

参赛队员姓名：俞琳薇

中学：福建省厦门第一中学

省份：福建省

国家/地区：中国

指导教师姓名：陈海强，王君黛

论文题目：

双十一购物节价格折扣如何影响线上商家  
销售表现—基于大数据的实证分析

Title:

How Price Discount Affects the Sales of Online  
Retailers during Double 11 Shopping Festival:  
an Empirical Analysis Using Big Data

# 双十一购物节价格折扣如何影响线上商家销售表现

## 一 基于大数据的实证分析

俞琳薇

福建省厦门第一中学

**摘要:** 双十一购物狂欢节是国内影响最大的网上购物节日, 商家往往通过打折等促销方式来吸引消费者购物。然而, 在大部分商家均采用打折的同质策略下, 折扣如何影响销售表现是一个非常具有现实意义的问题。而已有媒体也报道部分商家在双十一期间利用虚假折扣吸引消费者购物, 损害消费者利益, 本文利用数据分析来探究虚假折扣是否存在, 并分析消费者是否对虚假折扣行为有所甄别从而减少相关消费。具体而言, 本文利用网络爬虫技术获取了2018年双十一前后天猫、京东以及苏宁三大网购平台近40万商品大数据, 通过回归分析来研究折扣与商品销售增长率之间的关系。主要发现如下: 1) 商家普遍在双十一期间采取打折促销策略, 平均折扣为0.961, 其中男装、女装及户外运动提供折扣最大; 2) 折扣对销售具有显著提升作用, 平均而言, 折扣增加10%能带来22.23%的销售增长, 户外运动最为受益(增长29.56%), 手机最小(12.43%); 3) 部分商家采用了不同形式的虚假折扣行为, 消费者经过价格对比能甄别部分双十一提前涨价制造的虚假折扣, 然而往往掉入“折扣时限”陷阱, 无法预测商家双十一之后是否延长甚至采用更优惠的折扣; 4) 销量增长与同品牌下其他产品销量以及相似产品折扣力度存在正相关, 说明双十一带来的巨大网络流量使得商品销售之间更多体现互补效应; 5) 大牌产品以及评论关注较多的热门产品更受益于双十一折扣营销策略。本文研究发现对商家如何优化双十一营销策略, 消费者如何在双十一期间理性消费具有重要意义。

**关键词:** 双十一网络购物节; 虚假折扣; 词云分析; 互补效应

### 一、 引言

双十一购物狂欢节指每年11月11日的网络购物促销日, 最早源于阿里巴巴淘宝商城2009年11月11日举办的网络促销活动。经过10年的发展, 双十一已成为中国乃至全球电商行业的年度盛事。据阿里巴巴公开的统计数据, 其旗下天猫和淘宝两个电商平台从2009年的0.52亿交易量, 仅27个品牌的参与, 发展到2018年2135亿交易额, 全球多达18万个品牌和商家参与的巨大规模。如果计算其他平台, 2018年双十一期间, 国内22家电商平台的交易总额达3143.2亿, 其中天猫占比67.9%, 京东17.3%, 苏宁4.7%。消费者青睐的商品主要包括手机、家电以及服装等商品, 而参与双十一购物最活跃的地区也是国内经济最发达、互联网经济最活跃的地区, 分别是广东, 江苏, 浙江, 上海和北京。

为吸引消费者购物, 线上商家往往在双十一当天采取提供折扣、新品上市以及举办各种优惠促销活动。线上商家一方面通过促销活动来提高商品销售量, 另一方面也寄希望通过双十一巨大网络流量来提高品牌知名度。线下调查数据表明, 消费者参与双十一购物的最大驱动力依然还是商家提供的价格优惠力度, 消费者在面对巨大折扣诱惑下, 容易产生冲动性消

费, 尤其是线上购物的环境, 消费者往往更容易产生冲动性消费<sup>1</sup>。Hausman (2000) 利用美国的数据发现零售产业总体规模的 30%~50% 来自于冲动购物, 且 90% 的消费者存在冲动购物行为。Brashear et al. (2009) 利用包括中国在内的多个国家和地区的调查数据发现网络购物者更容易出现冲动性购物行为。Madhavaram and Laverie (2004) 发现网络购物刺激消费者冲动性消费行为的一个重要原因是购物者通过网页浏览商品产生身临其境的感觉, 而网络购物环境也使得消费者更快速的做出购物决定。也有研究发现, 线上商家提供的价格折扣也是刺激消费者网上冲动性消费的一个主要原因。Biseas 和 Burton (1993) 认为价格折扣往往会让消费者产生“赚到了”的心理, 刺激其作出购物决定。Zhou 和 Wong (2003) 通过研究发现商品价格是刺激网络消费者冲动购买的最重要因素。最后, 铺天盖地的广告信息、热烈的购物氛围、商家故意制造的折扣时间限制以及同伴购物行为等因素也刺激了消费者线上的冲动性消费 (Agee 和 Martin, 2001; Beatty 和 Ferrell, 1998)。以上研究大多数基于欧美数据, 而中国零售业, 尤其是网络零售业, 近十年来也快速发展, 已经成为经济增长的重要驱动力之一。双十一作为国内最重要的网上购物节, 给我们提供了一个非常独特的视角来分析中国消费者的购物行为。

除了对消费者的分析, 双十一也能够让我们对线上商家行为展开研究。近年来, 不少媒体报道, 部分商家通过在双十一之前提高价格来在购物节当天造成巨大折扣的假象, 同时也有商家提供的折扣与其宣传并不相符。2015 年, 中国消费者协会调查报告显示, 在双十一节日之后, 以 11 月 18 日和 12 月 5 日的价格作为基准, 发现仍然有接近 30% 的商品能够以双十一价格甚至更低的价格购买。虚假折扣宣传误导消费者, 损害了消费者权益。中国商务部在 2016 年警告相关商家, 反对这些商家虚假宣传对消费者造成的错误引导。因此, 本文也利用爬取的大数据来分析商家的行为。首先, 我们针对不同类型的商家, 包括产品类型、价格高低、品牌吸引力等, 分析其双十一商品折扣如何影响其商品销量。此外, 本文根据消费者协会的描述, 首次量化定义了狭义和广义的虚假折扣行为, 并分析消费者是否能够识别商家的虚假折扣行为而避免受骗。

国内关于网络购物的研究大多关注在线评论对消费者购物决策产生影响, 并最终如何影响商家的销售 (Berger et al., 2010, 郝媛媛等, 2009; 卢向华和冯越, 2009; 殷国鹏, 2012)。关于双十一购物节的研究并不多见, 谢琦、邵晓峰 (2018) 指出折扣力度越大双十一当日销量越高, 且消费者对价值较高 (如珠宝类) 的商品的折扣力度更敏感, 同时商品的热门程度对双十一当天的销量有显著的正向影响。殷猛等 (2019) 实证研究发现, 双十一购物节期间消费者的消费行为具有羊群效应, 从众行为会显著影响消费者的冲动购物倾向。

本文利用网络爬虫获取了 2018 年双十一前后天猫、京东以及苏宁购物平台畅销商品类别的价格和销售大数据, 并通过回归分析来研究商家折扣行为与双十一商品的销售表现之间的关系。具体发现如下: 首先, 商家普遍在双十一期间采取打折促销策略, 所有观察的商品中平均折扣为 0.96 左右, 而男装、女装及户外运动提供折扣最大; 其次, 价格折扣对销售具有显著提升作用, 平均而言, 折扣增加 10% 能够带来 22.23% 消费量增长, 而户外运动最为受益 (增长 29.56%), 手机受益最小 (12.43%); 第三, 部分商家存在虚假折扣行为, 然而, 对双十一提前涨价而造成的虚假折扣, 本文定义为狭义的虚假折扣, 消费者能够有效识别, 该类虚假折扣并未带来额外的销售增长, 然而对于双十一之后一段时间依然采用同等甚至更低的虚假折扣行为, 消费者难以预期, 因此, 该类长期折扣的行为带来了额外销售增长; 第四, 商品销量增长与同一品牌下的销量以及相似产品提供的折扣存在正相关, 说明双十一活动带来的巨大流量使得商品销售之间存在互补效应; 第五, 大牌产品以及评论关注较多的

---

<sup>1</sup>冲动性购物主要指的是消费者购买之前并没有详细的购买计划, 而是由于一些外生或内生的因素刺激了其不计后果的消费 (Rook, 1987; 李秀荣、梁承磊)。

热门产品更受益于双十一折扣营销策略, 说明国内消费者在双十一购物决策时, 同时注重产品质量与价格优惠。

本文研究具有一定的理论和现实意义。首先, 我们发现虚假折扣行为显著存在于双十一购物节中, 部分虚假折扣确实能够带来销售表现的增加, 说明消费者在双十一购物过程中存在“抢时限”、“占便宜”的冲动购物行为。此外, 本文研究对商家如何优化双十一的营销策略也有实践意义, 折扣对销售表现的影响取决于商品类型, 是否大牌以及商品受消费者欢迎程度等因素, 而基于本文建立的销售增长模型, 商家可以进一步优化其营销策略。

本文余下的结构如下: 第二部分为数据描述与初步分析, 第三部分为回归分析结果, 最后一部分为结论与建议。

## 二、数据描述和初步分析

### 一、数据来源与描述

本文所使用的线上商品零售大数据是利用 pypspider 爬虫技术在 2018 年 11 月 3 日起至 12 月 3 日期间对天猫商场 (Tmall), 京东商城 (JD), 苏宁易购 (Suning) 三大电商平台的商品信息进行爬取得到的。根据市场调研公司 eMarketer 在 2018 年发布的调查报告与各公司财报数据显示, 截止至 2018 年, 这三家电子商务公司分别包揽了中国电商市场的 58.2%, 16.3%, 1.9% 的销售份额; 从消费活动的参与情况出发, 阿里巴巴集团旗下的天猫与淘宝一体化销售平台拥有近 6 亿活跃消费者与 15 万家以上的合作品牌, 而京东与苏宁易购分别拥有 3.21 亿活跃消费者与 10 万以上的签约商户、4.07 亿注册用户数与遍布全国的 1 万 1 千家自营与加盟店; 天猫商城在 2018 年度双十一购物节当天成交额达到 2135 亿元, 京东总成交额为 1598 亿元, 而苏宁易购全渠道销售同比增长 132%, 拼购订单突破 8000 万单, 三平台均在双十一购物节期间表现强势。因此, 本文选取的三大电商平台具有交易体量大, 商品丰富, 使用群体广泛, 能够代表国内线上购物平台的整体情况。

本文数据记录了三大平台近 40 万件商品在 2018 年 11 月 3 日至 2018 年 12 月 3 日之间商品价格和销售情况, 观测对象来自 21613 家不同线上商家, 涉及 13706 个品牌<sup>2</sup>, 涵盖了双十一最热门的八大类商品大类, 即彩妆、大家电、男装、女装、手机、玩具、小家电、户外运动, 并按照平台的分类方式进一步细分至 25 个中类以及 836 个小类。具体来说, 商品样本赛选标准如下, 首先根据各个平台网页识别出 8 个大类里面所有小类, 然后基于各个小类根据销量或评论数选择出排名前 50 的品牌, 并在每个品牌里面选择出排名前 30 的商品。图 1 展示了采集的商品在各个大类中的分布, 观测商品数量最多的前三类依次是户外运动 (30.1%)、女装 (26.7%)、男装 (21.4%)。

本文数据集包含了在线销售商品的一系列特征变量, 包括商品在各个平台的唯一识别码、商品名称、最近 30 日累计销售量<sup>3</sup>、价格、累计评论数以及商品文本描述信息。其中, 价格为折扣之后的价格<sup>4</sup>。本文所使用的数据主要有以下优点: 首先, 本文所使用的数据集涵盖商品数量教大, 覆盖了双十一当天以及前后一段时间的商品价格与销售信息, 能较好的捕捉购物节所带动的消费增长; 其次, 数据集包含了商品丰富的文字描述, 使得本文可以通过自然语言处理, 即文本分析方法对商品进行聚类, 寻找相似商品, 从而可以研究相似商品折扣对销量的交叉影响。最后, 数据集包含了丰富的商品特征信息, 可以帮助我们在个体层面上

<sup>2</sup> 由于网络不稳定以及网上商城的并发容量问题, 爬取的商品信息在某些时段存在一定缺失。

<sup>3</sup> 数据集中只有天猫平台提供了最近 30 日销售额, 因此关于销售表现的回归分析主要基于天猫的商品, 关于价格及文本分析则基于三个平台的商品。

<sup>4</sup> 本文在计算折扣时并没有考虑平台提供的其他折扣类型, 比如买 500 送 100 的门槛性折扣, 因此, 本文计算的折扣与消费者真实获得的折扣存在一定低估。

进行异质性分析。

图 2 和图 3 分别展示了男装和女装在热点词汇方面的异同, 其中男装比较热门的词汇是秋季、休闲、商务、流行、青春等, 而女装则包含秋季、通勤、长袖、修身、纯色等, 反映了双十一购物节期间男装和女装不同的流行特征。

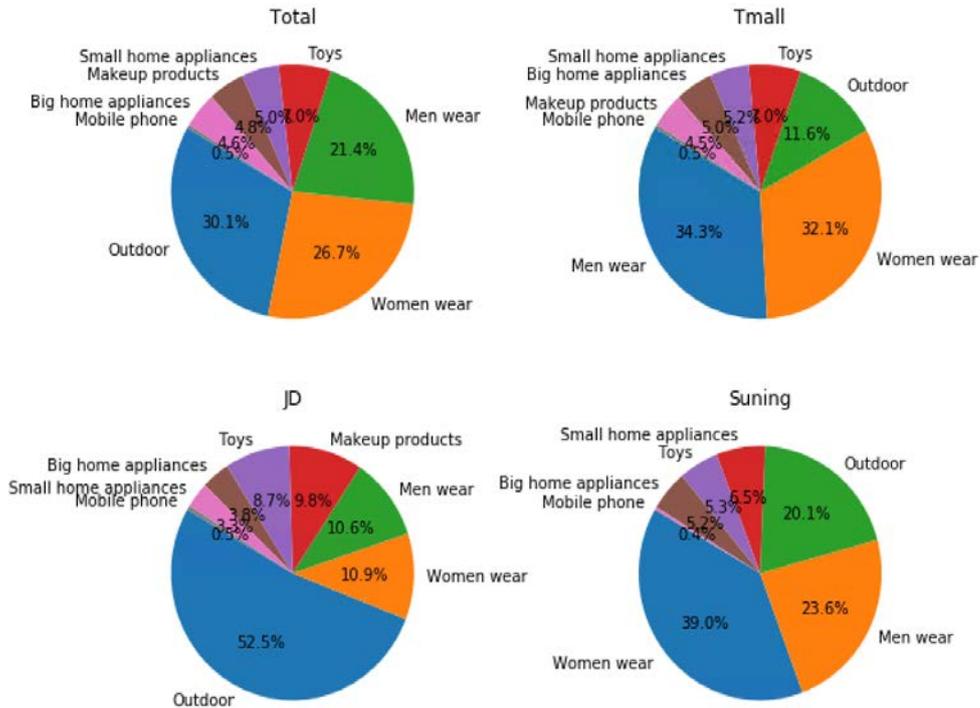


图 1 各平台商品数量分布图



图 2 男装大类商品词云分析



图 3 女装商品大类词云分析

## 二、变量构建及初步统计分析

### (一) 销量增长率:

天猫商城商品销量数据为访问网页当日起前三十日累积销售量, 且天猫后台对该信息按照分钟级别进行刷新, 考虑到双十一购物节影响存在滞后效应, 本文关注双十一当天以及随后一周 (11.11-11.17) 的销售增长率, 具体定义为 11.17 日当天的累积 30 日销量去对数, 减去双十一前一天 (11.10) 的累积 30 日销量, 即双十一当周相对于上一个月同一周的同比销量增长率。图 4 展示了天猫商城爬取的额商品销量增长率直方分布图。表 1 进一步展示了相关统计描述。从表 1 可以看出, 双十一当周销量平均增长达 80.6%, 中位数为 65%, 各大类表现各异, 其中最高的是女装和男装, 分别达到 105.7% 和 98.6%, 增长最少的是手机为 26.4%。

也有部分商品增速为负值, 说明缺乏竞争力的产品销售并没有获得双十一购物节效应的正面带动。

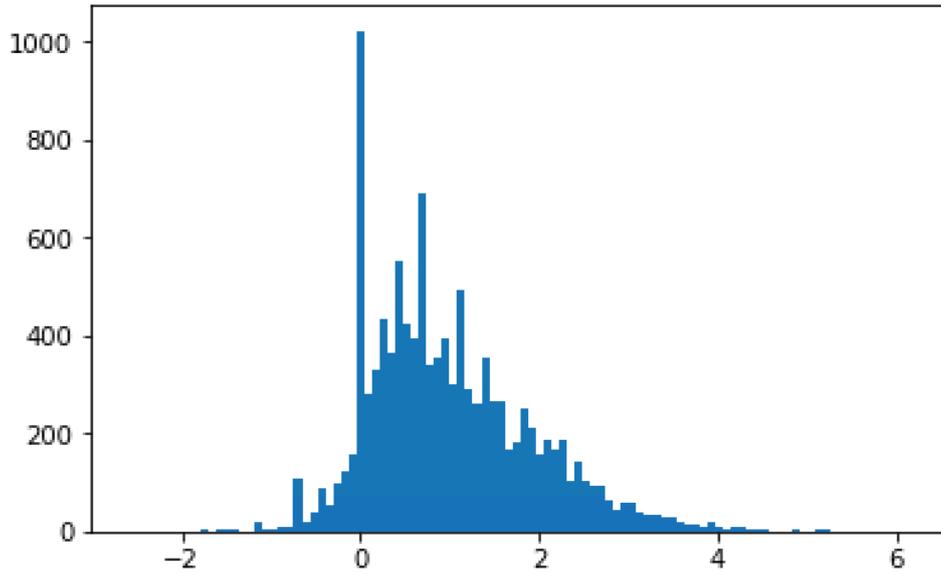


图 4: 双十一销售增长率分布图

表 1: 天猫购物平台商品销售量增长率描述统计

商品类别	(1) 商品样本量	(2) 均值	(3) 标准差	(4) 中位数	(5) 最小值	(6) 最大值
彩妆	10,628	0.783	0.791	0.606	-1.609	7.969
大家电	9,261	0.695	0.730	0.581	-3.336	6.230
男装	8,687	0.986	0.948	0.811	-2.580	6.153
女装	10,637	1.057	0.916	0.881	-1.558	5.862
手机	632	0.264	0.419	0.223	-1.754	3.045
玩具	13,212	0.662	0.719	0.504	-2.457	6.731
小家电	11,158	0.673	0.731	0.560	-4.557	7.110
户外运动	15,896	0.853	0.881	0.693	-3.100	7.022
所有样本	80,111	0.806	0.831	0.650	-4.557	7.969

## (二) 商品折扣

本文关注的解释变量为商品折扣, 而商品折扣的计算取决于基准价。常见的折扣主要以商品挂牌价来计算折扣, 然而线上购物在没有获取商品之前很难爬取到相关挂牌价, 因此我们在计算商品折扣的时候, 更多通过采用商品价格历史数据作为基准价来计算折扣。采用该方法也能真实反映双十一商家让利幅度以及消费者双十一购物获取的实惠。

基于稳健性考虑, 本文采取了两种折扣计算方法: 第一种折扣采用双十一前一天价格作为基准价, 计算得到即时折扣 (Discount), 这种折扣定义方法反映了双十一购物节各个线上平台要求商家在 11 日零点统一调价的规则; 第二种折扣方法使用双十一之前一段时间平均价格作为基准价来计算得到平均折扣 (Mean\_discount)。本文数据爬取的最初时间为 11.3 日, 因此采用 11.3 至 11.10 日平均价格来计算折扣。表 1 展示了三大购物平台价格数据完整的商品的即时折扣统计描述, 平均而言, 折扣为 0.961 (即打折 3.9%)。折扣在各个商品大类中

存在较大差异。其中,折扣力度最大的为男装,其按照最后一天来计算的及时折扣为 0.883,即打折 11.7%左右,而女装平均折扣为 0.897,即打折 10.3%左右。如果将折扣与表 1 销售表现联系起来,男装、女装以及户外运动折扣力度最大,而销售增长也最显著。

图 5 进一步展示各大类商品即时折扣和平均折扣对比,两个折扣存在一定的差距。

表 2: 三大平台商品即时折扣统计描述

商品类别	(1) 商品样本量	(2) 均值	(3) 标准差	(4) 中位数	(5) 最小值	(6) 最大值
彩妆	37,045	0.963	0.257	1.000	0.110	33.933
大家电	59,813	0.985	0.878	1.000	0.010	193.760
男装	23,606	0.883	0.190	0.940	0.109	3.773
女装	25,680	0.897	0.407	0.900	0.090	7.502
手机	4095	0.984	0.067	1.000	0.004	2.144
玩具	62,127	0.975	0.169	1.000	0.050	18.000
小家电	62,403	0.982	0.128	1.000	0.023	13.243
户外运动	121,499	0.957	0.141	1.000	0.057	16.040
所有样本	396,268	0.961	0.386	1.000	0.004	193.760

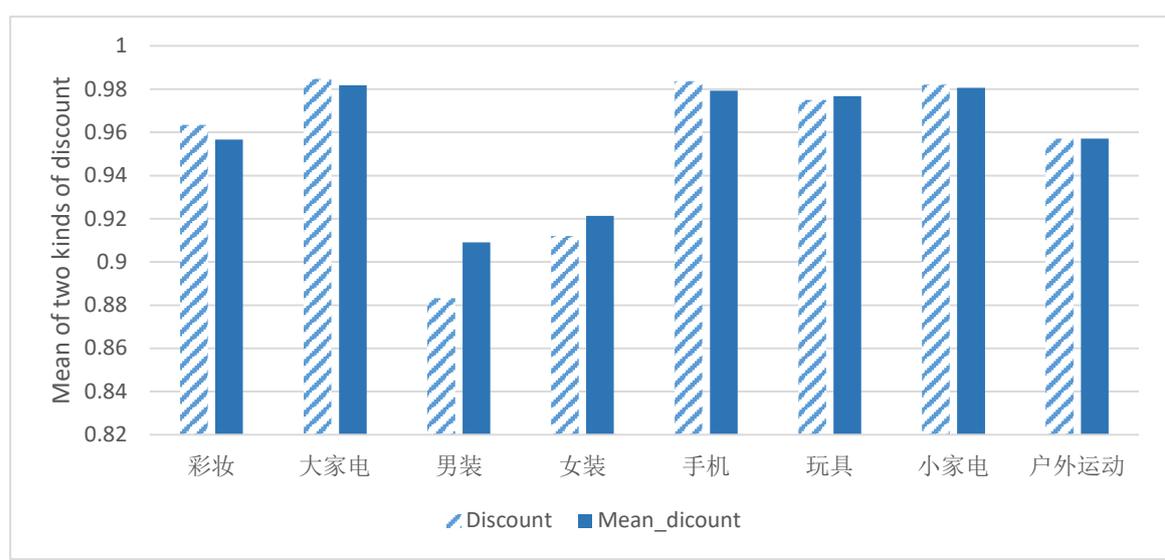


图 5: 各大类商品两种折扣均值比较

### (三) 虚假折扣虚拟变量

本文也关注商家是否存在虚假折扣行为。然而虚假折扣并没有一致定义,根据 2015 年中国消费者协会监测报告描述,部分商家以商品“专柜价”或“厂家建议价”为基准价进行大幅打折,但是其专柜价、建议价(原价)存在随意标注的情况,涉嫌虚假折扣;此外双十一期间,一些电商平台以“清仓价”、“限时抢购”、“史无前例,仅此一天”、“血拼 24 小时”、“逆天低价”、“双十一最省钱”等各种价格宣传形式误导消费者,然而相关产品在双十一之后能够继续以相关价格甚至更低价格能够购买到,消费者所谓的“双十一”实惠并不存在,涉嫌虚假折扣。监测报告以“双十一”当天价格和折扣为基准,追踪了 11 月 18 日和 12 月 5 日两个时间节点的商品价格,发现被监测的 94 种商品中,有 32 款能以不高于“双十一”的价格购买。而部分媒体报道的虚假折扣描述为部分商家在双十一之前提高价格从而达到双十一当天大幅打折的假象。因此,本文采取了两种方式来说明定义虚假折扣。

**狭义的虚假折扣:** 商品在 11.10 日或之前的一段时间内采取提价来制造双十一当天大幅

打折, 几位虚假折扣; 其虚假折扣的程度为前文描述的平均折扣与即时折扣的差值, 如果即时折扣大于平均折扣, 则认为该商品在双十一之前有明显的提价行为, 商品被判定为存在虚假折扣。

**广义的虚假折扣:** 双十一活动结束后之后商品的价格并没有回到原有正常水平, 如果 11.18 日后的商品平均价格 (正常价格水平) 小于 11.11 日前的平均价格, 本文则认为该商品存在虚假折扣, 其虚假折扣程度为二者之差。

基于上述定义, 图 6 展示了爬取商品中存在虚假折扣行为的 4 个典型案例。

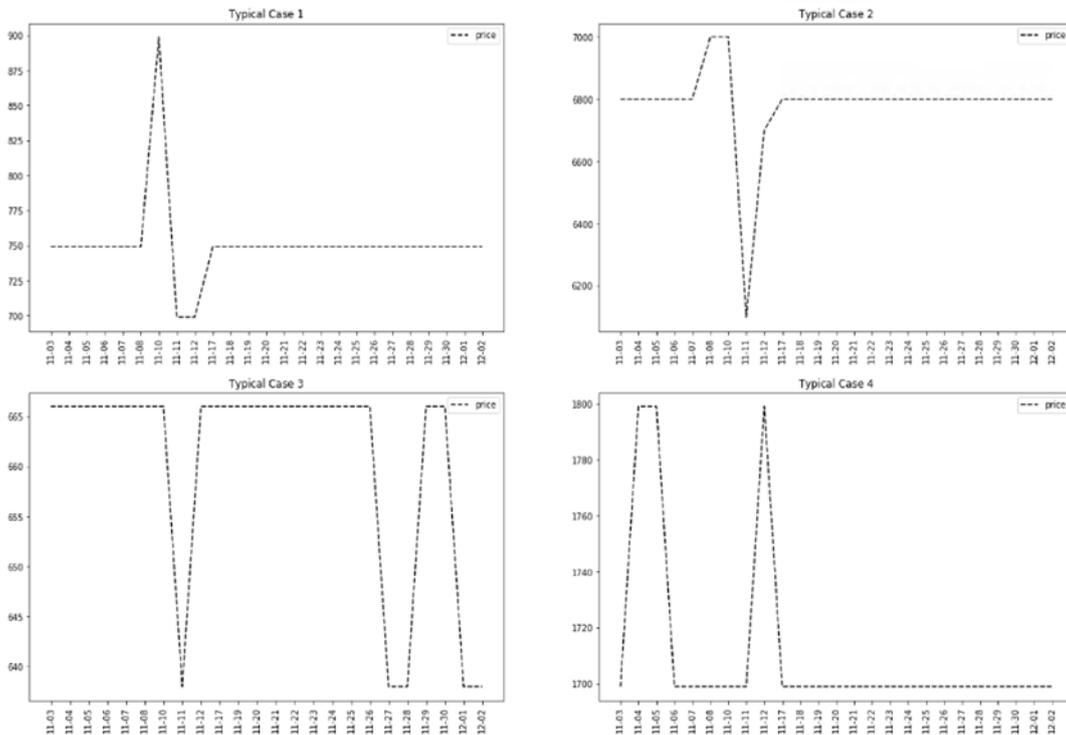


图 6: 虚假折扣的案例

其中案例 1 为厨卫电器中某品牌燃气灶, 其采取了在 11.10 当天提高价格的行为来制造大幅打折来吸引消费者。案例 2 为某品牌酒柜, 其采取了类似的首发, 在 11.8 日就开始提高价格来制造虚假折扣, 而案例 3 为某品牌冰箱, 其采取的策略是双十一打折, 但是在 11 月月底之后继续采用了同样的折扣, 而案例 4 为某品牌燃气灶, 其 11.6-11.11 期间打折, 在双十一之后短暂取消折扣之后, 在 11.17 开始继续使用双十一购物节的价格来销售产品。

表 4 和表 5 分别报告两种方法得到的虚假折扣虚拟变量 (即存在虚假折扣行为的商品定义为 1, 否则为 0) 的描述统计。按照狭义定义方式, 大概有 8.7% 的商品存在虚假折扣的行为, 最突出的是女装, 男装和彩妆三大类。按照广义定义方式, 大概有 20% 的商品存在虚假折扣的行为, 最突出的是女装, 男装和手机三大类<sup>5</sup>。本文在回归分析中同时采用了狭义和广义的虚假折扣定义。

表 4: 狭义虚假折扣虚拟变量的描述统计

	(1)	(2)	(3)	(4)	(5)	(6)
商品类别	样本量	均值	标准差	中位数	最小值	最大值
彩妆	35,369	0.131	0.337	0	0	1
大家电	58,190	0.085	0.278	0	0	1

<sup>5</sup>两种定义方式下样本量不一样主要是因为双十一期间网络不稳定导致数据存在一定的遗漏。

男装	22,402	0.093	0.290	0	0	1
女装	31,600	0.151	0.358	0	0	1
手机	3,865	0.073	0.260	0	0	1
玩具	58,134	0.079	0.270	0	0	1
小家电	60,695	0.063	0.243	0	0	1
户外运动	104,215	0.073	0.259	0	0	1
全样本	374,470	0.087	0.282	0	0	1

表 5: 广义虚假折扣虚拟变量的描述统计

商品类别	(1) 样本量	(2) 均值	(3) 标准差	(4) 中位数	(5) 最小值	(6) 最大值
彩妆	43,605	0.195	0.396	0	0	1
大家电	63,311	0.157	0.364	0	0	1
男装	31,035	0.314	0.464	0	0	1
女装	42,335	0.319	0.466	0	0	1
手机	4,206	0.321	0.467	0	0	1
玩具	69,573	0.197	0.398	0	0	1
小家电	68,174	0.168	0.374	0	0	1
户外运动	128,658	0.172	0.377	0	0	1
全样本	450,897	0.200	0.400	0	0	1

#### (四) 相似商品定义

商品折扣对商品销量的影响可能受到相似商品的影响。双十一期间,相似商品可能会从以下两方面影响商品销售表现。首先是替代效应,消费者货比三家,选择最优惠的相似产品购买;其次是流量效应,相似商品的折扣吸引更多消费者购买同类产品。为检验商品之间的折扣的交叉影响,本文对每一个中类下的产品之间建立相似性测度,并按照相似测度选取排名前 10 的商品的平均折扣作为该商品面对的相似产品价格折扣。

具体而言,针对爬取获得的商品描述文本,本文使用基于规则的词袋模型来定义商品之间的相似度,词袋模型也称文本空间向量模型(VSM),是由 Salton (1974)提出的将文本数据结构化的经典方法,后来经过 Tsai(2012)进一步拓展到分析图片等复杂情形。词袋模型在建模过程中忽略了文本顺序,适合对数据集中无序的商品描述进行建模。本文

首先使用 Python 提供的分词算法将每条商品的基本描述进行分词处理,生成词-矩阵位置字典,并处理得到每条文本描述的数值化向量。而商品 1 与商品 2 之间的相似度定义如下:

$$\text{相似度} = (\text{商品 1 与商品 2 重复分词数目}) / (\text{商品 1 丰富度} + \text{商品 2 丰富度})$$

根据上述定义,如果两商品重复词出现次数相对比例越高,则认为两件商品相似度越高。表 6 也展示了两个商品的相似度计算示例。

表 6 相似度计算示例

	商品 1	商品 2
初始商品描述	免洗牛仔裤, 卡其色做旧, 乞丐风格	通勤工装裤, 涤纶化纤, 卡其色
分词组合	免洗、牛仔裤、卡其色、做旧、乞丐、风格	通勤, 工装裤, 涤纶, 化纤, 卡其色
词袋向量	[1, 1, 1, 1, 1, 1, 0, 0, 0, 0]	[0, 0, 1, 0, 0, 0, 1, 1, 1, 1]

丰富度	6	5
重复词数目	1	1
相似度	1/11	

注：生成的词袋模型词典为：{"免洗":1, "牛仔裤":2, "卡其色":3, "做旧":4, "乞丐":5, "风格":6, "通勤":7, "工装裤":8, "涤纶":9, "化纤":10}, 并由此生成描述向量。

### （五）商品热度

本文基于爬虫数据中双十一前商品评论数排名来构造商品热度变量, 如果该商品的评论数为 0, 则定义其热度值为 0。然后, 在同一个商品中类下, 如果商品的评论数小于该商品所属中类所有商品评论数的下三分位数, 则其热度变量定义为 1, 并依次定义热度值 2 (上下三分位数之间), 热度值 3 (上三分位之上), 商品热度随数值递增。

我们也基于每一个前面所述的词云分析, 寻找出每个中类前 50 的关键词, 然后计算出每一个商品描述中包含前 50 关键词的个数作为商品热度的另一个测度。

## 三、回归分析

为了分析商品折扣对商品销售表现的影响, 本文做如下回归分析

$$sales\_growth_{ij} = \beta discount_{ij} + X_{ij}\gamma + \delta_j + e_{ij} \quad (1)$$

其中  $sales\_growth_{ij}$  表示第  $j$  类商品中第  $i$  种产品双十一销售增长率,  $discount_{ij}$  为双十一商品折扣,  $X_{ij}$  是影响销量变动其他产品相关特征, 如产品网络评论总数<sup>6</sup>, 产品双十一之前的平均价格和平均销量, 产品所属品牌的平均价格和平均销量, 相似产品平均折扣力度, 热门词汇数量等。  $\delta_j$  是商品中类固定效应, 用于捕捉为观察到的商品类型层面的综合因素对产品销量的影响。  $e_{ij}$  是不可观测的异质性。由于同类产品间具有相关性, 我们将标准误计算在商品类别层面进行聚类调整, 相关变量的定义描述见表 7。

表 7 变量定义和计算方法

特征对象	特征维度	变量名字	变量解释
商品	增长	$sales\_growth$	$\ln(11.17 \text{ 的月销量}) - \ln(11.10 \text{ 的月销量})$
	折扣	$discount$	11.11 的价格/11.10 的价格
	销量	$sale\_bef$	商品在 11.3 到 11.10 所有观测的平均月销量
	价格	$price\_bef$	商品在 11.3 到 11.10 所有观测的平均价格

<sup>6</sup> 由于无法获得具体每条评论的细节, 因此本文仅考虑了评论的条数。

	虚假折扣	<i>fake</i>	广义虚假折扣定义为: (商品在 11.3 到 11.10 之间的均价 - 商品在 11.17 到 12.3 所有观测的商品均价)/商品在 11.17 到 12.3 所有观测的商品均价; 狭义虚假折扣定义为: 11.3 - 11.10 的平均价格为基准计算的折扣 - 11.10 价格为基准计算的即时折扣。
	相似商品	<i>sim_disc</i>	与该商品相似度排前 10 的商品的折扣均值
	评论	<i>comment</i>	商品在 11.3 到 11.10 所有观测的评论数均值
	热词	<i>topword</i>	该商品文字描述中出现的热门词总数
	热度	<i>ave_bef</i>	按照商品评论变量的在中类商品所处分位数定义 1, 2, 3, 热度依次递增; 如果商品 11 月 10 日之前商品评论的均值为 0, 则该变量等于 0;
品牌	价格	<i>brand_price</i>	该品牌下属所有商品在 11.3 到 11.10 的价格平均值
	销量	<i>brand_sale</i>	该品牌下属所有商品在 11.3 到 11.10 的月销量的均值
	大牌	<i>brand</i>	该品牌所有商品在 11.3 到 11.10 的平均价格在其所属中类中高于在该类下所有品牌平均价格的中位数, 则定义为 1, 否则为 0

表 8 报告了全样本的回归结果。其中, 折扣采用的是 11.10 价格作为基准计算的即时折扣。回归结果显示, 在所有回归中, 折扣对应的估计系数显著为负, 说明商家提供的折扣越大 (即 *discount* 越小), 商品销售增长率越高。根据第 (8) 列完整模型的回归结果, 平均而言, 折扣每增加 10%, 销售增长率提高 22.23%。本文也发现商品销售增长与同一品牌在双十一之前的销量和平均价格显著正相关, 其原因可能来自于同品牌下的辐射效应。同时, 商品本身或者品牌平均价格越高, 双十一销量增长率更高, 其原因是同等程度的折扣在价格较高的情形下给消费者带来更多“占便宜”的感觉。此外, 本文也发现商品销量与代表商品热度的评论数有正向关系, 说明受到关注的商品在双十一期间销售增长更快。最后, 相似产品折扣力度对应的系数显著为负, 说明相似产品之间的交叉影响更多受益于互联网巨大流量带来的“互补效应”, 而非经典消费理论预测的相似产品之间的“替代效应”。

表 8: 商品折扣与销售增长率回归分析 (全样本)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Sales_growth</i>							
<i>discount</i>	-2.519*** (0.130)	-2.505*** (0.135)	-2.296*** (0.144)	-2.220*** (0.159)	-2.336*** (0.154)	-2.262*** (0.146)	-2.220*** (0.138)	-2.223*** (0.138)
<i>brand_sale</i>		0.011 (0.008)	0.037*** (0.009)	0.041*** (0.008)	0.091*** (0.011)	0.089*** (0.012)	0.086*** (0.011)	0.086*** (0.011)
<i>brand_price</i>			0.158*** (0.017)	0.084*** (0.018)	0.096*** (0.018)	0.087*** (0.018)	0.088*** (0.017)	0.088*** (0.017)

<i>price_bef</i>				0.090***	0.060***	0.081***	0.075***	0.075***
				(0.020)	(0.018)	(0.017)	(0.016)	(0.016)
<i>sale_bef</i>					-0.091***	-0.146***	-0.145***	-0.145***
					(0.009)	(0.011)	(0.010)	(0.010)
<i>comment</i>						0.061***	0.062***	0.062***
						(0.008)	(0.008)	(0.008)
<i>sim_disc</i>							-0.141**	-0.142**
							(0.069)	(0.069)
<i>topword#</i>								-0.017
								(0.014)
<i>Constant</i>	2.933***	2.868***	1.845***	1.711***	1.989***	1.788***	1.910***	1.927***
	(0.118)	(0.138)	(0.192)	(0.209)	(0.200)	(0.188)	(0.213)	(0.213)
<i>Sample#</i>	79,608	79,608	79,608	79,608	79,608	76,274	73,935	73,935
<i>R<sup>2</sup></i>	0.204	0.205	0.236	0.241	0.276	0.299	0.297	0.297
<i>midclass</i>	X	X	X	X	X	X	X	X

注：回归变量解释和计算方法见表 7，系数下方括号报告了标准误，显著性注释为\*\*\* p<0.01, \*\* p<0.05, \* p<0.1。

为进一步刻画各个商品类型折扣与销售增长之间关系的不同，表 9 报告了分 8 大商品类型回归的结果，可以发现价格折扣带来的销售增长表现差异较大。其中，运动户外更加受益于打折力度，其次是男装、玩具等，而女装受益于打折力度反而较小。其原因一方面是因为运动户外、男装以及玩具等，更加耐用，且产品质量更加透明稳定，所以消费者对价格折扣更敏感。女装消费者可能更重视体验，更加受流行时尚的影响，价格透明度较低，因此价格折扣带来的销售增长相对较少。还有一个可能原因是，女装存在虚假折扣较多，双十一折扣相对于平时并没有太多额外的优惠。

表 9：商品折扣与销售增长率回归分析（按商品类别分组）

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	男装	彩妆	大家电	女装	手机	玩具	小家电	运动户外
<i>discount</i>	-2.109***	-1.713***	-1.954***	-1.485***	-1.243***	-2.332***	-1.849***	-2.956**
	(0.322)	(0.090)	(0.061)	(0.097)	(0.305)	(0.302)	(0.260)	(0.299)
<i>brand_sale</i>	0.083**	0.057**	0.120***	0.201**	0.034	0.053***	0.054*	0.093**
	(0.028)	(0.016)	(0.017)	(0.051)	(0.025)	(0.011)	(0.023)	(0.011)
<i>brand_price</i>	0.105**	0.088	-0.181	0.072	0.009	0.138**	0.043	0.101**
	(0.025)	(0.056)	(0.086)	(0.047)	(0.029)	(0.037)	(0.025)	(0.020)
<i>price_bef</i>	0.109**	0.069	0.457***	0.207**	0.043*	0.026	0.029**	0.061**
	(0.026)	(0.037)	(0.078)	(0.041)	(0.025)	(0.022)	(0.007)	(0.013)
<i>sale_bef</i>	-0.126***	-0.134***	-0.152***	-0.172**	0.009	-0.134***	-0.116***	-0.195***
	(0.019)	(0.016)	(0.013)	(0.034)	(0.014)	(0.010)	(0.024)	(0.006)
<i>comment</i>	0.055**	0.105***	0.033**	0.033	-0.029***	0.066***	0.050**	0.097***
	(0.012)	(0.017)	(0.010)	(0.018)	(0.011)	(0.012)	(0.018)	(0.008)
<i>sim_disc</i>	0.225	0.29	-0.04	-0.075	-0.351	-0.007	-0.573	-0.569*

	(0.182)	(0.157)	(0.152)	(0.091)	(0.584)	(0.058)	(0.346)	(0.187)
<i>topword#</i>	-0.071	0.047	-0.039	-0.036*	0.119*	-0.017	0.038	-0.012
	(0.052)	(0.020)	(0.024)	(0.014)	(0.069)	(0.031)	(0.046)	(0.023)
<i>Constant</i>	1.130**	0.499	0.969*	0.367	1.237**	1.980***	2.388***	2.950***
	(0.252)	(0.674)	(0.426)	(0.327)	(0.495)	(0.360)	(0.331)	(0.157)
<i>Sample#</i>	9693	7060	7672	11035	292	12737	10219	15519
<i>R<sup>2</sup></i>	0.299	0.168	0.453	0.344	0.118	0.281	0.174	0.302
<i>midclass</i>	X	X	X	X		X	X	X

注：回归变量解释和计算方法见表 7，系数下方括号报告了标准误，显著性注释为\*\*\* p<0.01, \*\* p<0.05, \* p<0.1。由于手机没有中类，由此手机没有控制中类的固定效应

为研究虚假折扣对销售表现的影响，本文在回归方程（1）中加入虚假折扣变量以及与折扣本身的交叉变量。关于虚假折扣的测度，本文同时考虑了表 7 中定义的广义虚假折扣和狭义虚假折扣两个测度方式，表 10，表 11 分别报告了相关回归结果。从表 10 可以发现，虚假折扣对销售增长的影响可能存在非线性关系，从第（1）列可以看出，整体而言，广义虚假折扣对应的系数显著为正，说明广义虚假折扣能刺激商品销售量增长，消费者很难预测未来商品价格是否会进一步优惠，因此受双十一优惠时限的影响，在双十一期间抓紧购物。然而，从第（2）列到第（7）类回归结果来看，广义虚假折扣与折扣的交叉变量系数为负，说明广义虚假折扣程度确实也降低了双十一折扣对销售增长的刺激作用，即消费者能够识别出经常打折的产品，但是不能确认这些折扣出现的时间和力度，因此只是减少了部分冲动性消费。

表 11 的结果与表 10 存在差别，从第（1）列可以看出，整体而言，狭义虚假折扣的系数显著为负，即商家采用的双十一之前提价制造的狭义虚假折扣降低了商品销售量增长。而从第（2）列到第（7）类回归结果来看，狭义虚假折扣程度与折扣的交叉变量也为负，说明狭义虚假折扣的存在也降低了双十一折扣对销售增长的正面带动作用，消费者能有效识别狭义虚假折扣行为，减少了冲动性消费。

表 10：广义虚假折扣与销售增长率回归分析

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Sales_growth</i>						
<i>discount</i>	-2.415***	-2.467***	-2.454***	-2.239***	-2.178***	-2.314***	-2.216***
	(0.131)	(0.130)	(0.134)	(0.148)	(0.159)	(0.152)	(0.135)
<i>fake</i>	0.108**	-0.200**	-0.199**	-0.204***	-0.206***	-0.220***	-0.201***
	(0.047)	(0.074)	(0.074)	(0.073)	(0.072)	(0.064)	(0.061)
<i>fake* discount</i>		0.516***	0.515***	0.529***	0.510***	0.498***	0.430***
		(0.175)	(0.174)	(0.170)	(0.171)	(0.153)	(0.142)
<i>brand_sale</i>			0.011	0.037***	0.041***	0.090***	0.085***
			(0.008)	(0.009)	(0.008)	(0.011)	(0.011)
<i>brand_price</i>				0.159***	0.088***	0.099***	0.091***
				(0.018)	(0.019)	(0.018)	(0.017)
<i>price_bef</i>					0.086***	0.058***	0.073***
					(0.020)	(0.018)	(0.016)
<i>sale_bef</i>						-0.091***	-0.144***

						(0.008)	(0.010)
<i>comment</i>							0.061***
							(0.008)
<i>sim_disc</i>							-0.145**
							(0.067)
<i>topword#</i>							-0.017
							(0.014)
<i>Constant</i>	2.835***	2.881***	2.819***	1.784***	1.666***	1.963***	1.919***
	(0.118)	(0.117)	(0.138)	(0.201)	(0.214)	(0.201)	(0.209)
<i>Sample#</i>	79,128	79,128	79,128	79,128	79,128	79,128	73,494
<i>R<sup>2</sup></i>	0.207	0.209	0.209	0.240	0.246	0.280	0.300
<i>midclass</i>	X	X	X	X	X	X	X

注：回归变量解释和计算方法见表7，系数下方括号报告了标准误，显著性注释为\*\*\* p<0.01, \*\* p<0.05, \* p<0.1。

表 11：狭义虚假折扣与销售增长率回归分析

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Sales_growth</i>						
<i>discount</i>	-2.714***	-3.154***	-3.146***	-2.941***	-2.867***	-2.957***	-2.781***
	(0.151)	(0.159)	(0.163)	(0.162)	(0.178)	(0.165)	(0.154)
<i>fake</i>	-0.999***	-6.942***	-6.930***	-6.985***	-6.928***	-6.761***	-6.330***
	(0.282)	(0.663)	(0.665)	(0.702)	(0.717)	(0.656)	(0.685)
<i>fake* discount</i>		6.348***	6.336***	6.370***	6.329***	6.206***	5.849***
		(0.629)	(0.632)	(0.661)	(0.671)	(0.623)	(0.653)
<i>brand_sale</i>			0.006	0.032***	0.036***	0.083***	0.078***
			(0.008)	(0.009)	(0.008)	(0.010)	(0.010)
<i>brand_price</i>				0.159***	0.094***	0.105***	0.097***
				(0.017)	(0.018)	(0.018)	(0.017)
<i>price_bef</i>					0.080***	0.053***	0.066***
					(0.018)	(0.015)	(0.013)
<i>sale_bef</i>						-0.086***	-0.133***
						(0.007)	(0.008)
<i>comment</i>							0.053***
							(0.008)
<i>sim_disc</i>							-0.169**
							(0.063)
<i>topword#</i>							-0.022
							(0.015)
<i>Constant</i>	3.111***	3.530***	3.494***	2.468***	2.342***	2.584***	2.520***
	(0.137)	(0.146)	(0.166)	(0.180)	(0.199)	(0.188)	(0.204)
<i>Sample#</i>	78131	78131	78131	78131	78131	78131	72636
<i>R<sup>2</sup></i>	0.219	0.247	0.247	0.279	0.284	0.315	0.33
<i>midclass</i>	X	X	X	X	X	X	X

注: 回归变量解释和计算方法见表 7, 系数下方括号报告了标准误, 显著性注释为\*\*\* p<0.01, \*\* p<0.05, \* p<0.1。

商品折扣对商品销量的影响也可能取决于商品品牌特征。消费者购买决策往往与“占便宜”的心态有关, 尤其是青睐双十一提供的低价购买到平时很少打折的产品。因此, 本文将同一个中类下按照价格高低将各个品牌分为高档和抵挡, 高档对应的虚拟变量 *brand* 为 1, 否则为 0。按照价格高低来区分高档和低档, 还有一个原因是同样的折扣, 高价商品带来更多的消费者“占便宜”的感受。表 12 报告了相关回归结果。可以发现, *brand* 对应的系数显著为正, 同时与折扣的交叉变量对应的系数也为负, 说明高价商品更容易得到消费者的青睐, 打折带来的销售增长率的正面影响更明显。

表 12: 商品品牌对折扣与销售增长率的影响

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Sales_growth</i>						
<i>discount</i>	-2.320*** (0.130)	-2.068*** (0.133)	-1.990*** (0.146)	-1.846*** (0.146)	-1.743*** (0.153)	-1.994*** (0.141)	-1.870*** (0.126)
<i>brand</i>	0.259*** (0.036)	0.571*** (0.149)	0.621*** (0.155)	0.563*** (0.160)	0.593*** (0.158)	0.427*** (0.148)	0.473*** (0.135)
<i>brand* discount</i>		-0.338** (0.161)	-0.375** (0.164)	-0.483*** (0.164)	-0.526*** (0.159)	-0.349** (0.156)	-0.393*** (0.141)
<i>brand_sale</i>			0.030*** (0.008)	0.041*** (0.009)	0.045*** (0.008)	0.093*** (0.011)	0.089*** (0.011)
<i>brand_price</i>				0.120*** (0.019)	0.051*** (0.019)	0.063*** (0.019)	0.054*** (0.018)
<i>price_bef</i>					0.089*** (0.019)	0.059*** (0.017)	0.074*** (0.014)
<i>sale_bef</i>						-0.090*** (0.009)	-0.143*** (0.010)
<i>comment</i>							0.061*** (0.008)
<i>sim_disc</i>							-0.118* (0.068)
<i>topword#</i>							-0.019 (0.014)
<i>Constant</i>	2.588*** (0.116)	2.354*** (0.119)	2.137*** (0.155)	1.495*** (0.192)	1.326*** (0.199)	1.738*** (0.182)	1.644*** (0.180)
<i>Sample#</i>	74460	74460	74460	74460	74460	74460	69092
<i>R<sup>2</sup></i>	0.221	0.222	0.225	0.235	0.241	0.274	0.296
<i>midclass</i>	X	X	X	X	X	X	X

注: 回归变量解释和计算方法见表 3, 系数下方括号报告了标准误, 显著性注释为\*\*\* p<0.01, \*\* p<0.05, \* p<0.1。

最后, 我们也对不同商品热度的商品进行分类回归来研究商品热度的影响。我们按照商品评论数来定义变量 *ave\_bef*。回归结果表明, 在商品热度较高的分类, 如 *ave\_bef*=3 的分组

下评论数对应的系数为 0.088, 高于 ave\_bef=1 分组下的系数, 说明评论数增加能刺激更多销售增长。

表 13: 商品折扣与销售增长率回归分析 (按商品热度分组)

	(1) <i>ave_bef</i> = 0	(2) <i>ave_bef</i> = 1	(3) <i>ave_bef</i> = 2	(4) <i>ave_bef</i> = 3
<i>discount</i>	-2.143*** (0.283)	-2.335*** (0.182)	-2.329*** (0.154)	-2.077*** (0.143)
<i>brand_sale</i>	0.177*** (0.022)	0.108*** (0.013)	0.080*** (0.012)	0.044*** (0.010)
<i>brand_price</i>	0.160*** (0.046)	0.072*** (0.019)	0.081*** (0.016)	0.090*** (0.023)
<i>price_bef</i>	-0.037 (0.052)	0.063*** (0.021)	0.081*** (0.016)	0.109*** (0.018)
<i>sale_bef</i>	-0.247*** (0.025)	-0.086*** (0.014)	-0.136*** (0.010)	-0.153*** (0.011)
<i>sim_disc</i>	-0.099 (0.132)	-0.098 (0.068)	-0.158* (0.084)	-0.144* (0.083)
<i>topword#</i>	-0.052* (0.027)	-0.011 (0.019)	-0.028* (0.014)	-0.022 (0.018)
<i>comment</i>		0.005 (0.011)	0.073*** (0.015)	0.088*** (0.010)
<i>Constant</i>	1.931*** (0.370)	1.899*** (0.213)	1.999*** (0.221)	1.797*** (0.227)
<i>Sample#</i>	4172	16342	23590	29831
<i>R</i> <sup>2</sup>	0.251	0.259	0.315	0.369
<i>midclass</i>	X	X	X	X

注: 回归变量解释和计算方法见表 7, 系数下方括号报告了标准误, 显著性注释为\*\*\* p<0.01, \*\* p<0.05, \* p<0.1。

#### 四、结论和政策建议

双十一是国内影响最大的网络购物节, 商家往往通过商品打折的方式来吸引消费者购物。本文利用网络爬虫技术获取的 2018 年双十一前后天猫、京东以及苏宁购物平台八大畅销商品类别近 40 万商品价格和销售大数据, 分析线上商家折扣与双十一销售增长关系。基于大数据的研究发现, 线上商家普遍在双十一采取打折促销策略, 其中男装、女装及户外运动提供折扣较大。同时, 价格折扣对销售产生了显著提升作用, 平均而言, 折扣增加 10% 能够带来 22.23% 消费量增长, 其中户外运动最为受益, 而手机受益最小, 研究也首次证实部分商家采用了不同形式的虚假折扣行为, 消费者经过价格对比能够甄别双十一提前涨价制造的虚假折扣从而减少其消费, 但是无法识别商家是否在未来依然采取同样甚至更低折扣的行为。我们的研究也发现, 相似产品价格折扣对销售的影响更多的是受互联网巨大流量带来的互补效应, 而并不支持传统理论的替代效应。而大牌产品以及评论关注较多的热门产品更受益于双十一折扣营销策略。

本文研究具有一定的理论和现实意义, 虚假折扣行为显著存在于双十一购物节中, 而部分虚假折扣也正面带动了商家销售的增长, 说明消费者在双十一购物过程中存在“抢时限”、

“占便宜”的冲动购物行为。基于本文的销售增长模型，线上商家也可以进一步优化双十一的营销策略。本文还存在不少缺陷，首先是分析结果目前只关注了 2018 年的情况，需要进一步扩大时间范围；此外，对商品价格爬取的过程中由于网络技术问题还存在一些数据遗漏问题；最后，折扣与销售表现直接还存在内生性问题，需要通过工具变量或自然实验的方法来降低内生性问题带来的估计偏差。

## 参考文献

1. 郝媛媛、叶强、李一军, 2010, 基于影评数据的在线评论有用性影响因素研究,《管理科学学报》, 第8期。
2. 郝媛媛、邹鹏、李一军、叶强, 2009, 基于电影面板数据的在线评论情感倾向对销售收入影响的实证研究,《管理评论》, 第10期。
3. 卢向华、冯越, 2009, 网络口碑的价值——基于在线餐馆点评的实证研究,《管理世界》, 第7期。
4. 殷国鹏, 2012, 消费者认为怎样的在线评论更有帮助?—社会性因素的影响效应,《管理世界》, 12, 115-124。
5. 中商产业研究院, 2018年“双十一”网购大数据分析报告, <http://www.askci.com/news/chanye/20181116/1139281136843.shtml>。
6. Agee, T. and Martin, B., 2001. Planned or Impulse Purchases? How to Create Effective Infomercials, *Journal of Advertising Research* 41(6), 35-42.
7. Beatty, S.E. and Ferrell, M.E., 1998. Impulse buying: Modeling its precursors. *Journal of Retailing*, 74(2): 161-167.
8. Berger, J., Sorensen, A. T. and Rasmussen, S. J, 2010. Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29( 5), 815-827.
9. Brashear, T. G., Kashyap, V., Musante, M. D. and Donthu, N., 2009. A Profile of the Internet Shopper: Evidence from Six Countries. *Journal of Marketing Theory and Practice*, 17:3, 267-282.
10. Biseas, A. and Burton, S., 1993, Consumer perceptions of tensile price claims in advertisements: an assessment of claim types across different discount level. *Journal of Academy of Marketing Science*, 21(3): 217-229.
11. Hausman, A., 2000. A multi-method investigation of consumer motivations in impulse buying behavior. *Journal of Consumer Marketing*, 17(5):403-426.
12. Madhavaram S.R. and Laverie, D.A., 2004. Exploring impulse purchasing on the Internet. *Advances in Consumer Research*, 31(1):59-66.
13. Rook, D. W. (1987). The Buying Impulse. *Journal of Consumer Research*, 14(2), 189-197.
14. Salton G. A., 1974. Vector space model for automatic indexing. *Communications of the Acm*, 18(11):613-620.
15. Tsai, C. F., 2012. Bag-of-words representation in image annotation: A review. *Artificial Intelligence*, 1-19.
16. Zhou, L. and Wong, A., 2003. Consumer Impulse Buying and in-Store Stimuli in Chinese Supermarkets. *Journal of International Consumer Marketing*, 16, 37-53.

## 致谢页

非常荣幸能够参加本次竞赛并顺利完成论文的写作和提交。首先我想感谢丘成桐中学科学奖竞赛组委会能够提供这么一个宝贵的机会让我能够通过学习提高自己,也非常感谢我的两位指导老师,厦门大学王亚南经济研究院的陈海强教授以及福建厦门第一中学(厦门一中)的王君黛老师,从两位老师的辛苦指导下,我学习了非常多的知识,提高了自己编程数据分析能力,也进一步增进了对经济和金融学的了解。下面我想借此机会向各位评委介绍一下项目完成的一些细节。

### 选题来源:

厦门一中为了提升学生视野,专门举办了厦一大讲堂活动。每周利用周末时间邀请一些专家学者、大学教授来学校举办讲座,与学生进行交流互动从而拓展学生的知识面。在一次厦一大讲堂上,厦门大学陈海强教授为我们做了经济学方面的讲座,我对陈教授讲的经济金融知识很感兴趣,讲座之后和他进行了沟通交流,表达了自己想向他请教学习经济金融知识想法。同时我也和学校教我编程的王君黛老师进行了交流,探讨了用编程方法对数据处理分析的技术可行性。跟两位老师交流之后,我决定报名参加丘成桐中学科学奖经济金融建模大赛。我选定双十一商家打折行为与销售表现作为项目的选题,一个重要启发也是来自曾经看到的一个介绍双十一期间消费者争相抢购打折商品的新闻,给我留下了深刻印象,激发了我去思考:为什么消费者会在双十一期间去抢购商品?而打折到底给商家带来了多少销售量增长?跟陈海强教授交流之后,也得知他正好在关注电子商务发展对居民消费的影响的研究,同时也有想法来收集相关数据,因此最终我在老师的指导下确定了这个选题。

### 项目背景:

通过资料,我了解到双十一购物狂欢节目前是国内影响最大的网上购物节日,而商家为了吸引消费者购物,也往往提供不少打折优惠。然而,在大部分商家都打折的时候,折扣究竟如何影响线上商家销售表现并不清楚。同时也有媒体报道一些商家在双十一期间存在虚假折扣行为,损害了消费者利益,但是消费者是否对这些虚假折扣有所甄别,我也非常关注这个问题。在陈教授的指导和帮助下,通过团队合作,利用网络爬虫技术获取了2018年双十一前后天猫、京东以及苏宁三大购物平台近40万商品的价格和销售数据,并通过经济学建模分析来回答上述问题。

之所以关注双十一购物节,一方面因为其来自于我们生活场景,具有中国特色,同时,在购物节期间,商家如何利用促销来吸引消费者,而消费者如何在这么多选择和信息下做出购物决策,也是重要的理论和实践问题。在老师的指导下,我也利用了一些前沿分析方法,比如对商品描述文字进行文本分析来计算各个商品之间的相似程度,通过词云分析来识别不同商品类别下的热门词汇来定义该商品是否为热门商品等。而关于虚假折扣的定义,我也参考了消费者协会2015年调查报告的相关描述,提出了两个定义方式,并发现了一些有趣的结果。

### 项目完成过程:

项目从寻找选题到论文提交大概经历了1年时间,主要分为两个阶段。前期,我在老师指导下对一些经济金融建模知识进行学习,同时开始收集数据,查看相关文献;后期开始动手实践,通过由浅入深,由理论到实践的过程逐步推进研究进展。在这个过程中,我主要利

用课余以及假期时间来进行学习和研究工作。具体时间大致如下：2018年9月-10月：了解相关经济金融背景知识，与指导老师讨论确定项目选题以及研究框架；2018年11月-2019年1月：利用课余时间学习统计、计量建模知识以及数据处理编程方法；在双十一期间，参加了厦门大学陈海强教授指导的数据科学俱乐部 WISER Club 组织的线上线下消费数据收集活动，通过团队合作，利用爬虫技术抓取2018年双十一数据；2019年2月-4月：利用假期和课余时间学习编写数据分析处理程序；2019年5月-6月：调整优化程序，获得分析结果，并在老师指导下撰写论文草稿；2019年7月-8月，基本完成了论文草稿，并进行修改完善；2019年9月上旬，完成论文修改并提交。

### 指导老师角色：

在我确定选题之后，两位老师都很支持，利用他们的专业知识给我分享了一些项目所需的知识和分析手段，在项目研究过程中给予了非常多的无偿指导帮助。我在两位老师的指导下制定了项目计划，厦门大学陈海强教授向我介绍了经济金融建模理论知识，并且邀请我参与他们团队的一些讨论，向我传授了一些经济金融建模方法，同时也教我如何使用 Stata 语言进行回归分析。厦门一中王君黛老师向我讲解了数据处理过程中会用到编程知识并且给我提供了一些 python 学习资料。在两位老师的指导下，我以经济金融理论作为依据，查阅相关文献，并以计量回归方法作为分析手段，稳步推进项目的研究进展。最后，我将研究过程中获得的结果进行整理，撰写成为论文，并根据陈海强教授的意见进行了修改。

通过这次研究工作我受益匪浅，首先这个研究让我接触到了前沿的经济学知识以及数据分析方法，视野得到了很大的开拓。同时这个研究让我动手能力得到很大的提高，经过多次反复调试之后，终于达到预期的设计目的，真正感受到了程序开发的乐趣，让我今后利用编程解决实际问题的信心倍增。最后，我的意志力也得到了锻炼，我真正理解到，真实的快乐不仅来自于梦想的实现，还在于成功路上的体验，只有亲身经历，克服艰难，成功才更有价值。再次感谢组委会以及帮助我的两位指导老师，还有家人的支持！

### 附录：指导老师简历

**陈海强**（1980.10-），经济学博士，教授，博士生导师，目前任职于厦门大学王亚南经济研究院以及厦门大学经济学院金融系，现担任教育部“计量经济学”重点实验室副主任，厦门大学经济学科数据科学与决策咨询中心主任，国家自然科学基金《防范和化解金融风险》应急管理重点项目首席专家，厦门大学经济学院金融系金融工程教研室负责人，福建省高等学校新世纪优秀人才，厦门市高层次引进人才。2003年毕业于北京大学，获经济学与统计学双学士学位，2005年毕业于香港中文大学，获经济学硕士，2011年毕业于美国康奈尔大学，获经济学博士。主要研究方向为金融大数据、金融工程、量化投资、风险管理等，先后在《经济研究》、《金融研究》、《管理科学学报》，*Econometric Theory*, *Journal of Empirical Finance*, *Journal of International Money and Finance* 等国内外顶尖期刊发表论文几十篇。陈海强教授作为负责人和主要参与者承担了十余项国家级、省部级和企业科研课题研究。

**王君黛**（1989.03-），硕士，毕业于华中师范大学国家数字化学习工程技术研究中心，研究方向为教育大数据、教育云服务。目前任职于福建省厦门第一中学，现担任厦门一中信息学奥赛教练及科技活动辅导员。多次指导学生在全国信息学奥林匹克竞赛、全国青少年科技创新大赛中获奖，其中袁泽清同学入围参加了第70届英特尔国际科学与工程大奖赛。

## 学术诚信声明

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知,除了文中特别加以标注和致谢中所罗列的内容以外,论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处,本人愿意承担一切相关责任。

参赛队员: 俞琳薇

指导老师: 王君黛

2019年 9月 12日