

另类数据的信息含量研究*

——来自电商销售的证据

廖理 崔向博 孙琼

摘要: 本文使用中国消费行业上市公司2015~2018年的电商销售数据,研究消费者线上消费活动是否能向金融市场传递与公司价值相关的信息。研究结果表明,电商销售数据含有与公司基本面相关的信息,可以预测公司的未预期收入、未预期盈余和盈余公告累计超额收益。进一步分析发现,电商销售数据的预测能力在线上销售收入占比高、投资者关注度低、成长型、信息透明度高的公司中更强。本文也基于这些异质性检验深入分析了电商销售数据具有预测作用的原因。电商销售数据与消费者的消费行为及上市公司的销售活动直接相关,是一类很有代表性的另类数据,研究电商销售数据的信息含量和投资价值有助于丰富另类数据文献,同时也可厘清另类数据在新兴资本市场中的作用提供新的经验证据。

关键词: 电商销售数据 信息含量 另类数据

DOI:10.19744/j.cnki.11-1235/f.2021.0139

一、引言

近年来,另类数据(Alternative Data)的信息含量问题正成为金融和财务会计领域的前沿与热点研究问题,日益受到业界和学术界的重视^①。随着金融市场的发展和市场有效性的增强,传统公开渠道披露的信息大部分会反映在股票价格中^②,以往定价因子的有效性逐渐减弱(Mclean and Pontiff, 2016),导致市场投资者越来越难获取超额收益(Baker and Wurgler, 2007; Bai et al., 2016)。此外,投资者的信息获取和解读能力不一,对企业盈余预测的质量也会存在差别,从而影响投资决策的有效性(Hirshleifer and Teoh, 2003; DellaVigna and Pollet, 2009; Kempf et al., 2017; Kacperczyk et al., 2016)。因此,投资者会积极通过各类信息渠道来获取新的信息,并尝试利用新技术来提高信息获取效率(Jame et al., 2016; Jagtiani and Lemieux, 2019)。当前,互联网技术和数字技术发展迅猛,大量以前无法收集的信息得以积累和留存,如在线搜索数据、社交媒体数据、卫星图像数据、信用卡交易数据等,这些另类数据可以拓宽投资者的信息获取渠道,同时,信息技术的进步也有助于提高投资者的信息解读能力和信息处理效率(Loder, 2019)。资本市场中的投资者,特别是寻求超额回报最积极的买方参与者,比如对冲基金、量化基金、资产管理公司等,已经开始探索利用另类数据中的信息来获取超额收益。

学者们也对另类数据的信息含量进行了初步的探索研究。大部分研究认为另类数据包含基本面信息,对未来的股票收益具有预测能力(Chen et al., 2014; Froot et al., 2017; Huang, 2018; Zhu, 2019)。但也有部分研究对另类数据的信息含量提出了质疑,他们认为另类数据中的噪音成分太大(Das and Chen, 2007; Chen et al., 2020; Goldstein and Yang, 2019; Cookson and Niessner, 2020)。由此可见,有关另类数据的研究仍存在争议,尚无定论。此外,目前关于另类数据的研究主要集中在资本市场发达的国家,国内鲜有文献涉及。那么,在新兴资本市场中,另类数据是否也包含了有效信息,是否对公司基本面信息和未来股票收益有预测能力,还尚未可知。而这正是本文的研究价值所在。

在中国,有一类重要的另类数据是电商销售数据。自2013年起,中国的网络零售市场规模已经连续6年稳居全球第一。2019年上半年,中国网络零售额高达4.82万亿元,占社会消费品零售总额的比重为24.7%,其中7家主要电商平台店铺数量为1853万家^③。在全球十大电商公司中,中国公司占了四席,阿里巴巴位列全球

*本项研究得到国家自然科学基金项目(71472100、71790605、71790591)和中国博士后科学基金第63批面上项目(2018M630128)的资助。感谢中国金融科技学术年会参会专家学者的宝贵意见,当然文责自负。廖理为本文通讯作者。

第一。截至2019年6月,中国互联网普及率为61.2%,网络购物用户规模达6.39亿,占网民整体的74.8%^④。尤其是消费行业,有些公司的线上销售额占总营收的比重已经超过了线下销售额的占比。由于公司的成本结构相对稳定,在短期内易于预测,所以,许多公司的盈利波动主要由销售收入的波动驱动(Da et al., 2011)。由此可知,中国消费类上市公司的线上销售情况很可能与公司未来的销售和盈利能力密切相关。那么,能否根据电商销售数据构造公司盈利和股价的领先指标,预测公司的基本面信息和未来股票收益,将是一个值得研究的重要问题,对理论界研究是重要的补充,对实务界投资也能提供重要的参考。

基于此,本文使用Wind资讯在2018年12月推出的“线上销量”数据来检验另类数据的信息含量。该数据覆盖了170家中国消费行业上市公司近700个品牌的电商销售数据,样本区间为2015年1月至2018年9月。本文数据具有以下几点优势:第一,电商销售数据是一种聚合网上商店零售额的数字信息,而非衍生的观点或情绪,受伪信息、假信息及行为偏差等噪音的影响较小(Boudoukh et al., 2019);第二,上市公司披露财报信息的频率至少为季度,相对于传统数据,电商销售数据每月更新一次,实时性更强、颗粒度更细,可以对公司的未来业绩进行更及时的预测;第三,电商销售额与公司潜在的销售收入和现金流之间的相关性比现有文献所使用的公司销售代理指标或文本类信息更紧密,是一种更直接有效的领先指标;第四,已有研究还存在样本量少、数据观察期较短的问题,而本文的研究样本量较为充足。

本文主要有以下研究结论。第一,季度的线上销售额与季度的营业收入具有正相关性。第二,线上销售额的增长率对标准化未预期收入和标准化未预期盈余具有显著的预测作用,这一结论在控制了可能对未预期收入和未预期盈余产生影响的变量后仍然显著,这意味着线上销售额传达了有关公司基本面的信息。第三,线上销售额的增长率对盈余公告期间股票的累计超额收益也具有显著预测作用,这表明线上销售额包含的信息会在盈余公告时反映在股价中。即使控制当期的未预期收入和未预期盈余,这种预测作用仍然显著,说明线上销售额包含了当前季度的营业收入和利润之外的增量信息。第四,线上销售额所含有的有价值信息在盈余公告后会充分反映到股价中,不存在盈余公告后的漂移现象。第五,进一步分析发现,电商销售数据的预测能力在线上销售收入占比高、投资者关注度低、成长型、信息透明度高的公司中更强。本文也基于这些异质性检验深入分析了电商销售数据具有预测作用的原因。此外,经过一系列稳健性检验,本文的研究结论依然成立。这些结论表明,电商销售数据包含关于公司基本面和未来股票收益的信息。

本文的研究主要有以下学术贡献。首先,由于另类数据挖掘成本高昂,现有文献主要关注了少量较易获取的另类数据,还有大量较难获取但可能含有更高信息含量的另类数据有待探究。而本研究所使用的电商销售数据正是这样一类数据,关于这类数据的信息价值问题尚未有相关文献。因此,研究电商销售数据的信息含量和投资价值有助于丰富另类数据文献。其次,学术界对于另类数据是否能够传递与价值相关的信息,还存在较大的争议,一些学者认为认知偏差、行为偏差可能会导致聚合的另类数据的信噪比较低,信息含量也较低。对电商销售数据而言,它不仅与公司的营业收入高度相关,而且是一种实时高频的数字信息,信噪比更高,本研究为另类数据包含有价值信息的结论提供了有力的支持,也补充和发展了另类数据在资产定价中的作用的文献研究。最后,目前消费者线上活动数据受到了学术界的广泛关注,但由于数据可得性问题,现有文献中使用的另类数据大多是衡量消费者购买的间接数据类型,如消费者搜索、消费者评论等。这些数据虽然可以在一定程度上反映出消费者对公司产品的潜在需求、满意度或未来的活动倾向,但无法充分和准确地反映实际发生的交易活动。而本文的数据直接来自消费者的网上购买记录,是由消费者最终的购买行为汇聚而成的。相比之前文献使用的数据类型,电商销售数据与公司的实际销售活动和消费者的实际购买行为的关联性更强。此外,基于中国数据探讨消费者行为与股票收益的文献研究还很匮乏,而本文的研究有助于丰富该领域的文献。

本文其余部分的结构安排如下:第二部分是文献回顾与研究假设;第三部分介绍本文的样本选取、数据来源与研究设计;第四部分为实证检验结果的分析与讨论;第五部分围绕主要假设开展进一步研究;第六部分是稳健性检验;最后一部分总结全文。

二、文献回顾与研究假设

(一)文献回顾

近年来,另类数据正逐渐成为金融科技研究的热点和前沿问题。现有文献主要对另类数据的信息含量及其对资本市场的影响进行了探索,相关研究尚处于起步阶段,亟待进一步拓展。

另类数据的类型和来源非常广泛,不同类型数据的信息价值和预测能力存在一定的差别。已有文献表明,与消费者消费行为相关的另类数据在一定程度上能够预测公司的未来业绩和股票收益。因为消费者是公司的重要利益相关者,他们的购买行为和消费支出与公司销售活动密切相关,是构成公司现金流和潜在销售收入的重要来源(Shapiro, 1983; Huang, 2018)。所以,与消费者行为相关的另类数据正日益受到国内外学者和专业投资者的重视。消费者的消费或购买行为有可能发生在线上,也有可能发生在实体商店,消费者的在线搜索、在线评论、在线浏览、在线支付和地理位置等都可能反映出他们未来的活动倾向和消费动态。

首先,在线搜索数据在一定程度上能反映消费者未来的消费倾向,如消费者在做出购买决策之前,会搜索商家或产品信息(Beatty and Smith, 1987)及与其相关的在线评论或评价。Da等(2011)发现,公司产品的谷歌搜索量指数(Search Volume Index, SVI)具有前瞻性,能有效预测营业收入和股票收益。Froot等(2017)发现消费者对大型零售公司的地址搜索量包含了有关公司基本面和盈余以外的重要信息。其次,互联网和社交媒体上的消费者在线评价数据可能具有信息含量。一方面,在线评价代表了广大消费者对产品或品牌的态度,消费者越满意,他们未来购买特定的产品或品牌的意向就越大,公司同期销售额或未来销售额就越高(Ittner et al., 2003);另一方面,在线评论也可能通过口碑效应影响其他消费者对公司产品或品牌的印象而刺激或抑制更多的销售。Huang(2018)通过研究亚马逊网站上2004~2015年的640万消费者的1450多万条消费者评价信息发现,消费者的评价包含了有关公司未来现金流和股票回报的信息。Tang(2018)和Bartov等(2018)通过汇总个人的twitter观点,发现这些聚合信息能显著预测公司即将发布的季度营业收入和盈余公告超额收益。再次,卫星图像数据也能在一定程度上反映零售商店中的消费者交易情况。Katona等(2018)利用美国零售店停车场的卫星图像数据,发现汽车数量的增长能够预测公司基本面和股价。最后,消费者交易行为数据直接反映了消费者的购买行为,具备的信息含量更加丰富。Agarwal等(2020)利用信用卡持有者的交易数据获得有关消费者分类支出和消费者特征的信息,发现消费者支出可以预测公司未来现金流,上市公司客户的购买能力越强、客户的人口统计特征或地理特征越多元化,指标的预测能力越强。Zhu(2019)结合消费者在线交易数据和汽车卫星图像数据发现,另类数据的可用性降低了信息获取成本,有助于提高股票价格的信息含量。

从以上研究不难发现,大多数学者认为涉及消费者行为的另类数据含有与公司基本面和股价相关的增量信息。然而,也有一些研究对这类数据的信息价值提出了质疑,他们认为,尽管另类数据可能包含了传统数据之外的新信息,但信噪比可能较低,使得股价信息含量的变化方向模糊(Agarwal et al., 2020; Chen et al., 2020)。首先,社交媒体类数据大多以文本信息为主,主观性较强,可能会有偏差(曾庆生等, 2018; Huang, 2018)。比如,消费者或投资者可能缺乏提供关于产品或服务的真实信息的动机,使得发表的信息是虚假或伪造的。他们也可能不具备准确评估产品或服务专业的能力,导致评论或观点存在偏差。其次,在线搜索数据和卫星图像数据可能无法真实反映公司的未来现金流。例如,谷歌搜索指数虽然可以在一定程度上反映出消费者对产品的潜在需求或未来的活动倾向,但产品搜索难以区分消费者的正面情绪或负面情绪(Huang, 2018),也无法反映实际发生的交易活动。最后,市场参与者容易产生羊群效应。社交媒体上的个体往往会受他人信息的诱惑而搭便车,降低整体信息效率(Banerjee et al., 2018; Da and Huang, 2020; Goldstein and Yang, 2019)。因此,另类数据的信息含量和真实投资价值还需要学术界进一步探索和研究。

本文主要利用中国电商平台的销售数据来研究另类数据的信息含量问题,电商销售数据不仅是直接的数字信息,信噪比高于文本类信息。而且与现有文献中使用的涉及消费者信息的另类数据相比,电商销售数据与公司未来业绩之间的相关性更高,因为线上销售记录直接反映了整个线上市场消费者的购买情况。与本文

最紧密相关的文献是Zhu(2019)的研究,但由于数据供应商提供的线上交易数据较少,导致该文的数据样本量较小且代表性不足。因此,有关电商销售数据的信息含量问题还有待进一步研究。如果实时的电商销售数据能提供关于上市公司未来业绩和股价的信息,那么,它不仅可以作为资本市场有用的补充信息来源,还能为金融领域的学术研究长期以来关注的传统问题提供全新的研究视角(Froot et al., 2017; Zhu, 2019; Green et al., 2019)。

(二)研究假设

本文需要回答的第一个问题是电商销售数据能否预测公司的财务业绩,即电商销售数据是否含有与公司基本面相关的信息。本文所指的预测是同期预测,即在公司还未公布季度财务报告时,使用该季度的电商销售数据指标对该季度财务报告的公司业绩进行预测。本文认为线上销售情况可能对公司经营业绩有预测作用的原因主要有3个方面。

第一,对消费行业上市公司而言,消费者是重要的利益相关者,消费者对产品和服务的偏好程度直接影响公司获取现金流的能力。随着大数据技术的发展,越来越多的消费类公司基于电商平台的销售情况洞察消费者需求、判断同行竞争者的情况,也有分析师通过电商平台的销售情况判断公司产品业绩的变化趋势。但是,如果只观测某类消费者对产品的购买情况,容易受到消费者群体消费特征和偏好的影响无法全面反映消费者对公司产品的购买情况。如果只观测某个品牌或某类产品的销售情况,也会存在偏差,无法衡量公司销售的整体情况。而电商销售数据的直接来源是消费者在电商平台上的每一笔消费。将国内主要电商平台和公司自设平台的每个品牌、每件商品的每一笔消费记录收集和汇总,就能够全面反映消费者在线上对公司产品的购买情况。因此,线上销售情况很可能包含公司当前经营状况的信息。

第二,消费类上市公司大多采用了线上线下融合的商业模式。公司的销售业绩由线上线下两部分构成,而电商销售数据只包含线上的销售情况,这可能影响电商销售数据的有效性。但从我国消费行业实际情况来看,这种影响相对较小。我国绝大多数消费行业上市公司已经开设了独立的电商销售平台或在淘宝网、京东商城等主要电商平台开设了公司直营的旗舰店。并且,线上线下渠道在公司营销渠道中的重要程度越来越接近,基于线上渠道预测公司整体情况具有合理性和可靠性。另外,影响电商销售数据有效性的另一个因素是线上线下渠道的产品品牌、产品类别有差异。本文从一家国内领先的金融数据平台获得了消费行业上市公司在线下商超渠道的产品销售情况。经过对比,发现线上线下的产品品牌和产品类别有超过90%的重合度,这在一定程度上也能降低电商销售数据反映公司实际业绩的偏差程度。

第三,从单个上市公司的线上销售额与实际营业收入的变动趋势来看,电商销售数据覆盖的上市公司的线上销售额与营业收入有很高的相关性。

基于上述三方面原因,本文提出第一个研究假设。

研究假设H1:线上销售额与公司营业收入具有正相关性。

本文更关心的是线上销售情况能否预测公司的其他基本面指标。具体而言,我们关注公司的未预期营业收入(*SUR*)、未预期盈余(*SUE*)。未预期营业收入和未预期盈余反映的是公司业绩的超预期部分。

线上销售额的变化可能反映两方面信息,一是反映公司现金流的变动,二是反映消费者对公司产品整体购买意愿的变化。这两类信息都与公司的营业收入相关。那么,线上销售额的变动很可能反映了公司现金流和营业收入的未预期变动。因此,本文提出第二个研究假设。

研究假设H2:线上销售额的增长率对未预期营业收入有预测能力。

另外,从我国消费行业整体来看,受到近几年消费升级和消费品行业稳定增长的影响,消费类公司的线上和线下的成本结构总体趋于平稳,甚至有些特定细分行业由于规模效应出现了成本下降的趋势。相对于收入因素,成本因素对企业营业利润的影响程度要低。因此,线上销售额的变化在一定程度上可以反映公司营业利润变动的信息,即线上销售额的变化与公司未预期盈余具有一定的相关性。鉴于此,本文提出第三个研究假设。

研究假设 H3:线上销售额的增长率对未预期盈余有预测能力。

本文需要回答的第二个问题是电商销售数据能否预测公司盈余公告期间的累计超额收益。

在金融学文献中,标准化未预期盈余(*SUE*)被认为是引起盈余公告效应的重要因素。Jegadeesh 和 Livnat (2006)发现,标准化未预期收入(*SUR*)也具有信息含量,同样可以预测盈余公告累计超额收益。如果基于电商销售数据构造的增长率指标对标准化未预期盈余(*SUE*)和标准化未预期收入(*SUR*)都有预测能力,则线上销售增长率更高的企业有可能获得更高的未预期盈余或未预期收入。在公司发布盈余公告时,投资者会根据盈余公告发布的当期财务业绩买入或卖出公司股票,未预期盈余和未预期收入更高的企业的股票价格上涨的概率就越大。而由于电商销售数据与未预期盈余和未预期收入包含一些相同的信息,因此,增长率指标可能对盈余公告超额收益有预测能力。

另外,电商销售数据除了含有未预期盈余和未预期收入中的信息,可能还含有与产品市场相关的信息。线上销售增长率越高的公司,消费者对其产品的购买意愿比过去更高。在实际的投资过程中,无论是机构投资者还是个人投资者,除了关注上市公司披露的财务数据,也会重视一些非财务信息。上市公司在业绩发布会上介绍的业务发展、产品销售等方面信息,以及财务报告中的管理层讨论与分析等部分也都可能包含与产品销售目标、产品竞争力相关的信息。当这些信息被投资者获取时,就有可能引起股价的变化。而电商销售数据由于直接来自线上产品的每一笔销售记录,很可能也含有这些与产品市场相关的信息。所以,电商销售数据也可能因为含有这类与产品相关的非财务信息而对盈余公告累计超额收益有预测能力。基于上述分析,本文提出第四个假设。

研究假设 H4:线上销售额的增长率能够预测盈余公告累计超额收益。

三、样本选取与研究设计

(一)样本来源

本文使用的电商销售数据来自 Wind 资讯。Wind 资讯在 2018 年 12 月推出了“线上销量”数据产品,即另类投资数据。这套数据覆盖超过 170 个 A 股公司的近 700 个品牌的电商销售数据。本文也从其他金融数据平台获得了对应的 A 股公司在线下渠道的品牌信息,发现线上品牌与线下品牌的重合度超过 90%,即这些 A 股上市公司在线上销售和线下销售的品牌几乎一致。这套数据收集于天猫、淘宝、京东、一号店等主流电商平台,范围覆盖了 95% 以上的中国电商市场。Wind 依据每个品牌所属的上市公司对线上销售额数据进行月度汇总,从而得到各公司月度线上销售额数据。本文的实证分析基于月度线上销售数据进行,样本期间为 2015 年 1 月~2018 年 9 月。

在具体公司的选择上,由于本文需要使用标准化未预期收入(*SUR*)和标准化未预期盈余(*SUE*),这两个指标的计算需要向前滚动 8 个季度,又由于本文的主要控制变量需要前一期的 *SUR* 和 *SUE*,因此,本文最终选取了在 2013 年之前上市的 132 个消费行业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、文体娱、建材家具、家化等 7 个阿里电商一级品类。

本文使用的财务数据、股票交易数据、分析师关注数据均来自 CSMAR 数据库,机构持股数据来自 Wind 资讯。

(二)变量定义

1. 核心解释变量

本文的核心解释变量为线上销售增长率(*SG*),考虑到同比指标能避免季节性差异造成的干扰,本文使用线上销售额的同比增长率来衡量线上销售变化程度。具体计算公式如下:

$$SG_{i,t} = \log(Sales_{i,t}) - \log(Sales_{i,t-4})$$

其中, $SG_{i,t}$ 是公司 i 在季度 t 的线上销售增长率, $Sales_{i,t}$ 是公司 i 在季度 t 的线上销售额, $Sales_{i,t-4}$ 是公司 i 在季度 $t-4$ 的线上销售额。

2. 被解释变量

本文的被解释变量包括标准化未预期营业收入(*SUR*)、标准化未预期盈余(*SUE*)和盈余公告累计超额收益(*CAR*)。

标准化未预期营业收入(*SUR*)。该指标的构建主要借鉴了 Jegadeesh 和 Livnat (2006) 的研究,具体计算公式如下:

$$SUR_{i,t} = \frac{REV_{i,t} - REV_{i,t-4}}{\delta_{i,t}}$$

其中, $REV_{i,t}$ 是公司*i*在季度*t*的营业收入, $REV_{i,t-4}$ 是公司*i*在去年同季度的营业收入, $\delta_{i,t}$ 是*t*季度之前8个季度未预期营业收入的标准差^⑤。

标准化未预期盈余(*SUE*)。本文参考 Jegadeesh 和 Livnat (2006)、张然等 (2017) 的做法来构建 *SUE* 指标,具体计算公式如下:

$$SUE_{i,t} = \frac{Earnings_{i,t} - Earnings_{i,t-4}}{\xi_{i,t}}$$

$Earnings_{i,t}$ 是公司*i*在季度*t*的营业利润, $Earnings_{i,t-4}$ 是公司*i*在去年同季度的营业利润, $\xi_{i,t}$ 是*t*季度之前8个季度未预期营业利润的标准差。

盈余公告累计超额收益(*CAR*)。借鉴 Froot 等 (2017) 的研究,将盈余公告累计超额收益定义为[-1,+3]窗口相对于市场收益率的累计超额收益。

3. 其他变量

控制变量主要包括资产收益率(*ROA*)、账面市值比(*BM*)、总市值的自然对数(*Log_Size*)和股票的过去收益(*Pastreturn*)。各变量的具体定义如表1所示。

(三) 研究设计

本文的主要实证由两部分组成,一是检验线上销售额指标是否可以预测公司的基本面,二是检验线上销售额指标的信息是否会反映到股票市场。

在第一部分检验中,本文首先需要对线上销售额与营业收入之间的关系展开探讨,因为判断线上销售额能否成为预测公司财务信息的领先指标的前提是线上销售额应当与公司的营业收入、营业利润等指标具有高度相关性。为了检验线上销售额和营业收入是否存在正相关关系,本文采用以下模型:

$$\text{Log_Revenue}_{i,t} - \text{Log_Revenue}_{i,t-1} = \beta_0 + \beta_1 \times (\text{Log_Sales}_{i,t} - \text{Log_Sales}_{i,t-1}) + \varepsilon_{i,t} \quad (1)$$

其中, $\text{Revenue}_{i,t}$ 是公司*i*在*t*季度的营业收入, $\text{Log_Sales}_{i,t}$ 是公司*i*在季度*t*的线上销售额的自然对数。为了避免公司个体因素和时间因素的影响,在实证过程中控制公司固定效应和时间固定效应。

表1 主要变量定义

变量符号	变量名称	变量定义
<i>SUR</i>	标准化未预期收入	<i>t</i> 季度未预期营业收入与 <i>t</i> 季度之前的8个季度未预期营业收入的标准差的比值;其中, <i>t</i> 季度未预期营业收入是 <i>t</i> 季度营业收入减去 <i>t-4</i> 季度的营业收入
<i>SUE</i>	标准化未预期盈余	<i>t</i> 季度未预期营业利润与 <i>t</i> 季度之前的8个季度未预期营业利润的标准差的比值;其中, <i>t</i> 季度未预期营业利润是 <i>t</i> 季度营业利润减去 <i>t-4</i> 季度的营业利润
<i>CAR</i> [-1,+3]	盈余公告[-1,3]窗口的累计超额收益	窗口期是盈余公告前1天至盈余公告后3天;超额收益等于股票收益率减去市场收益率
<i>CAR</i> [+4,+60]	盈余公告[+4,+60]窗口的累计超额收益	窗口期是盈余公告后4天至盈余公告后60天;超额收益等于股票收益率减去市场收益率
<i>SG</i>	线上销售额同比增长率	<i>t</i> 季度线上销售额的自然对数减去 <i>t-4</i> 季度线上销售额的自然对数
<i>ROA</i>	资产收益率	<i>t</i> 季度的营业利润除以 <i>t</i> 季度末的总资产
<i>BM</i>	账面市值比	<i>t</i> 季度的总资产除以 <i>t</i> 季度的总市值
<i>Log_Size</i>	总市值的自然对数	<i>t</i> 季度的总市值的自然对数
<i>Pastreturn</i>	股票的过去收益	盈余公告前30天至公告前3天的累计超额收益
<i>Analyst</i>	分析师关注人数	<i>t</i> 季度所在年份前一年分析师关注人数的总和
<i>Smooth</i>	净利润的平滑度	<i>t</i> 季度之前12个季度的净利润的标准差除以 <i>t</i> 季度之前12个季度的经营性现金流的标准差
<i>Stdta</i>	应计利润的波动率	<i>t</i> 季度之前12个季度的应计利润的标准差,其中,应计利润等于净利润减去经营性现金流
<i>Log_Sales</i>	线上销售额的自然对数	<i>t</i> 季度的线上销售额的自然对数

接下来,本文检验线上销售额的增长率是否对未预期营业收入有预测作用。参考Da等(2011)、Froot等(2017)、Huang(2018)的做法,本文构建如下检验模型:

$$SUR_{i,t} = \beta_0 + \beta_1 \times SG_{i,t} + \beta_2 \times SUR_{i,t-1} + \beta_3 \times X_{i,t-1} + \varepsilon_{i,t} \quad (2)$$

其中, $SUR_{i,t}$ 是公司*i*在季度*t*的标准化未预期营业收入, $SG_{i,t}$ 是公司*i*在季度*t*的线上销售额增长率, $SUR_{i,t-1}$ 是*t-1*季度的标准化未预期营业收入, $X_{i,t-1}$ 是控制变量,是公司*i*在*t-1*季度的其他一些可能影响 $SUR_{i,t}$ 的变量。由于另类数据的预测作用是在财务报告未公布时,对本季度的财务指标和基本面信息进行预测,预测所依赖的信息是*t*季度的电商销售数据与*t*季度之前的财务信息,因此,在该模型中需要使用*t*季度的线上销售增长率和*t-1*季度的控制变量。 $X_{i,t-1}$ 包括*t-1*季度的总市值(Lag_Size)、*t-1*季度的账面市值比(Lag_BM)、*t-1*季度的资产收益率(Lag_ROA)、盈余公告前30天至公告前3天的累计收益率(Pastreturn)。

然后,本文检验线上销售额的增长率是否可以预测标准化未预期盈余。回归方程如下:

$$SUE_{i,t} = \beta_0 + \beta_1 \times SG_{i,t} + \beta_2 \times SUE_{i,t-1} + \beta_3 \times SUR_{i,t-1} + \beta_4 \times X_{i,t-1} + \varepsilon_{i,t} \quad (3)$$

其中, $SUE_{i,t}$ 是公司*i*在季度*t*的标准化未预期盈余, $SUE_{i,t-1}$ 是上一季度的标准化未预期盈余。根据Jegadeesh和Livnat(2006),前一期的未预期盈余和未预期收入对本期的未预期盈余有显著影响。因此,在模型(3)中,本文同时加入可能对 $SUE_{i,t}$ 造成影响的前一期未预期收入和未预期盈余。 $X_{i,t-1}$ 与模型(2)相同,是可能影响 $SUE_{i,t}$ 的控制变量。

另外,本文也在模型(3)中加入*t*季度的标准化未预期营业收入 $SUR_{i,t}$ 。Froot等(2017)发现同期的标准化未预期营业收入和标准化未预期盈余有很高的相关性。本文也需要探讨线上销售额指标对 $SUE_{i,t}$ 的预测作用是否是来自 $SUR_{i,t}$ 。如果 $SG_{i,t}$ 可以预测 $SUR_{i,t}$,而 $SUR_{i,t}$ 对 $SUE_{i,t}$ 也有预测作用,则 $SG_{i,t}$ 对 $SUE_{i,t}$ 的预测作用有可能来自 $SUR_{i,t}$,即 $SG_{i,t}$ 和 $SUR_{i,t}$ 有一些相同的信息,正是这些相同的信息使线上销售额有预测作用。但也可能是,线上销售额与未预期营业收入所含的信息不完全相同,线上销售额指标还含有未预期营业收入之外的信息,这些信息对未预期盈余也有预测作用。因此,在模型(3)中加入同季度的线上销售额增长率和未预期营业收入可以对 $SG_{i,t}$ 预测作用的来源进行判断,并可以对 $SG_{i,t}$ 中信息含量的特征做进一步分析。

模型(1)至模型(3)主要检验电商销售数据是否含有公司基本面信息,之后本文检验这些信息能否反映到股票价格中,即线上销售额的增长率是否对本季度盈余公告累计超额收益有预测作用。电商销售数据中的信息对盈余公告超额收益的预测作用可能来源于两方面,一是电商销售数据含有标准化未预期收入和标准化未预期盈余中的信息,如果 $SG_{i,t}$ 对盈余公告超额收益有预测作用,可能是 $SUR_{i,t}$ 和 $SUE_{i,t}$ 中的信息在发挥作用。即投资者在盈余公告期间,对未预期盈余和未预期收入做出了反应。另一种预测作用的来源是电商销售数据中含有标准化未预期收入和标准化未预期盈余之外的信息。比如,投资者可能会对财务报告或业绩报告会中的非财务信息做出反应,而这些非财务信息可能与线上销售额的变动有相关性。因此,本文采用两个回归模型来检验 $SG_{i,t}$ 对盈余公告超额收益的预测作用。回归方程如下:

$$CAR[-1,+3] = \beta_0 + \beta_1 \times SG_{i,t} + \beta_2 \times X_{i,t} + \varepsilon_{i,t} \quad (4)$$

$$CAR[-1,+3] = \beta_0 + \beta_1 \times SG_{i,t} + \beta_2 \times SUE_{i,t} + \beta_3 \times SUR_{i,t} + \beta_4 \times X_{i,t} + \varepsilon_{i,t} \quad (5)$$

其中, $CAR[-1,+3]$ 是公司*i*在季度*t*的盈余公告前1天至公告后3天的5天窗口期的累计超额收益, $SG_{i,t}$ 是公司*i*在季度*t*的线上销售额的增长率指标, $X_{i,t}$ 是控制变量,与模型(2)和模型(3)中的控制变量一致。但与模型(2)和(3)不同的是,此处的控制变量均为*t*季度的变量,而非*t-1*季度的变量。

模型(4)与模型(5)的不同之处是,模型(5)同时控制了*t*季度的未预期收入和未预期盈余,以检验 $SG_{i,t}$ 的预测作用是因为含有未预期收入和未预期盈余中的信息,还是因为含有未预期收入和未预期盈余之外的信息。

本文还检验了电商销售数据的信息是否也存在盈余公告后的漂移现象(PEAD)[®]。通过检验线上销售额指标对盈余公告后超额收益的预测作用,可以判断 $SG_{i,t}$ 中的信息是否能够快速反映到股票市场中。回归方程如下:

$$CAR[+4, +60] = \beta_0 + \beta_1 \times SG_{i,t} + \beta_2 \times SUE_{i,t} + \beta_3 \times SUR_{i,t} + \beta_4 \times X_{i,t} + \varepsilon_{i,t} \quad (6)$$

模型(6)的解释变量、控制变量都与模型(5)相同,被解释变量是盈余公告后4天至公告后60天窗口的累计超额收益。

(四)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。 SG 的均值是0.426,中值为0.381。 SUR 的均值、中值、波动率都比 SUE 要高, SUR 的均值是1.194,标准差是1.785, SUE 的均值是0.62,标准差是1.592。另外,公告期超额收益和公告后超额收益的均值都高于0,其中, $CAR[-1, +3]$ 的均值是0.24%,中值小于0,是-0.1%, $CAR[+4, +60]$ 的均值是0.962%,中值是0.524%。

四、回归结果与分析

(一)电商销售数据对基本面指标的预测

表3是模型(1)的回归结果。为了避免公司不可观测因素和时间因素的影响,列(2)控制公司固定效应,列(3)控制季度固定效应,列(4)同时控制两种固定效应。从列(1)~列(4)的结果表明,每个季度的线上销售额与该季度的营业收入高度相关。这说明线上销售额含有与营业收入相关的信息,这是电商销售数据可以作为预测公司基本面的先行指标的基础。

表4是标准化未预期营业收入对线上销售额变动率的回归结果。列(1)不加入控制变量, $SG_{i,t}$ 的系数在1%水平上显著,另外, $SG_{i,t}$ 的系数是0.297,即1单位线上销售增长率的增加对应0.297单位标准化未预期收入的增长。从经济意义看, $SG_{i,t}$ 对 $SUR_{i,t}$ 的预测作用也比较明显。列(2)和列(3)加入控制变量, $SG_{i,t}$ 的系数仍显著为正,这表明在控制了可能影响未预期营业收入的主要变量后,线上销售额的增长率仍然对未预期的营业收入有很强的预测能力,这也进一步证明了电商销售额确实含有与公司营业收入相关的信息。

从Beaver(1968)、Ball和Brown(1968)开始,大量研究发现,盈余公告能向市场传达重要的增量信息。那么,能否在财报未公布时,使用电商销售数据对本季度的未预期盈余做出预测就显得尤为重要。

表5是标准化未预期盈余对线上销售额变动率的回归结果。(1)~(3)列结果显示 $SG_{i,t}$ 的系数为正,且在1%水平上显著。列(2)和列(3)在加入控制变量后, $SG_{i,t}$ 的系数仍显著为正。这表明电商销售数据对标准化未预期盈余有很强的预测作用,且预测作用很稳定。另外,从列(3)可以发现,上一期的未预期盈余和未预期收入都对

表4 电商销售数据对SUR的预测

	SUR		
	(1)	(2)	(3)
SG	0.297*** (2.92)	0.307*** (3.03)	0.270*** (2.66)
Lag_SUR			0.135*** (5.76)
Lag_Size		0.442** (2.25)	0.280 (1.42)
Lag_BM		0.304 (1.26)	0.139 (0.58)
Lag_ROA		-0.075** (-2.23)	-0.102*** (-3.05)
Pastreturn		1.222*** (2.66)	1.057** (2.32)
Constant	0.668*** (4.61)	-9.866** (-2.11)	-6.027 (-1.29)
Firm fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
N	1338	1338	1327
R ²	0.053	0.064	0.087

表2 主要变量描述性统计

Variables	Obs	Mean	Std	p25	p50	p75
SG	1338	0.426	0.474	0.148	0.381	0.648
SUR	1338	1.194	1.785	0.0667	0.901	2.087
SUE	1338	0.62	1.592	-0.139	0.402	1.26
CAR[-1, +3]	1338	0.0024	0.0498	-0.0239	-0.001	0.0264
CAR[+4, +60]	1338	0.00962	0.15	-0.0714	0.00524	0.0956
Size	1338	23.2	0.966	22.58	22.96	23.7
BM	1338	0.692	0.691	0.312	0.494	0.757
ROA	1338	0.041	0.063	0.011	0.031	0.067
Pastreturn	1338	0.00212	0.0939	-0.0531	-0.0044	0.0566

表3 线上销售额与营业收入的相关性

	Revenue Growth			
	(1)	(2)	(3)	(4)
$\Delta \log_Sales$	0.267*** (15.75)	0.269*** (15.09)	0.300*** (15.06)	0.305*** (14.51)
Constant	0.010 (1.15)	0.010 (1.10)	0.038 (1.28)	0.047 (1.50)
Firm fixed effects	No	Yes	No	Yes
Time fixed effects	No	No	Yes	Yes
N	1338	1338	1338	1338
R ²	0.159	0.159	0.196	0.196

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

表5 电商销售数据对SUE的预测

	SUE			
	(1)	(2)	(3)	(4)
SG	0.345*** (3.40)	0.348*** (3.43)	0.315*** (3.06)	0.229*** (2.33)
SUR				0.314*** (11.16)
Lag_SUE			0.044** (2.26)	0.049*** (2.63)
Lag_SUR			0.040* (1.68)	-0.003 (-0.12)
Lag_Size		0.563*** (2.86)	0.522*** (2.61)	0.435** (2.28)
Lag_BM		0.180 (0.75)	0.125 (0.51)	0.079 (0.34)
Lag_ROA		-0.083** (-2.46)	-0.113*** (-3.23)	-0.083** (-2.48)
Pastreturn		0.754 (1.64)	0.722 (1.56)	0.397 (0.90)
Constant	0.203 (1.40)	-13.072*** (-2.80)	-12.144** (-2.56)	-10.262** (-2.27)
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	1338	1338	1327	1327
R ²	0.037	0.049	0.055	0.145

本期的未预期盈余有显著的影响,而上一期末预期盈余的影响作用更大,这也表明未预期盈余有较高的持续性,这与 Jegadeesh 和 Livnat (2006)、Da 等 (2011)、Huang (2018) 的发现一致。

在表 5 的列 (4) 中,本文也加入了当期的未预期营业收入 $SUR_{i,t}$ 作为控制变量。从 $SUR_{i,t}$ 的系数可以看出,当期的未预期收入对当期的未预期盈余有显著正向影响,这和 Froot 等 (2017) 的发现一致。另外,虽然 $SG_{i,t}$ 的系数有一定程度的下降,但仍在 5% 水平上显著。这说明电商销售数据对未预期盈余的预测作用一方面是由于电商销售数据中含有营业收入、未预期收入中的信息,另一方面是由于电商销售数据中还含有未预期营业收入之外的信息,而这些信息也对未预期盈余有预测能力。

(二) 电商销售数据对股票收益的预测

表 6 是 $CAR[-1,+3]$ 对 $SG_{i,t}$ 的回归结果。列 (1) 不加入控制变量, $SG_{i,t}$ 的系数是 0.008, 且在 5% 水平上显著。该结果的经济意义同样显著,线上销售增长率增加一单位对应的累计超额收益增长 0.8%, 近似等于 66.7% 的年化收益率。列 (2) 和列 (3) 加入控制变量后结果保持不变。另外,从列 (3) 中 $SUE_{i,t}$ 和 $SUR_{i,t}$ 的系数来看, $SUE_{i,t}$ 的系数是 0.006, 在 1% 水平上显著。 $SUR_{i,t}$ 的系数是 0.002, 在 10% 水平上显著。说明当期的未预期盈余和当期的未预期收入都是造成盈余公告超额收益的重要因素,而市场投资者对未预期盈余中所含信息的反应要超过对未预期收入中所含信息的反应,这与以往相关文献的研究相符。另外, Agarwal 等 (2020) 也得出了类似结论,他们发现在控制当期的 SUE 和 SUR 后,信用卡消费指标也对股票收益有很强的影响。总之,上述结论说明电商销售额包含了未预期收入和未预期盈余之外的增量信息,这些信息也能预测未来的股票收益。

本文已证明电商销售数据包含有价值信息,可以对盈余公告前一天至公告后 3 天的累计超额收益有预测作用。本文关心的另一个问题是,线上销售增长率中的信息是在盈余公告窗口内就充分反映到股价中,还是也存在公告后的漂移现象? 为了研究该问题,本文用盈余公告后的超额收益 $CAR[+4,+60]$ 对本季度的线上销售增长率回归,检验 $SG_{i,t}$ 是否对 $CAR[+4,+60]$ 有预测作用。表 7 是模型 (6) 的回归结果。虽然表 7 中 $SG_{i,t}$ 的系数始终大于 0,但并不显著。这说明市场投资者对电商销售数据中的增量信息在盈余公告窗口期内有充分的反应,不存在公告后的漂移现象。

五、进一步研究

在第四部分,本文的研究逻辑是通过检验电商销售数据对标准化未预期收入 (SUR)、标准化未预期盈余 (SUE)、盈余公告累计超额收益的预测作用研究电商销售数据这种另类数据是否具有信息含量。第四部分的研究结果表明,电商销售数据对基本面指标和股价的市场反应都有预测作用。

然而,本文未曾深入探讨为什么电商销售数据对基本面指标具有预测能力,即电商销售数据为什么含有公司基本面信息的问题。虽然,本文的理论假设部分对电商销售数据预测作用的内在机理进行了分析,认为是由于电商销售数据直接来自公司的经营和销售活动,从而具有了公司基本面相关的信息含量。但是,本文并未对这一猜测进行深入的实证检验。因此,在第五部分,本文首先通过异质性检验分析电商销售数据对公司基本面指标具有预测作用的内在机理。具体来说,本文通过研究指标噪音对电商销售数据预测作用的影响来进行分析。

表 6 电商销售数据对 $CAR[-1,+3]$ 的预测

	$CAR[-1,+3]$		
	(1)	(2)	(3)
SG	0.008** (2.31)	0.008** (2.40)	0.007** (1.97)
SUE			0.006*** (4.83)
SUR			0.002* (1.90)
$Size$		-0.022*** (-3.26)	-0.026*** (-3.82)
BM		0.004 (0.47)	0.002 (0.29)
ROA		0.003** (2.46)	-0.001 (-0.43)
$Pastreturn$		0.002 (0.14)	-0.004 (-0.29)
Constant	0.016*** (3.23)	0.522*** (3.27)	0.612*** (3.85)
Firm fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
N	1338	1338	1338
R ²	0.035	0.052	0.078

表 7 电商销售数据对 $CAR[+4,+60]$ 的预测

	$CAR[+4,+60]$		
	(1)	(2)	(3)
SG	0.002 (0.26)	0.007 (0.80)	0.007 (0.77)
SUE			0.004 (1.26)
SUR			-0.002 (-0.75)
$Size$		-0.118*** (-6.85)	-0.118*** (-6.78)
BM		0.022 (1.12)	0.023 (1.17)
ROA		0.004 (1.23)	0.002 (0.61)
$Pastreturn$		0.079** (2.00)	0.079** (1.98)
Constant	0.062*** (4.84)	2.767*** (6.85)	2.763*** (6.78)
Firm fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
N	1338	1338	1338
R ²	0.067	0.129	0.130

另外,本文也未曾深入探讨为什么电商销售数据对盈余公告期间的股票累计超额收益有预测作用。虽然在理论假设部分本文猜测是由于投资者对电商销售数据中的信息做出反应造成了股票价格的上涨或下跌,但对这一猜测本文缺少深入的实证检验。因此,本文也通过异质性检验分析投资者在电商销售数据预测能力中发挥的作用。具体来说,本文设计了3个异质性检验,分别在投资者关注度、投资者对公司特征的不同反应、投资者对电商销售数据的使用方式等3个方面开展研究。

(一)指标噪音与信息含量

本文在理论假设部分提出,电商销售数据可能具有信息含量的原因是这些直接来自公司销售活动的数据既包含了与公司收入、利润相关的现金流信息,也包含了公司产品市场层面的信息。正是基于这一逻辑,本文用电商销售数据预测公司的基本面指标,以此反推电商销售数据含有基本面信息。然而,如果这一理论假设本身是错误的,即电商销售数据可能并不含有基本面信息,或者即使含有部分基本面信息,但也不足以作为公司基本面信息的衡量时,则本文的研究结果将不再成立。

因此,为了更严谨的论证电商销售数据的信息含量问题,本文根据线上销售额占公司营业收入的比重,将研究样本分为两组进行异质性检验。从线上和线下销售渠道来看,如果线上销售额占公司营业收入的比重越高,则电商销售数据能够反映公司实际经营状况的准确性就越高,即根据电商销售数据构造的预测指标的噪音就越少。反之,如果线上销售额占公司营业收入的比重越低,则电商销售数据反映的公司经营信息与实际状况的偏离程度就越大。

如果本文在理论假设部分的分析是正确的,则在噪音越低,即线上销售收入占总销售收入比重越大的公司样本中,线上销售额同比增长率能够反映真实公司经营状况的准确度就越高,电商销售数据对标准化未预期收入和标准化未预期盈余等基本指标预测作用就越强。反之,噪音越高,电商销售数据的预测作用越弱。

因此,在两类样本组中,本文重新使用线上销售额的同比增长率 $SG_{i,t}$ 对标准化未预期收入和标准化未预期盈余进行预测。表8结果表明,无论是对未预期收入的预测还是对未预期盈余的预测,在线上销售额占营业收入比重高的样本组中的预测作用都要优于线上销售额占比低的样本组,即电商销售数据在线上占比高的样本组的信息含量更多。这也直接证明了电商销售数据的预测作用和信息含量的来源确实是因为电商销售数据直接来自公司的销售记录,本身就包含了公司经营活动相关的基本面信息。

(二)投资者关注与信息含量

为了深入分析电商销售数据对盈余公告超额收益有预测作用的原因,本文首先探讨投资者关注度是否会影响电商销售数据的预测作用。电商在中国已经发展了十余年时间,一些证券分析师已将电商数据作为分析公司基本面的重要信息源。因此,如果本文在理论假设部分的分析是正确的,即是由于投资者能够直接获取电商销售数据中的信息或间接获取与电商销售数据包含的信息相同的信息,并基于这些信息买卖股票,从而使电商销售数据对盈余公告超额收益有预测作用,则投资者对这些信息获取的程度和使用的程度会直接影响这种预测作用的大小。其中的逻辑是,被投资者关注更多的公司,电商销售数据所含的信息更可能在盈余公告前就被投资者使用。Huang(2018)发现购物网站的消费者评论信息会被对冲基金等机构投资者使用,导致投资者关注度高的公司的超额收益要比投资者关注度低的公司超额收益小。因此,本文希望通过检验投资者关注度对电商销售数据预测效果的影响来验证电商销售数据具有预测作用的原因。

表8 噪音高低与基本面预测

	SUR		SUE	
	线上占比大	线上占比小	线上占比大	线上占比小
SG	0.310* (1.70)	0.219* (1.70)	0.632*** (3.22)	0.197 (1.63)
Lag_SUE			0.004 (0.17)	0.121*** (3.50)
Lag_SUR	0.191*** (4.67)	0.164*** (4.14)	0.073* (1.66)	0.019 (0.66)
Lag_Size	0.175 (0.56)	0.136 (0.51)	0.741** (2.20)	0.381 (1.49)
Lag_BM	-0.178 (-0.50)	0.189 (0.53)	0.514 (1.34)	-0.269 (-0.80)
Lag_ROA	-0.117** (-2.39)	-0.096** (-1.99)	-0.107** (-2.22)	-0.135*** (-2.89)
Pastreturn	1.311** (2.08)	0.837 (1.23)	1.014 (1.49)	0.585 (0.91)
Constant	-3.534 (-0.47)	-2.794 (-0.44)	-17.716**	-8.528 (-1.41)
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	668	659	668	659
R ²	0.128	0.080	0.078	0.072

本文用两个指标衡量投资者关注度,一个是发布过关于公司报告的分析师人数,另一个是公司市值。这两个指标在以往文献中都被广泛的用作投资者关注度的代理变量。之后,本文分别在每个组中做 $CAR[-1,+3]$ 对 $SG_{i,t}$ 的回归分析。表 9 是回归结果,结果显示,在分析师关注度低的公司和小市值的公司中,线上销售增长率对盈余公告超额收益的预测作用更强。该结果验证了本文的猜测,即对投资者关注度高的公司而言,电商销售额中的信息会更充分的在盈余公告前被市场使用。这也直接证明了本文理论分析的正确,即电商销售数据对股票收益预测作用的来源正是由于投资者对这些信息做出反应造成的。

(三)公司成长性与信息含量

在从投资者关注度的角度研究了电商销售数据预测作用的来源之后,本文也关心投资者是否有效或理性地使用电商销售数据中的信息。考虑到电商销售额是由消费者的每笔线上消费金额聚合而成的,电商销售数据很可能包含与公司未来经营前景,特别是消费者对公司的需求或购买意愿相关的信息。而对于不同类型的公司而言,投资者对这些信息的反应可能会有差异。对于成长型公司,投资者可能对这些反映公司前景的信息反应更为强烈,因为成长型公司的市场估值会对公司未来的潜在现金流和增长前景更为敏感。因此,本文根据账面市值比将样本分为两组,即低账面市值比的公司和高账面市值比的公司,分别代表成长型公司和价值型公司。本文猜测电商销售数据中的这些可能代表公司未来销售前景的信息在成长型公司中的预测作用更强,即投资者对电商销售数据的信息的反应程度更大。

表 10 是 $CAR[-1,+3]$ 对 $SG_{i,t}$ 的回归结果。第一列是在成长型公司中的结果,第二列是在价值型公司中的回归结果。结果显示,线上销售增长率对盈余公告超额收益的预测作用在成长股中更强。这也从投资者对成长型公司和价值型公司的信息反应程度的层面证明了电商销售数据对股票收益预测作用的来源确实是投资者反应造成的。

(四)信息透明度与信息含量

通过投资者关注度和投资者对成长型和价值型公司的不同反应等两个方面的研究,本文已证明正是由于投资者对电商销售数据中的信息做出反应,才使得电商销售数据对股票收益具有预测作用。为了研究问题和研究逻辑的完整性,本文继续探讨投资者究竟如何使用电商销售数据中的信息。

电商销售数据在本质上是一种非财务信息。在会计和审计实务中,非财务信息是判断公司是否有盈余管理行为的重要工具。如 Allee 等(2019)发现用电力大数据可以判断公司是否有盈余管理;Chiu 等(2018)用 Google 搜索指数判断公司是否有营业收入操纵的行为。另外,公司信息的透明度也会直接影响投资者对信息的使用行为。在该部分,本文研究公司的财务信息透明度是否影响投资者对信息的使用行为,并进而影响电商销售数据的预测能力。即如果公司存在盈余管理,或者公司的财务信息透明度较差,则电商销售数据的预测作用是更强还是更弱?

本文认为有两种可能。一种可能是,如果公司财务信息的透明度低,则市场投资者、分析师会更依赖于非财务信息做出投资决策,即在盈余管理程度高的公司中,电商销售数据的信息含量要相对高于在盈余管理程度低的公司中的信息含量。第二种可能是,投资者和分析师更依赖非财务信息对公司的财务信息

表 9 投资者关注度与 $CAR[-1,+3]$ 预测

	$CAR[-1,+3]$			
	分析师关注度高	分析师关注度低	大市值	小市值
<i>SG</i>	0.005 (0.92)	0.009** (2.21)	0.004 (0.74)	0.008* (1.84)
<i>SUE</i>	0.007*** (3.51)	0.003** (2.06)	0.004** (2.03)	0.007*** (4.16)
<i>SUR</i>	0.001 (0.69)	0.002 (1.54)	0.003* (1.86)	0.000 (0.27)
<i>Size</i>	-0.032** (-2.20)	-0.035*** (-3.71)	-0.021 (-1.46)	-0.037*** (-3.24)
<i>BM</i>	0.005 (0.31)	-0.007 (-0.76)	0.004 (0.23)	0.007 (0.64)
<i>ROA</i>	0.001 (0.36)	0.000 (0.20)	0.001 (0.54)	-0.001 (-0.63)
<i>Pastreturn</i>	0.006 (0.22)	-0.015 (-0.77)	0.022 (0.88)	-0.040* (-1.85)
Constant	0.753** (2.19)	0.803*** (3.76)	0.500 (1.45)	0.848*** (3.27)
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	674	664	675	663
R ²	0.098	0.095	0.068	0.114

表 10 公司成长性与 $CAR[-1,+3]$ 预测

	$CAR[-1,+3]$	
	成长股	价值股
<i>SG</i>	0.009* (1.85)	0.001 (0.24)
<i>SUE</i>	0.005*** (2.90)	0.004** (2.47)
<i>SUR</i>	0.002 (1.23)	0.002 (0.97)
<i>Size</i>	0.002 (0.09)	-0.045*** (-5.31)
<i>BM</i>	0.058 (0.74)	-0.013** (-2.29)
<i>ROA</i>	-0.003* (-1.70)	0.003* (1.79)
<i>Pastreturn</i>	-0.004 (-0.14)	0.001 (0.05)
Constant	-0.054 (-0.10)	1.069*** (5.44)
Firm fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
N	666	672
R ²	0.093	0.128

进行印证,并在财务信息和非财务信息可以相互印证的基础上做出投资决策。在这种情况下,在盈余管理程度低的公司中,非财务信息和财务信息更容易得到相互印证,从而表现为电商销售数据对股票收益的预测作用要比在盈余管理程度高的公司中更强。

目前,国内外学者对这一问题的探讨仍不充分,因此本文希望对这一问题进行检验。本文选用了两种衡量公司盈余管理程度或信息透明度的指标,分别是盈余平滑度和总应计利润的波动率。并分别根据这两个指标将公司分为两组,在每个组中检验 $SG_{i,t}$ 对 $CAR[-1,+3]$ 的预测效果。回归结果如表 11 所示。结果显示, $SG_{i,t}$ 在应计波动性低的组和盈余平滑度高的组中的预测作用都高于其他两组。这表明

电商销售额这种非财务信息的预测能力会受到公司财务信息操纵程度的影响,在盈余管理程度低的公司中,这种非财务信息的预测作用更好。

六、稳健性检验

为了确保本文研究结论的可靠性,本文进行了以下稳健性检验:(1)调整基本面指标的定义,将前文中不含季节漂移项的标准化未预期收入和标准化未预期盈余改为含季节漂移项的标准化未预期收入和标准化未预期盈余;(2)更换线上销售额增长率指标的计算方法,参考 Froot 等(2017)和 Huang(2018)的处理方法,本文将线上销售额的未预期增长率的定义由同比增长率改为 t 季度线上销售额相比于前四个季度线上销售额均值的对数增长率;(3)加入其他控制变量,将机构持股比例、换手率、分析师关注度等可能对盈余公告超额收益产生影响的变量和同季度营业收入这一可能干扰电商销售数据预测作用的变量也都作为控制变量加入回归方程;(4)采用行业固定效应,为避免行业因素对预测效果的影响,本文也做了同时控制行业固定效应和季度固定效应的回归检验;(5)更换盈余公告窗口,在对盈余公告超额收益进行预测时,本文也参考 Katona 等(2018)的窗口设定,并考虑投资者对信息的充分反应,检验了 $[-2,+2]$ 和 $[-3,+3]$ 两个窗口;(6)更换累计超额收益率的计算方法,在对盈余公告超额收益进行预测时,本文也更换了累计超额收益率的计算方法,在前文的实证分析中,本文使用股票收益率减去市场收益率作为超额收益率,为了避免不同的超额收益率计算方法对实证结果的影响,本文重新使用 CAPM 和 Fama-French 三因子模型计算超额收益率。在上述稳健性检验中,主要结果均保持不变。限于篇幅,此处不列示表格^⑦。

七、总结

本文基于 Wind 资讯提供的电商销售数据,通过构造消费行业上市公司线上销售同比增长率指标,研究了电商销售数据是否具有信息含量的问题。研究表明,电商销售额的增长率可以显著预测未预期收入和未预期盈余,同时对盈余公告累计超额收益也有显著预测能力。本文进一步通过异质性检验分析了不同公司特征对预测效果的影响。指标噪音、投资者关注度、公司成长性和公司财务信息透明度都会影响电商销售数据的预测效果和在股价中的反映程度。线上销售和实际销售差距越小,即指标的噪声越小,电商销售数据的预测效果越好。成长型公司的估值更容易受到公司发展前景的影响,在这类公司中电商销售数据的预测效果更好。另外,对于财务信息透明度更高和投资者关注度更低的公司,电商销售数据在预测公司基本面信息和股价反应时具有更强预测能力。本文也基于这些异质性检验深入分析了电商销售数据对基本面指标和盈余公

表 11 会计信息透明度与 $CAR[-1,+3]$ 预测

	$CAR[-1,+3]$			
	应计波动性高	应计波动性低	盈余平滑度高	盈余平滑度低
<i>SG</i>	0.004 (0.69)	0.013*** (2.85)	0.008* (1.71)	0.006 (0.98)
<i>SUE</i>	0.005*** (3.33)	0.007*** (3.79)	0.005*** (3.15)	0.006*** (3.65)
<i>SUR</i>	0.001 (0.62)	0.002 (1.17)	0.001 (0.59)	0.003* (1.75)
<i>Size</i>	-0.027** (-2.20)	-0.024*** (-2.59)	-0.039*** (-3.34)	-0.022** (-2.43)
<i>BM</i>	0.003 (0.29)	0.006 (0.46)	0.007 (0.53)	0.000 (0.01)
<i>ROA</i>	0.002 (1.16)	-0.004** (-2.04)	0.006* (1.92)	-0.002 (-1.48)
<i>Pastreturn</i>	0.002 (0.08)	-0.010 (-0.46)	-0.036 (-1.56)	0.015 (0.66)
Constant	0.635** (2.18)	0.557*** (2.63)	0.900*** (3.29)	0.519** (2.49)
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	677	661	667	671
R ²	0.100	0.078	0.113	0.091

告累计超额收益具有预测作用的原因。

本研究在理论层面具有一定的贡献。首先,本研究利用新颖的与公司营业收入直接相关的电商销售数据来研究另类数据的信息含量,为另类数据是否具有信息含量提供了新的证据,丰富了另类数据领域的文献,也克服了以往非财务数据存在的不足,如聚合观点或意见的社交媒体文本信息容易受认知偏差和行为偏差的影响而有较高噪音。其次,本研究关注消费者线上消费活动是否能向金融市场传递与公司价值相关的信息。将消费者行为与公司价值、股票价格和股价信息含量相关联,是近年来学术研究的热点。而本文的研究提供了消费者行为可以传递有价值信息的新的经验证据。最后,已有研究主要探讨了成熟资本市场中另类数据的信息含量,但鲜有文献探讨新兴资本市场中另类数据的价值相关性,而本研究提供了有益补充。

本研究在实践层面也具有一定的启示意义。第一,随着另类数据的传播和大规模使用,新的资本市场信息日益增多,资产价格的波动可能会更剧烈,不同类别资产价格对各类信息的反应机制可能也会更加复杂,这都会导致资本市场潜在风险的增加。而如何应对另类数据造成的监管风险,需要引起相关监管机构的重视。第二,另类数据能够为投资者提供前瞻性的信息,有助于市场参与者及时了解上市公司当前与未来的发展状况。但是,另类数据普遍含有噪声,并不是所有的另类数据都有投研价值。即使是有投研价值的另类数据,也未必可以获取超出市场回报的收益。因此,在另类数据的收集和清洗过程中,市场参与者应当注重另类数据噪声的剥离和投资信号的测试。

(作者单位:廖理,清华大学五道口金融学院;崔向博,中国工商银行博士后科研工作站;孙琼,北京工商大学经济学院)

注释

①另类数据是指投资研究中使用的非传统来源的新型数据。相较于传统数据,其数据体量更大、时效性和真实性更强、颗粒度更细。另类数据最初并非应用于投资领域,而是用于不同行业和其他需求,比如,在大众点评上给餐厅评价、在淘宝网上给商品评论、在浏览器或导航软件中搜索某些商场的地址等等。单个的这些数据点可能是没有规律的,但是,当这些数据被汇聚到一起,就可能反映群体活动的特点和趋势。这些平常的、原始的数据经过金融数据公司的整理、清洗和测试,就可能成为有投资价值的另类投资数据。

②包括上市公司年度报告、资本市场交易数据、银行的借贷数据、管理层自我陈述及分析师报告等。

③数据来源于国家统计局。

④数据来源:中国互联网络信息中心(CNNIC)完成的第44次《中国互联网络发展状况统计报告》,报告中详细分析了中国网民规模情况。

⑤在计算标准差时,之所以使用过去8个季度,是因为用更少的季度可能会导致误差较大。在计算标准化未预期收入和标准化未预期盈余时,国内外文献通常采用过去8个季度,如Jegadeesh和Livnat(2006)、张然等(2017)。

⑥盈余公告后价格漂移现象(Post-Earnings-Announcement Drift)是指由于投资者的注意力有限,对新信息的反应不足,在盈余公告后,未预期盈余为正的股票累计超额收益会持续上扬,呈现盈余惯性现象。该现象首次由Ball和Brown(1968)证明发现,之后被受到广泛的关注。

⑦为了节省文章篇幅,稳健性检验结果未在文中汇报,留存备案。

参考文献

- (1)曾庆生、周波、张程、陈信元:《年报语调与内部人交易:“表里如一”还是“口是心非”?》,《管理世界》,2018年第9期。
- (2)张然、汪荣飞、王胜华:《分析师修正信息、基本面分析与未来股票收益》,《金融研究》,2017年第7期。
- (3)Agarwal, S., Qian, W. L. and Zou, X., 2020, “Disaggregated Sales and Stock Returns”, *Management Science*, Forthcoming.
- (4)Allee, K. D., Baik, B. and Roh, Y., 2019, “Detecting Financial Misreporting with Real Production Activity”, SSRN Working Paper, No.2856472.
- (5) Bai, J., Philippon, T. and Savov, A., 2016, “Have Financial Markets Become More Informative?”, *Journal of Financial Economics*, Vol.122(3), pp.625~654.
- (6) Baker, M. and Wurgler, J., 2007, “Investor Sentiment in the Stock Market”, *Journal of Economic Perspectives*, Vol.21(2), pp.129~152.
- (7) Ball, R. and Brown, P., 1968, “An Empirical Evaluation of Accounting Income Numbers”, *Journal of Accounting Research*, Vol.6(2), pp.159~178.
- (8) Banerjee, S., Davis, J. and Gondhi, N., 2018, “When Transparency Improves, Must Prices Reflect Fundamentals Better?”, *The Review of Financial Studies*, Vol.31(6), pp.2377~2414.
- (9) Bartov, E., Faurel, L. and Mohanram, P. S., 2018, “Can Twitter Help Predict Firm-Level Earnings and Stock Returns?”, *The Accounting Review*, Vol.93(3), pp.25~57.
- (10) Beatty, S. E. and Smith, S. M., 1987, “External Search Effort: An Investigation across Several Product Categories”, *Journal of Con-*

sumer Research, Vol.14(1), pp.83~95.

(11) Beaver, W., 1968, "The Information Content of Annual Earnings Announcements", *Journal of Accounting Research*, Vol.6, pp.67~92.

(12) Boudoukh, J., Feldman, R., Kogan, S. and Richardson, M., 2019, "Information, Trading, and Volatility: Evidence from Firm-Specific News", *The Review of Financial Studies*, Vol.32(3), pp.992~1033.

(13) Chen, H., De, P., Hu, Y. J. and Hwang, B. H., 2014, "Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media", *The Review of Financial Studies*, Vol.27(5), pp.1367~1403.

(14) Chen, Y., Kelly, B. and Wu, W., 2020, "Sophisticated Investors and Market Efficiency: Evidence from a Natural Experiment", *Journal of Financial Economics*, Vol.138(2), pp.316~341.

(15) Chiu, P. C., Teoh, S. H., Zhang, Y. and Huang, X., 2018, "Using Google Searches of Firm Products to Assess Revenue Quality and Detect Revenue Management", SSRN Working Paper, No.3252314.

(16) Cookson, J. A. and Niessner, M., 2020, "Why Don't We Agree? Evidence from a Social Network of Investors", *The Journal of Finance*, Vol.75(1), pp.173~228.

(17) Da, Z. and Huang, X., 2020, "Harnessing the Wisdom of Crowds", *Management Science*, Vol.66(5), pp.1847~1867.

(18) Da, Z., Engelberg, J. and Gao, P., 2011, "In Search of Fundamentals", Unpublished Working Paper.

(19) Das, S. R. and Chen, M. Y., 2007, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web", *Management Science*, Vol.53(9), pp.1375~1388.

(20) DellaVigna, S. and Pollet, J. M., 2009, "Investor Inattention and Friday Earnings Announcements", *The Journal of Finance*, Vol.64(2), pp.709~749.

(21) Froot, K., Kang, N., Ozik, G. and Sadka, R., 2017, "What do Measures of Real-time Corporate Sales Say about Earnings Surprises and Post-announcement Returns?", *Journal of Financial Economics*, Vol.125(1), pp.143~162.

(22) Goldstein, I. and Yang, L., 2019, "Good Disclosure, Bad Disclosure", *Journal of Financial Economics*, Vol.131(1), pp.118~138.

(23) Green, T. C., Huang, R., Wen, Q. and Zhou, D., 2019, "Crowdsourced Employer Reviews and Stock Returns", *Journal of Financial Economics*, Vol.134(1), pp.236~251.

(24) Hirshleifer, D. and Teoh, S. H., 2003, "Limited Attention, Information Disclosure, and Financial Reporting", *Journal of Accounting and Economics*, Vol.36(1~3), pp.337~386.

(25) Huang, J., 2018, "The Customer Knows Best: The Investment Value of Consumer Opinions", *Journal of Financial Economics*, Vol.128(1), pp.164~182.

(26) Ittner, C. D., Larcker, D. F. and Meyer, M. W., 2003, "Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard", *The Accounting Review*, Vol.78(3), pp.725~758.

(27) Jagtiani, J. and Lemieux, C., 2019, "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform", *Financial Management*, Vol.48(4), pp.1009~1029.

(28) Jame, R., Johnston, R., Markov, S. and Wolfe, M. C., 2016, "The Value of Crowdsourced Earnings Forecasts", *Journal of Accounting Research*, Vol.54(4), pp.1077~1110.

(29) Jegadeesh, N. and Livnat, J., 2006, "Revenue Surprises and Stock Returns", *Journal of Accounting and Economics*, Vol.41(1~2), pp.147~171.

(30) Kacperczyk, M., Van Nieuwerburgh, S. and Veldkamp, L., 2016, "A Rational Theory of Mutual Funds' Attention Allocation", *Econometrica*, Vol.84(2), pp.571~626.

(31) Katona, Z., Painter, M., Patatoukas, P. and Zeng, J., 2018, "On the Capital Market Consequences of Alternative Data: Evidence from Outer Space", SSRN Working Paper, No.3222741.

(32) Kempf, E., Manconi, A. and Spalt, O., 2017, "Distracted Shareholders and Corporate Actions", *The Review of Financial Studies*, Vol.30(5), pp.1660~1695.

(33) Loder, A., 2019, "Goldman Rolls Out New ETFs Focused on Artificial Intelligence", *Wall Street Journal*, 2019-03-07.

(34) Mclean, R. D. and Pontiff, J., 2016, "Does Academic Research Destroy Stock Return Predictability?", *The Journal of Finance*, Vol.71(1), pp.5~32.

(35) Shapiro, C., 1983, "Premiums for High Quality Products as Returns to Reputations", *The Quarterly Journal of Economics*, Vol.98(4), pp.659~679.

(36) Tang, V. W., 2018, "Wisdom of Crowds: Cross-sectional Variation in the Informativeness of Third-Party-Generated Product Information on Twitter", *Journal of Accounting Research*, Vol.56(3), pp.989~1034.

(37) Zhu, C., 2019, "Big Data as a Governance Mechanism", *Review of Financial Studies*, Vol.32(5), pp.2021~2061.

The Information Content of Alternative Data: Evidence from E-commerce Sales

Liao Li^a, Cui Xiangbo^b and Sun Qiong^c

(a. PBC School of Finance, Tsinghua University; b. Postdoctoral Research Center, Industrial and Commercial Bank of China; c. School of Economics, Beijing Technology and Business University)

Summary: Alternative data refers to new types of data from non-traditional sources used in investment research, such as satellite image data and consumer behavior data. Compared with traditional financial data, alternative data is updated more timely, and can be updated daily or monthly. With the enhancement of capital market transparency and the increase of information value, alternative data has become an important channel for timely access to market information. Some institutional investors have begun to explore the use of information in alternative data to obtain excess returns.

Alternative data has also attracted widespread attention in the academic field. Some scholars have studied the issue of whether alternative data has information content. Most studies believe that alternative data contains fundamental information of listed companies and has predictive power for future stock returns. However, some studies believe that there is too much noise in alternative data and it is difficult to extract valuable information. Therefore, whether alternative data can convey valuable information is still uncertain. In addition, the existing researches on alternative data are mainly concentrated in countries with developed capital markets, and there are few researches based on emerging markets. This article hopes to conduct research based on alternative data in the Chinese market to fill this gap.

E-commerce sales data is one of the most widely used alternative data. The Internet has become an important channel for retailing social goods in China, and the online sales of some listed companies in the consumer industry have exceeded offline sales. So, can online sales information reflect the actual operating conditions of listed companies? Can e-commerce sales data collected in real time be used to predict the performance of listed companies before the financial report is disclosed? This article studies these issues based on the e-commerce sales data provided by Wind Information.

The main conclusions are as follows. First, online sales have a positive correlation with operating income. Second, the growth rate of online sales can significantly predict standardized unexpected income, standardized unexpected earnings, and earnings announcement returns. This indicates that online sales convey information about future fundamentals of listed companies. In addition, we also prove that online sales contain incremental information beyond earnings surprises and revenue surprises. Third, the predictive power of e-commerce sales data is stronger among companies with a high proportion of online sales, companies with low investor attention, growth companies, and companies with high information transparency. Based on these heterogeneity tests, this article also analyzes the reasons why e-commerce sales data have predictive power.

The contributions of our study are as follows. Firstly, there is still a big controversy in academia over whether alternative data can convey valuable information. This research provides empirical evidence that alternative data contains valuable information based on e-commerce sales data. Secondly, consumers' online purchase behavior has received widespread attention from scholars. However, due to the availability of data, most existing studies use data types that indirectly measure consumer purchases, such as product search indexes and product reviews. Although these data can reflect consumers' preference and potential demand for the company's products to a certain extent, they cannot fully and accurately reflect actual transaction activities. The e-commerce sales data used in this article is directly obtained by summing up the final purchase records of consumers, which is more closely related to the actual purchase behavior of consumers. Finally, there are few literatures on the relationship between consumer behavior and stock returns based on Chinese data, and this research helps to enrich the literature in this field.

Keywords: E-commerce data; information content; alternative data

JEL Classification: D12, G12, G14