Internet Appendix

of

"Measuring accounting quality based on real-transaction online sales data"

Figure IA.1: Structure of online sales data

As an example, the structure of online sales data for Midea Group Co., Ltd. on Tmall are shown. There are three levels of data: firm-level, brand- or store-level, and product-level. Product-level data include each product's price, sales volume, and sales value. Brand- or store-level data are aggregated from product-level sales data. Sales data at the brand- or store-level are further aggregated to obtain firm-level data.

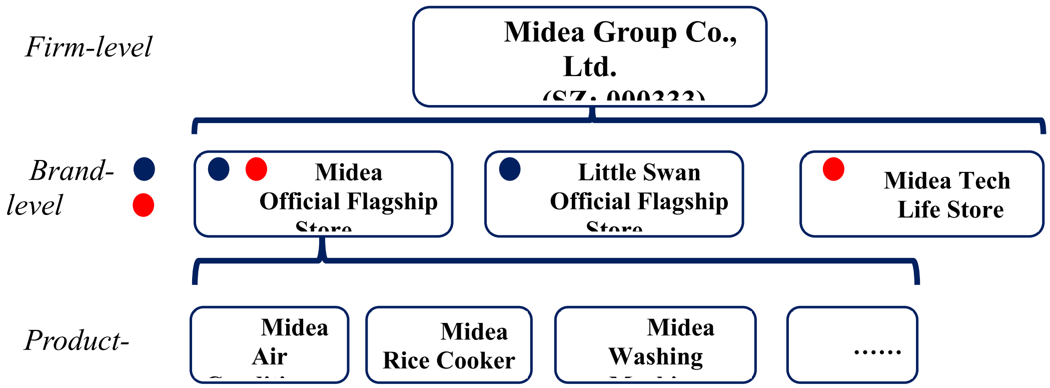

Figure IA.2: Example pages on the Tmall website

The homepage of the Tmall website (left) and the pages of a typical store (center) and a typical product (right) are shown.

Tmall website Typical store page Typical product

Table IA.1: Sample list of the stock codes and names of 201 publicly listed firms with available online sales data

Food and beverages (79)															
568	泸州老窖	895	双汇发展	2570	贝因美	2852	道道全	600186	莲花健康	600519	贵州茅台	600882	妙可蓝多	603589	口子窖
596	古井贡酒	2216	三全食品	2582	好想你	2891	中宠股份	600197	伊力特	600543	莫高股份	600887	伊利股份	603697	有友食品
639	西王食品	2304	洋河股份	2646	青青稞酒	2910	庄园牧场	600199	金种子酒	600559	老白干酒	601579	会稽山	603711	香飘飘
716	黑芝麻	2329	皇氏集团	2650	ST 加加	300146	汤臣倍健	600238	ST 椰岛	600597	光明乳业	603043	广州酒家	603777	来伊份
729	燕京啤酒	2330	得利斯	2661	克明面业	300673	佩蒂股份	600298	安琪酵母	600600	青岛啤酒	603156	养元饮品	603866	桃李面包
799	酒鬼酒	2461	珠江啤酒	2695	煌上煌	300783	三只松鼠	600300	维维股份	600616	金枫酒业	603198	迎驾贡酒	603886	元祖股份
848	承德露露	2481	双塔食品	2702	海欣食品	600059	古越龙山	600305	恒顺醋业	600702	ST 舍得	603288	海天味业	2387	维信诺
858	五粮液	2507	涪陵榨菜	2719	*ST 麦趣	600073	上海梅林	600365	ST 通葡	600779	水井坊	603345	安井食品	2515	金字火腿
860	顺鑫农业	2557	洽洽食品	2732	燕塘乳业	600084	ST 中葡	600419	天润乳业	600809	山西汾酒	603369	今世缘	603719	良品铺子
869	张裕 A	2568	百润股份	2847	盐津铺子	600132	重庆啤酒	600429	三元股份	600872	中炬高新	603517	绝味食品		

Textiles and apparel (37)							
2029	七匹狼	2569	ST 步森	600177	雅戈尔	603365	水星家纺
2127	南极电商	2574	明牌珠宝	600272	开开实业	603555	*ST 贵人
2154	报喜鸟	2612	朗姿股份	600295	鄂尔多斯	603587	地素时尚
2269	美邦服饰	2687	乔治白	600398	海澜之家	603608	天创时尚
2291	星期六	2731	萃华珠宝	600400	红豆股份	603808	歌力思
2293	罗莱生活	2875	安奈儿	600439	瑞贝卡	603877	太平鸟
2327	富安娜	300005	探路者	600884	杉杉股份	603908	牧高笛
2345	潮宏基	300577	开润股份	601566	九牧王		
2397	梦洁股份	600107	美尔雅	603001	奥康国际		
2563	森马服饰	600137	浪莎股份	603157	*ST 拉夏		

Medicine and biology (22)					
423	东阿阿胶	300147	香雪制药	603579	荣泰健康
538	云南白药	600085	同仁堂	688363	华熙生物
915	山大华特	600329	中新药业		
989	九芝堂	600332	白云山		
999	华润三九	600351	亚宝药业		
2044	美年健康	600436	片仔癀		
2223	鱼跃医疗	600572	康恩贝		
2275	桂林三金	600664	哈药股份		
2626	金达威	600750	江中药业		
2907	华森制药	600993	马应龙		

Electronic appliances (34)							
16	深康佳 A	2035	华帝股份	2705	新宝股份	603366	日出东方
100	TCL 科技	2076	*ST 雪莱	2959	小熊电器	603515	欧普照明
333	美的集团	2242	九阳股份	300625	三雄极光	603868	飞科电器
521	长虹美菱	2403	爱仕达	300824	北鼎股份	688169	石头科技
533	顺钠股份	2508	老板电器	600261	阳光照明		
541	佛山照明	2543	万和电气	600336	澳柯玛		
651	格力电器	2614	奥佳华	600690	海尔智家		
921	海信家电	2668	奥马电器	600839	四川长虹		
2005	ST 德豪	2677	浙江美大	600983	惠而浦		
2032	苏泊尔	2681	*ST 奋达	603355	莱克电气		

Furniture and building materials (13)			
910	大亚圣象	603818	曲美家居
2043	兔宝宝	603833	欧派家居
2081	金螳螂	603898	好莱客
2572	索菲亚		
2718	友邦吊顶		
300616	尚品宅配		
600337	美克家居		
603008	喜临门		
603600	永艺股份		
603816	顾家家居		

Household chemicals (9)					
523	广州浪奇	300740	御家汇	603605	珀莱雅
2094	青岛金王	600249	两面针	603630	拉芳家化
2511	中顺洁柔	600315	上海家化	603983	丸美股份

Culture, education, sports, and crafts (7)					
2103	广博股份	2678	珠江钢琴	603899	晨光文具
2301	齐心集团	300329	海伦钢琴		
2605	姚记科技	603398	邦宝益智		

Section IA.1: Online sales and future stock returns

In this section of the Internet Appendix, we examine whether online sales can inform investors about firms' future stock returns. We conduct both portfolio and Fama–MacBeth regression analyses. Specifically, we follow Katona et al. (2021) and construct the variable online sales growth (*OSG*), which is calculated as the logarithm of online sales in quarter *q* minus the logarithm of online sales in the same quarter of the previous year (*q-4*). For the portfolio analysis, we sort stocks into portfolios by *OSG* and track the returns in the subsequent period. We conduct regression analyses of the relationship between *OSG* and firms' stock returns in the subsequent period, controlling for firm characteristics and common risk factors. The summary statistics of variables used are reported in Table 1 of this appendix.

[Insert Table 1 of this appendix about here]

1. Portfolio analysis

To conduct the portfolio analysis, for each quarter *q* during the Q1 2016–Q4 2021 period, we sort our sample stocks into portfolios based on the *OSG* quintile. We additionally create a zero-cost hedge portfolio that buys stocks on the top quintile (Q5) of *OSG* and sells stocks on the bottom quintile (Q1). We then track the monthly returns of these portfolios over the next quarter (*q+1*). The portfolios are rebalanced every quarter. Finally, we regress the portfolios' monthly excess returns on the Fama–French–Carhart four-factor model (Carhart, 1997)[1]; the excess stock returns, calculated as the raw returns minus the risk-free rate, are used for portfolios Q1–Q5. The estimated intercept from the regression is the alpha, or average monthly abnormal return, of each portfolio. The results are reported in Table 2 of this appendix.

The alphas in Panel A are computed according to the portfolios' value-weighted returns. Rows (1) and (2) report the alpha and *t*-value of each portfolio based on the full sample. The alphas monotonically increase with *OSG*. In Q1, the alpha is 0.40% ($t = 2.65$), and this increases to 1.26% in Q5 ($t = 8.70$). The hedge portfolio, as shown in the last column, has a positive alpha of 0.86%, which is significant at the 1% level ($t = 6.86$). This result means that by constructing a portfolio in which stocks are bought in Q5 and sold in Q1, an average monthly future abnormal return of 0.86%, or 10.32%, can be generated per year.

[1] The results are robust when using the capital asset pricing model and the Fama–French three-factor model (Fama and French, 1993).

Next, we divide the full sample into two equal periods. Rows (3) and (4) of Table 2 show the results for the 2016–2018 period (the first part of our sample period). Rows (5) and (6) show the results for the 2019–2021 period (the second part of our sample period). Although we find similar monotonical relationships in both periods, the relationship is more salient in the early period than in the later period. The alpha of the hedge portfolio is 1.10% (t = 4.39) for the early period and 0.69% (t = 3.7) for the recent period. These results suggest that trading on *OSG* generates substantial abnormal returns over time.

We also divide the full sample into large and small stocks according to the median of market capitalization. Rows (7) and (8) of Table 3 report the results for small stocks, while Rows (9) and (10) report the result for large stocks. Surprisingly, and interestingly, we find that the results are more pronounced for large stocks than for small stocks. For example, the alpha of the hedge portfolio is 0.73% (t = 3.37) for large stocks and 0.51% (t = 2.13) for small stocks; both alphas are statistically significant. These results suggest that our baseline findings are not driven by microcap stocks that often have high transaction costs and volatile returns (Novy-Marx and Velikov, 2016).

We repeat the portfolio analysis according to equal-weighted returns. The results are reported in Panel B of Table 2. We observe a similar pattern as described above, although the level of significance is decreased. As shown in Rows (1) and (2), the alpha of the hedge portfolio based on the full sample is 0.83% (t = 3.71), similar to the return under the value weighting approach. We also obtain significant positive alphas in analyses of subsamples stratified by period (Rows (3)–(6)) and market capitalization (Rows (7)–(10)).

Overall, the results indicate that an investment strategy based on online sales growth can generate significant abnormal returns, supporting our view that online sales data contain information about firms' future stock performance.

[Insert Table 2 of this appendix about here]

2. *Fama–MacBeth regression*

Next, we apply Fama–MacBeth regression to achieve a more comprehensive analysis. For each month m during Q1 2016–Q4 2021, we run a cross-sectional regression, as follows:

$$Ret_{i,q+1,m} = \alpha + \beta \times OSG_{i,q} + \gamma \times X_{i,q} + \varepsilon_{i,q+1,m}, \tag{IA.R1}$$

where the dependent variable is the monthly excess stock returns (raw returns minus the risk-free rate) of firm i in month m of quarter $q+1$ ($Ret_{i,q+1,m}$). $OSG_{i,q}$ is the quarterly online sales growth in quarter q. X is a battery of firm i's characteristic variables in quarter q, including the total reported sales growth rate (*TSG*), firm capitalization (*MV*), the book-to-market ratio (*BTM*), momentum (*MOM*), the asset growth rate (*TAG*), operating profitability (*OP*), and *R&D*. Studies show that these variables affect firms' stock returns (e.g., Carhart, 1997; Jegadeesh and Titman, 2001; Cooper et al., 2008).

We calculate the time-series average of the cross-sectional regression coefficients. The standard deviation of the coefficients is adjusted using the Newey–West estimator (up to six lags).

The regression results are reported in Table 3. Column (1) shows the results obtained with Model IA.R1 when the growth rate of reported total sales (*TSG*) is included as a control. We find that the coefficient on *OSG* is 0.98, which is significant at the 1% level ($t = 3.61$). That is, a one-standard-deviation increase in *OSG*, on average, is associated with a 0.30% (0.31 × 0.98) increase in the monthly return in the following quarter, or a 3.65% increase in the annualized return.

Column (2) reports the estimates when we further control for *MV*, *BTM*, and *MOM*, as in Carhart (1997). The coefficient on *OSG* remains highly significant ($t = 4.12$). The magnitude of the coefficient decreases to 0.71, meaning that a one-standard-deviation increase in *OSG* is associated with a 0.22% (0.31 × 0.71) increase in the monthly return, or a 2.64% increase in the annualized return.

Column (3) reports the estimates when we further control for the one-period lag of the dependent variable ($Ret_{i,\ q+1,\ m-1}$) and additional firm characteristics, including *TAG*, *OP*, and *R&D*. In this column, the coefficient on *OSG* is similar to that in Column (2).

Overall, the Fama–MacBeth regression results are consistent with the portfolio analysis results. The evidence indicates that online sales can predict firms' future stock performance.

[Insert Table 3 of this appendix about here]

3. Return predictability and the role of online sales

The return predictability of online sales is supported by the assumptions that online sales are a measure of firm performance and that the financial market does not realize value-relevant information before it is made public. If these assumptions are correct, then arguably, the return predictability of online sales should be stronger when online sales are a more crucial part of a firm's total sales.

As online sales data are a summation of online consumers' purchases, they reflect the aggregated demand or opinions of various consumers. From the perspective of the wisdom of the crowd, an indicator has a higher information content and a lower noise level when it is aggregated from a larger and more diverse customer base (Huang et al., 2020). Therefore, we expect the return predictability of online sales to be greater in firms with more (versus fewer) online customers.

To test the above predictions, we conduct a double-sort portfolio analysis according to the share of online sales (or the number of online customers) and *OSG*. Specifically, we sort our sample stocks into tercile portfolios based on the share of online sales (*OS/TS*), which is calculated as the total online sales divided by the total sales in quarter *q* (or the number of customers of online shops in quarter *q*, *OCN*). Within each tercile portfolio, we further sort the stocks into quintile portfolios based on *OSG*; additionally, we form a zero-cost hedge portfolio that buys stocks on Q5 of *OSG* and sells stocks on Q1. We then calculate the alpha of the hedge portfolio by regressing the monthly returns of the hedge portfolios on the Fama–French–Carhart four-factor model.

The results are reported in Table 4 of this appendix. Panel A reports the results when the sample stocks are first sorted by *OS/TS*. In the top tercile of *OS/TS*, the alpha is 1.00% (*t*=3.43) in Q5 of *OSG* and -0.22% (*t*=-0.75) in Q1. The hedge portfolio that buys (sells) stocks in Q5 (Q1) has an alpha of 1.22% (*t* = 3.72). In the middle tercile of *OS/TS*, the alpha of the hedge portfolio is 0.83% (*t* = 2.52). The alpha is nonsignificant in the bottom tercile of *OS/TS*. The difference in alpha between the high and low terciles of *OS/TS*, as shown in Row (7), is 1.45%, which is significant at the 1% level. These results suggest that the return predictability of online sales is stronger when online sales constitute a greater part of a firm's total sales.

Panel B of Table 4 reports the results when the sample stocks are first sorted by *OCN*. In the top tercile of *OCN*, the alpha of the hedge portfolio is 1.25% (*t* = 4.15). The alpha is nonsignificant in the bottom tercile of *OCN*. The difference in alpha between the high and low terciles of *OCN* is 1.51%, which is significant at the 1% level. The results suggest that the return predictability of online sales is pronounced when firms' online sales are distributed among a large number of online customers.

[Insert Table 4 of this appendix about here]

We also run Fama–MacBeth regressions. Here, we divide the sample based on the median of *OS/TS* or *OCN* and re-estimate Equation (IA.R1) using the subsamples. The results are reported in Table 5 of this Appendix. Consistent with the double-sorted portfolio analysis reported above, we find that the coefficients on *OSG* are significant and positive only in the high *OS/TS* and high *OCN* subsamples.

Overall, our findings indicate that the relationship between *OSG* and future stock returns increases with both the share of online sales and the number of online customers, signifying the relevance of online sales to firms' fundamentals.

[Insert Table 5 of this appendix about here]

4. Earnings surprises

We further examine whether investors can realize the information contained in online sales data. Investors who observe such information or other indicators related to online sales are likely to incorporate the information when making decisions and thus will not be surprised when firms' revenues are publicly released. On the contrary, investors who are unable to realize the information contained in online sales data are likely to be surprised by earnings announcements, and their reaction to earnings announcement should be positively correlated to online sales.

To test investors' ability to realize information contained in online sales data, we construct two measures of earnings surprises: *SUE* and *AnnRet*[-1,3] (Froot et al., 2017; Huang, 2018). *SUE* represents unexpected earnings and is calculated as $(AE–FE)/P$, where *AE* is the earnings per share (EPS) announced for quarter q, *FE* is the analysts' forecast consensus regarding the EPS, and *P* is the stock price at the end of quarter q. *AnnRet*[-1,3] represents the buy-and-hold return over the period from 1 day before to 3 days after the earnings announcement.

We then regress *SUE* and *AnnRet*[-1,3] on the *OSG* of the quarter to which the announced earnings apply. We control for firm characteristics as in the previous analysis. We also include a one-period lag of the dependent variable (SUE_{q-1} or $AnnRet[-1, 3]_{q-1}$), which is constructed using the announcement of earnings for quarter $q-1$. We also include quarter and firm fixed effects to control for unobserved time and firm factors.

The regression results are presented in Table 6 of this appendix. Columns (1) and (2) present the results when *SUE* is the dependent variable. We find that *OSG* is significantly and positively related to *SUE*, suggesting that financial analysts do not incorporate information from online sales data in their earnings forecasts. Columns (3) and (4) present the results when *AnnRet*[-1,3] is the dependent variable. The coefficients on *OSG* are also significant and positive. Furthermore, these results have economic significance. As shown in Column (4), a one-standard-deviation increase in *OSG* is associated with a 0.2% (0.31×0.005) increase in *AnnRet*[-1,3]. These results suggest that investors may not anticipate the information contained in online sales data before earnings announcements.

Overall, the evidence supports our argument that online sales data contain value-relevant information that is not revealed to the public before an earnings announcement.

[Insert Table 6 of this appendix about here]

5. Future long-term returns

We further examine the persistence of the return predictability of online sales. If online sales provide information about firms' fundamentals, the relationship between online sales and future stock returns should be persistent. In contrast, if the relationship is driven by short-term mispricing (e.g., investors chase tech stocks and buy e-commerce stocks, which drives up the prices of stocks associated with high online sales), it should be temporary, and its direction should reverse in the long term. To disentangle these two channels, we study the long-term returns of portfolios sorted by online sales.

We sort our sample stocks into quintile portfolios based on *OSG* and create a zero-cost hedge portfolio as described in Section 1 of this appendix. We next calculate the monthly returns (either value-weighted or equal-weighted) of each portfolio. We then compute the buy-and-hold cumulative returns (raw returns minus the risk-free rate) of the quintile and hedge portfolios over the next 12 months.

Figure 1 of this appendix shows the cumulative value-weighted return of the hedge portfolio, which increases steadily over the 12-month period after portfolio formation. Specifically, the return is about 2% in the first month, and increases to about 5% in the sixth month and about 10% in the twelfth month. The evidence does not indicate a reversal in stock returns for portfolios constructed using *OSG*. Moreover, the increased cumulative returns over a longer holding period imply that online sales data contain information about firms' earnings not only in the current quarter but also in the future.

We then test the statistical significance of each portfolio's returns. Specifically, we calculate the portfolios' cumulative returns for the period from quarter $q+1$ to $q+k$ (k = 1, 2, 3, and 4). We regress each portfolio's returns on a constant to obtain the average return and significance level. The results are reported in Table 7 of this appendix.

Panel A reports the returns of portfolios based on the monthly value-weighted returns. Consistent with previous results, the average returns of the quintile portfolios monotonically increase with *OSG* (across the row) and k (over the column). When k = 1, the average return is 0.45% (t = 0.46) in Q1 and increases to 3.10% (t = 2.39) in Q5. The average return of the hedge portfolio is 2.66% (t = 3.67). A similar pattern is observed when k = 2, 3, or 4. The average returns of the hedge portfolio are highly significant and increase with k: when $k = 4$, the average return of the hedge portfolio is about 9.55%. Panel B reports the returns of portfolios based on the monthly equal-weighted returns. We find a similar pattern of results from this analysis.

Overall, the results suggest that online sales are predictive of firms' long-term stock returns.

[Insert Table 7 of this appendix about here]

Figure 1: Cumulative returns on a hedge portfolio based on online sales growth

This figure shows the cumulative return on a zero-cost hedge portfolio based on online sales growth data. Specifically, for each quarter *q* from Q1 2016 to Q4 2021, we sort our sample stocks into quintile portfolios based on the variable online sales growth (*OSG*), which is calculated as the logarithm of online sales in quarter *q* minus the logarithm of online sales in the same quarter of the previous year (*q-4*). We then create a zero-cost hedge portfolio that buys stocks in the top quintile of *OSG* and sells stocks in the bottom quintile and calculate this hedge portfolio's value-weighted returns in each month. We finally compute the hedge portfolio's buy-and-hold cumulative returns over the next 12 months.

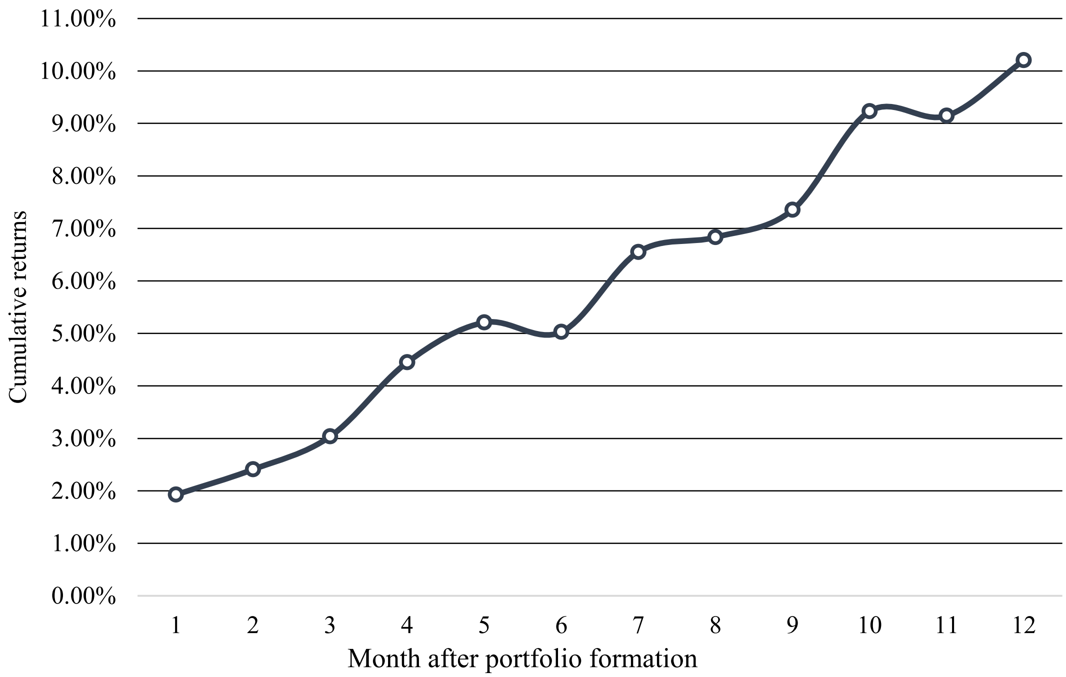

Table 1: Descriptive statistics

This table shows the descriptive statistics of the variables used in the stock return analysis of this appendix. The sample period is from Q1 2016 to Q4 2021. All variables are defined in Appendix 1 of the paper.

	N	Mean	Std. dev.	P25	Median	P75
Ret	10,068	0.400	10.127	-6.164	-0.440	6.050
OSG	10,068	0.180	0.308	-0.030	0.179	0.384
TSG	10,068	0.087	0.224	-0.024	0.096	0.204
MV	10,068	23.101	1.078	22.290	22.911	23.799
BTM	10,068	0.397	0.244	0.211	0.337	0.536
MOM	10,068	0.073	0.389	-0.188	-0.007	0.252
TAG	10,068	0.110	0.156	0.010	0.084	0.176
OP	10,068	0.080	0.044	0.043	0.075	0.108
R&D	10,068	0.072	0.069	0.005	0.062	0.120

Table 2: Portfolios sorted by online sales growth

This table reports the abnormal returns (i.e., alphas) of portfolios sorted by online sales growth. For each quarter *q* from Q1 2016 to Q4 2021, we sort our sample stocks into quintile portfolios based on online sales growth (OSG_q), which is calculated as the logarithm of online sales in quarter *q* minus the logarithm of online sales in the same quarter of the previous year (*q-4*). We also create a zero-cost hedge portfolio that buys stocks on the top quintile of *OSG* and sells stocks on the bottom quintile. The performance of the portfolios is tracked over the following quarter (*q+1*). Monthly returns are computed for each portfolio in *q+1* (value and equal weighting are used in Panels A and B, respectively). The alphas of the portfolios are obtained by regressing the portfolios' returns on the Fama–French–Carhart four-factor model. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Value weighting

	Online sales growth (*OSG*)						
	Row	Q1 Bottom	Q2	Q3	Q4	Q5 Top	Q5–Q1 Long/Short
Full Sample	(1)	0.403**	0.462**	0.692***	0.846***	1.259***	**0.856*** **
	(2)	(2.65)	(2.01)	(5.20)	(4.59)	(8.70)	**(6.86)**
2016-2018	(3)	0.389*	0.571**	0.827***	1.223***	1.483***	**1.094*** **
	(4)	(1.88)	(2.59)	(4.69)	(6.38)	(6.55)	**(4.39)**
2019-2021	(5)	0.501***	0.473**	0.675***	0.485*	1.192***	**0.692*** **
	(6)	(3.85)	(2.28)	(3.28)	(2.01)	(6.03)	**(3.70)**
Small stocks	(7)	0.148	0.347**	-0.108	0.229	0.662***	**0.513** **
	(8)	(0.96)	(2.26)	(-0.42)	(0.80)	(3.02)	**(2.13)**
Large stocks	(9)	0.115	0.582***	0.580***	0.764***	0.848***	**0.733*** **
	(10)	(0.70)	(3.38)	(3.64)	(4.45)	(5.71)	**(3.37)**

Panel B: Equal weighting

	Online sales growth (*OSG*)						
	Row	Q1 Bottom	Q2	Q3	Q4	Q5 Top	Q5–Q1 Long/Short
Full Sample	(1)	-0.165	0.240	0.076	0.498*	0.664**	**0.828*** **
	(2)	(-0.79)	(1.07)	(0.33)	(1.95)	(2.49)	**(3.71)**
2016-2018	(3)	0.321	0.708**	0.622***	1.025***	1.084***	**0.763*** **
	(4)	(0.98)	(2.18)	(3.12)	(4.00)	(3.97)	**(5.21)**
2019-2021	(5)	-0.346	-0.120	-0.216	0.117	0.585	**0.931** **
	(6)	(-1.37)	(-0.50)	(-0.59)	(0.45)	(1.28)	**(2.50)**
Small stocks	(7)	-0.559**	-0.026	-0.700**	0.089	0.204	**0.763*** **
	(8)	(-2.58)	(-0.17)	(-2.51)	(0.24)	(0.72)	**(3.16)**
Large stocks	(9)	-0.804**	-0.046	0.182	0.389	0.664***	**1.468*** **
	(10)	(-2.21)	(-0.17)	(0.85)	(1.32)	(2.66)	**(3.84)**

Table 3: Fama–MacBeth regressions

This table reports the estimates of Fama–MacBeth regressions according to the following model:

$$Ret_{i,q+1,m} = \alpha + \beta \cdot OSG_{i,q} + \gamma \cdot X_{i,q} + \varepsilon_{i,q+1,m}$$

The dependent variable $Ret_{i,q+1,m}$ is the excess return (raw return minus the risk-free rate) in each month m of quarter $q+1$. The key independent variable is OSG_q, which is calculated as the logarithm of online sales in quarter q minus the logarithm of online sales in the same quarter of the previous year (q-4). The control variables represented by $X_{i,q}$ include *TSG*, *MV*, *BTM*, *MOM*, *TAG*, *OP*, *R&D*, and the monthly lag of the dependent variable ($Ret_{i,q+1,m-1}$). All variables are defined in Appendix 1. We run a cross-sectional regression in each quarter from Q1 2016 to Q4 2021 and report the time-series averages of the cross-sectional regression coefficients. Numbers in parentheses are the *t*-statistics calculated using Newey–West standard errors with 12 lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable		$Ret_{i,q+1,m}$	
	(1)	(2)	(3)
$\boldsymbol{OSG_{i,q}}$	**0.976*****	**0.713*****	**0.688*****
	(3.61)	**(4.12)**	**(5.26)**
$TSG_{i,q}$	1.586**	0.939*	0.954**
	(2.28)	(1.94)	(2.17)
$MV_{i,q}$		0.381	0.222
		(1.28)	(0.92)
$BTM_{i,q}$		0.082	0.702
		(0.14)	(1.23)
$MOM_{i,q}$		0.623	0.746
		(1.50)	(1.60)
$TAG_{i,q}$			-1.739
			(-1.57)
$OP_{i,q}$			11.721***
			(3.42)
$R\&D_{i,q}$			-0.170
			(-0.17)
$Ret_{i,q+1,m-1}$			-0.003
			(-0.15)
Observations	10,068	10,068	10,068
Average R^2	0.027	0.121	0.184

Table 4: The role of online sales (portfolio sort analysis)

This table reports the alpha of portfolios sorted by the share of online sales (or the number of online sales customers) and online sales growth. For each quarter *q* from Q1 2016 to Q4 2021, we sort portfolios in a two-step process. First, we sort our sample stocks into tercile portfolios based on the share of online sales (*OS/TS*), which is calculated as the total online sales over the total sales in quarter *q* (or the number of customers of online shops in a quarter, *OCN*). Second, within each tercile portfolio, we further sort the stocks into quintile portfolios based on online sales growth (*OSG*), which is calculated as the logarithm of online sales in quarter *q* minus the logarithm of online sales in the same quarter of the previous year (*q-4*), and form a zero-cost long–short portfolio that buys stocks on the top quintile of *OSG* and sells stocks on the bottom quintile. The performance of the portfolios is tracked over the following quarter (*q+1*). Monthly returns (value-weighted) are computed for each portfolio in *q+1*. The alphas of portfolios are obtained by regressing the portfolios' returns on the Fama–French–Carhart four-factor model. Panel A (B) reports the alphas when the portfolios are first sorted using *OS/TS* (*OCN*). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Share of online sales

		OSG					
OS/TS	Row	Q1 Bottom	Q2	Q3	Q4	Q5 Top	Q5–Q1 Long/Short
High	(1)	-0.219	-0.113	-0.396	0.552*	1.001***	**1.220*** **
	(2)	(-0.75)	(-0.41)	(-1.44)	(1.97)	(3.43)	**(3.72)**
Medium	(3)	-0.010	0.417**	0.535***	0.912***	0.815***	**0.825** **
	(4)	(-0.04)	(2.34)	(3.24)	(6.37)	(3.51)	**(2.52)**
Low	(5)	0.576*	0.126	0.930***	0.719***	0.342*	**-0.234**
	(6)	(1.90)	(0.50)	(4.41)	(2.71)	(1.84)	**(-0.73)**
High-Low	(7)						**1.454*** **
(Long/Short)	(8)						**(2.66)**

Panel B: Number of online customers

		OSG					
OCN	Row	Q1 Bottom	Q2	Q3	Q4	Q5 Top	Q5–Q1 Long/Short
High	(1)	-0.106	0.341	-0.118	0.708***	1.141***	**1.246*** **
	(2)	(-0.56)	(1.11)	(-0.65)	(3.59)	(5.49)	**(4.15)**
Medium	(3)	-0.136	-0.094	0.591**	0.234	0.705***	**0.841*** **
	(4)	(-1.00)	(-0.33)	(2.47)	(1.02)	(3.39)	**(3.08)**
Low	(5)	0.797***	0.963***	0.729***	0.689***	0.535**	**-0.262**
	(6)	(4.78)	(6.04)	(2.74)	(4.10)	(2.61)	**(-1.21)**
High-Low	(7)						**1.508*** **
(Long/Short)	(8)						**(4.03)**

Table 5: The role of online sales (Fama–MacBeth regressions)

This table reports the estimates of Fama–MacBeth regressions in subsamples constructed according to the role of online sales, using the following model:

$$Ret_{1,q+1,m} = \alpha + \beta \cdot OSG_{i,q} + \gamma \cdot X_{i,q} + \varepsilon_{i,q+1,m}$$

The dependent variable $Ret_{i,q+1,m}$ is the excess return (raw return minus the risk-free rate) in each month m of quarter $q+1$. The key independent variable is OSG_q, calculated as the logarithm of online sales in quarter q minus the logarithm of online sales in the same quarter of the previous year (q-4). The control variables represented by $X_{i,q}$ include *TSG*, *MV*, *BTM*, *MOM*, *TAG*, *Profitability*, *R&D*, and the monthly lag of the dependent variable ($Ret_{i,q+1,m-1}$). All variables are defined in Appendix 1. We run a cross-sectional regression in each quarter from Q1 2016 to Q4 2021 and report the time-series averages of the cross-sectional regression coefficients. Panel A reports the estimates for the subsamples of firms with high and low shares of online sales (*OS/TS*; based on the sample median), which is calculated as the online sales scaled by the total sales in a quarter. Panel B reports the estimates for the subsamples of firms with high and low numbers of online customers (*OCN*; based on the sample median), which is defined as the number of customers of online shops in a quarter. Numbers in parentheses are the *t*-statistics calculated using Newey–West standard errors with 12 lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Share of online sales

Dependent variable	$Ret_{i,\,q+1,\,m}$	
Samples	High *OS/TS*	Low *OS/TS*
	(1)	(2)
OSG_q	**1.317*** **	**0.211**
	(3.52)	**(0.80)**
Control	Yes	Yes
Observations	5,174	4,893
Average R^2	0.234	0.257

Panel B: Number of online customers

Dependent variable	$Ret_{i,\,q+1,\,m}$	
Samples	High *OCN*	Low *OCN*
	(1)	(2)
OSG_q	**1.226*** **	**0.276**
	(2.90)	**(1.05)**
Control	Yes	Yes
Observations	5,283	4,785
Average R^2	0.247	0.24

Table 6: Online sales and earnings surprise

This table reports the results of the regression of earnings surprises on online sales growth (*OSG*). We measure earnings surprises using *SUE* and *AnnRet*[-1, 3]. *SUE* represents unexpected earnings and is calculated as $(AE_q - FE_q)/P_q$, where AE_q is the earnings per share (EPS) announced for quarter q, FE_a is analysts' forecast consensus on the EPS, and P_q is the stock price at the end of quarter q. *AnnRet*[-1, 3] is a four-day buy-and-hold return over the period from one day before to three days after an earnings announcement. *OSG* is calculated as the logarithm of online sales in quarter q minus the logarithm of online sales in the same quarter of the previous year (q-4). All variables are defined in Appendix 1. Quarterly data from Q1 2016 to Q4 2021 are used. The *t*-statistics of robust standard errors clustered at the firm level are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable	SUE_q		$AnnRet[-1, 3]_q$	
	(1)	(2)	(3)	(4)
$OSG_{i,q-1}$	**0.002****	**0.002*****	**0.006*****	**0.005****
	(2.13)	**(2.62)**	**(2.90)**	**(2.30)**
$SUE_{i,q-1}$	0.483***	0.214***		
	(15.27)	(7.48)		
$AnnRet[-1, 3]_{i,q-1}$			0.099***	0.047**
			(5.35)	(2.35)
$MV_{i,q-1}$	0.001**	-0.003**	0.005***	-0.006
	(2.45)	(-2.48)	(4.07)	(-1.10)
$BTM_{i,q-1}$	0.002	-0.016***	0.006	-0.002
	(0.78)	(-3.40)	(0.96)	(-0.12)
$MOM_{i,q-1}$	0.002***	0.002**	-0.005	-0.007
	(2.91)	(2.10)	(-1.20)	(-1.37)
$TAG_{i,q-1}$	-0.003**	-0.000	-0.007	-0.005
	(-2.00)	(-0.06)	(-1.03)	(-0.53)
$OP_{i,q-1}$	0.022**	0.079***	0.084***	-0.018
	(2.46)	(5.29)	(2.82)	(-0.35)
$R\&D_{i,q-1}$	-0.018***	0.006	-0.007	0.007
	(-3.05)	(0.76)	(-0.32)	(0.19)
$Ret_{i,q-1}$	0.004**	0.003	0.005	0.004
	(2.56)	(1.65)	(0.69)	(0.43)
Fixed effects	Quarter	Quarter & Firm	Quarter	Quarter & Firm
Observations	3,377	3,375	4,166	4,166
Adj. R^2	0.504	0.576	0.075	0.087

Table 7: Future long-term returns

This table reports the future returns of portfolios sorted by online sales growth (*OSG*). For each quarter q from Q1 2016 to Q4 2021, we sort our sample stocks into quintile portfolios based on OSG_q, which is the logarithm of online sales in quarter q minus the logarithm of online sales in the same quarter of the previous year (q-4). We also create a zero-cost hedge portfolio that buys stocks on the top quintile of *OSG* and sells stocks on the bottom quintile. We calculate each portfolio's monthly returns (value and equal weighting are used in Panels A and B, respectively). We then calculate the buy-and-hold cumulative returns on each portfolio from $q+1$ to $q+k$ (k = 1, 2, 3, and 4). Finally, we regress each portfolio's returns on a constant to obtain the average return and significance level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Value weighting

		OSG_q					
Cumulative returns from $q+1$ to $q+k$	Row	Q1 Bottom	Q2	Q3	Q4	Q5 Top	Q5–Q1 Long/Short
$k = 1$	(1)	0.447	1.666	1.455	1.878*	3.103**	**2.656*** **
	(2)	(0.46)	(1.66)	(1.35)	(1.96)	(2.39)	**(3.67)**
$k = 2$	(3)	1.317	2.845	3.317	3.622	6.213**	**4.896*** **
	(4)	(0.62)	(1.39)	(1.64)	(1.66)	(2.12)	**(3.14)**
$k = 3$	(5)	3.149	3.309	4.215	4.607	8.803*	**5.655*** **
	(6)	(0.93)	(1.15)	(1.33)	(1.32)	(1.92)	**(2.72)**
$k = 4$	(7)	3.105	5.659	4.175	8.157	12.656**	**9.552*** **
	(8)	(0.66)	(1.48)	(1.09)	(1.59)	(2.11)	**(3.94)**

Panel B: Equal weighting

		OSG_q					
Cumulative returns from $q+1$ to $q+k$	Row	Q1 Bottom	Q2	Q3	Q4	Q5 Top	Q5–Q1 Long/Short
$k = 1$	(1)	-1.312	-0.211	-0.504	0.535	1.107	**2.419*** **
	(2)	(-1.15)	(-0.18)	(-0.43)	(0.49)	(0.82)	**(3.37)**
$k = 2$	(3)	-0.848	0.812	0.212	1.965	3.649	**4.497*** **
	(4)	(-0.37)	(0.33)	(0.09)	(0.90)	(1.28)	**(3.01)**
$k = 3$	(5)	0.488	1.397	0.356	2.898	5.188	**4.700*** **
	(6)	(0.14)	(0.46)	(0.10)	(0.86)	(1.35)	**(2.77)**
$k = 4$	(7)	1.286	1.372	0.118	5.171	7.675	**6.388*** **
	(8)	(0.26)	(0.37)	(0.03)	(1.09)	(1.48)	**(3.05)**

Reference

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1):57-82.

Cooper, M. J., H. Gulen, and M. J. Schill. 2008. Asset growth and the cross-section of stock returns. *The Journal of Finance* 63 (4):1609-1651.

Fama, E. F., and K. R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1):3-56.

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

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

Huang, K., M. Li, and S. Markov. 2020. What do employees know? Evidence from a social media platform. *The Accounting Review* 95 (2):199-226.

Jegadeesh, N., and S. Titman. 2001. Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance* 56 (2):699-720.

Katona, Z., M. Painter, P. N. Patatoukas, and J. Zeng. 2021. On the capital market consequences of alternative data: Evidence from outer space. *Review of Financial Studies, Forthcoming*.

Novy-Marx, R., and M. Velikov. 2016. A taxonomy of anomalies and their trading costs. *The Review of Financial Studies* 29 (1):104-147.