 1 份即食谷物。而我们考虑消费者实际的消费预算具有随机性，根据其他研究者的研究方法，设定每个市场中消费者的总预算为 $U \sim (2, 5)$ 的均匀分布。相应地，每种商品在每个市场的购买量由（市场份额 * 市场消费者总预算 / 市场所有商品加权价格）计算得到。数据中各产品包含价格、市场份额、所属公司编号、城市、年份和季度信息。除此之外，数据中含有两种商品特性，包括：糖分含量、是否在牛奶中成糊状（虚拟变量，如果在牛奶中成糊状则为 1）。人口统计数据来自当时的人口调查（Current Population

Survey)，包括收入的对数、收入平方的对数、年龄、是否是孩子（若该消费者年龄小于 16，则等于 1）。未观测到的人口统计学变量，从标准正态分布中抽样得到，其中对每个市场随机抽取了 20 个人的特性。此外，各市场中的每种商品有 20 个和价格变量有关的工具变量。

Nevo 在这里采用的工具变量可以分成两大类：

第一，类似于 Hausman（1996 年）^[108]运用的方法，假设在控制和商品有关的均值和人口统计变量的情况下，和城市特别相关的各数值在不同城市间相互独立，而允许在该城市内部相关。在这种假设下，这个产品在其他城市各季度的价格是有效的工具变量。

第二，考虑到当市场有全国性或地方性的需求冲击时，第一类工具变量不再满足城市间独立性假设，也使用以下第二类变量。假设商品特性在决策空间具有外生性，即在消费者透露产品偏好之前就已存在，可以使用观测到的产品特性、同公司其他产品的相同特性之和、其他公司产品的相同特性之和。

首先计算商品价格和其市场份额的相关性，得到其相关性系数为-0.14，意味着各商品的市场份额一般会随市场价格升高而降低，符合通常规律。接着对数据整体做描述性统计。关于数据集市场份额、价格、糖度、是否糊状特性的描述性统计表，及所用到的 20 个工具变量的箱线图如下：

表 4.1 数据集产品特性描述性统计

	市场份额	价格	糖度	是否糊状
观察数	2256	2256	2256	2256
均值	0.020	0.126	8.625	0.333
标准差	0.026	0.029	5.788	0.472
最小值	0.000	0.045	0.000	0.000
25%	0.005	0.105	3.000	0.000
50%	0.011	0.124	8.500	0.000
75%	0.025	0.143	13.250	1.000
最大值	0.447	0.226	20.000	1.000

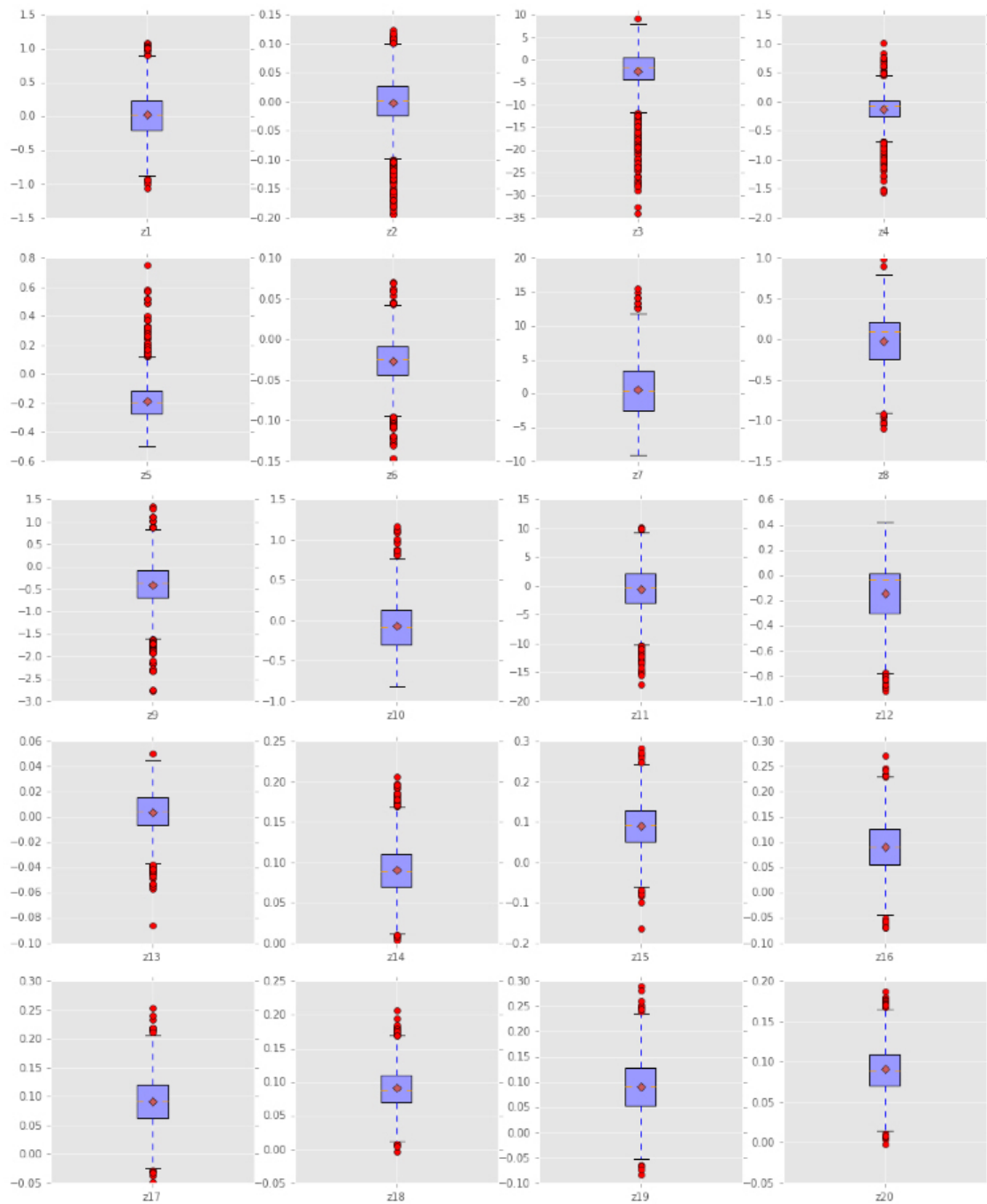


图 4.1 各工具变量箱线图

4.2 模型结果

根据之前推导的模型，我采用 Python 对前文所述离散选择模型与机器学习模型进行了编程实现。



图 4.2 程序主要功能结构块

我们根据相应数据获得了选择模型的模型估计结果。首先，对于 Nevo 的经典 logit 模型估计结果如下表。Nevo 的模型非显性地假设了对于每件商品，消费者只会购买一份。

表 4.2 经典 logit 模型估计结果

	均值 β	标准偏差 σ	和人口统计学变量的交互作用			
			收入	收入平方	年龄	是否小孩
常数	-2.01 (-0.33)	0.56 (-0.16)	2.29 (-1.21)	- (-0.63)	1.28 (-0.63)	- (-0.63)
价格	-62.73 (-14.80)	3.31 (-1.34)	588.33 (-270.44)	-30.19 (-14.10)	- (-0.63)	11.06 (-4.12)
糖度	0.12 (-0.02)	-0.01 (-0.01)	-0.39 (-0.12)	- (-0.03)	0.05 (-0.03)	- (-0.03)
是否糊状	0.50 (-0.20)	0.09 (-0.19)	0.75 (-0.80)	- (-0.67)	-1.35 (-0.67)	- (-0.67)
GMM 函数值			4.56			

从以上结果可以看出，由均值这一栏估计结果，对大多数消费者而言，糖度和是否糊状的系数为正，意味着更多的糖会提高这种麦片的消费者效用，而大多数消费者会偏好可溶于牛奶的即食谷物。价格的系数为负，说明消费者更偏好价格相对更低廉的商品。

标准偏差这一栏表示未观测到的人口变量带来的影响，除价格对应的标准偏差估计值以外其系数值较小（远小于 1），表明它对除价格外的产品特性给消费者带来的影响效果非常小。这意味着在这个案例里，模型中所包含的人口统计学变量基本上可以解释绝大部分模型系数的异质性。

后四栏表示这几种产品特性和人口统计学变量的交互作用。对于糖度这一行，由收入和年龄的系数符号可知糖度的边际值随着消费者收入增加而降低，随着年龄增加而增加。对于价格这一栏，可见标准偏差系数值的估计较后四栏的估计值很小，说明大部分消费者关于价格的偏好异质性由后面四栏的人口统计学变量来解释。价格的均值为负，表明更高的价格会降低消费者对产品效用的评价，由后面的交互项系数也可以看出小孩和收入较高的家庭对价格会相对更加不敏感。

如第 2 章 2.3 节模型所述，考虑到对某一种产品，消费者可能会购买多件，我们在 Nair 的模型基础上对 Nevo 的模型进行了改进。改进后的考虑消费者购买量的模型估计结果如下表：

表 4.3 考虑消费者购买量的 logit 模型估计结果

	均值 β	标准偏差 σ	和人口统计学变量的交互作用			
			收入	收入平方	年龄	是否小孩
常数	-0.88	0.26	1.58	-	0.48	-
	(0.43)	(0.13)	(1.77)	-	(0.46)	-
价格	-53.39	5.04	712.51	-38.11	-	7.57
	(17.55)	(2.03)	(406.02)	(22.02)	-	(4.80)
糖度	-0.07	0.04	-0.17	-	0.06	-
	(0.03)	(0.02)	(0.10)	-	(0.05)	-
是否糊状	1.92	0.25	-0.78	-	-0.73	-
	(0.30)	(0.18)	(2.20)	-	(0.78)	-
GMM 函数值			13.23			

从以上结果可以看出，估计结果与之前不考虑消费者一次性购买多种商品的模型得到的消费者行为解释性倾向大体相同。但常数、价格、糖度的系数绝对值估计都比之前更小，说明之前的模型系统性夸大了这些因素对消费者需求的影响作用。而这里估计的是否糊状系数绝对值较上一模型大，说明消费者对可溶于牛奶的麦片的偏好度可能较之前估计的更高。收入越高和年龄越大的消费者越倾向于喜欢不溶于牛奶的即食谷物。

接下来，调用 `scikit-learn` 库^[109]进行回归模型训练。我运用了岭回归、随机森林和支持向量回归这三种机器学习模型对数据进行回归，并运用线性结合和 `stacking` 两种方式进行模型集成。在训练模型之前，对数据进行了正则化处理，并使用主成分分析法 (principal component analysis)^[110]进行了维度缩减，保留了 90% 的方差解释性。对于各模型最优参数，我通过经验确定初始值，并通过网格搜索方法确定最优参数。为了避免有的模型（比如线性回归模型）可能会有很好的样本内拟合度而使得模型集成时产生偏差，我们将数据集划分为训练集、验证集和测试集三部分进行交叉验证，模型的权重在交叉集中产生。

所得各类模型的表现总结如下表：

表 4.4 机器学习模型及集成模型比较

	均方误差 MSE	线性模型权重
随机森林	0.001257	0.618224
岭回归	0.296318	0.338181
SVR	0.001619	-0.035238
线性集成模型	0.000289	
stacking 集成模型	0.000926	

均方误差是衡量回归模型表现的重要方法。由上表结果可以比较看到，就单个模型而言，随机森林和支持向量回归的回归效果优于岭回归。而从线性模型的权重分配结果可以看到，在模型的简单线性集成时，也带来了过拟合问题，使得在有随机森林这样较强的模型存在时，无约束的线性回归模型会给这个强模型赋予较大的权值，却给另外相似但较弱的模型赋上负权值。而此时这个较弱的模型的预测效果可能与真实结果并不是负相关的。从这一点上来看，stacking 集成模型同样能带来很好的集成效果，避免了线性集成模型的过拟合问题，是更理想的选择。

在以后的研究中，可以考虑选择模型和机器学习模型的混合模型。根据 Cui 和 Curry 的研究^[11]，选择模型和机器学习模型的结构能力差别很大。一方面，选择模型较机器学习模型有更好的系数解释性。另一方面，机器学习模型跳过函数估计步骤，直接计算回归结果，即使样本较少的情况下也可以有很鲁棒的预测效果。而目前研究者们希望获得结构准确、有解释性且预测效果好的模型。这里集成学习的方法，也可以用于之后离散选择模型和机器学习模型的模型集成。

在这个案例中，模型的显著系数数量还较为有限，在实际情况中可以通过使用更多的数据、增加仿真次数、改进仿真模型来改进这些模型的估计效果。此处我们仅在选取工具变量时考虑了供应侧，跟据 BLP 的研究，可以在模型中同时考虑需求侧和供应侧来获得更有效的模型估计。根据 Bajari 的研究，在选取商品特征时还可以运用自然语言处理方法，从消费者对商品的评论中挖掘得到更多消费者关心的商品特征。

第 5 章 总结与展望

如今，消费者日益看重购物过程中的自然交互体验，而各种电商平台、移动应用平台中的智能推荐与消费者需求预测愈发重要。本研究通过对消费者需求预测模型进行理论综述推导，并结合实例进行分析，希望为以后提出具有更强解释性、更高效精准的消费者模型提供思路。本章简述了本课题的研究脉络，点明研究创新点，并从模型结合、离散选择模型拓展、数据拓展三方面进行了未来研究展望。

5.1 研究总结

本文在综述了利用聚合数据估计消费者需求的领域研究及面临的挑战之后，推导了一种离散选择模型，并分别采用该离散选择模型和三种机器学习模型及它们的集成模型来估计了一个商品水平的即食谷物需求系统。该离散选择模型引入了随机系数和工具变量，能处理消费者偏好异质性和商品价格内生性问题，并和传统 logit 模型不同，在模型构建中考虑了消费者对某种产品可能发生多件购买的情况。在该选择模型的估计中，我运用了数据的面板特性，并假设各城市之间的需求冲击互不相关，使得选取的工具变量有效。另一方面，对于机器学习预测，我使用了岭回归、随机森林、支持向量机回归三种机器学习模型，并运用网格搜索确定了单个模型的最优系数。对于模型的集成，我们采用了两种经典机器学习模型集成的方法，运用 stacking 集成方法取得了很好的模型改进效果，而发现简单的模型线性组合却可能带来过拟合等问题。本文最后提出，在之后的研究中，还可考虑通过模型集成方法直接结合离散选择模型和机器学习模型的估计预测结果。

5.2 创新点

本课题针对离散选择模型与机器学习模型在消费者需求预测领域做了综述性梳理和部分模型的理论推导，并进行了编程实现与实证研究，在离散选择模型改进、机器学习模型集成之后改善了模型表现。

在我调研的文献中，在消费者需求预测领域采用机器学习的研究还比较少。现有的工作大多是单独使用机器学习和选择模型进行研究，或是在选择模型中运用机器学习方法以便模型估计，而暂时没有将两者结合起来推出统一算法的研究。本课题同时对这两领域方法进行了理论推导与实例研究，并提出运用集成学习方法进行模型结合，这是本课题最大的创新点。

5.3 研究展望

由于各产业性质的复杂性、消费者和市场间各种因素的复杂交互与不确定性，很难找到一个详尽的模型完整描述所有情况。数据的高维度、选择集的庞大、一些商品和销售者消费者特征的不可区分或不可识别，会使离散选择模型的应用受到限制，而机器学习能更适合这种复杂关系中的预测工作，可以和离散选择模型优势互补。目前相关领域的理论研究仍然缺乏。

此外，这里的离散选择模型还有更多延伸点。如 Nevo 所提到，我们可以加入更多供应侧的数据信息^[112]。除本课题讨论的离散选择模型以外，根据具体的数据和市场及消费者情况，还可以考虑多层次模型（multilevel demand model）、动态模型（dynamic model）等。

数据方面，可以使用更多、维度更丰富、时空粒度更细致的数据进行研究。在商品特征提取方面，除传统方法外还可关注消费者对商品的评价，通过自然语言处理提取商品特性。此外，还可以结合消费者自我汇报等数据，研究如囤货等消费者行为和心理，探究消费者社会关系互动对其购买行为的影响。

插图索引

图 1.1	本课题研究路线框图	9
图 3.1	Stacking 方法应用原理图	32
图 4.1	各工具变量箱线图	35
图 4.2	程序主要功能结构块	36

表格索引

表 1.1	各类离散选择模型概览	5
表 1.2	聚合型消费者需求数据集类别定义及相关研究	6
表 1.3	使用 BLP 估计聚合数据的部分文章总结.....	7
表 2.1	模型涉及变量及相应描述	16
表 3.1	常用集成学习方法比较	31
表 4.1	数据集产品特性描述性统计	34
表 4.2	经典 logit 模型估计结果.....	37
表 4.3	改进 logit 模型估计结果.....	38
表 4.4	机器学习模型及集成模型比较	39

参考文献

- [1] Chintagunta, P. K., & Nair, H. S. (2011). Structural workshop paper—Discrete-choice models of consumer demand in marketing. *Marketing Science*, 30(6), 977-996.
- [2] Chiang, J. (1991). A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science*, 10(4), 297-315.
- [3] Chiang, J., & Lee, L. F. (1992). Discrete/continuous models of consumer demand with binding nonnegativity constraints. *Journal of Econometrics*, 54(1-3), 79-93.
- [4] Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, 9(4), 513-548.
- [5] McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
- [6] Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89, 57-78.
- [7] Train, K. E. (2009). Discrete choice methods with simulation. *Cambridge university press*.
- [8] Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89, 57-78.
- [9] Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, 26(6), 805-818.
- [10] Mishra, V. K., Natarajan, K., Padmanabhan, D., Teo, C. P., & Li, X. (2014). On theoretical and empirical aspects of marginal distribution choice models. *Management Science*, 60(6), 1511-1531.
- [11] Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242-262.
- [12] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.
- [13] Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2), 307-342.
- [14] Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4), 705-729.
- [15] Goolsbee, A., & Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica*, 72(2), 351-381.

- [16] Chintagunta, P., Dubé, J. P., & Goh, K. Y. (2005). Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Science*, 51(5), 832-849.
- [17] Wales, T. J., & Woodland, A. D. (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21(3), 263-285.
- [18] Lee, L. F., & Pitt, M. M. (1986). Microeconomic demand system with binding nonnegativity constraints: the dual approach. *Econometrica: Journal of the Econometric Society*, 1237-1242.
- [19] King, M. A. (1980). An econometric model of tenure choice and demand for housing as a joint decision. *Journal of public Economics*, 14(2), 137-159.
- [20] Dubin, J. A., & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 345-362.
- [21] Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing research*, 342-355.
- [22] Chiang, J., & Lee, L. F. (1992). Discrete/continuous models of consumer demand with binding nonnegativity constraints. *Journal of Econometrics*, 54(1-3), 79-93.
- [23] Krishnamurthi, L., & Raj, S. P. (1988). A model of brand choice and purchase quantity price sensitivities. *Marketing Science*, 7(1), 1-20.
- [24] Chiang, J. (1991). A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science*, 10(4), 297-315.
- [25] Chintagunta, P. K. (1993). Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12(2), 184-208.
- [26] Besanko, D., Gupta, S., & Jain, D. (1998). Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science*, 44(11-part-1), 1533-1547.
- [27] Sudhir, K. (2001). Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Science*, 20(3), 244-264.
- [28] Nair, H., Dubé, J. P., & Chintagunta, P. (2005). Accounting for primary and secondary demand effects with aggregate data. *Marketing Science*, 24(3), 444-460.
- [29] Erdem, T., Imai, S., & Keane, M. P. (2003). Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and economics*, 1(1), 5-64.
- [30] Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615.
- [31] Van Gestel, T., Suykens, J. A., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., ... & Vandewalle, J. (2004). Benchmarking least squares support vector machine classifiers. *Machine Learning*, 54(1), 5-32.

- [32] Platt, J. C., Cristianini, N., & Shawe-Taylor, J. (2000). Large margin DAGs for multiclass classification. *In Advances in neural information processing systems* (pp. 547-553).
- [33] Chapelle, O., & Vapnik, V. (2000). Model selection for support vector machines. *In Advances in neural information processing systems* (pp. 230-236).
- [34] Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615.
- [35] Andriy Norets. (2012). Estimation of dynamic discrete choice models using artificial neural network approximations. *Econometric Reviews*, 31(1), 84-106.
- [36] Benson, A. R., Kumar, R., & Tomkins, A. (2018). A Discrete Choice Model for Subset Selection. *the Eleventh ACM International Conference*(pp.37-45). ACM.
- [37] Chiong, K. X., & Shum, M. (2016). Random projection estimation of discrete-choice models with large choice sets. *Social Science Electronic Publishing*.
- [38] McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
- [39] 王媛媛. (2005). 基于个人决策行为的居住地选择模型及应用研究 (Doctoral dissertation, 清华大学).
- [40] Luce, R. D. (2005). Individual choice behavior: A theoretical analysis. Courier Corporation.
- [41] 王媛媛. (2005). 基于个人决策行为的居住地选择模型及应用研究 (Doctoral dissertation, 清华大学).
- [42] <https://faculty.fuqua.duke.edu/econometrics/presentations/2013/Sudhir-Demand%20Estimation-Aggregate%20Data%20Workshop-updated%202013.pdf>
- [43] Train, K. E. (2009). Discrete choice methods with simulation. *Cambridge university press*.
- [44] Gourieroux, C., & Monfort, A. (1993). Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics*, 59(1-2), 5-33.
- [45] Lee, L. F. (1995). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 11(3), 437-483.
- [46] Hajivassiliou, V. A., & Ruud, P. A. (1994). Classical estimation methods for LDV models using simulation. *Handbook of econometrics*, 4, 2383-2441.
- [47] McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995-1026.
- [48] Hajivassiliou, V. A., & McFadden, D. L. (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, 863-896.
- [49] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.

- [50] Sudhir, K. (2001). Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Science*, 20(3), 244-264.
- [51] Chintagunta, P. K. (2001). Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science*, 20(4), 442-456.
- [52] Levis, A. A., & Papageorgiou, L. G. (2005). Customer demand forecasting via support vector regression analysis. *Chemical Engineering Research and Design*, 83(8), 1009-1018.
- [53] Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Demand estimation with machine learning and model combination (No. w20955). *National Bureau of Economic Research*.
- [54] Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451-468.
- [55] Train, K. E., McFadden, D. L., & Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*, 109-123.
- [56] McFadden, D., & Reid, F. (1975). Aggregate travel demand forecasting from disaggregated behavioral models. Institute of Transportation and Traffic Engineering, University of California.
- [57] McFadden, D. (2001). Economic choices. *American economic review*, 91(3), 351-378.
- [58] Amemiya, Takeshi. (1981) Qualitative Response Models A Survey. *Journal of Economic Literature* 19 (December,1981); 1483-1536.
- [59] 王媛媛. (2005). 基于个人决策行为的居住地选择模型及应用研究 (Doctoral dissertation, 清华大学).
- [60] 离散选择模型的基础.
<http://wiki.mbalib.com/wiki/%E7%A6%BB%E6%95%A3%E9%80%89%E6%8B%A9%E6%A8%A1%E5%9E%8B>
- [61] Bierlaire, M. (1998). Discrete choice models. In *Operations research and decision aid methodologies in traffic and transportation management* (pp. 203-227). Springer, Berlin, Heidelberg.
- [62] Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3), 229-254.
- [63] Marschak, M. (1960). A method for evaluating child-parent interaction under controlled conditions. *The Journal of genetic psychology*, 97(1), 3-22.
- [64] Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3), 229-254.
- [65] Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3), 229-254.

- [66] Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American economic review*, 49(3), 253-283.
- [67] Liao, T. F. (1994). Interpreting probability models: Logit, probit, and other generalized linear models (No. 101). *Sage*.
- [68] Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, Journal of the Econometric Society*, 178-191.
- [69] McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics*, 3(4), 303-328.
- [70] 杨丽丹. (2011). 基于离散选择模型的多式联运路径选择(Master's thesis, 清华大学).
- [71] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.
- [72] Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American journal of agricultural economics*, 66(3), 332-341.
- [73] Chiang, J. (1991). A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science*, 10(4), 297-315.
- [74] Chintagunta, P. K. (1993). Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12(2), 184-208.
- [75] Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, 9(4), 513-548.
- [76] Nair, H., Dubé, J. P., & Chintagunta, P. (2005). Accounting for primary and secondary demand effects with aggregate data. *Marketing Science*, 24(3), 444-460.
- [77] Akerberg, D. A., & Botticini, M. (2002). Endogenous matching and the empirical determinants of contract form. *Journal of Political Economy*, 110(3), 564-591.
- [78] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.
- [79] Berry, S., Levinsohn, J., & Pakes, A. (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review*, 400-430.
- [80] Bajari, P., & Benkard, C. L. (2001). Discrete choice models as structural models of demand: Some economic implications of common approaches. Graduate School of Business, Stanford University.
- [81] Nair, H., Dubé, J. P., & Chintagunta, P. (2005). Accounting for primary and secondary demand effects with aggregate data. *Marketing Science*, 24(3), 444-460.
- [82] Chiang, J. (1991). A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science*, 10(4), 297-315.

- [83] Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, 9(4), 513-548.
- [84] Nair, H., Dubé, J. P., & Chintagunta, P. (2005). Accounting for primary and secondary demand effects with aggregate data. *Marketing Science*, 24(3), 444-460.
- [85] Besanko, D., Gupta, S., & Jain, D. (1998). Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science*, 44(11-part-1), 1533-1547.
- [86] Sudhir, K. (2001). Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Science*, 20(3), 244-264.
- [87] Chintagunta, P. K., Dubé, J. P., & Singh, V. (2003). Balancing profitability and customer welfare in a supermarket chain. *Quantitative Marketing and Economics*, 1(1), 111-147.
- [88] Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, 26(6), 805-818.
- [89] Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615.
- [90] Levis, A. A., & Papageorgiou, L. G. (2005). Customer demand forecasting via support vector regression analysis. *Chemical Engineering Research and Design*, 83(8), 1009-1018.
- [91] Tikhonov, A. N., Arsenin, V. I., & John, F. (1977). *Solutions of ill-posed problems* (Vol. 14). Washington, DC: Winston.
- [92] Wahba, G. (1990). *Spline models for observational data*. Society for industrial and applied mathematics.
- [93] Tibshirani, R., Hastie, T., Narasimhan, B., & Chu, G. (2003). Class prediction by nearest shrunken centroids, with applications to DNA microarrays. *Statistical Science*, 104-117.
- [94] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- [95] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [96] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- [97] 闫国华, & 朱永生. (2009). 支持向量机回归的参数选择方法. *计算机工程*, 35(14), 218-220.
- [98] Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS technical report*, 14(1), 5-16.
- [99] Liu, Y., & Yao, X. (1999). Ensemble learning via negative correlation. *Neural networks*, 12(10), 1399-1404.
- [100] Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451-468.

- [101] Granger, C. W., & Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, 5(1), 35-47.
- [102] Diebold, F. X., & Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, 6(1), 21-40.
- [103] Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Demand estimation with machine learning and model combination (No. w20955). *National Bureau of Economic Research*.
- [104] <https://stats.stackexchange.com/questions/18891/bagging-boosting-and-stacking-in-machine-learning>
- [105] 雷杨. (2011). 基于集成学习的垃圾短信多级分类技术研究(Master's thesis, 电子科技大学).
- [106] Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, 9(4), 513-548.
- [107] Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat cereal industry. *The RAND Journal of Economics*, 395-421.
- [108] Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods*(pp. 207-248). *University of Chicago Press*.
- [109] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.
- [110] Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.
- [111] Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615.
- [112] Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy*, 9(4), 513-548.

致 谢

首先，衷心感谢邓天虎老师对我的悉心指导和鼓励。邓老师不仅在学术研究中给了我极大的帮助，更是教育我以严谨求实的态度对待科学研究。此外，邓老师曾和我讲述自己的学习方法、读博期间的成长经历与当年申请出国时的未来选择，这些都给了我极大的思考与启发，也更坚定了我选择学术道路的决心。在此课题研究过程中，我深深体会到了邓老师对运筹学研究的热情与热爱，也同时深切体会到了运筹学前沿研究的魅力所在。

衷心感谢饶培伦老师和喻纯老师在我面临学术科研和人生抉择时给我的鼓励和建议，在我最迷茫的时候指引我，帮助我分析如何为自己所感兴趣的研究方向来准备自己，让我对科研充满热情与干劲。

感谢朋友们的关心、支持和爱，并且给了我毕设思路和方法上的启发。也感谢一个朋友为我中期汇报提出的建议，我也为他希望变更好。感谢我的父母为我提供了无忧的物质生活条件，鼓励我闯荡世界追逐梦想，并永远是我的坚强后盾。

此外，还要向毕业论文的评审老师们和参考文献的作者们表示由衷敬意。特别感谢我所景仰的研究者 Chintagunta 和 Nair 教授邮件答复了我的疑问。

最后，感谢我自己。这半年来虽经历过许多艰难痛苦时刻，我却始终在坚持追逐自己的梦想与热情，并一直怀着浪漫而坚定的态度对待自己的生活。这是终点，更是起点，希望自己今后不负所有的期望和鼓励，继续勇敢探索世界，用自己的学术研究让这个世界变得更好。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

附录 外文资料调研阅读报告

In the field of marketing, sociological analysis and consumer demand analysis have always occupied a very important position. Consumer needs analysis not only can be used to analyze market intervention and policy implementation to the market impact, but also can be used to predict sales, describe and test consumer behavior, and analyze the company's specific marketing strategy to achieve the company's interests and consumer welfare maximization goal [1]. In recent years, the application of consumer demand forecast has also experienced rapid development, from the traditional sales data analysis and forecasting needs, now to effective advertising, video recommendations and many other user recommendation intelligent algorithms.

For different marketing strategies, consumers may make a variety of reactions, such as changing brands, changing the number of purchases, adjusting both the brand choice and purchase quantity, and taking no action. The various demand models developed to study consumer needs are mainly used to quantitatively analyze the welfare responses of consumers and markets under different marketing strategies. With the advent of the big data era, more and more sources of consumer data, breakthroughs in data processing methods, continuous updating of hardware performance and increasing storage capacity of computing devices become available. As a result, the type and performance of relevant demand forecasting models and their respective application features keep developing. These models obtain the corresponding optimal conditions by solving the functions to maximize the customer's utility, which can also reflect the mutual influence and trade-off between different choices of consumers [2].

Usually, consumers make the following three kinds of decisions when making a purchase: whether or not to buy the product, which one or more of the products to buy, and the quantity of each purchase. These consumer behavior data are usually recorded separately as a behavioral characteristic of each purchase [3]. In the earliest models of consumer choice, most of the researchers studied separately whether or not the consumer made the purchase or what type and quantity of the product is purchased by the consumer. For example, Wales and Woodland [4], Lee and Pitt [5] carried out their studies in 1983 and 1986 about when consumers did not buy. In 1984, King [6], Dubin

and McFadden [7] studied the discrete and continuous choice behavior of consumers under the conditions of buying behavior. Gupta established a model in 1988 to independently study the consumer brand, purchase quantity and purchase time of these three decisions [8]. Later, Chiang et al. Developed the model, taking into account the three types of decision-making processes of whether to purchase, the brand of the specific choice, and the corresponding number of purchases, and the relationship between the effect of the non-purchase on the probability and quantity of the final purchase [6].

Scholars such as Krishnamurthi and Raj (1988) [10], Chiang (1991) [11] and Chintagunta (1993) [12] have proposed some discrete choice models that use household-level sales data. However, in reality, this kind of non-aggregated and precise data is hard to observe and obtain. In contrast, we can easily find aggregated data of all kinds at store level, chain level or market level. Therefore, it is necessary to establish a demand forecasting model for the corresponding aggregated data. Besanko (1998) [13] and Sudhir (2001) [14] proposed a systematic analysis of Logit requirements for aggregated data. The Logit choice model and the traditional discrete choice model are very concise and can apply to the broad product category which has many subdivisions. However, this model has its drawbacks. It is based on the assumption that consumers are buying single items under the Independence of Irrelevant Alternatives (IIA) condition. Under the assumption of this model, the calculated market share of various commodities will deviate from the true value, and the estimated market sales and the price elasticity of demand of commodities may also be biased, which may mislead the pricing strategy and marketing profit measures. Nair then proposed a discrete/continuous demand model based on aggregated data that would provide better estimates of store-level data than previous logit models while addressing consumer personal tastes heterogeneity and marketing strategies' endogeneity [15]. They deal with endogenous problem from Berry et al.'s study [16]. However, their models do not distinguish between the behavior of consumers who buy large quantities of goods voluntarily and the behavior of hoarding in anticipation of market conditions. Erdem and others made use of the data at the household level in 2003, and conducted preliminary research on consumer stockpiling behavior [17].

As studied by Allenby and Rossi (1999), in many econometric studies, the heterogeneity of consumer preferences is offset from the integral of the objective function as a distraction factor [18]. In the realm of demand forecasting, on the

contrary, a correct and reasonable model to fit the heterogeneity of consumer tastes is important to improve the accuracy of the choice forecast. Train (2009) [19], Allenby and Rossi (1999) [20] used a mixed logit model to describe consumer taste heterogeneity and estimate the choice probabilities and correlation coefficients through simulation. Another way to deal with the heterogeneity of consumer tastes is by regularization, which is often used in machine learning. In 2007, Evgeniou et al. used convex optimization to establish the corresponding equation [21]. In 2014, Mishra et al. studied the characteristics of the marginal distribution model (MDM) and found that the family of generalized extreme value distribution models (including multinomial logit, nested logit, etc.) can be derived from this model. Unlike the extreme value distribution model family which requires error entries to be independent, the marginal distribution model can be used to infer a more flexible and representative choice model. In addition, the combination of marginal distribution models and machine learning methods can address the compositionality and scale heterogeneity of consumers and products [22].

The resulting statistical predictive models, such as machine learning models, will play an increasingly important role in the real marketplace, represented by automated modeling and intelligent software assistants. Cui and Curry systematically studied the prediction of support vector machine (SVM) models under different forecasting environments in the marketing field in their article published in 2005 and pointed out that the machine learning methods and the traditional consumer choice models can complement each other's strengths [23]. Although currently SVM has relatively few applications in the field of marketing, Viaene et al. have found that one of the simplest OLS class from SVMs can get better performance than other more complicated algorithms such as the artificial neural network (ANN), the k-nearest neighbor algorithm (KNN), the decision tree (DT) and Sri Lanka multi-layer perceptrons [24]. Experiments by Cui and Curry show that the choice model is more susceptible to the size of the choice set and the dimensions of customer characteristics than the SVM model. Contrary to the traditional customer choice model, larger choice sets and higher feature dimensions can improve the prediction accuracy of machine learning models, which makes them more suitable for predicting complex relationships. However, the SVM model has many shortcomings, too. Besides, its further development needs further theoretical inference support before it is put into practical use. On the other hand, this kind of model makes use of direct data, but it cannot be directly constructed

and predicted by the classic choice probability theories. Platt (2000) [25], Vapnik and Chapelle (2000) [26] have done some research on how to get these probability estimates. How to get the structural prediction of other models except for the simple result prediction from the SVM model is still a future research topic to explore [27].

参考文献

- [1] Chintagunta, P. K., & Nair, H. S. (2011). Structural workshop paper—Discrete-choice models of consumer demand in marketing. *Marketing Science*, 30(6), 977-996.
- [2] Chiang, J. (1991). A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science*, 10(4), 297-315.
- [3] Chiang, J., & Lee, L. F. (1992). Discrete/continuous models of consumer demand with binding nonnegativity constraints. *Journal of Econometrics*, 54(1-3), 79-93.
- [4] Wales, T. J., & Woodland, A. D. (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics*, 21(3), 263-285.
- [5] Lee, L. F., & Pitt, M. M. (1986). Microeconomic demand system with binding nonnegativity constraints: the dual approach. *Econometrica: Journal of the Econometric Society*, 1237-1242.
- [6] King, M. A. (1980). An econometric model of tenure choice and demand for housing as a joint decision. *Journal of public Economics*, 14(2), 137-159.
- [7] Dubin, J. A., & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 345-362.
- [8] Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing research*, 342-355.
- [9] Chiang, J., & Lee, L. F. (1992). Discrete/continuous models of consumer demand with binding nonnegativity constraints. *Journal of Econometrics*, 54(1-3), 79-93.
- [10] Krishnamurthi, L., & Raj, S. P. (1988). A model of brand choice and purchase quantity price sensitivities. *Marketing Science*, 7(1), 1-20.
- [11] Chiang, J. (1991). A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science*, 10(4), 297-315.
- [12] Chintagunta, P. K. (1993). Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12(2), 184-208.

- [13] Besanko, D., Gupta, S., & Jain, D. (1998). Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science*, 44(11-part-1), 1533-1547.
- [14] Sudhir, K. (2001). Structural analysis of manufacturer pricing in the presence of a strategic retailer. *Marketing Science*, 20(3), 244-264.
- [15] Nair, H., Dubé, J. P., & Chintagunta, P. (2005). Accounting for primary and secondary demand effects with aggregate data. *Marketing Science*, 24(3), 444-460.
- [16] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841-890.
- [17] Erdem, T., Imai, S., & Keane, M. P. (2003). Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and economics*, 1(1), 5-64.
- [18] Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89, 57-78.
- [19] Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- [20] Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89, 57-78.
- [21] Evgeniou, T., Pontil, M., & Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, 26(6), 805-818.
- [22] Mishra, V. K., Natarajan, K., Padmanabhan, D., Teo, C. P., & Li, X. (2014). On theoretical and empirical aspects of marginal distribution choice models. *Management Science*, 60(6), 1511-1531.
- [23] Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615.
- [24] Van Gestel, T., Suykens, J. A., Baesens, B., Viaene, S., Vanthienen, J., Dedene, G., ... & Vandewalle, J. (2004). Benchmarking least squares support vector machine classifiers. *Machine Learning*, 54(1), 5-32.
- [25] Platt, J. C., Cristianini, N., & Shawe-Taylor, J. (2000). Large margin DAGs for multiclass classification. In *Advances in neural information processing systems* (pp. 547-553).
- [26] Chapelle, O., & Vapnik, V. (2000). Model selection for support vector machines. In *Advances in neural information processing systems* (pp. 230-236).
- [27] Cui, D., & Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615.