

# 线上销售、市场反应与未来股票收益

## 引言

另类数据的兴起是近年来金融市场最引人注目的变化之一。另类数据泛指在投资研究过程中使用的非传统数据，如文本数据、传感器数据、卫星影像数据等。随着市场透明度的提高，历史财务数据、交易数据等公开渠道信息会很快纳入到股价中，一些因子对未来收益率的预测效果减弱（McLean and Pontiff, 2016），导致基于公开数据获取超额收益越来越困难（Green et al., 2011; Chordia et al., 2014）。因此，市场投资者普遍开始寻求另类数据，希望利用这些数据提供的前瞻信息和独特的非财务信息获取新的阿尔法。在这样的背景下，另类数据在全球迅速兴起。

我国另类数据行业起步较晚但发展很快。与国外相比，我国的网络平台和移动互联网发展程度较高，网络平台上沉淀了大量的用户行为记录。单个用户的行为可能是无意义的，然而，如果将大量用户的数字行为记录汇总，就可能形成反映线上用户群体行为、包含特定信息的另类数据。从用户数字记录中挖掘有价值信息已成为资本市场新的重要信息渠道。在这种类型的另类数据中，有一个非常典型也很有中国特色的数据是线上销售数据。消费者的每次购买行为都会在电商平台上留下购买记录，如果能获取整个线上零售市场的购买记录，并在个体公司层面汇总，就可能获得所有线上消费者对公司产品的购买情况，也就是公司产品在线上渠道的销售数据。

投资者了解上市公司信息的传统渠道主要是财务数据，与传统财务数据相比，线上销售数据的时效性更强，大多为每月更新，即每月月底就能获取当月的线上销售信息。而且，线上销售数据属于计算营业收入等财务指标的中间环节数据，上市公司一般不会详细披露这些中间数据。这就意味着，线上销售数据不仅可能向资本市场更加及时地传递关于公司销售和经营的信息，也可能提供财务报告中所没有的信息。当这

些信息纳入到股票价格时，就可能对股票收益率与投资者行为产生影响。然而，国内外关于线上销售数据对投资者和股票收益影响的文献还很匮乏。本文的研究目的即希望填补这一空白，并尝试回答以下问题：在我国 A 股市场，投资者能否利用这些与销售活动相关但更加及时的数据先于市场发现上市公司的价值，预测未来股票收益，并获取超额收益？线上销售数据产生超额收益的机制是怎样的？如果线上销售数据能给投资者带来超额收益，说明线上销售数据包含了影响股票收益的信息，那么，这些信息是影响股票收益的传统财务信息之外的新信息吗？

本文使用从 Wind 资讯获取的线上销售数据对上述问题进行研究。这套数据来源于国内主要电商平台的销售记录，覆盖超过 95% 的电商市场，包括约 200 家 A 股上市公司的月度线上销售数据，数据每月更新。本文构造了线上销售额同比增长率指标，基于该指标，主要得到以下研究结论。第一，根据线上销售额同比增长率构建的投资组合可以获取超额收益，经过风险调整后，超额收益仍显著存在；第二，线上销售额同比增长率可以显著预测未来的股票收益；第三，线上销售额增长率能提供财务因素之外的影响股票收益的增量信息；第四，线上销售额增长率能够预测营业收入的增长和资产收益率的变化，说明线上销售数据包含公司经营业绩信息，但股票市场对这些信息的反应缓慢，因此，能及时获取线上销售信息的投资者可以利用信息优势获取超额收益。

本文的学术贡献体现在以下几个方面。第一，本文丰富了消费者信息与股票收益率预测的研究。有关消费者信息、产品信息的数据是连接消费市场、产品市场和股票市场的重要数据。随着电商平台的发展，大量消费者的线上行为记录开始被投资者获取。国外已有学者研究消费者的线上评论、线上搜索等消费者信息对股票收益率的影响。然而，受制于数据可得性，国内还很少有文献深入探讨消费者信息对股票定价的

影响，本文是第一篇从投资组合收益率、组合收益率的变化趋势等角度探讨线上销售对股价影响的研究。

第二，本文丰富了另类数据的投资价值与信息含量的研究。另类数据能否提供超额收益和前瞻性信息，不仅是投资者所关心的，也已成为学术研究的重要话题。本文基于中国 A 股市场的消费行业上市公司，提供了线上销售数据这种另类数据能为投资者带来超额收益的经验证据，并证明线上销售数据所包含的影响股票收益的信息与传统的财务信息不完全一致，而是含有财务信息之外的新信息。

第三，本文与投资者有限关注的研究相关。线上销售数据能预测未来股票收益的原因是股票市场对线上销售信息的反应有滞后性，而这种滞后性又源于大多数市场投资者无法及时关注或处理线上销售信息。在以往文献中，有一类研究正是从有限关注的角度探讨股票市场缺乏有效性的问题 (Huberman and Regev, 2001; DellaVigna and Pollet, 2009; Hirshleifer et al., 2009)。本文发现的市场对信息反应延迟的结论为该领域文献提供了新的证据。

## 1 文献回顾与理论假

### 1.1 文献回顾

本文主要与两类文献相关。第一类是关于消费者信息对股票价格影响的文献。在早期研究中，由于涉及消费者信息的数据较少，一些文献主要通过问卷调查数据研究消费者满意度对企业经营业绩和股票估值的影响 (Ittner and Larcker, 1998; Anderson et al., 2004; Fornell et al., 2006; Aksoy et al., 2008; Luo et al., 2010)，并发现消费者满意度越高，企业的未来经营业绩越好，企业估值越高。2010年之后，随着大数据、云计算等技术的发展，一些关于消费者行为的数据开始出现，并逐渐应用到学术领域。Da et al. (2011) 通过谷歌获取了美国上市公司主要商品的

搜索指数，发现消费者对主要商品的搜索指数能作为公司盈余和股票价格的先行指标。Froot et al. (2017) 获取了大量手机和电脑用户的搜索记录与地理定位数据，如果用户搜索某个商场或用户的定位地点在某个商场，则用户有一定概率在该商场购物。基于这种逻辑，他们构造了美国消费行业上市公司的销售活跃度指标，发现该指标对上市公司的营业收入、盈余、盈余公告期间的股票收益都有显著预测能力。随着卫星应用的日益成熟，衡量消费者数量或销售活跃度的数据也不断发展，有文献从高分辨率的商场停车场卫星影像数据中识别汽车流量，用汽车流量的变化衡量消费者人数或公司销售活跃度的变化 (Katona et al., 2018; Zhu, 2019)。也有文献研究消费者观点对上市公司股价的影响，如 Huang (2018) 使用了亚马逊网站消费者对上市公司主要商品的评价数据，Tang (2018) 使用推特网站上用户对上市公司商品或品牌的评论数据。与上述研究相比，本文的线上销售数据比消费者的评论、搜索等间接衡量消费者潜在购买的数据更加直接地衡量购买行为。另外，上述研究往往只选取了部分消费者群体的数据，而本文的线上销售数据包含了几乎整个中国电商市场，覆盖所有中国零售行业上市公司。近年来，国内也有文献基于线上销售数据展开研究，廖理等 (2021) 发现季度线上销售能够显著预测盈余公告期间的累计超额收益。但该文献未探讨线上销售数据的投资价值和超额收益的产生机制。本文用月度频率的消费数据从股票市场月度反应和周度反应的角度分析投资者对线上消费信息的反应特征和超额收益的来源。

第二类是关于大量用户行为汇总和群众智慧的文献。随着网络平台的发展，用户各类平台上沉淀了很多历史行为的数字记录，如社交媒体上的评论与转发，投资平台的观点分享等。单个用户的数字记录可能是无规律的，然而，将大量用户的行为记录汇总，就可能反映群体的一些行为规律。越来越多的文献开始探讨这种行为信息汇

总和群众智慧对股票价格的影响。Kelley and Tetlock (2013) 使用近两亿个人投资者的股票买卖订单数据,发现个人投资者群体的股票净买入能够正向预测上市公司的月度股票收益。Chen et al. (2014)基于美国投研社交平台 Seeking Alpha 上投资者在平台社区发布的近十万个帖子进行研究,发现将大量投资者观点汇总可以预测未来股票收益。Jame et al. (2016) 将投资分析平台 Estimize 上超过 3000 个用户的五万多分析预测进行汇总,发现汇总后的 Estimize 一致预期可以预测上市公司的未来盈余、市场预期和市场反应。Green et al. (2019) 获取了美国企业点评网站 Glassdoor 上一百多万条公司雇员对上市公司的评价数据,发现汇总后的公司评价数值能显著预测公司盈余和股票市场反应。Sheng (2019)基于 Glassdoor 平台的用户评论数据也得到类似发现。在国内,从大量用户数字记录中挖掘信息、研究群众智慧的文献也日益增多,这些文献大多基于股吧投资者的发帖或评论数据,将帖子或评论的情感、数量等加总,研究投资者群体的情感或关注度对股票价格的影响(杨晓兰等, 2016; 段江娇等, 2017; 金德环和李岩, 2017; 孙鲲鹏和肖星, 2018; 尹海员和吴兴颖, 2019; 关静怡等, 2020; 钱宇等, 2020; 黄创霞等, 2020)。与上述文献不同,本文使用的是消费者线上购物信息的汇总,虽然线上购物可能受到促销等因素的影响,但是,将长时间、大范围的购买记录汇总,就可能反映消费者对上市公司产品的整体消费情况。

## 1.2 理论假设

另类数据出现的重要动因是投资者追求超额收益,而本文要探讨的第一个问题是线上销售数据这种另类数据能否为投资者带来超额收益。线上销售数据如果能够预测未来股票收益,需要满足两个条件:一是线上销售数据具有信息含量,且所含信息能够有效传递到股票市场,反映到股票价格中;二是这些信息不能在数据发布后迅速反

映到股票价格中，即大部分投资者不能及时获取线上销售数据中的信息并做出投资决策。本文将详细阐述这两个条件。

首先，线上销售数据可能含有与公司销售和经营相关的信息。廖理等（2021）发现线上销售能显著预测季度盈余意外和收入意外，证明了线上销售含有基本面信息。从数据来源看，本文与廖理等（2021）的数据来源相同，这套数据覆盖了几乎所有的消费品行业 A 股上市公司，能提供每个被覆盖公司的线上真实销售信息。从产品的销售渠道来看，线上平台已成为公司销售的重要渠道，一些公司的线上销售量已超过线下。随着线上线下渠道逐渐融合，线上销售信息就可能反映公司整体销售状况。另外，对消费品公司而言，消费者是重要的利益相关者，产品销售是公司经营的重要内容。因此，销售数据也很可能与公司的营业收入、营业利润等经营业绩有相关性，即线上销售数据不仅含有公司的销售信息，还可能含有与公司经营业绩相关的信息。

其次，线上销售数据中的信息不能及时充分地反映到股票价格中。线上销售数据等另类数据的价格普遍高昂，其价格能达到个人投资者和机构投资者日常使用的传统数据库价格的十几倍甚至几十倍。多数的个人投资者和一部分机构投资者出于成本考虑可能不会购买另类数据的使用权限，也就无法及时获取数据。此外，与传统的结构化的数据相比，另类数据的使用更加复杂。在上市公司基本面分析、股票价格预测等投研分析中，财务数据、股票市场交易数据等传统数据的使用方法已较为成熟。而另类数据是新型数据，且数据结构形式多样，从另类数据中挖掘有价值信息需要有相应的数据处理和数据分析技术。因此，真正能够有效使用另类数据的可能只有少部分领先的机构投资者，如对冲基金、中大型公募基金和证券公司等。虽然这些机构投资者能够及时获取和使用信息并先于大多数市场投资者在股价未充分反应前进行投资决策，但它们的交易行为可能仍不足以推动股票价格及时充分地对线上销售等另类数据

中的信息进行反应。如 Huang (2018) 和 Green 等 (2019) 都基于美国市场, 发现虽然对冲基金会使用另类数据进行投资决策, 但股票价格对另类数据中信息的反应仍有滞后性。我国的线上销售数据出现较晚, 且价格至少为普通数据库的十几倍, 购买权限的机构仍是少数。同时, 我国资本市场的发达程度和机构投资者的发展程度与欧美发达的资本市场仍有一定的差距。因此, 线上销售数据中的信息也很可能无法及时充分地反映到股票价格中。

因此, 基于上述分析, 本文提出第一个研究假设:

**假设 1:** 线上销售数据可以预测未来股票收益, 为投资者提供超额收益。

无论在投资实务界还是学术研究中, 对影响股票定价的信息的挖掘都是重要的课题。在另类数据出现之前, 大多数投资者和学者只能基于财务报告、公司公告等数据展开研究, 如分析财务因素对股票价格的影响和预测作用, 根据财务指标构造定价因子等。随着另类数据的发展, 越来越多财务数据之外的数据类型开始在资本市场出现, 如线上销售数据。这些新的非财务数据也为分析股票定价信息提供了新的视角, 那么, 非财务数据是否含有能够影响股价的非财务信息? 这一问题还很少有文献探讨。本文认为, 非财务数据中含有传统财务信息之外的影响股价的信息, 主要有两个原因: 一是财务报告主要提供经过会计师处理过的营业收入、利润等财务指标, 而很少详细披露销售量等非财务数据, 即财务指标与非财务数据中所含的公司信息本身就可能存在差异; 二是投资者和分析师在筛选上市公司时既可能基于财务指标也可能基于非财务信息, 即他们投资决策使用的信息可能也存在差异。

基于上述分析, 本文提出第二个研究假设。

**假设 2:** 线上销售数据含有传统财务信息之外的影响股票定价的新信息。

## 2 样本选取与研究设计

## 2.1 样本来源

本文使用的线上销售数据来自 Wind 资讯。Wind 资讯在 2018 年 12 月推出了“线上销量”另类数据产品，该产品包括超过 200 家 A 股公司的上千个品牌的线上销售数据，提供了具体商品品类、商品品牌和上市公司所有商品的月度线上销售额、销售量、平均价格等指标。这套数据的数据源是天猫、京东等电商平台，Wind 资讯通过网络爬虫和人工智能技术，实时监测和爬取电商销售情况，覆盖超过 95% 的中国电商市场。为了避免线上渠道和线下渠道的品牌差异对研究造成干扰，本文也对比了 A 股公司在线下渠道的品牌信息，发现线上与线下品牌的重合度超过 90%，即线上和线下品牌几乎一致。

本文选取公司月度线上销售额指标展开实证分析，样本区间为 2015 年 1 月至 2020 年 6 月，共 54 个交易月。在上市公司选取方面，为了避免上市年份的干扰，本文剔除了每个公司上市年份的数据，最终获得 181 个 A 股上市公司作为研究样本。这些公司分属电子电器、食品饮料、医药生物、纺织服装、文教职娱、建材家具、家化 7 个阿里电商一级品类。财务数据、股票交易数据和因子数据来自 CSMAR 数据库。

## 2.2 变量定义

本文的关键变量是上市公司的线上销售额同比增长率。无论在财务分析还是市场营销中，同比增长率都是分析业绩变化的重要指标。由于同比指标能减弱季节性差异的干扰，本文用线上销售额的同比增长率衡量上市公司线上销售变化的程度。具体计算公式如下：

$$\text{SalesGrowth}_{i,t} = \ln(\text{Sales}_{i,t}) - \ln(\text{Sales}_{i,t-12})$$

其中,  $SalesGrowth_{i,t}$  是  $i$  公司  $t$  月份的线上销售额增长率,  $Sales_{i,t}$  是  $i$  公司  $t$  月份的线上销售额,  $Sales_{i,t-12}$  是  $i$  公司在去年同期的线上销售额。

本文主要变量的具体定义如表 1 所示。

表 1 主要变量定义

变量符号	变量名称	变量定义
Sales Growth	线上销售额同比增长率	$t$ 月线上销售额的自然对数减去 $t-12$ 月线上销售额的自然对数
Ret	月度股票收益率	$t+1$ 月的月度股票收益率
Ln_Size	总市值的自然对数	最近一个季度总市值的自然对数
BM	账面市值比	最近一个季度的总资产除以同季度末的总市值
ROA	盈利能力	最近一个季度的资产收益率
Ln_Volume	累计交易金额	$t-12$ 月至 $t-1$ 月股票累计交易金额的自然对数
CV_Volume	交易金额的变异系数	$t-12$ 月至 $t-1$ 月交易金额变异系数的自然对数
Reversal	价格反转	最近一个月, 即 $t-1$ 月的月度股票收益率
Momentum	价格动量	$t-12$ 月至 $t-2$ 月的股票累计收益
Turnover	换手率	$t-1$ 月日均换手率的平均值
Analyst	分析师关注人数	$t$ 月所在年份前一年分析师关注人数的总和
Insti	机构持股比例	最近一个季度的机构持股比例

## 2.3 研究设计

本文的主要实证由三部分组成, 一是检验投资者是否可以利用线上销售数据先于市场发现公司价值从而获取超额收益; 二是检验线上销售数据是否含有财务信息之外的影响股票价格的信息; 三是分析线上销售数据能预测股票收益和获取超额收益的机制。

在实证中第一部分, 本文构造日历投资组合并检验是否存在超额收益。具体来说, 在第  $t$  月末计算每个上市公司  $t$  月线上销售额的同比增长率, 并根据  $t$  月末所有公司线上销售额增长率  $SalesGrowth_{i,t}$  的分位数将公司样本分为三组。之后, 在  $t+1$  月将套利组合进行投资, 每月调仓一次, 从 2016 年 2 月开始持有, 到 2020 年 7 月结束, 共 54 个月。套利组合即买入处于最高分组的投资组合, 卖出处于最低分组的投资组合。本文按照等权重加权和流通市值加权两种方法计算投资组合的月度收益率。

在第二部分，为了检验线上销售额增长率是否含有影响股票收益的增量信息，本文使用横截面回归的方法，在控制可能影响股票收益的财务指标的基础上，研究线上销售额增长率对未来股票收益的预测效果。本文参考以往文献的控制变量设定(张然等, 2017; Huang, 2018; Green et al., 2019), 采用以下模型进行 Fama-MacBeth 回归:

$$\begin{aligned} \text{Ret}_{i,t+1} = & \beta_0 + \beta_1 \times \text{SalesGrowth}_{i,t} + \beta_2 \times \text{Ln\_Size}_{i,t} + \beta_3 \times \text{BM}_{i,t} + \\ & \beta_4 \times \text{ROA}_{i,t} + \beta_5 \times \text{Ln\_Volume}_{i,t} + \beta_6 \times \text{CV\_Volume}_{i,t} + \\ & \beta_7 \times \text{Reversal}_{i,t} + \beta_8 \times \text{Momentum}_{i,t} + \beta_9 \times \text{Turnover}_{i,t} + \\ & \beta_{10} \times \text{Analyst}_{i,t} + \beta_{11} \times \text{Insti}_{i,t} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

在第三部分，本文首先检验线上销售额增长率是否包含公司经营业绩信息，参考 Green et al. (2019)，使用营业收入的同比增长率和 ROA 的同比变化率衡量企业经营业绩的变动，并设定如下模型：

$$\Delta \text{Ln\_Revenue}_{i,t} = \beta_0 + \beta_1 \times \text{QSG}_{i,t} + \beta_2 \times X_{i,t} + \varepsilon_{i,t} \quad (2)$$

$$\Delta \text{ROA}_{i,t} = \beta_0 + \beta_1 \times \text{QSG}_{i,t} + \beta_2 \times X_{i,t} + \varepsilon_{i,t} \quad (3)$$

这两个模型的目的是验证同期的线上销售额增长率可以预测同期的经营业绩变动，以此证明线上销售数据中包含企业经营业绩信息。由于财务数据都是季度数据，因此，本文将模型 (2) 和 (3) 中的变量均调整为季度变量。其中， $\Delta \text{Ln\_Revenue}_{i,t}$  是  $i$  公司  $t$  季度营业收入的自然对数与去年同季度营业收入自然对数的差值， $\Delta \text{ROA}_{i,t}$  是  $i$  公司  $t$  季度 ROA 与去年同季度 ROA 的差值， $\text{QSG}_{i,t}$  是  $i$  公司  $t$  季度线上销售额的自然对数与去年同季度线上销售额自然对数的差值， $X_{i,t}$  是  $t$  季度的控制变量。

其次，本文检验市场对线上销售信息反应的变化趋势。分别检验  $t+1$  月内每周市场的反应情况和  $t+2$  月至  $t+7$  月每月的市场反应情况。仍采用日历组合方法检验，构造组合的方法与第一部分相同，分别检验  $t+1$  月内每周的套利组合超额收益和  $t+2$

至 t+7 月每月的套利组合超额收益，通过超额收益的变化趋势分析市场对经营业绩信息反应的时效性，进而验证线上销售数据产生超额收益的机制。

## 2.4 描述性统计

表 2 是主要变量的描述性统计，包括样本数量、均值、标准差和四分位数等统计量。Sales Growth 的均值是 0.325，中值为 0.282，且 p25 分位数的值仍大于 0，这表明多数消费行业上市公司的线上销售额都有正向增长，这也与近几年我国消费品行业持续增长的趋势相符。Ret 是股票月度收益率，均值是 0.012，标准差是 0.106。

表 2 主要变量描述性统计

Variables	Obs	Mean	Std	p25	p50	p75
Sales Growth	8793	0.325	0.556	0.012	0.282	0.579
Ret	8793	0.012	0.106	-0.055	0.000	0.066
QSG	2596	0.309	0.433	0.050	0.278	0.528
Revenue Growth	2596	0.082	0.277	-0.025	0.090	0.208
$\Delta$ ROA	2596	-0.001	0.020	-0.008	0.000	0.007
Ln_Size	8793	23.210	1.090	22.510	23.010	23.770
BM	8793	0.539	0.235	0.355	0.510	0.708
Ln_Volume	8793	10.420	0.416	10.150	10.380	10.670
CV_Volume	8793	-0.641	0.381	-0.906	-0.633	-0.388
Reversal	8793	0.005	0.111	-0.060	-0.004	0.063
Momentum	8793	0.038	0.390	-0.227	-0.038	0.231
Turnover	8793	1.713	1.942	0.620	1.048	1.977
Analyst	8793	13.920	13.460	2.000	9.000	23.000
Insti	8793	51.770	23.790	37.200	54.860	70.940

## 3 回归结果与分析

### 3.1 线上销售数据与投资组合收益

表 3 展示了投资组合风险调整收益率的结果。本文根据线上销售额增长率构造投资组合，套利组合 Q3-Q1 的构造方法为买入线上销售额增长率最高分组 Q3 中的股票，卖空线上销售额增长率最低分组 Q1 中的股票。本文使用 Fama-French-Carhart 四因子模型对套利组合的收益率进行风险调整，将套利组合的月度收益率对每个风险因子的月度收益率进行回归，得到风险调整后的超额收益率 Alpha。

Panel A 是按照等权重加权计算的月度收益率，Q3-Q1 行 Alpha 列的等权重组合月度收益率为 1.4% (t 值是 3.47)，Panel B 是按照流通市值加权计算的月度收益率，Q3-Q1 行 Alpha 列的市值加权组合月度收益率为 1.5% (t 值是 2.09)。表 3 的结果证实了假设一，即线上销售数据具有投资价值，能够为市场投资者提供超额收益。

表 3 投资组合的收益率检验

Group	Excess Returns / %	Alpha / %	Market	SMB	HML	UMD
Panel A: Equal Weights						
Q1 (Low)	0.299 (0.33)	-0.174 (-0.57)	1.020 *** (14.12)	0.647 *** (5.88)	-0.175 * (-1.77)	0.004 (0.05)
Q2	1.060 (1.24)	0.507 (1.03)	0.915 *** (10.43)	0.480 ** (2.47)	-0.126 (-0.56)	0.066 (0.61)
Q3 (High)	1.833 ** (2.11)	1.181 ** (2.21)	0.943 *** (9.63)	0.339 (1.63)	-0.155 (-0.72)	0.091 (0.67)
Q3-Q1	1.535 *** (3.87)	1.355 *** (3.47)	-0.077 (-0.74)	-0.308 * (-1.98)	0.02 (0.11)	0.087 (0.93)
Panel B: Value Weights						
Q1 (Low)	1.327 (1.46)	0.255 (0.45)	0.841 *** (6.68)	-0.606 ** (-2.10)	-1.039 *** (-3.86)	0.166 (1.04)
Q2	1.669 * (2.08)	0.827 (1.34)	0.811 *** (6.76)	-0.210 (-0.78)	-0.364 (-1.13)	0.150 (1.07)
Q3 (High)	2.840 *** (3.15)	1.799 ** (2.51)	0.825 *** (5.60)	-0.543 * (-1.71)	-0.594 (-1.50)	0.194 (1.19)
Q3-Q1	1.513 ** (2.25)	1.544 ** (2.09)	-0.016 (-0.11)	0.063 (0.18)	0.445 (1.15)	0.029 (0.15)

注：括号中数字为 t 值，\*、\*\*、\*\*\* 分别表示在 10%、5%、1% 水平显著。下表同，不再赘述。

### 3.2 线上销售数据对未来股票收益的预测

线上销售数据能为市场提供超额收益，说明线上销售数据中含有影响股票定价的信息。在线上销售数据出现之前，资本市场的投资者无法获取这类高频率、覆盖范围广的电商数据。那么，这类新数据中含有的影响股票定价的信息与以往发现的影响股票定价的信息一样吗？这些新数据能否提供影响股票定价的增量信息？通过 Fama-MacBeth 回归并控制同期的财务变量，本文对该问题进行检验。

表 4 展示了线上销售额增长率对横截面股票收益的预测。列 (1) 不加入控制变量, 分析线上销售额增长率单独对股票收益的预测, Sales Growth 的系数为 0.008, 在 1%水平上显著。列 (2) 至列 (4) 依次加入其他控制变量, 结果显示, 线上销售额增长率的系数有一定程度下降, 但仍然显著。以列 (4) 为例, 在控制了其他可能影响股票价格的因素后, Sales Growth 的系数从列 (1) 的 0.008 下降到了 0.005, 但仍在 5%水平上显著。表 4 的结果表明, 虽然有些财务指标和公司指标也对横截面股票收益有影响, 但控制这些变量后并没有显著影响线上销售额增长率与未来股票收益之间的关系, 这说明线上销售数据为资本市场提供的信息很可能是一些新的增量信息, 是难以从财务报表、公司公告等传统的数据源中获取的。这一结果也表明线上销售数据等另类数据的出现对增加资本市场的信息类型和发掘影响股票定价的新信息都有重要意义。

表 4 线上销售额增长率与未来股票收益

	Ret			
	(1)	(2)	(3)	(4)
Sales Growth	0.008 *** (2.89)	0.006 *** (2.79)	0.005 ** (2.51)	0.005 ** (2.14)
Ln_Size		0.003 (1.18)	0.002 (0.81)	-0.003 (-1.23)
BM		-0.001 (-0.12)	0.005 (0.97)	0.011 * (1.98)
ROA		0.266 *** (3.39)	0.269 *** (3.42)	0.249 *** (3.49)
Reversal			-0.006 (-0.37)	-0.001 (-0.08)
Momentum			0.010 ** (2.10)	0.012 ** (2.30)
Ln_Volume			-0.003 (-0.61)	0.006 (1.07)

CV_Volume			-0.003 (-0.69)	0.001 (0.17)
Turnover				-0.002 *** (-3.18)
Insti				0.000 (1.10)
Analyst				0.000 ** (2.12)
Constant	0.009 (0.97)	-0.056 (-1.03)	-0.019 (-0.31)	0.007 (0.12)
N	8793	8793	8793	8793
R <sup>2</sup>	0.012	0.104	0.153	0.184

## 4 进一步研究

在理论假设部分，本文分析了线上销售预测股票收益和提供超额收益的机制。在该部分，本文实证检验这一机制，即分别证明线上销售数据含有与公司经营相关的信息、股票价格不能对线上销售数据中的信息做出及时反应两个条件。

### 4.1 线上销售数据对经营业绩的预测

表 5 展示了线上销售额增长率对经营业绩预测的结果，经营业绩指标选择营业收入的同比增长衡量收入的变动，选择资产收益率的同比变动衡量盈利能力的变化。每个指标都进行了不含控制变量和含控制变量的检验。结果表明，线上销售额增长率对营业收入的增长和盈利能力的变动都有正向预测作用，并且，在控制相关变量，特别是控制滞后一期的营业收入增长率和资产收益率变动后，QSG 的系数仍然显著为正。这些结果说明，线上销售数据中含有与企业的营业收入、盈余等经营业绩相关的信息。

表 5 线上销售额增长率对经营业绩的预测

	Revenue Growth		ΔROA	
	(1)	(2)	(3)	(4)
QSG	0.056 *** (4.26)	0.040 *** (3.25)	0.003 ** (2.53)	0.003 ** (2.51)
Ln_Size		0.055 * (1.95)		-0.001 (-0.59)
BM		-0.004 (-0.06)		-0.023 *** (-4.48)
Lag_ROA				-0.041 *** (-3.25)
Lag_ΔROA				0.123 *** (9.32)
Lag_Revenue		-0.074 *** (-4.46)		
Lag_Revenue Growth		0.363 *** (19.23)		
Pastreturn		0.151 *** (4.90)		0.015 *** (5.81)
Ln_Volume		0.013 (0.90)		0.002 (1.59)
Turnover		-0.004 (-0.32)		-0.001 (-0.53)
Insti		0.198 ** (2.08)		0.018 ** (2.28)
Analyst		-0.001 (-1.44)		-0.000 *** (-6.94)
	Revenue Growth		ΔROA	
	(1)	(2)	(3)	(4)
Constant	0.016 (0.75)	-0.140 (-0.24)	-0.003 * (-1.88)	-0.008 (-0.17)
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	2596	2532	2596	2532
R <sup>2</sup>	0.175	0.315	0.031	0.122

#### 4.2 组合周度收益率变化趋势

在证明线上销售数据含有企业经营业绩信息后, 本文进一步分析股价对线上销售信息的反应。如果股价对信息反应迅速, 当最新的线上销售数据产生后, 信息就迅速纳入到股价中, 则投资者也无法获取超额收益。

前文的表 3 检验的是投资组合 t+1 月的收益率, 为了分析股价对信息反应的变化趋势, 本文检验投资组合在 t+1 月的前四周的周度收益率。表 6 是投资组合在第一周至第四周每周的收益率, 图 1 是对应的柱状图。结果显示, 第一周投资组合的收益率为 0.1%, 而第二周的收益率为 0.7%, 第三周和第四周的收益率有所下降但也高于第一周的收益率。该结果表明, 市场对线上销售数据中的信息并非及时反应, 而是有一定滞后性。因此, 能够及时获取线上销售数据的投资者, 根据数据中的信息做出投资决策, 就能够先于市场大多数投资者发现未来的公司价值, 并获取超额收益。

**表 6 投资组合周度收益率变化趋势**

	Week1	Week2	Week3	Week4
Mean/%	0.134	0.651 ***	0.230	0.475 **
t-stat	0.89	3.64	1.11	2.22

#### 4.3 组合月度收益率变化趋势

为进一步分析市场对线上销售信息反应的特征, 本文也对 t+2 月至 t+7 月的月度收益率进行了检验。表 7 是投资组合的月度收益率, 图 2 是对应的柱状图。结果显示, 虽然 t+2 月至 t+6 月这六个月的组合收益率都小于 t+1 月的收益率, 但仍在统计意义上显著。该结果表明, 股票价格没能及时对线上销售数据中的信息做出反应。表 7 和表 6 的结果共同说明, 股票市场对线上销售数据中信息的反应确实存在滞后性。

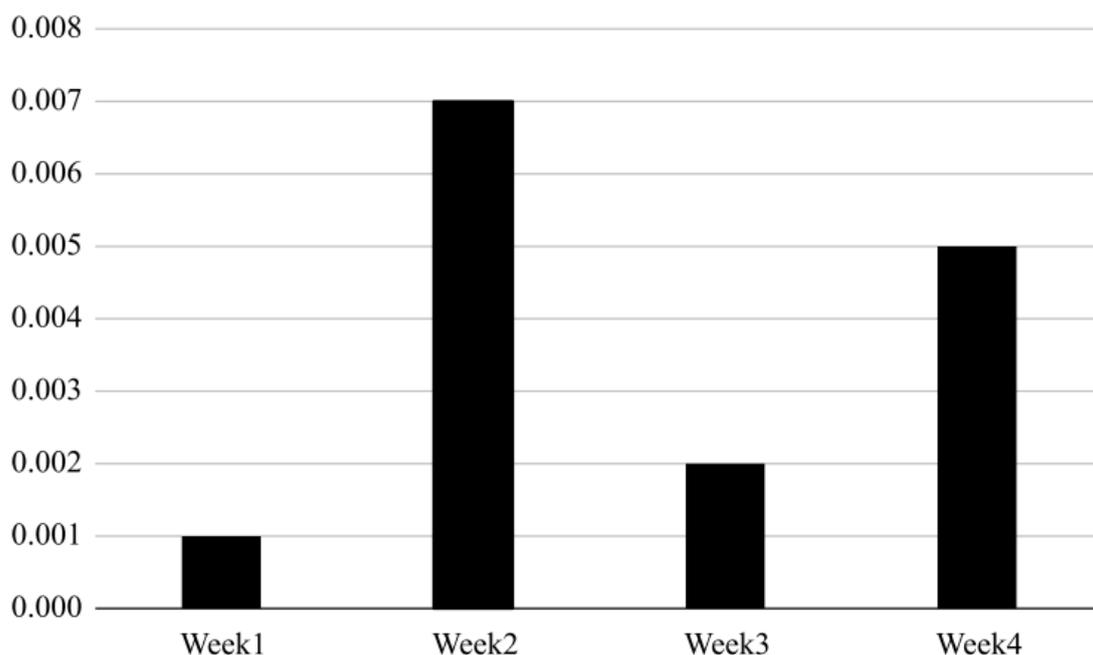


图 1 投资组合周度收益率变化趋势

表 7 投资组合月度收益率变化趋势

	$t+1$ 月	$t+2$ 月	$t+3$ 月	$t+4$ 月	$t+5$ 月	$t+6$ 月	$t+7$ 月
Mean/%	1.553 ***	0.825 **	1.029 ***	0.917 ***	1.195 ***	0.886 **	0.284
$t$ -stat	4.12	2.35	2.72	3.30	3.73	2.64	0.83

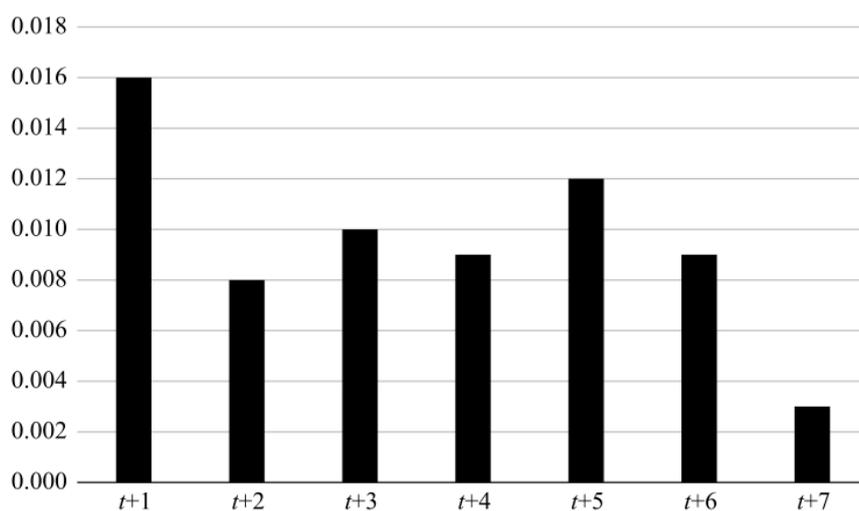


图 2 投资组合月度收益率变化趋势

## 5 稳健性检验

为了证明研究结论的稳健性, 本文围绕日历投资组合的超额收益进行稳健性检验, 分别检验剔除盈余公告效应后的收益率和用其他因子模型风险调整后的收益率。

## 5.1 剔除盈余公告月份的检验

本文需要证明投资组合有显著的月度超额收益,以此说明线上销售数据等另类数据能够通过更加及时地向获取数据的投资者传递公司销售和经营的信息为投资者带来超额收益。然而,在表 3 中,本文未剔除盈余公告效应的影响,这会对研究结果产生干扰,即存在这样的可能性:表 3 的月度超额收益率是公司披露财务报告时,股价大幅变动带来的,而并非这些高频的线上销售数据通过每月向市场传递信息引起月度股价变化带来的。因此,需要将盈余公告效应剔除后再进行检验。

本文首先检验线上销售额增长率是否与盈余公告期间的股票收益具有相关性。根据每个季度线上销售额增长率将样本公司分为三组,追踪三组公司股票在盈余公告[-2,+2]和[-3,+3]窗口的盈余公告累计超额收益,检验结果如表 8 所示。Panel A 和 Panel B 的最后一列 Q3-Q1 组合的收益率都显著为正,说明线上销售额增长率较高的公司,其股票价格在盈余公告期间也有更高幅度的增长,即盈余公告期间的累计超额收益也更高。该结果确实会干扰前文结论,因此,本文将样本中每个公司发布盈余公告的月份剔除,重新检验组合的月度超额收益率。

表 8 盈余公告窗口累计收益率的差异性检验

	Q1 (Low)	Q2	Q3 (High)	Q3-Q1
Panel A: [-2, +2] 窗口				
Mean / %	-0.008	0.365 **	0.720 ***	0.728 ***
t-stat	-0.04	2.02	3.79	2.78
Panel B: [-3, +3] 窗口				
Mean / %	0.184	0.538 ***	1.100 ***	0.916 ***
t-stat	0.92	2.71	5.07	3.11

表 9 是剔除盈余公告月份后,投资组合 Q3-Q1 月度收益率的检验结果。组合 Q3-Q1 的平均收益率为 0.014,比表 3 中的收益率略有下降,但仍在统计上显著。该结果表明,上市公司发布财务报告带来的盈余公告效应对组合收益率有一定程度的

干扰，但将盈余公告效应剔除后，月度日历组合的收益率仍然显著为正，即本文的结论仍成立。

表 9 剔除盈余公告样本的组合收益率

	Q1 (Low)	Q2	Q3 (High)	Q3-Q1
Mean/%	1.221	2.112 **	2.632 **	1.411 ***
<i>t</i> -stat	1.14	2.22	2.68	2.89

## 5.2 其他因子模型调整

表 3 的投资组合收益率是使用 Fama-French-Carhart 四因子模型进行的风险调整。为避免风险调整模型差异对本文结果带来影响，与以往文献保持一致，本文也分别使用三因子模型、五因子模型以及在五因子模型基础上加入动量因子的六因子模型对投资组合的收益率进行风险调整。表 10 的 Panel A 和 Panel B 分别是等权重加权和流通市值加权的组合收益率，结果显示，经过几种风险模型调整后，组合 Q3-Q1 的超额收益率都显著为正。该结果表明本文的研究结论非常稳健，即投资者通过线上销售数据可以获取超额收益。

表 10 投资组合的收益率检验

Panel A: Equal Weights	Q1	Q2	Q3	Q3-Q1
Excess Returns/%	0.299 (0.33)	1.060 (1.24)	1.833 ** (2.11)	1.535 *** (3.87)
Three-Factor Alpha/%	-0.168 (-0.51)	0.603 (1.20)	1.312 ** (2.37)	1.480 *** (3.89)
Five-Factor Alpha/%	-0.415 (-1.44)	0.263 (0.56)	0.996 * (1.85)	1.411 *** (3.36)
Six-Factor Alpha/%	-0.285 (-0.99)	0.375 (0.78)	1.064 * (1.93)	1.349 *** (3.28)
Panel B: Value Weights	Q1	Q2	Q3	Q3-Q1
Excess Returns/%	1.327 (1.46)	1.669 * (2.08)	2.840 *** (3.15)	1.513 ** (2.25)
Three-Factor Alpha/%	0.493 (0.92)	1.042 * (1.72)	2.079 *** (2.96)	1.585 ** (2.46)
Five-Factor Alpha/%	0.041 (0.08)	0.570 (0.96)	1.683 ** (2.26)	1.642 ** (2.26)
Six-Factor Alpha/%	0.092 (0.18)	0.657 (1.11)	1.675 ** (2.22)	1.582 ** (2.10)

## 6 结论

本文采用构造日历投资组合的研究方法，基于资本市场上出现的线上销售数据，证实我国 A 股市场的线上销售数据具有投资价值，能够及时获取这些数据的投资者可以利用数据中的信息预测未来股票收益并获取超额收益。本文还发现，线上销售数据含有影响股票定价的传统财务信息之外的新信息。此外，本文进一步指出了线上销售数据能提供超额收益的机制：线上销售数据包含与上市公司经营业绩相关的信息，能预测上市公司的营业收入和盈利能力；股票价格对线上销售数据中信息的反应有滞后性，线上销售数据中的信息不能及时被纳入股票价格中。因此，及时获取这些数据的投资者会比股票市场的其他投资者有信息优势，能基于这一信息优势先于市场了解上市公司的潜在价值和进行投资决策。

近年来, 另类数据迅速兴起并逐渐应用于经济分析和投资研究中。无论对于实务界还是学界, 另类数据都是新事物, 基于另类数据的学术研究也还很少。特别是, 围绕另类数据能否为资本市场提供增量信息、能否预测股票收益等问题也存在较多争议。本文的研究不仅围绕中国 A 股市场的线上销售数据这一典型的另类数据为上述问题提供了新的经验证据, 也有助于资本市场的参与者深入了解另类数据的信息含量特征和信息传递方式。本文希望监管部门和投资者能够意识到另类数据的信息价值, 并以实际行动促进另类数据在市场中的传播, 提高市场效率。同时, 也应注意防范另类数据可能造成的股价大幅波动的风险。

## 参考文献

- 段江娇, 刘红忠, 曾剑平. 2017. 中国股票网络论坛的信息含量分析[J]. 金融研究, (10): 178-192.
- Duan J J, Liu H Z, Zeng J P. 2017. Analysis on the information content of China, s internet stock message boards[J]. Journal of Financial Research, (10): 178-192. (in Chinese)
- 关静怡, 朱恒, 刘娥平. 2020. 股吧评论、分析师跟踪与股价崩溃风险——关于模糊信息的信息含量分析[J]. 证券市场导报, (3): 58-68.
- Guan J Y, Zhu H, Liu E P. 2020. Stock BBS reviews, analyst coverage and stock price crash risk: Analysis on the information content of ambiguous information[J]. Securities Market Herald, (3): 58-68. (in Chinese)
- 黄创霞, 温石刚, 杨鑫, 等. 2020. 个体投资者情绪与股票价格行为的互动关系研究[J]. 中国管理科学, 28(3): 191-200.
- Huang C X, Wen S G, Yang X, et al. 2020. The interactive relationship between individual investor sentiment and stock price behaviors[J]. Chinese Journal of Management Science, 28(3): 191-200. (in Chinese)
- 金德环, 李岩. 2017. 群体智慧: 同伴观点与价值发现——来自社交媒体的经验证据[J]. 经济管理, 39(12): 157-173.

- Jin D H, Li Y. 2017. Wisdom of crowds: Peer opinions and value discovery—Empirical evidences of social media[J]. *Business Management Journal*, 39(12): 157-173. (in Chinese)
- 廖理, 崔向博, 孙琼. 2021. 另类数据的信息含量研究——来自电商销售的证据[J]. *管理世界*, 37(9): 90-103.
- Liao L, Cui X B, Sun Q. 2021. The information content of alternative data: Evidence from E-commerce sales[J]. *Journal of Management World*, 37(9): 90-103. (in Chinese)
- 钱宇, 李子饶, 李强, 等. 2020. 在线社区支持倾向对股市收益和波动的影响[J]. *管理科学学报*, 23(2): 140-154.
- Qian Y, Li Z R, Li Q, et al. 2020. Impact of online community support tendencies on returns and volatility in Chinese stock market[J]. *Journal of Management Sciences in China*, 23(2): 140-154. (in Chinese)
- 孙鲲鹏, 肖星. 2018. 互联网社交媒体、投资者之间交流与资本市场定价效率[J]. *投资研究*, 37(4): 140-160.
- Sun K P, Xiao X. 2018. Social media, communication among investors, and capital market pricing efficiency[J]. *Review of Investment Studies*, 37(4): 140-160. (in Chinese)
- 杨晓兰, 沈翰彬, 祝宇. 2016. 本地偏好、投资者情绪与股票收益率: 来自网络论坛的经验证据[J]. *金融研究*, (12): 143-158.
- Yang X L, Shen H B, Zhu Y. 2016. The effect of local bias in investor attention and investor sentiment on stock markets: Evidence from online forum[J]. *Journal of Financial Research*, (12): 143-158. (in Chinese)
- 尹海员, 吴兴颖. 2019. 投资者高频情绪对股票日内收益率的预测作用[J]. *中国工业经济*, (8): 80-98.
- Yin H Y, Wu X Y. 2019. Predictive effect of high-frequency investor sentiment on the intraday stocks return[J]. *China Industrial Economics*, (8): 80-98. (in Chinese)
- 张然, 汪荣飞, 王胜华. 2017. 分析师修正信息、基本面分析与未来股票收益[J]. *金融研究*, (7): 156-174.
- Zhang R, Wang R F, Wang S H. 2017. Analysts, revisions, fundamental analysis and future stock returns[J]. *Journal of Financial Research*, (7): 156-174. (in Chinese)
- Aksoy L, Cooil B, Groening C, et al. 2008. The long-term stock market valuation of customer satisfaction[J]. *Journal of Marketing*, 72(4): 105-122.
- Anderson E W, Fornell C, Mazvancheryl S K. 2004. Customer satisfaction and shareholder value[J]. *Journal of Marketing*, 68(4): 172-185.

Chen H L, De P, Hu Y, et al. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media[J]. *The Review of Financial Studies*, 27(5): 1367-1403.

Chordia T, Subrahmanyam A, Tong Q. 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?[J]. *Journal of Accounting and Economics*, 58(1): 41-58.

Da Z, Engelberg J, Gao P J. 2011. In search of fundamentals[R]. Working Paper.

DellaVigna S, Pollet J M. 2009. Investor inattention and Friday earnings announcements[J]. *The Journal of Finance*, 64(2): 709-749.

Fornell C, Mithas S, Morgeson III F, et al. 2006. Customer satisfaction and stock prices: High returns, low risk[J]. *Journal of Marketing*, 70(1): 3-14.

Froot K, Kang N, Ozik G, et al. 2017. What do measures of real-time corporate sales say about earnings surprises and post-announcement returns[J]. *Journal of Financial Economics*, 125(1): 143-162.

Green J, Hand J R M, Soliman M T. 2011. Going, going, gone? The apparent demise of the accruals anomaly[J]. *Management Science*, 57(5): 797-816.

Green T C, Huang R Y, Wen Q, et al. 2019. Crowdsourced employer reviews and stock returns[J]. *Journal of Financial Economics*, 134(1): 236-251.

Hirshleifer D, Lim S S, Teoh S H. 2009. Driven to distraction: Extraneous events and underreaction to earnings news[J]. *The Journal of Finance*, 64(5): 2289-2325.

Huang J K. 2018. The customer knows best: The investment value of consumer opinions[J]. *Journal of Financial Economics*, 128(1): 164-182.

Huberman G, Regev T. 2001. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar[J]. *The Journal of Finance*, 56(1): 387-396.

Ittner C D, Larcker D F. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction[J]. *Journal of Accounting Research*, 36(S1): 1-35.

Jame R, Johnston R, Markov S, et al. 2016. The value of crowdsourced earnings forecasts[J]. *Journal of Accounting Research*, 54(4): 1077-1110.

Katona Z, Painter M, Patatoukas P, et al. (2018-07-30). On the capital market consequences of alternative data: Evidence from outer space[EB/OL]. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3222741](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3222741).

Kelley E K, Tetlock P C. 2013. How wise are crowds? insights from retail orders and stock returns[J]. *The Journal of Finance*, 68(3): 1229-1265.

Luo X M, Homburg C, Wieseke J. 2010. Customer satisfaction, analyst stock recommendations, and firm value[J]. *Journal of Marketing Research*, 47(6): 1041-1058.

McLean R D, Pontiff J. 2016. Does academic research destroy stock return predictability?[J]. *The Journal of Finance*, 71(1): 5-32.

Sheng J F. (2019-01-31). Asset pricing in the information age: Employee expectations and stock returns[EB/OL].

[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3321275](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3321275).

Tang V W. 2018. Wisdom of crowds: Cross-sectional variation in the informativeness of third-party-generated product information on twitter[J]. *Journal of Accounting Research*, 56(3): 989-1034.

Zhu C. 2019. Big data as a governance mechanism[J]. *The Review of Financial Studies*, 32(5): 2021-2061.