

文章编号:1003-207(2011)06-0092-10

购买金额与时间间隔相关条件下 商店消费总金额的分解

董松挺¹, 赵平¹, 王高², 刘瑞明³

(1. 清华大学经济管理学院, 北京 100084; 2. 中欧国际工商学院, 上海 201206;
3. 美国波士顿麻州大学管理学院, MA 02125)

摘要: 单次购买金额和购买时间间隔相互独立的假设不符合实际购买行为, 看似不显著的相关系数掩盖了两者的真实的相互影响; 据此而得到的消费总金额分解不能准确地反映顾客店内消费总金额的变化规律。文章在对消费者的购买行为进行微观分析的基础上提出了分解商店消费总金额的模型。该模型刻画了单次购买金额和购买时间间隔的相互作用关系, 将商店营销决策中常见的变量引入到模型当中, 并把这些变量对于消费总金额的影响分解开来, 为准确地分析顾客在商店消费的总金额提供了模型框架和实证方法。

关键词: 消费总金额分解; 购买金额; 购买时间间隔

中图分类号: F203.9 **文献标识码:** A

1 引言

从商店赢利的角度来说, 顾客在一定时间内的店内消费总金额是商店管理者最关心的问题。为了了解顾客消费总金额是如何变化的、受到什么因素的影响, 研究者往往把总金额分解为两个部分的乘积: 平均每次的购买金额和一定时间内的光顾次数。因此关于消费总金额的研究问题就可以分解为两个子问题: 关于单次购买金额的研究和关于购买时间间隔(购买频率)的研究。正确地分解消费总金额是研究消费总金额的变化规律及其影响因素的前提条件。

前人的研究假设单次购买金额与购买时间间隔是两个相互独立的子问题^[1-3], 我们经过分析, 发现这个假设是值得置疑的, 根据这个假设建立的消费总金额分解模型没有考虑单次购买金额与购买时间间隔的相互作用, 因而不能恰当地反应消费总金额的变化规律, 也不能准确地体现各种影响因素的效果。

本文在单次购买金额与购买时间间隔相互作用

的假设下, 提出了商店消费总金额的分解模型。该模型除了刻画单次购买金额与购买时间间隔的相互关系以外, 还将商店营销决策中常见的变量引入到模型当中, 把这些变量对于消费总金额的影响进行分解, 说明这些变量如何分别影响单次购买金额和购买时间间隔, 进而对消费总金额产生影响。最后, 本文应用带协变量的随机模型对该分解模型进行了实证分析。

2 文献综述

Schmittlein 和 Peterson(1994)^[3]在他们的文章中明确提出了(单次)购买金额与交易频率(购买频率)以及客户保留率相互独立的假设。他们指出, 这样假设的原因是为了克服模型构建时数学计算上的困难, 而这个假设成立与否, 需要根据具体的应用情况而定。

在后来的研究当中, 研究者或是直接使用了该假设^[1], 或者在检验了单次购买金额与购买频率的相关系数之后应用了该假设^[2]。由于 Colombo 和 Jiang(1999)、Fader 和 Hardie 等(2005)^[1,2]等模型的广泛应用, 独立假设也被广泛地应用。

Fader 和 Hardie 等(2005)^[2]使用 CDNOW 网上商店 946 个第一次购买的消费者在 1997 年第一季度的购买数据, 对于单次购买金额与购买频率的相关系数进行了检验, 相关系数为 0.11, 排除了一

收稿日期: 2009-01-12; 修订日期: 2011-07-04

基金项目: 国家自然科学基金资助项目(70632003/G0206)

作者简介: 董松挺(1982-), 男(汉族), 浙江人, 清华大学经济管理学院, 博士研究生, 研究方向: 营销模型。

个消费者(outlier)的数据之后,得到相关系数为0.06(p -value=0.08)。该文章据此认定“单次购买金额与购买时间间隔相互独立”的假设成立。我们进行了相同的检验,利用某大型超市70个家庭顾客在2005年第三季度的消费数据(数据说明见后文的实证分析),得到两者的相关系数为-0.30(p -value=0.012),单次购买金额与购买时间间隔显著地负相关。

我们试图从一个微观的角度,用消费者的购买行为来解释相关系数的不同。消费者在某个特定商店的消费需求总是有一个上限的,我们称之为“总需求”。当这个总需求不变或者变化不大的时候,单次购买金额与购买频率应该表现出负相关关系:每次买得越多,购买次数就会越少,购买时间间隔越长;反之,每次买得越少,购买次数就会越多,购买时间间隔越短。我们把这种现象称为“总需求约束效应”。然而,当总需求明显变化的时候,就有可能引起单次购买金额与购买频率出现同方向的变化(正相关关系):消费者越来越多地把他的购买集中到某个商店,则单次购买金额和购买频率可能同时增加;反之,消费者越来越多把他的购买从某家商店转移到其他商店,则单次购买金额和购买频率可能同时减少。我们把这种现象称为“总需求变化效应”。当然,不论总需求如何变化,作为消费者需求的上限,它总是对消费者的购买总金额有一个约束。所以我们认为,总需求约束效应总是存在的,而总需求变化效应的发生需要特定的条件(促使消费者在该商店的总需求发生变化的条件)。当这些条件发生的时候,总需求变化效应所引起的单次购买金额与购买时间间隔的正相关关系有可能削弱、掩盖、甚至超过总需求约束效应所引起的两者的负相关关系,从而导致相关系数看上去不显著、或者得到正值。

Fader和Hardie等(2005)^[2]用于检验相关关系的数据来源是946个第一次实施购买的消费者,他们在CDNOW商店的需求处于一个上升的阶段,利用这个数据得到的不显著的相关系数,是总需求约束效应和总需求变化效应共同起作用的结果。而我们的数据来源于某大型超市,消费者的总需求相对稳定,总需求约束效应带来的结果就是显著的负相关系数。

因此,我们认为,相关系数的检验不能用来判断单次购买金额与购买时间间隔是否相互独立,看似不显著的相关关系掩盖了单次购买金额与购买时间间隔之间多重的相互作用。

我们无意说明前人的研究是完全错误的。Schmittlein和Peterson(1994)、Colombo和Jiang(1999)和Fader,Hardie和Lee(2005)^[1-3]等的研究主要是对于数据趋势的分析,这类研究的着眼点是对数据的拟合和预测。因此,如果独立假设能够简化数学模型,同时使得模型的拟合精度和预测效果都比较满意,那么这种假设也有可取之处。但是,当我们应用基于独立假设的模型时,必须检验单次购买金额和购买频率的相关系数^[2]。只有在相关系数不显著的情况下,应用这些模型才是安全的,否则就可能使结果出现难以预料的偏差。因此,前人给我们留下了这样一个问题:在一个更一般的情况下,特别是当我们不能忽略单次购买金额和购买频率的相互作用时(两者相关系数显著不为0时),应该如何构建消费总金额分解模型?

我们之所以提出了与前人不同的假设,是因为我们从微观角度考察了消费者购买行为。我们可以看到:从消费者个人行为的角度,其单次购买金额和购买时间间隔有着十分明显的相互依赖关系,两者不可能是相互独立的。Ho和Lim等(2006)^[4]指出,一个好的模型不仅要有通用性(generality)、精确性(precision)、实证准确性(empirical accuracy),还应该符合消费者的心理过程(psychological plausibility)。同样地,我们认为一个好的购买行为模型应该符合消费者实际的决策行为,一个合理的消费总金额分解模型理应将单次购买金额和购买时间间隔的相互关系刻画出来。

此外,管理实践的要求迫使我们不能满足于数据拟合和趋势预测,我们还希望知道商店营销决策中常见的各种变量对于消费总金额起了什么样的作用。例如,大量研究表明,促销等营销变量对于产品选择、购买时间和购买数量都有重要的影响^[5-10],很自然地,这些营销变量也会影响消费者在某一商店的消费总金额。Roslow和Li等(2000)^[11]的研究表明情境变量(situational variables)、人口统计变量(demographic variables)以及季节因素对于消费者购买行为的影响是客观存在的;Jones和Zufryden(1980,1982)^[12,13]也强调要把人口统计变量引入到营销模型当中,以体现消费者特征对于其购买行为的影响。

这些变量对于消费总金额的影响可能是非常复杂的,比如一次促销活动可能增加了单次购买金额却延长了下次购买的时间^[14],那么这次促销活动在多大程度上增加了单次购买金额,多大程度上延长

了购买时间,进而对未来一段时间内的消费总金额产生了什么样的影响?要回答这些问题就要求我们把营销决策中常见的变量引入到模型中来,并且对消费总金额作出正确的分解。

本文试图回答前人留下的问题,构建一个能够刻画单次购买金额与购买时间间隔的相互关系、符合消费者微观购买行为的模型;同时把营销决策中常见的变量引入模型,为研究这些变量对消费者产生的影响提供依据。

3 问题建模

把消费总金额分解为单次购买金额和购买时间间隔,实际上是从商店的视角来看待顾客的店内消费总金额,即:商店观测到消费者每次在本店购买了多少金额,多长时间来本店光顾一次。但是从消费者的角度来说,一次购买的实施包括三个阶段:首先是要不要出门购物,然后是去哪个商店购物,最后是购买多少。商店观测到的“购买时间间隔”实际上是消费者两个阶段决策的结果:要不要出门购物、选择哪家商店。只有当消费者决定出门,然后选择这一家商店的时候,商店才观测到了顾客在本店的一次购买。因此要搞清单次购买金额和购买时间间隔的相互作用,了解各种营销变量对于两者的影响,需要具体分析消费者实施购买的三个决策步骤,以及各个步骤中对消费者产生影响的因素。当然,对消费者购买行为产生影响的因素非常多,我们在分析中不可能穷尽这些因素(这也不是本文研究的目的),所以只列举了一些我们认为最常见的,对分析消费总金额最有影响的变量。

“要不要出门购物”,这个问题的产生往往是因为消费者发现要用的东西家里没有,或者快用完了。这一状态与前一次购买了多少(前次购物量)、用了多长时间(购买时间间隔)有密切的关系,同时也受到目标商店或者竞争商店营销努力的影响(比如消费者听说目标商店正在进行某产品的促销)。“选择哪家商店”是商店竞争、消费者选择的结果。消费者选择商店的决策不仅受到目标商店和竞争商店营销努力的影响,还与其他众多的竞争因素有关。比如商店的硬件设施、店内商品的陈设、商店的服务质量等等。我们可以用消费者对某一商店的忠诚来反映这些影响结果的总和。“购买多少”的问题,从消费者的角度来说用“购买数量”来反映比较准确。但对于商店来说,由于消费者购买的商品种类一般比较多,单价各不相同,利用商品个数不能准确反映消费

者“买了多少”,因此常用消费金额来反映一次消费的总量。消费者的购买量受到消费者所处状态的影响:目前还剩多少(前次买了多少,用了多长时间)、未来需要多少。同时,虽然消费者已经走进了一家商店,他/她仍然会考虑什么东西在这家商店买,什么东西到其他地方去买,商店之间的竞争仍然存在。因此,购买金额还受到目标商店和竞争商店的营销努力的影响,也受到消费者忠诚程度的影响。

另外,所有这三个阶段的决策都要受到消费者特征以及各种不确定因素的影响。比如收入水平和家庭规模的不同会引起使用数量的不同,从而影响“要不要出门购物”以及“购买多少”的决策;而收入水平的不同会影响消费者对于不同档次商店的偏好,从而影响对商店的选择。比如突然出现的需求以及出门逛街的冲动会影响“要不要出门购物”的决策;逛街过程中商店的遍历增加了“选择商店”的随机性;而选购商品的过程中意外地发现其他产品可能会改变“购物多少”的计划。

从以上的分析中,我们看到:前次购买量会影响“要不要出门购物”的决策,从而影响购买时间间隔;而购买时间间隔和前次购买量会一起影响当期的购买量。另外,我们可以根据影响因素的来源将它们分为三类。第一类是目标商店对消费者施加的影响,主要通过营销变量(如促销)来体现;第二类是消费者自身的特点,比如收入水平、家庭状况等特征变量;第三类是情境变量,即购买发生时消费者所处的状态和各种环境因素,比如消费者对于目标商店的忠诚、周围商店(竞争对手)的促销措施等等。这三类变量分别影响着单次购买金额和购买时间间隔,从而对消费者在某一商店内的消费总金额产生影响。

根据前面的分析,我们提出了以下一个分解模型的框架(图1):

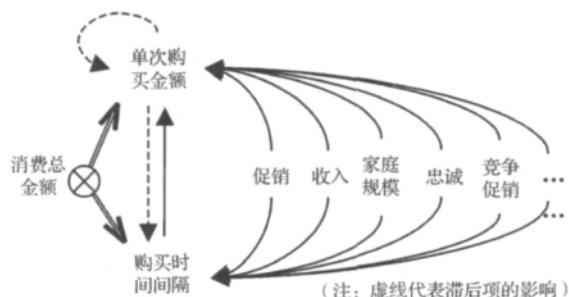


图1 消费总金额的分解模型框架

在这个模型中,单次购买金额和购买时间间隔就不再是相互独立的了。购买时间间隔影响单次购买金额,而前次购买金额会影响到下一次购买的发生时间和金额大小。这个模型体现了促销等营销决策中常见的变量对于单次购买金额和购买时间间隔分别的影响;我们可以分析这些变量对两者产生的影响方向与大小,将这些变量对于消费总金额的影响分解开来。除了模型中提到的促销等变量,还可能有一些其他的变量以及一些不确定因素对消费总金额存在影响。由于我们的模型框架是可拓展的,这些变量或因素同样可以在这个模型框架中得到分解,进而分析他们对于消费总金额的影响方向与大小。

4 估计方法

根据分解模型框架(图1)的特点,我们用一对联立模型来为消费总金额分解模型提供估计方法。联立模型由单次购买金额和购买时间间隔两个子模型组成。这两个子模型要分别体现营销决策中常见的各种变量对于单次购买金额和购买时间间隔的影响(即分解这些变量对于消费总金额的影响),同时反映单次购买金额和购买时间间隔的相互作用关系。模型的设定应该是开放,允许应用者引入更多可能的变量,对他们所产生的影响进行分解。另外,对于购买行为中的不确定因素,模型也应有所体现。基于这些要求,我们用带协变量的随机模型(Stochastic model with covariates)为两个子模型建模。下面我们首先回顾一下前人对购买时间间隔和单次购买金额的模型研究。

关于购买时间间隔的研究有着比较长的历史,模型也比较多样。自从 Ehrenberg (1959)^[15] 将 NBD 模型(Negative Binomial Distribution)引入营销领域,NBD 和它的一些拓展模型就被广泛地应用到对购买时间间隔的研究当中。NBD 模型假设购买频率服从泊松(Poisson)分布(即购买时间间隔服从指数分布,Exponential),而分布参数 λ 服从 Gamma 分布。虽然 NBD 的前提假设常常被质疑,但模型的论证结果却一再地证实了 NBD 的鲁棒性(Robustness)^[16,17],并被广泛地应用。与此同时,学者们还提出了其他一些模型方法,如用爱尔朗(Erlang)分布来代替 NBD 的指数分布^[16,18],得到改进的 Erlang2-Gamma 分布(CNBD);用逆高斯(Inverse Gaussian)分布来描述购买时间间隔^[19];或者用对数正态分布^[20]来描述购买时间间隔;等等。后来的研究^[21]认为这些描述时间间隔的分布各有适

用范围。而 Gupta(1991)^[22]则为指数分布、二阶爱尔朗分布、NBD、CNBD 等模型引入了协变量,把购买时间间隔与营销决策中常见的变量联系起来,说明了这些变量对于购买时间间隔的影响。

相比而言,单次消费金额的研究起步比较晚,是在 20 世纪 90 年代中期以后,随着顾客价值的概念而兴起的,到目前的研究也不算多。比较早的研究^[3]曾经用正态分布来拟合消费者的购买金额数据。但是金额的分布常常是偏态的(不像正态分布是对称的),而金额的数值又必须保证是正数,因此后来的研究就用 Gamma 分布来拟合购买金额数据。Colombo 和 Jiang (1999)^[1]在 RFM 模型(Recency, Frequency, and Monetary value)中对金额 Gamma 分布的尺度参数进一步引入了 Gamma 分布,来捕捉消费者个体之间的差异性,形成了 Gamma-Gamma 随机模型。这两位作者用该随机模型来研究消费者单次购买金额的规律,拟合效果很好,被后来者广泛采用。Fader 和 Hardie 等(2005)^[2]在 Gamma-Gamma 随机模型的基础上作了一点拓展,推导了平均消费金额的 Gamma-Gamma 随机模型。该文章推导的公式中虽然出现了“购买次数”这一变量,但这个变量只能表示“平均消费金额出现的次数”,还是没有把购买时间间隔与单次购买金额联合起来考虑。董松挺等(2006)^[23]在 Gamma-Gamma 随机模型的基础上,把营销决策中常见的变量作为协变量引入了模型,将单次购买金额与这些变量联系起来,说明了这些变量如何影响单次购买金额的变化。

在本研究中,我们用带协变量的 Gamma-Gamma 模型^[23]为单次购买金额建立子模型,用带协变量的 Exponential-Gamma 模型^[22]为购买时间间隔建立子模型。这样做从三个方面满足了分解模型框架对于估计模型的要求:(1)消费者的购买行为存在着大量的不确定因素(即随机因素),随机模型能够很好地捕捉这些随机因素,其拟合能力与预测能力已经得到了广泛的认可。(2)协变量可以把营销决策中常见的变量引入到模型中来,反映这些变量对于单次购买金额和购买时间间隔的影响,实现分解的目标;而且模型具备可拓展性,我们可以引入任意多个相关的变量进行分析。(3)我们可以用协变量的方式把单次购买金额与购买时间间隔的相互影响反映到模型当中,而不是假设两个子模型的随机系数相互影响,这样就避免了随机系数相互影响时数学处理上的困难^[3]。

4.1 带协变量的 Gamma-Gamma 单次购买金额模型

Gamma-Gamma 随机模型最早是由 Colombo 和 Jiang(1999)^[1] 提出来的,假定:

(1) 对于每个消费者个体来说,其单次购买金额 m 是一个随机变量,服从参数为 p, v 的 Gamma 分布,即 $m \sim \text{Gamma}(p, v)$ 。其中 p 是形状参数 (Shape parameter), v 是尺度参数 (Scale parameter)。

(2) 每个消费者的参数 v 服从参数为 q, r 的 Gamma 分布,即 $v \sim \text{Gamma}(q, r)$ 。其中, q 是形状参数, r 是尺度参数。

在这样的假设下,单次消费金额 (m) 的条件概率可以表示为:

$$f(m | p, q, r) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{m}{r+m}\right)^p \left(\frac{r}{r+m}\right)^q \frac{1}{m} \quad (1)$$

董松挺等 (2006)^[23] 为了把协变量引入 Gamma-Gamma 随机模型,增加了一条假设:

(3) 每个消费者的参数 p 由方程 $p = p_0 \exp(X\beta_1)$ 决定,其中 X 是影响单次购买金额的协变量 (购买时间间隔、忠诚、促销、前次购买金额、收入、家庭规模等), β_1 是 X 的系数。

根据假设 1, 单次购买金额 m 服从参数为 p, v 的 Gamma 分布, m 的均值就可以表示为 p/v ; 假设 2 认为 v 是一个随机变量, 通过令 v 服从 Gamma 分布来捕捉不同消费者个体在 m 的均值上的差异; 而根据假设 3, m 的均值可以进一步表示为 $p_0/v \times \exp(X\beta_1)$, 实际上是以 p_0/v 为基准, 对 m 的均值引入协变量所引起的差异性 $\exp(X\beta_1)$ 。这样就把协变量引入到了模型当中。

根据这三个假设, 我们可以得到单次购买金额 m 的条件概率公式 (公式 2, 推导和说明见附录); 利用这个条件概率公式, 我们就可以用极大似然估计方法估计出各个参数的数值。

$$f(m | p_0, q, r, \beta_1, X) = \frac{\Gamma(p_0 e^{X\beta_1} + q)}{\Gamma(p_0 e^{X\beta_1})\Gamma(q)} \left(\frac{m}{r+m}\right)^{p_0 e^{X\beta_1}} \left(\frac{r}{r+m}\right)^q \frac{1}{m} \quad (2)$$

其中,

$$X = \begin{pmatrix} \text{购买时间间隔, 前次购买金额, 促销,} \\ \text{收入, 家庭规模, 忠诚, 竞争促销} \dots \end{pmatrix}$$

$$\beta_1 = (c, a_1, a_2, a_3, a_4, a_5, a_6, a_7, \dots)$$

$$X\beta_1 = c \times \text{购买时间间隔} + a_1 \times \text{前次购买金额} + a_2 \times \text{促销} + a_3 \times \text{收入} + a_4 \times \text{家庭规模}$$

$$+ a_5 \times \text{忠诚} + a_6 \times \text{竞争促销} + a_7 \times \dots$$

这个模型利用 Gamma-Gamma 随机分布来捕捉各种不确定因素 (随机因素), 通过在协变量 X 中引入购买时间间隔、前次购买金额来体现单次购买金额与购买时间间隔的相互影响, 通过引入促销、收入等变量来反映营销决策中常见变量对单次购买金额的影响。

4.2 带协变量的 Exponential-Gamma 购买时间间隔模型

根据 Gupta(1991)^[22], 我们对消费者的购买时间间隔 t 作如下三个假设:

(1) 购买时间间隔 t 服从参数为 λ 的指数分布, 即 $t \sim \text{Exponential}(\lambda)$

(2) $\lambda = \lambda_0 \exp(Y\beta_2)$, 其中 Y 是影响购买时间间隔的协变量 (忠诚、促销、前次购买金额、收入、家庭规模等), β_2 是 Y 的系数。

(3) λ_0 服从参数为 s, g 的 Gamma 分布, 即 $\lambda_0 \sim \text{Gamma}(s, g)$ 。其中 s 是形状参数, g 是尺度参数。

根据假设 1, 购买时间间隔 t 服从参数为 λ 的指数分布, t 的均值就可以表示为 $1/\lambda$ 。假设 2 将参数 λ 进一步分解为两个部分, 其中 λ_0 被视为一个随机变量, 通过令 λ_0 服从 Gamma 分布来捕捉不同消费者个体在 t 的均值上的差异 (假设 3); 而 $\exp(Y\beta_2)$ 则代表了协变量所引起的差异性对于均值的影响。这样就把协变量引入到了模型当中。

根据这三个假设, 我们可以得到购买时间间隔 t 的条件概率公式 (公式 3, 推导和说明见附录); 利用这个条件概率公式, 我们就可以用极大似然估计方法估计出各个参数的数值。

$$f(t | s, g, \beta_2, Y) = \frac{s}{g} \left(\frac{g}{g + te^{Y\beta_2}}\right)^{s+1} \times e^{Y\beta_2} \quad (3)$$

其中:

$$Y = \begin{pmatrix} \text{前次购买金额, 促销,} \\ \text{收入, 家庭规模, 忠诚, 竞争促销} \dots \end{pmatrix}$$

$$\beta_2 = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, \dots)$$

$$Y\beta_2 = b_1 \times \text{前次购买金额}$$

$$+ b_2 \times \text{促销} + b_3 \times \text{收入} + b_4 \times \text{家庭规模}$$

$$+ b_5 \times \text{忠诚} + b_6 \times \text{竞争促销} + b_7 \times \dots$$

这个模型利用 Exponential-Gamma 随机分布来捕捉各种不确定因素 (随机因素), 通过在协变量 X 中引入前次购买金额来体现单次购买金额与购买时间间隔的相互影响, 通过引入促销、收入等变量来反映营销决策中常见变量对购买时间间隔的影响。

5 分析讨论

5.1 数据说明

我们应用以上的分解模型和估计方法,对某大型连锁超市的顾客购买情况进行分析。数据来源于央视市场调查公司固定样本组数据,该样本组数据以家庭为单位,记录了这些家庭在一段时间内的所有消费情况。其中,从 2005 年 7 月 1 日至 2005 年 9 月 29 日的 13 周当中,88 个家庭曾经在这家超市中有过购物记录,平均每个家庭的消费总金额是 224.36 元。我们将消费总金额不足 50 元的 18 个家庭剔除,这是因为这些家庭客户的总购买量太少,很难算作该超市的常规客户群体,从他们身上难以反映出该超市顾客在消费总金额上的规律。这样,我们筛选出了 70 个家庭 2005 年第三季度(13 周)的购买数据,这些家庭在该超市的购买记录共有 720 条,平均每个家庭的消费总金额是 276.72 元。另外,这 70 个家庭在 2005 年 7 月 1 日之前和 2005 年 9 月 29 日之后都有在这家超市消费的记录,所以不存在新顾客到达或者老顾客流失的情况。

除了家庭消费者每次购买金额和购买时间间隔的数据,我们还选取了促销、月收入、家庭人口规模和忠诚的数据,来分解这些变量对于消费总金额的影响。由于该超市在北京市有多家连锁店,而央视市场调查公司的固定样本组也是从遍布北京市的广大市民中征集的,所以记录的购买情况是在该超市不同的连锁店中发生的。由于不同连锁店所处的竞争情况非常复杂,我们很难描述竞争商店的营销努力,因此我们不将该变量作为协变量引入模型;而竞争对手的营销努力将作为一个随机因素由模型的随机部分来捕捉。此外,为了体现季节变化对于销售金额的影响,我们把购买发生的月份作为哑变量,也加入到两个模型中。因此,两个子模型的协变量部分可以表示为:

$$X\beta_1 = c \times \text{购买时间间隔} + a_1 \times \text{前次购买金额} + a_2 \times \text{促销} + a_3 \times \text{收入} + a_4 \times \text{家庭规模} + a_5 \times \text{忠诚} + a_6 \times \text{8 月} + a_7 \times \text{9 月}$$

$$Y\beta_2 = b_1 \times \text{前次购买金额} + b_2 \times \text{促销} + b_3 \times \text{收入} + b_4 \times \text{家庭规模} + b_5 \times \text{忠诚} + b_6 \times \text{8 月} + b_7 \times \text{9 月}$$

其中,我们对变量“忠诚”是这样定义的:购买发生前的一个月(30 天)内,该消费者(家庭)在该超市的购物金额占该消费者(家庭)在该月内所有消费支出的比重。即用过去一段时间内观测到的实际购买

行为来定义忠诚^[24,25]。这样定义的忠诚,实际上是用消费者对于目标商店总需求的变化来反映商店之间众多竞争因素影响的总和。

由于数据的限制,“促销”是用 0、1 变量来表示的,如果消费者的购物篮中有一样是促销产品,我们就认为商店的“促销”在这个消费者的这次购物中是存在的。由于商店的促销活动一般都能够落实到产品上,我们认为这样的表示方法虽有局限,却也能够代表商店所做的促销努力对于消费者某次购买的影响。

此外,收入分为 1~19 共 19 档,依次递增。家庭规模是指家庭里的人口数量。购买发生的月份以 7 月为基准,时间间隔以天为单位。

5.2 估计结果分析

我们应用极大似然估计的方法,利用 Matlab 程序估计模型参数,结果如下页表 1 所示。

表 1 模型系数估计结果

变量	单次购买金额(<i>m</i>)	购买时间间隔(<i>t</i>)
	(Gamma-Gamma) ^a	(Exponential-Gamma) ^b
<i>p</i> ₀	1.39***	
<i>q</i>	3.32***	
<i>r</i>	24.57***	
<i>s</i>		4.56×10 ⁴ ***
<i>g</i>		9.93×10 ⁵ ***
时间间隔	0.0118***	
前次购买金额	0.0029***	-0.0067***
促销	0.3961***	0.0055
收入	0.0029	0.0542***
家庭规模	0.0373	0.0059
忠诚	0.2523*	1.8868***
8 月	-0.0374	-0.0435
9 月	-0.0752	0.0107
-2LL	6022.38	4334.05

注:(1)^a $m \sim \text{Gamma}(1.39e^{X\beta_1}, v)$, $v \sim \text{Gamma}(3.32, 24.57)$
 $\text{Exp}(v) = q/r = 0.135$, $\text{Var}(v) = q/r^2 = 0.0055$;
(2)^b $t \sim \text{Exponential}(\lambda_0 e^{Y\beta_2})$, $\lambda_0 \sim \text{Gamma}(4.56 \times 10^4, 9.93 \times 10^5)$
 $\text{Exp}(\lambda_0) = s/g = 0.046$, $\text{Var}(\lambda_0) = s/g^2 = 4.63 \times 10^{-8}$;
(3)***:p-value<0.001; **:p-value<0.01; *:p-value<0.05

从估计结果中我们可以看到, *v* 和 λ_0 的方差(0.0055 和 4.63×10^{-8})相对于均值(0.135 和 0.046)来说都很小,这说明除了忠诚等协变量引起的差异以外,这群顾客的差异性是比较小的。因此我们把这 70 个家庭顾客作为一个消费者群体是可行的。

由于单次购买金额的均值(\bar{m})和购买时间间隔的均值(\bar{t})分别是:

$$\bar{m} = \frac{p}{v} = \frac{p_0 \exp(X\beta_1)}{q/r} = \frac{p_0 r}{q} \exp(X\beta_1) \quad (6)$$

$$\bar{t} = \frac{1}{\lambda} = \frac{1}{\lambda_0 \exp(Y\beta_2)} = \frac{g}{s \exp(Y\beta_2)} \quad (7)$$

因此在单次购买金额模型中,正的系数表示对应的变量与单次购买金额呈正相关关系,变量数值的增长将提高单次购买金额;而在购买时间间隔模型中,正的系数表示对应的变量与购买时间间隔呈负相关关系,变量数值的增长将缩短购买时间间隔,提高购买频率。

结合表 1 的结果,我们可以看到,两次购买之间的时间间隔越长,之后的一次购买金额就会越多(0.0118, $p < 0.001$);而前次购买金额越多,下一次购买的时间间隔就越长(-0.0067, $p < 0.001$)。这个结果与我们的分析是一致的:在不考虑忠诚变化的情况下,消费者在一家商店的总需求存在一个上限,在这个限制的约束之下,单次购买金额和购买时间间隔不会同时向商店管理者希望的方向变化(既增加单次购买金额,又缩短了购买时间间隔)。总需求约束效应是存在的。

所有的协变量中,只有忠诚在单次购买金额模型和购买时间间隔模型中都产生显著的正向影响(单次购买金额模型:0.2523, $p = 0.021$;购买时间间隔模型:1.8868, $p < 0.001$);即忠诚上升,单次购买金额增加,购买时间间隔缩短。而根据我们的定义,这里忠诚变化实际上就是顾客在该商店中总需求变化的反应,“总需求变化效应”也是存在的。当然,“总需求变化效应”是否起作用,要看忠诚的变化(总需求的变化)是否明显。只有当忠诚的变化很明显,它所引起的单次购买金额和购买频率的正相关关系比较强烈的时候,才会削弱、掩盖、甚至超过总需求约束效应所引起的两者的负相关关系,从而使得单次购买金额和购买频率的相关系数不显著、或者得到正值。

5.3 协变量的影响

下面我们把分解模型的结果作一个合成,示意协变量是如何分别影响单次购买金额和购买时间间隔,进而对消费总金额产生影响的。

我们知道一段时间内消费总金额可以表示成:

消费总金额 = 平均单次购买金额

$$\times \frac{\text{时间长度}}{\text{平均购买时间间隔}} \quad (8)$$

那么 T 时间内(T 的单位应与购买时间间隔的单位一致,这里是“天”)的消费总金额就可以用单次

购买金额与购买时间间隔的平均值得到:

$$M = \bar{m} \times \frac{T}{\bar{t}} = \frac{p_0 r}{q} \exp(X\beta_1) \times \frac{T}{g / (s \exp(Y\beta_2))} \\ = \frac{p_0 r s T}{g q} \exp(X\beta_1 + Y\beta_2) \quad (9)$$

因此各个变量对于消费总金额的综合影响就可以用 $X\beta_1 + Y\beta_2$ 来表示。我们把 X 和 Y 中共同的变量列示出来,将合成以后的系数与合成前两个模型的系数加以比较(表 2):

表 2 合成前后变量系数比较

变量 ^a	单次购买金额	购买时间间隔	消费总金额
	系数(β_1)	系数(β_2)	系数($\beta_1 + \beta_2$) ^b
前次购买金额	0.0029**	-0.0067***	-0.0038
促销	0.3961***	0.0055	0.4016
收入	0.0029	0.0542***	0.0571
家庭规模	0.0373	0.0059	0.0432
忠诚	0.2523*	1.8868***	2.1391
8 月	-0.0374	-0.0435	-0.0809
9 月	-0.0752	0.0107	-0.0645

注:(1)^a 变量没有包括购买时间间隔

(2)^b 合成系数为两数之和,没有检验其显著性

(3)***: $p\text{-value} < 0.001$; **: $p\text{-value} < 0.01$; *: $p\text{-value} <$

0.05

从这个表中,我们就可以看到各个变量对消费总金额的影响是如何被分解开来的:“前次购买金额”的作用表面在两个方面:一方面,前一次购买得比较多就会推迟下一次的购买时间(-0.0067, $p < 0.001$);另一方面,它反映了消费者的购买习惯,有的消费者喜欢每次多买一点(少买几次),而有的则习惯于每次少买一点(多买几次),前后的购买金额因而表现出正的相关关系(0.0029, $p = 0.0013$)。促销的作用是刺激单次购买(0.3961, $p < 0.001$),从而提高消费总金额,但对于缩短购买时间间隔没有明显的作用(0.0055, $p = 0.466$)。收入高的家庭在这家超市有更多的消费,但这些消费是通过比较高的购买频率体现出来的(0.0542, $p < 0.001$),单次购买金额并没有随收入增加(0.0029, $p = 0.371$)。对于这些家庭消费者来说,家庭规模对于单次购买金额和购买时间间隔的影响都不显著,因而对消费总金额影响不大。忠诚通过单次购买金额和购买时间间隔对消费总金额产生的影响都是显著的,但通过购买时间间隔所产生的影响更大(1.8868 > 0.2523),也更显著。而这些客户在 7 月、8 月、9 月的单次购买金额和购买时间间隔没有明显的区别。

通过以上分解,我们就对这些消费者在该超市的购买行为有了一个比较全面的了解,可以据此采

取针对性的措施。但如果我们只看到一个总的结果,而不把它分解开来,就可能丢失一些重要的信息,失去改进的机会。比如“促销”一项,从总的结果来看,该商店的促销提高了消费总金额,但如果我们把它分解开来,就会发现这个“促销”虽然起到了刺激顾客临时提高购买金额的作用,但对于提高顾客的光顾频率却作用不大。那么如果我们要在“促销”方面作一些改进,可能就要在促销活动的宣传方面加大努力,让顾客知道商店在进行的促销活动,吸引他们进入商店,从而提高购物频率。进一步地,如果我们能够得到不同形式促销的数据,就可以判断哪些促销手段起了多大的作用,通过什么样的途径在起作用。

6 结语

在前人关于消费总金额分解的研究中,单次购买金额和购买时间间隔被假定为相互独立。但这并不意味着前人的研究认为两者在消费者购买行为上没有相互影响,作这样的假设只是出于研究问题特性(趋势分析)以及数学处理需要的考虑。然而,这毕竟是在一定条件下对问题的简化(相关系数不显著时);在更一般的情况下,如何考虑单次购买金额和购买时间间隔的相互作用,如何在两者不独立的假设之下对消费总金额进行准确地分解?这是前人给我们留下的问题。

本文从微观的角度分析了消费者购买行为中单次购买金额和购买时间间隔的相互作用,并在两者相互影响的假设下构建了消费总金额的分解模型。这个分解模型以分析消费者的微观购买行为为基础,符合消费者的购买决策过程;以协变量的方式刻画单次购买金额和购买时间间隔的相互作用,避免了随机系数相互影响时数学处理上的困难;把营销决策中常见的变量引入模型,通过检查变量对于单次购买金额和购买时间间隔分别的作用,帮助研究者和管理者准确地了解这些变量对于消费总金额的影响机理,得到对管理决策有益的结果。

由于数据的限制,我们仅用 0、1 变量来表示促销的有无。这对于结果的丰富程度是一个很大的限制,如果我们能够得到更全面的促销数据(比如做了什么样的促销、促销强度有多大),就可以清楚地分析出各个促销手段是否起了作用、起了多大的作用,这无疑能为研究和管理实践提供更多有用的信息。由于本文的模型是可拓展的,未来的研究可以在本模型框架的基础上,引入更多的变量(促销手段、顾

客信息等等),得到更丰富的结果。

我们定义的忠诚是消费者对目标商店总需求的反应,是各种竞争因素对消费者购买产生影响的总和,是一个结果变量。那么是什么引起了忠诚的变化?本文用忠诚这一变量来体现各种因素的影响结果,是一个比较简化的做法,未来的研究可以将忠诚这一结果变量进一步分解,将各种竞争因素(如商店的硬件设施、店内商品的陈设、商店的服务质量等)引入到模型当中,从而对影响消费总金额的因素有一个更深入的了解。

另外,我们在研究中将 70 个家庭视为一群同质的顾客,然而在现实生活中,客户群体可能是多样的,管理者往往会把顾客区分为不同的细分市场,了解每个细分市场的特点,针对不同的细分市场施加营销影响。进一步的研究应该使模型能够满足区分不同细分市场的需求,进而研究营销决策中常见的变量在不同的细分市场中对于消费总金额的影响存在何种区别。

附录:条件概率公式推导和说明

单次购买金额模型的条件概率公式:

由于 Gamma - Gamma 随机模型的条件概率公式是:

$$f(m | p, q, r) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \left(\frac{m}{r+m}\right)^p \left(\frac{r}{r+m}\right)^q \frac{1}{m}$$

而 $p = p_0 \exp(X\beta_1)$, 所以我们只需要把 p 的表达式代入到上面的式子中,就可以得到带协变量的 Gamma - Gamma 随机模型的条件概率公式:

$$f(m | p_0, q, r, \beta_1, X) = \frac{\Gamma(p_0 e^{X\beta_1} + q)}{\Gamma(p_0 e^{X\beta_1})\Gamma(q)} \left(\frac{m}{r+m}\right)^{p_0 \exp(X\beta_1)} \left(\frac{r}{r+m}\right)^q \frac{1}{m}$$

这里需要指出的是,这个条件概率公式与董松挺等(2006)^[23]给出的条件概率公式略有不同。原因是:本文的条件概率公式基于单次购买金额的 Gamma-Gamma 随机模型得到;而董松挺等(2006)^[23]是在 Fader 和 Hardie 等(2005)^[2]的基础上,基于平均购买金额的 Gamma-Gamma 随机模型得到的。

购买时间间隔模型的条件概率公式:

Gupta(1991)^[22]的文章没有给出条件概率公式,只给出了对数似然函数(log-likelihood function)的推导。由于计算机技术的发展,我们没有必要去求出复杂的对数似然函数公式,只需要给出条件概率公式,使用计算机的极大似然估计程序就可以求解。下面我们给出条件概率公式的推导。

由假设:

$$t \sim \text{Exponential}(\lambda), \lambda = \lambda_0 \exp(Y\beta_2), \lambda_0 \sim \text{Gamma}(s, g)$$

因此:

$$f(t | \lambda_0, \beta_2, Y) = \lambda e^{-\lambda t} = \lambda_0 e^{Y\beta_2} \times e^{-\lambda_0 \exp(Y\beta_2) \times t}$$

$$= e^{Y\beta_2} \times \lambda_0 e^{-\lambda_0 \times \exp(Y\beta_2)}$$

由于 λ_0 的概率密度函数为 $g(\lambda_0 | s, g) =$

$$\frac{g^s}{\Gamma(s)} \lambda_0^{s-1} e^{-g\lambda_0}$$

$$f(t | s, g, \beta_2, Y)$$

$$= \int f(t | \lambda_0, \beta_2, Y) g(\lambda_0) d\lambda_0$$

$$= \int e^{Y\beta_2} \times \lambda_0 e^{-\lambda_0 \times \exp(Y\beta_2)} \times \frac{g^s}{\Gamma(s)} \lambda_0^{s-1} e^{-g\lambda_0} d\lambda_0$$

$$= \frac{g^s}{\Gamma(s)} e^{Y\beta_2} \int \lambda_0^s e^{-\lambda_0 \times (g + \exp(Y\beta_2))} d\lambda_0$$

$$= \frac{g^s}{\Gamma(s)} e^{Y\beta_2} \times \frac{\Gamma(s+1)}{(g + te^{Y\beta_2})^{s+1}} \cdot$$

$$\int \frac{(g + te^{Y\beta_2})^{s+1} \lambda_0^s e^{-\lambda_0 \times (g + \exp(Y\beta_2))}}{\Gamma(s+1)} d\lambda_0$$

$$= \frac{g^s}{\Gamma(s)} e^{Y\beta_2} \times \frac{\Gamma(s+1)}{(g + te^{Y\beta_2})^{s+1}}$$

$$= \frac{s}{g} \left(\frac{g}{g + te^{Y\beta_2}} \right)^{s+1} \times e^{Y\beta_2}$$

(注: 这里的 $\frac{(g + te^{Y\beta_2})^{s+1} \lambda_0^s e^{-\lambda_0 \times (g + \exp(Y\beta_2))}}{\Gamma(s+1)}$ 是一个

Gamma 分布的概率密度函数, 因此它对于 λ_0 的积分正好等于 1)

参考文献:

- [1] Richard, C., Jiang, W. N.. A stochastic RFM model [J]. Journal of Interactive Marketing, 1999, 13 (3): 2-12.
- [2] Fader, P. S., Bruce, G. S. H., Lee, K. L.. RFM and CLV: vsing Iso-value curves for customer base analysis [J]. Journal of Marketing Research, 2005, 42 (Nov.): 415-430.
- [3] Schmittlein, D. C., Peterson, R. A.. Customer base analysis: An industrial purchase process application [J]. Marketing Science, 1994, 13 (1): 41-67.
- [4] Ho, T. H., Lim, N., Camerer, C. F.. Modeling the psychology of consumer and firm behavior with behavioral economics [J]. Journal of Marketing Research, 2006, 43(3): 307-331.
- [5] Blattberg, R. C., Eppen, G. D., Lieberman, A. J.. Theoretical and empirical evaluation of price deals for consumer nondurables [J]. Journal of Marketing, 1981, 45(1): 116-129.
- [6] Guadagni, P. M., Little, J. D. C.. A logit model of brand choice calibrated on scanner data [J]. Marketing Science, 1983, 2(3): 203-238.
- [7] Sunil, G.. Impact of sales promotions on when, what, and how much to buy [J]. Journal of Marketing Research, 1988, 25(4): 342-355.
- [8] Yoram, K., Rao, A. G., Shakun, M. F.. A mathematical model for price promotions [J]. Management Science, 1974, 20(6): 948-59.
- [9] Massy, W. F., Frank, R. E.. Short term price and dealing effects in selected market segments [J]. Journal of Marketing Research, 1965, 2(2): 171-185.
- [10] Ward, R. W., Davis, J. E.. A pooled cross-section time series model of coupon promotions [J]. American Journal of Agricultural Economics, 1978, 60 (3): 393-401.
- [11] Sydney, R., Li, T., Nicholls, F.. Impact of situational variables and demographic attributes in two seasons on purchase behaviour [J]. European Journal of Marketing, 2000, 34(9/10): 1167-1180.
- [12] Jones, J. M., Zufryden, F. S.. Adding explanatory variables to a consumer purchase behavior model: An exploratory study [J]. Journal of Marketing Research, 1980, 17(3): 323-334.
- [13] Jones, J. M., Zufryden, F. S.. An approach for assessing demographic and price influences on brand purchase behavior [J]. Journal of Marketing, 1982, 46(1): 36-46.
- [14] Neslin, S. A., Henderson, C., Quelch, Consumer promotions and the acceleration of product purchases [J]. Marketing Science, 1985, 4(2): 147-165.
- [15] Ehrenberg, A. S. C.. The pattern of consumer purchases [J]. Applied Statistics, 1959, 8 (1): 26-41.
- [16] Chatfield, C., Goodhardt, G. J.. A consumer purchasing model with erlang inter-purchase time [J]. Journal of the American Statistical Association, 1973, 68 (Dec.): 828-835.
- [17] Dunn, R., Reader, S., Wrigley, N.. An investigation of the assumptions of the NBD model as applied to purchasing at individual stores [J]. Applied Statistics, 1983, 32 (3): 249-259.
- [18] Herniter, J. A probabilistic market model of purchase timing and brand selection [J]. Management Science, 1971, 18 (4), Application Series, Part 2, Marketing Management Models: 102-113.
- [19] Banerjee, A. K., Bhattacharyya, G. K.. A purchase incidence model with inverse gaussian interpurchase times [J]. Journal of the American Statistical Association, 1976, 71 (Dec.): 823-829.
- [20] Lawrence, R. J.. The lognormal distribution of buying frequency rates [J]. Journal of Marketing Research, 1980, 17 (2): 212-220.
- [21] Jain, D. C., Vilcassim, N. J.. Investigating household purchase timing decisions: A conditional hazard function approach [J]. Marketing Science, 1991, 10 (1): 1-23.
- [22] Sunil, G.. Stochastic models of interpurchase time with time-dependent covariates [J]. Journal of Marketing Research, 1991, 28(1): 1-15.
- [23] 董松挺, 王高, 赵平. 拓展的 Gamma-Gamma 随机模型

- 及其应用 [J]. 中国管理科学, 2006, 14(z1): 52—55.
- [24] Bucklin, R. E., Lattin, J. M.. A two-state model of purchase incidence and brand choice [J]. Marketing Science, 1991, 10(1): 24—39.
- [25] Dick, A. S., Basu, K.. Customer loyalty: toward an integrated conceptual framework [J]. Journal of the Academy of Marketing Science, 1994, 22(2): 99—113.

Decomposition of Total Store Purchase When Purchase Payment and Inter-Purchase Time are Interdependent

DONG Song-ting¹, ZHAO Ping¹, WANG Gao², LIU Rui-ming³

(1. School of Economics and Management, Tsinghua University, Beijing 100084, China;

2. China Europe International Business School, Shanghai 201206, China;

3. College of Management, University of Massachusetts Boston, MA 02125, American)

Abstract: The assumption that single purchase payment and inter-purchase time are independent is inconsistent with real purchase behaviors. The interdependence between the two variables may be concealed by a seemingly insignificant correlation coefficient. Therefore, the decomposition of the total store purchase based on the independence assumption could not represent the real purchase pattern accurately. Based on the analysis of consumer purchase behaviors, we develop a decompositional model which takes in the interdependence between single purchase payment and inter-purchase time, and the impact of marketing efforts. Our empirical analysis shows the effectiveness and managerial implications of the model in analyzing customers' total store purchases.

Key words: total store purchase decomposition; purchase payment; inter-purchase time