# Predicting Performance Using Consumer Big Data

KENNETH FROOT

NAMHO KANG

GIDEON OZIK

RONNIE SADKA*

June 2019

* Froot: Harvard Business School; email: kfroot@hbs.edu. Kang: Bentley University, Finance Department; email: nkang@bentley.edu. Ozik: EDHEC; email: gideon.ozik@edhec.edu. Sadka: Carroll School of Management, Boston College, Department of Finance; email: sadka@bc.edu. Froot, Ozik, and Sadka are affiliated with MKT MediaStats, LLC.

# Predicting Performance Using Consumer Big Data

June 2019


## Abstract

Big data on consumer behavior are useful in predicting firms' fundamentals and stock returns. We generate three distinct proxies for corporate sales: In-store foot traffic (IN-STORE), web traffic to companies' websites (WEB), and consumers' interest level in corporate brands and products (BRAND). Using a sample of 330 firms over the 2009–2017, we show that each sales proxy captures different information content, with the information being more pertinent for consumer firms. Analyses of returns around earnings announcements suggest that IN-STORE and BRAND information are not readily available to investors, while a sizeable portion of WEB information is incorporated into prior prices.

## 1. Introduction

The efficient market hypothesis in its semi-strong form posits that asset prices reflect all publicly available information (Fama, 1970). The recent development in data storage, cloud computing, and machine learning have greatly reduced costs of gathering data and led to the proliferation of new data and technologies to which investors may gain access. The mere fact that granular data is public does not, of course, guarantee its full/immediate incorporation into prices. It may take time to figure out how to map public information into complicated companies (Cohen and Lou, 2012). It may also take time and costly resources to develop the technology and analytics needed to collect and organize raw data, to determine its incremental impact on market prices, and to propagate the resulting data/information/knowledge across capital sources. Consequently, our granular data sources may not in the short run be incorporated into market prices, even if they are publicly available.

However, there are many anecdotes that resourceful investors have already developed investment strategies utilizing new data sources to obtain an informational edge. These investment strategies have real consequences on market efficiency and corporate policies. For example, Zhu (2018) shows that the availability of alternative data increases the efficiency of stock prices and reduces agency conflicts.

In this study, we ask whether certain alternative data sources can help to predict firms' fundamentals and stock returns. We note that the ability to predict fundamentals does not necessarily ensure our alternative data sources are not incorporated into expectations. Because other information may forecast fundamentals (which aren't traded), one would need to incorporate those sources alongside of ours to determine incremental forecasting over and above existing expectations. However, if the information associated with our sources can forecast short-term stock returns, that is sufficient to say that the information is currently unincorporated in prices.

We explore various sources of big data which measure aspects of consumer behavior. We focus on three distinct types of consumer activities that are related to firms' revenue: visits to retail stores; visits to firms' web sites; and measures of consumer interest in products or brand names. Based on these consumer activities, we develop variables that measure the firms' sales. Specifically, we develop the following corporate sales proxies: an in-store foot traffic proxy (IN-STORE), which measures consumers' likely intention to visit retail stores (Froot, Kang, Ozik, and Sadka, 2017); a web traffic proxy (WEB), which calculates the number of consumer visits to firms' websites; and a consumer brands awareness proxy (BRAND), which estimates the level of consumer interests in products and brand names. We study whether these proxies are useful in predicting firms' fundamentals and returns around earnings announcement dates.

All the above proxies are estimated as growth rates. We follow Froot, Kang, Ozik, and Sadka (2017) to derive two measures of each proxy, one measuring sales activity within the quarter covered by an earnings announcement — denoted by WQS — and the other measuring consumer activity between the end of the earnings quarter and the announcement date — denoted by PQS. For a given firm in a given quarter, WQS and PQS are proxies for growth rates of sales where we take the log difference between the number of events aggregated under each proxy measure over the given quarter and the quarterly average of the number of events over quarters *t*-1 to *t*-4.

First, we investigate whether these proxies contain value-relevant information. Since the sales proxies are constructed from consumer activities that are presumably related to corporate sales, we would expect that our proxies are strongly correlated with fundamentals. Indeed, we find that all three measures significantly predict standardized sales and earnings growth, as well as analyst forecast errors. Again, there may be other variables that predict these non-traded changes, so we cannot know whether the information in our proxies is already included in investor expectations for these changes.

We next examine whether the information in our sales proxies are incorporated into prices by examining earnings announcement returns. If investors are unaware of such sources of data and/or face costs and time barriers in extracting information from these sources, we expect that the proxies will significantly predict announcement returns.

We find, in fact, that all our proxies have predictive power for earnings announcement returns, though to a varying degree. For example, the average excess earnings-announcement return for stocks in the highest quintile of IN-STORE is 1.89%, while that for stocks in the lowest quintile is -1.34%, generating the return differential of 3.23% (with *t*-value of 3.70).

This predictive ability is much stronger for Consumer-Goods-sector firms, from which all IN-STORE firms are drawn. For BRAND, approximately 61% of the 250 sample firms are not included in Consumer Goods sector. The WEB proxy covers a broader group of 312 firms, of which 53% are included in Consumer Goods sector. The average return differential between the highest and the lowest quintiles for all BRAND (WEB) firms is 0.62% (0.76%) (*t*-values of 2.87 and 2.83, respectively). This spread increases to 1.29% (1.19%), however, for firms in the Consumer Goods sector.

However, the magnitudes of return predictability for IN-STORE and BRAND are larger than that of WEB, indicating the information in WEB is more accessible to investors. The regressions of the announcement returns on WEB implies that a one-standard-deviation increase in WEB is associated with a 0.17–0.26% increase in the announcement return. On the contrary, one-standard-deviation increases in IN-STORE and BRAND are associated about 1.1–1.3% increases in announcement returns. Clearly, a firm's inclusion is in the Consumer Goods sector is not the primary reason that WEB results are weaker. Perhaps is it that there is greater access to this information by investors, and so is the information better and more rapidly incorporated into prices across the board?

We test this hypothesis in our examination of pre- and post-earnings announcement returns. If WEB is incorporated in prices relatively quickly with no delay, the WEB PQS information should be reflected more or less fully in pre-announcement prices. And, because the information pertains only to post-accounting period information, it should be positively associated with pre-earnings-announcement returns. Likewise, if the PQS information in IN-STORE and BRAND are diffused more slowly, then their PQS information should leak less into pre-announcement and more into post-announcement returns.

We find that, indeed, only WEB PQS (and not IN-STORE or BRAND PQS) is positively related to pre-announcement returns. We also find a positive relation of PQS IN-STORE and BRAND with post-announcement returns, consistent with the delayed leakage hypothesis above. WEB information appears to be diffused into prices at a faster rate than other proxies, and investors may have incorporated a sizable portion of this information into prices before earnings-announcement dates.

The rest of this paper is organized as follow. In the next section, we review related literature. Section 3 describes the sales proxies and other variables. Section 4 examines the information contents of each sales proxies. In Section 5, we study returns around earnings announcement dates. In Section 6, we provide our concluding remarks.

## 2. Related Literature

Big data sources that describe aspects of consumer/investor behavior have become one of the fastest growing themes in many disciplines including empirical asset pricing. In what today already sounds quaint, a few years ago UBS analyst reportedly purchased satellite images of Walmart parking lots prior to the earnings announcement to gain information. (Ozik and Sadka, 2013). Beyond such anecdotes, there is now evidence to support the hypothesis that big data have begun to influence the efficiency of stock prices, in the sense of being better indicators of future earnings, of investment opportunities and of future investment spending (Zhu, 2018).

In addition to the publicly realized benefit of more informative stock prices and better corporate resource allocation, there is also evidence concerning privately realized benefits. That is, the evidence suggests that the information in these data sources is partially — but not yet fully — incorporated into market prices. Consumer-oriented firms, whose stock prices Zhu (2018) has shown to have become more informative since proliferation of big data over the last decade, are the same firms whose stock returns are recently predictable based on similar big data sources. For example, Froot, Kang, Ozik, and Sadka (2017) utilize big data to proxy for consumer-oriented firms' unobserved fundamentals (i.e., sales and consumer-store visits) and show these proxies help predict earnings announcement returns.

Selected uses of social media information have become a popular topic in financial studies. Chen, De, Hu, and Hwang (2015) study Seeking Alpha, a popular financial blog, and find that positive sentiment measures predict earnings announcements and future stock returns. Bartov, Faurel, and Mohanram (2015) use the Tweeter feed to extract aggregate sentiment before earnings announcements. Da, Engleberg, and Gao (2011) show that Google search volume is a momentum forecaster for near term stock returns and to later subsequent reversals. Froot, Lou, Ozik, Sadka, and Shen (2018) shows that sentiment measured from professional media sources reinforce and positively predict recent returns.

## 3. Data and Variables

### 3.1. Data sources

Exploring various sources of big data on consumer behavior, we estimate consumer activities with respect to visits, in person and on-line, to retail stores and company websites as well as consumer interest in products or brand names. We use these different consumer activities to develop proxies to describe each: IN-STORE, WEB, and BRAND. To develop these proxies, we utilize underlying data from multiple proprietary sources collected by MKT MediaStats, LLC. We expect, naturally, that these variables serve as proxies for sales.

IN-STORE measures consumers' activities at retail stores. Specifically, we attempt to estimate consumers' intention to visit or shop at a retail store from activities such as searches for driving directions to a store location, queries concerning store hours, or downloads of discount-purchase coupons. These consumer activities at retail stores are collected from various sources, including millions of consumer devices (for more detail, see Froot, Kang, Ozik, and Sadka, 2017). IN-STORE measures only large "big-box" retailers whose revenue comes mainly from their physical retail stores. It does not include e-commerce businesses or other types of retailers, such as telecommunication companies or restaurants. Consequently, IN-STORE coves only 66 retail firms.

WEB is an estimate of consumer visits to consumer-firm websites. In particular, it is estimated by observing Internet activities of a panel of a few tens of millions of individual Internet users. Visits are defined as the number of visitors to the websites of sample firms. Some companies have multiple websites for distinct brand names. For example, TripAdvisor operates TripAdvisor-branded websites, including tripadvisor.com in the United States. It also manages and operates 23 other websites of media brands that provide travel-planning resources across the travel sectors, such as airfarewatchdog.com, citymaps.com, cruisecritic.com, flipkey.com, gateguru.com, housetrip.com, and viator.com. In WEB we are trying to estimate overall firm-related consumer activities from online visits. Consequently, the WEB estimate of total events for TripAdvisor, for example, is an aggregate of activities across these brand-name sites. Our sample for the WEB data consists 312 firms in various industry sectors, including big-box retailers, online retailers, restaurants, hotels and entertainment.

BRAND estimates the level of consumer interest in product brand names. It is similar to WEB, however, rather than estimate using firm-site visits, it builds from consumer search and social-media activities for sample-company products and brand names. The sample, which consists of 250 companies in various sectors, is then aggregated across within-company brands.

Using the underlying information described above, we derive proxies to cover two distinct periods, one to measure within-quarter sales activity — denoted by WQS — and the other to measure post-quarter activity up until the earnings announcement date — denoted by PQS. Figure 1 helps to explain how we construct these proxies. The figure plots the time line around the announcement date for Quarter *q*. The post-quarter period is defined as the time period between the beginning of the fiscal-quarter *q*+1 and the announcement date of quarter-*q* earnings.

For both WQS and PQS, firm *i*'s consumer activities are aggregated over a given quarter then scaled by the average of previous four quarters, using log differences. While WQS covers a more or less constant temporal interval – one accounting quarter – PQS must cover the varying interval between the beginning of the new accounting quarter and the announcement date of prior-quarter earnings. Consequently, to measure growth rates in consistent annualized units, PQS is further scaled by the fraction of the full quarter elapsed prior to announcement.

### 3.2. Sample and Summary Statistics

The full sample includes 331 US public companies in various industries during the period 2009–2017. This is comprised of 178 companies belonging to the consumer sectors (Consumer Discretionary and Consumer Staples, 136 and 42 companies for each sector, respectively);[1] 76 companies belonging to Financials and Information Technology sectors (47 and 29 companies, respectively); and 77 companies spanning across Health Care, Real Estates, Communication, Industrial and Materials sectors. We use CRSP to obtain stock market variables, including stock returns and prices and Compustat to obtain the information on financial statements. Analyst forecasts are from IBES.

Table 1 shows the summary statistics of the three WQS sales proxies and other main variables. Panel A provides descriptive statistics, while Panel B provides correlations. SIZE is the natural logarithm of market capitalization at the end of fiscal quarter. LogBM is the natural logarithm of the book-to-market ratio, measured as of the most recent fiscal year ending at least three months prior to fiscal quarter-$q$ end. MOM is a buy-and-hold return during the 6-month period of $t$-$8$ to $t$-$2$, where $t$ is the month of fiscal quarter-$q$ end.

Panel A shows that the BRAND has higher mean and standard deviation than other proxies. For example, the mean of BRAND is 2.2% and the standard deviation is 29%. The mean of WEB is -2.4% and its standard deviation is 15.5%, while the IN-STORE has a mean of 1.2% and a standard deviation of 20.8%.

Panel B shows that the three proxies are all positively correlated and statistically significant. For example, IN-STORE has a correlation with BRAND of 0.47. The correlations between IN-STORE and WEB and between BRAND and WEB are 0.36 and 0.21, respectively. Panel B also shows

[1] We follow Global Industry Classification Standard (GICS) to define sectors. There are 11 sectors based on GICS specification. We define Sectors 25 (Consumer Discretionary) and 30 (Consumer Staples) as the consumer sectors.

that all three measures are highly correlated with MOM, indicating that firms with a high level of consumer activities have recently experienced high stock returns.

## 4. Firms' Fundamentals and Sales Proxies

Table 2 examines the predictive power of the sales proxies for revenue growth. Specifically, the table shows results from the regressions of standardized changes in revenue (SCR) on the sales proxies. The standardized changes in revenue (SCR) for stock *i* in quarter *t* is calculated as $[(S_{i,t} - S_{i,t\text{-}4}) - r_{i,t}]/\sigma_{i,t}$, where $S_{i,t}$ is the quarterly revenue as of fiscal quarter *t* for firm *i*, $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and average, respectively, of $(S_{i,t} - S_{i,t\text{-}4})$ over the preceding eight quarters. Panel A shows the results of the entire sample, while Panel B reports the analyses using the subsample of the firms in the consumer sectors. The IN-STORE results are identical for both Panels A and B, because it covers only consumer-sector firms.

Panel A shows that while all three proxies are significantly associated with the SCR, the coefficients of WEB tend to be larger and more significant than those of other proxies. For example, the regression of the SCR on WEB (Model (7)) yields the coefficient of 0.582, *t*-value of 9.30, and $R^2$ of 1.1%, all of which are bigger than the values of the corresponding models for other proxies.

Panel B reports the results using the subsample of firms in the consumer sectors. The results show that the information contained in the proxies are more relevant for consumer firms. For both BRAND and WEB, the coefficients are larger than those reported in Panel A. For example, the coefficient on BRAND in Model (4) is 0.656, compared to 0.433 of the corresponding model in Panel A. Also, the values of $R^2$ in Panel B are higher than those in Panel A, indicating the data sources that are used to construct sales proxies contain information more pertinent for consumer firms.

In Table 3, we further examine the informativeness of the sales proxies with respect to firms' earnings. In particular, we regress standardized changes in earnings (SCE) and analyst forecast errors (FE) on our within-quarter sales proxy, WQS. Panels A and B report the regression results of SCE on WQS, while Panels C and D show regressions of analyst FE on WQS.

Similar to SCR, SCE for stock *i* in quarter *t* is calculated as $[(E_{i,t} - E_{i,t-4}) - r_{i,t}]/\sigma_{i,t}$, where $E_{i,t}$ is quarterly earnings as of fiscal quarter *t* for firm *i*, $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and average, respectively, of $(E_{i,t} - E_{i,t-4})$ over the preceding eight quarters. Analyst FE is estimated as $(A_{i,q} - F_{i,q})/P_{i,q}$, where $A_{i,q}$ is quarterly earnings per share announced for quarter *q* of stock *i*, $F_{i,q}$ is mean analysts' forecasted EPS, and $P_{i,q}$ is quarter-end price. Analyst FE is standardized to have mean zero and unit standard deviation.

Panels A and B show that all three of sales proxies significantly predict SCE. Although BRAND is not significantly related to SCE when the entire sample is used (Panel A), subsample results in Panel B indicate that BRAND also contains significant information with respect to earnings growth for consumer companies. As with the SCR results, WEB is the strongest predictor for SCE in terms of the magnitude of coefficients as well as statistical significance.

Panels C and D show that the sales proxies also significantly predict analyst forecast errors. Similar to the SCE results, predictability is higher for consumer-sector firms. For example, a one-standard-deviation change in BRAND is associated with a change in analyst FE of about 2.8% of its standard deviation. For consumer-sector firms, the magnitude increases to about 7.8% of a standard deviation.

# 5. Returns Around Earnings Announcement Dates

In this section, we examine the relation between the sales proxies and returns around earnings announcement dates. First, we study whether the availability of alternative data sources helps to predict earnings-surprise returns. By examining earnings-announcement returns, we test whether the predictability from our information sources for previously-unannounced quarterly fundamentals is not observed/acted-upon by investors.

In addition to the predictive power for earnings-announcement returns, we also study how efficiently stock prices reflect the information in those proxies. Thus, we investigate pre- and post-earnings announcement returns to examine whether the sales proxies are incorporated in prices in a timely manner.

## 5.1. Event Studies

To examine the predictability of the sales proxies for earnings announcement returns, we use standard event-study methods. Table 4 shows the results. We use a five-day event window around the announcement, beginning one day prior to the announcement date and ending three days after[2]. Panel A reports the results of the entire sample, while Panel B uses the consumer-sector subsample.

First, we sort firms into quintiles based on the WQS and calculate the average announcement excess return for each quintile. We form quintiles as follows. At each month-end *t*, we rank all sample firms based on their sales proxies, calculated for their most recent fiscal-quarter, to obtain quintile cutoff values. We then use these values to assign quintile ranks for the firms whose fiscal quarter ends at month *t*. We follow this process to make sure that we use the full sample of firms when ranking

---

[2] A large portion of US companies makes after-hour earnings announcements (Berkman and Troung, 2009). For those companies, we observe earnings-related price changes on day one. Therefore, we use a slightly longer event window to capture the market's complete reaction to the announcement. Choosing different event window does not alter our inference.

them. Different methods of assigning quintile scores – for example, in each month *t*, ranking firms using only those that have their fiscal quarters end at *t* – do not change the inferences.

Panel A shows that while all three proxies significantly predict announcement returns, IN-STORE has the strongest predictive power. The average announcement excess return during the five-day period around earnings-announcement dates for stocks in the highest quintile of IN-STORE is 1.89%, while that for stocks in the lowest quintile is -1.34%. The return differential between stocks in the highest and lowest quintiles is 3.23% (with *t*-value of 3.70), which is statistically and economically significant.

The BRAND also strongly predicts announcement returns. The return differential between the highest and the lowest quintiles is 0.62% with *t*-value of 2.87, showing that the economic magnitude of predictive power of BRAND is smaller than that of IN-STORE. However, Panel B shows that when we restrict the sample to firms in the consumer sector, the spread between the highest and lowest quintile increases to 1.29%, indicating the predictive ability is more powerful for the firms in the consumer sectors.

Panel A shows that the return differential between the highest and lowest-quintiles sorted by WEB is 0.76%, which is also statistically significant with the *t*-value of 2.83. However, the average return of the second quintile is the same as that of the highest quintile (0.69%). This raises a concern about the significance of WEB's predictive ability for the announcement returns. Panel B shows that for the companies in the consumer sectors, the differential between the highest and lowest quintile increases to 1.19%, and the average return of the highest quintile is larger than all other quintiles. However, except for extreme quintiles, there is not much variation in the average returns by quintiles, indicating that this information may already be, to some extent, disseminated to market participants before announcement dates.

Figure 2 complements the results of Table 4. The figure plots the average buy-and-hold returns (in excess of the market) of the firms in the consumer-sector subsample during the event window from 10 days prior to the earnings announcement date (day 0) to 10 days afterward. On average, a sharp decline (resp., increase) in returns is observed for firms in Quintile 1 (resp, Quintile 5) of IN-STORE and BRAND around earnings announcement dates. This indicates that information contained in those proxies has not been disseminated before announcement dates, and investors are surprised by the release of the information.

However, announcement returns sorted by WEB display quite different patterns. While a relatively gradual increase in returns during announcement dates is observed for firms in Quintile 5, one can spot a slow increase in returns up to day one followed by a gradual decrease for firms in Quintile 1. Overall, the return pattern of WEB shown in Figure 2 is consistent with the weak predictability reported in Table 4. Therefore, despite the strong predictability of WEB for revenue and earnings, the results in Figure 2 suggest that a sizable portion of information contained in WEB is likely to be diffused before earnings announcement dates.

In sum, while all three proxies demonstrate power for predicting earnings-announcement returns, IN-STORE and BRAND display the stronger ability, suggesting the information included in these proxies is not readily available to a large cohort of investors before earnings announcements. In addition, the magnitude of predictability of each proxy is much larger for firms in the consumer sectors, where big data on consumer activities is most likely to be insightful about earnings. However, despite the strong predictability of WEB for firms' fundamentals, the proxy's predictability for announcement return is weak, indicating that the information in WEB is likely to be diffused prior to announcement dates.

### 5.2. Earnings-Announcements Returns

In Table 4, we study the predictions of our sales proxies in a univariate setting. In this section, we estimate multivariate regressions, controlling for various firm characteristics that may affect announcement returns.

Table 5 reports the regression results of the earnings-announcement returns on the sales proxies, and other control variables. Panel A shows the results of the entire sample, while Panel B reports the results of the consumer-sector subsample. The dependent variable is the returns around announcement dates for fiscal quarter-*q* earnings for the period of one day before and three day after the announcement date. The main explanatory variables are the sales proxies for within-quarter sales activity (WQS) and those for the post-quarter activity up until the announcement date (PQS).

The control variables are SIZE, Analyst FE, logBM, MOM, Loss, and PastRet. Loss is an indicator variable that equals one if the announced earnings is negative, and zero otherwise. So and Wang (2014) document that announcement returns typically incorporate at least some reversal of past returns. Thus, we include PastRet, which is the cumulative return in excess over the market from thirteen to three days prior to the announcement date. Definitions of other variables are provided in the previous section.

The multivariate results of Table 5 are consistent with the univariate results in Table 4. First, the coefficients on WQS are significantly positive for all three sales proxies, and they are larger for firms in the consumer sectors. However, similar to Table 4, the magnitude of WEB is smaller than other sales proxies. For example, Models (3) and (5) of Panel B show that the coefficients on IN-STORE and BRAND are comparable, being 0.053 and 0.042, respectively, while the coefficient on WEB is much smaller at 0.011, and it is statistically significant only at 10% with a *t*-value of 1.82 (Model (7)).

The coefficient in Model (7) implies that a one-standard-deviation increase in WEB is associated with a 0.17% increase in the earnings announcement return. The economic magnitude of IN-STORE and BRAND is much larger. One-standard-deviation increases in IN-STORE and BRAND are associated with 1.1% and 1.2% increases in the announcement returns, respectively. This result indicates that investors, at least to some extent, may have access to the information underlying WEB before announcement, and, therefore, are not so surprised on announcement day.

Next, we examine PQS. Consistent with Froot, Kang, Ozik, and Sadka (2017), we find a significant negative coefficient on IN-STORE.[3] Assuming that managers have the information behind PQS privately at announcement time, it appears that they use discretionary channels (such as tone and guidance) to understate the private information contained in PQS, thereby reducing any announcement return surprise. That is, managers do not fully disclose their private information, actually biasing their disclosures downward at announcements. The negative coefficient on PQS is also observed for BRAND, suggesting that the BRAND information for the post-quarter period is not fully disclosed at earnings announcements.

We do not observe the similar pattern in WEB with respect to PQS. The coefficient on PQS is slightly positive, but insignificant. This may suggest that investors have access to the PQS information in WEB before earnings announcement dates. Thus, stock prices are not affected by managers' disclosure behavior. We conclude that, regardless of whether managers themselves faithfully reveal the information in WEB PQS, that information is already incorporated into stock prices.

[3] The sample period of Froot, Kang, Ozik, and Sadka (2017) is 2009 to July 2014. This paper extends the sample period to 2017 and confirms their main findings.

### 5.3. Pre- and Post-Earnings-Announcement Returns

The observed pattern in Table 5 regarding WQS and PQS suggests that WEB information is likely to be reflected in stock prices in a timely manner. This leads to an additional test. If WEB is incorporated in prices without a delay, the WEB PQS information should be positively associated with pre-earnings-announcement returns. Likewise, if the information in IN-STORE PQS and BRAND PQS diffuse into prices with a delay, then these proxies less associated with pre-announcement and more associated with post-announcement returns.

Table 6 provides the results of regressions of pre-earnings-announcement returns on WQS and PQS of the sales proxies. The pre-earnings-announcement return is calculated as the excess return between 25 and two days prior to the announcement date. Because it takes 4–6 weeks after the accounting quarter ends for a typical firm to announce quarterly earnings, this period approximately coincides with the post-quarter period for which PQS is calculated.

First, except for some specifications, the coefficients on WQS are positive for all three proxies. This suggests that at least some portion of WQS information has become known to investors prior to announcement. For example, Panel B shows that the coefficients on WEB are between 0.013 and 0.015, similar to or even larger than the coefficients on WEB for announcement returns reported in Table 5. This suggests that, indeed, a large portion of this information is extracted by the market prior to announcement.

Next, we find some evidence that investors have timely access to the information in WEB. The WEB PQS coefficient is positive and marginally significant, suggesting investors obtain and act on the information quickly. The positive relation between PQS and pre-announcement returns is not observed for other proxies.

Table 7 provides the results of regressions of post-earnings-announcement returns. The post-earnings-announcement return is measured for the period beginning on 4 days after the quarter-*q* earnings announcement date and ending on 60 days after the announcement date. Overall, we observe negative coefficients on WQS for all three proxies. This suggests an overreaction to WQS information during the announcement periods.

More importantly, the coefficients of PQS of all proxies are positive and significant (for all but a few specifications). This demonstrates a delayed response of stock prices to the PQS information. Recall that we observe a negative relation between PQS and announcement returns for IN-STORE and BRAND. Therefore, a positive coefficient on PQS for post-announcement returns is expected due to the reversal that is caused by gradual release of private PQS information that managers withheld at announcement (see Froot, Kang, Ozik, and Sadka, 2017). Our results confirm this.

Notice that the WEB PQS coefficient for post-announcement returns is also strongly positive. In addition to a timely response, there is also a delayed response. The market process of absorbing WEB PQS information begins sooner, but is nevertheless far from complete at the time of the announcement.

In sum, our analyses show that all three proxies have predictive power for earnings-announcement returns. The predictive power of WEB is relatively weaker, because the information in WEB is dispersed faster than other proxies. The information in IN-STORE and BRAND are incorporated in prices at a slower pace than WEB, probably due to the difficulty in obtaining and analyzing the involved in constructing these variables.

## 6. Conclusion

To study the power of big data in predicting firms' fundamentals and earnings-surprise returns, we examine data sources that contain information on consumer activities, such as visits to retail stores, visits to firms' web sites, and consumers' interest in products and brand names. For the sample of about 330 firms and the time period of 2009–2017, we utilize these data sources to develop three sales proxies; IN-STORE, WEB, and BRAND.

Our analyses show that all three proxies predict the growth of revenue and earnings. This suggests that the proxies do indeed contain value-relevant contemporaneous information on firms' fundamentals. However, each sales proxy captures different information contents, with the information being more impactful for consumer firms.

We also find that the sales proxies have strong predictive power for earnings announcement returns, where, again, predictions are much stronger for companies in consumer-goods sectors. However, the magnitudes of IN-STORE and BRAND are larger than WEB, indicating the information in these variables is not immediately available to investors. Analysis of pre- and post-earnings announcement returns shows that a sizeable portion of WEB information is likely to be diffused to investors before earnings announcements dates but that all sources are positively predictive of post-announcement returns.

**References**


Bartov, Eli, Lucile Faurel, and Partha Mohanram, 2015, Can Twitter Help Predict Firm-Level Earnings and Stock Returns?, *The Accounting Review* 93, 25-57.

Berkman, Henk and Cameron Troung, 2009, Event Day 0? After-Hours Earnings Announcements, *Journal of Accounting Research* 47, 71-103

Chen, Hailiang, Prabuddha De, Yu Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media, *Review of Financial Studies* 27, 1367-1403.

Cohen, Lauren and Dong Lou, 2012, Complicated Firms, *Journal of Financial Economics* 104, 383-400.

Da, Zhi, Joseph Engleberg, and Pengjie Gao, 2011, In Search of Attention, *Journal of Finance* 66, 1461-1499.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383–417.

Froot, Kenneth, Namho Kang, Gideon Ozik, and Ronnie Sadka, 2017, What do measures of real-time corporate sales say about earnings surprises and post-announcement returns? *Journal of Financial Economics* 125, 143–162.

Froot, Kenneth, Xiaoxia Lou, Gideon Ozik, Ronnie Sadka, and Siyi Shen, 2018, Media Reinforcement in International Financial Markets, Working paper.

Ozik, Gideon and Ronnie Sadka, 2013, Big data and information edge, *Hedge Funds Review* December 2013/January 2014, 32–34.

So Eric C., and Sean Wang, 2014, News-driven Return Reversals: Liquidity Provision Ahead of Earnings Announcements, *Journal of Financial Economics* 114, 20-35

Zhu, Christina, 2018, Big Data as a Governance Mechanism, *Working paper.*

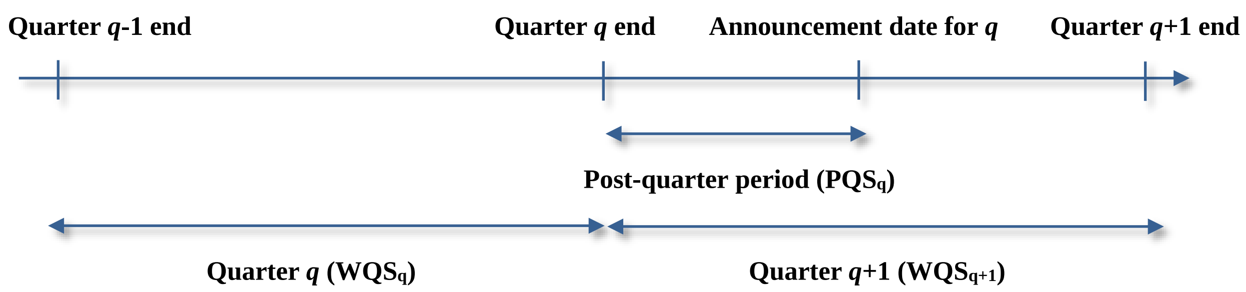


**Figure 1. Time periods around earnings announcements and corporate sales proxy.** The figure plots the time line around quarterly earnings announcement dates and describes the time periods for which corporate sales proxies are measured. WQS is the within-quarter measure, while PQS is the measure for the post-quarter period. The post-quarter period is defined as the time period beginning the fiscal-quarter *q+1* and ending prior to the announcement date for quarter-*q* earnings.

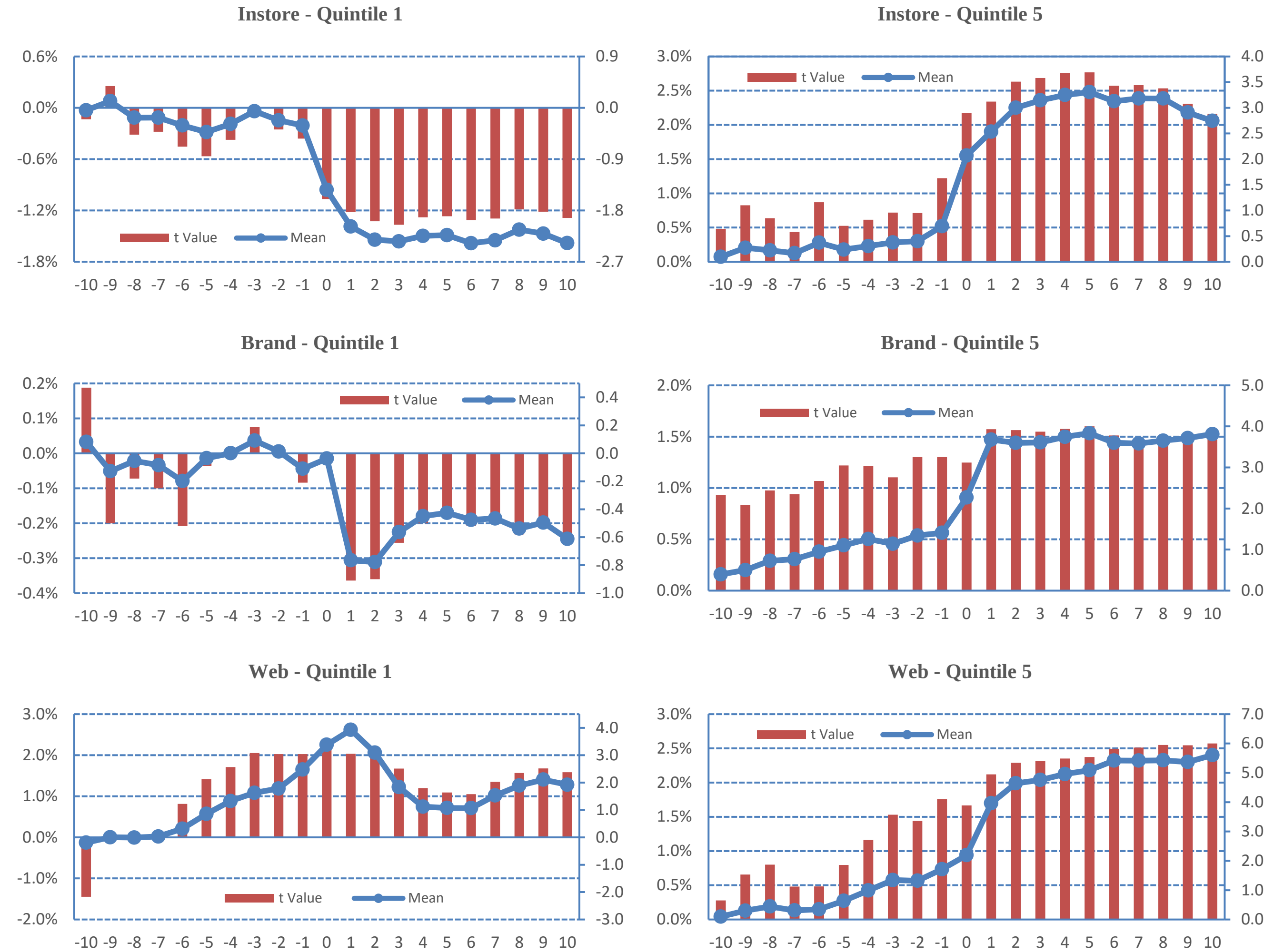


**Figure 2. Excess returns around earnings announcement dates for firms in the consumer sectors.** This figure plots the average buy-and-hold returns during the event window from 10 days prior to the earnings announcement date (day 0) to 10 days afterward for the subsample of firms in the consumer sectors. Returns are calculated in excess of the market returns of corresponding periods. The first column shows the average buy-and-hold return of firms in Quintile 1 of WQS, while the second column shows the results of firms in Quintile 5. The sample period is 2009 – 2017.

## Table 1: Summary Statistics

This table provides the summary statistics of main variables, including sales proxies calculated from various big-data sources. In-store is the foot traffic measure that is estimated based on the amount of consumer activities at retail stores. Brand awareness is estimated based on consumer web searches and interactions with respect to products and brand names in various social media channels of each firm. Web is the amount of web traffic that is calculated from the number of visitors to websites of each sample firm. The sales proxies are calculated from the growth rate of consumer activities during Fiscal Quarter $q$ from the quarterly average of consumer activities over the past four quarters ($q$-4 to $q$-1). Size is the natural logarithm of market capitalization at the end of fiscal quarter. LogBM is the natural logarithm of the book-to-market ratio as of the most recent fiscal year ending at least three month prior to fiscal quarter-$q$ end. MOM is a buy-and-hold return during the 6-month period of $t$-$8$ to $t$-$2$, where $t$ is the month of fiscal quarter-$q$ end. Panel A provides the descriptive statistics. The upper right corner of Panel B reports Pearson correlations and the lower left corner of the table provides Spearman rank correlations. The p-values of correlations are reported in the bracket. The sample period is 2009–2017.

**Panel A: Descriptive Statistics**

| Variables | N | Mean | Std Dev | 10th | 25th | Median | 75th | 90th |
|---|---|---|---|---|---|---|---|---|
| In-store | 1617 | 0.012 | 0.208 | -0.193 | -0.093 | -0.008 | 0.094 | 0.257 |
| Brand | 9114 | 0.022 | 0.290 | -0.259 | -0.105 | 0.020 | 0.159 | 0.315 |
| Web | 8466 | -0.024 | 0.155 | -0.172 | -0.087 | -0.020 | 0.043 | 0.117 |
| SIZE | 10817 | 16.152 | 1.576 | 14.084 | 15.238 | 16.213 | 17.115 | 18.194 |
| LogBM | 9635 | -1.050 | 0.962 | -2.163 | -1.567 | -1.004 | -0.452 | 0.035 |
| MOM | 10767 | 0.089 | 0.268 | -0.179 | -0.039 | 0.078 | 0.200 | 0.343 |

**Panel B: Correlations**

| | In-store | Brand | Web | SIZE | LogBM | MOM |
|---|---|---|---|---|---|---|
| In-store | | 0.456 | 0.356 | 0.032 | 0.046 | 0.107 |
| | | [0.000] | [0.000] | [0.205] | [0.072] | [0.000] |
| Brand | 0.577 | | 0.210 | 0.087 | -0.039 | 0.048 |
| | [0.000] | | [0.000] | [0.000] | [0.001] | [0.000] |
| Web | 0.399 | 0.191 | | 0.012 | 0.025 | 0.143 |
| | [0.000] | [0.000] | | [0.258] | [0.027] | [0.000] |
| SIZE | 0.066 | 0.102 | 0.003 | | -0.253 | 0.044 |
| | [0.008] | [0.000] | [0.807] | | [0.000] | [0.000] |
| LogBM | -0.024 | -0.091 | 0.019 | -0.268 | | 0.030 |
| | [0.336] | [0.000] | [0.095] | [0.000] | | [0.004] |
| MOM | 0.149 | 0.055 | 0.165 | 0.083 | -0.010 | |
| | [0.000] | [0.000] | [0.000] | [0.000] | [0.322] | |

## Table 2: Standardized Changes in Revenue and Sales Proxies

This table reports the regression results of the standardized changes in revenue (SCR) on the within-quarter sales proxies (WQS). Panel A reports the regression results using the entire sample, while Panel B shows the results of the subsample of consumer-sector firms. The SCR for stock *i* in quarter *t* is calculated as $[(S_{i,t} - S_{i,t-4}) - r_{i,t}]/\sigma_{i,t}$, where $S_{i,t}$ is the quarterly revenue as of fiscal quarter *t* for firm *i*, $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and average, respectively, of $(S_{i,t} - S_{i,t-4})$ over the preceding eight quarters. WQS is the sales proxy for the fiscal quarter *t*, which is obtained from the growth rate of consumer activities during Fiscal Quarter *t* from the quarterly average of consumer activities over the past four quarters (*t*-4 to *t*-1). *t* statistics are reported in the bracket. The sample period is 2009–2017.

**Panel A: All Firms**

| Sales Proxy | In-Store (Consumer Firms) | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Coefficient | 0.505 | 0.444 | 0.330 | 0.433 | 0.476 | 0.480 | 0.582 | 0.565 | 0.460 |
| t value | [2.98] | [2.64] | [1.99] | [2.76] | [2.99] | [3.10] | [9.30] | [8.68] | [6.83] |
| N | 1283 | 1283 | 1283 | 7065 | 7065 | 7065 | 7897 | 7897 | 7897 |
| (Avg) $R^2$ | 0.7% | 7.5% | 13.4% | 0.1% | 4.8% | 10.4% | 1.1% | 6.9% | 11.3% |
| Fixed Effects | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time |

**Panel B: Consumer Firms**

| Sales Proxy | In-Store | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Coefficient | 0.505 | 0.444 | 0.330 | 0.656 | 0.668 | 0.676 | 0.820 | 0.744 | 0.630 |
| t value | [2.98] | [2.64] | [1.99] | [2.69] | [2.69] | [2.84] | [8.81] | [7.74] | [6.33] |
| N | 1283 | 1283 | 1283 | 2871 | 2871 | 2871 | 4440 | 4440 | 4440 |
| (Avg) $R^2$ | 0.7% | 7.5% | 13.4% | 0.3% | 3.8% | 12.4% | 1.7% | 7.1% | 11.9% |
| Fixed Effects | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time |

**Table 3: Standardized Changes in Earnings, Analsyt Forecast Errors, and Sales Proxies**

This table reports the regression results of the standardized changes in earnings (SCE) and analyst forecast errors (FE) on the within-quarter sales proxies (WQS). Panels A and B report the regression results of SCE on WQS, while Panels C and D show the regressions of analyst FE on WQS. The SCE for stock *i* in quarter *t* is calculated as $[(E_{i,t} - E_{i,t-4}) - r_{i,t}]/\sigma_{i,t}$, where $E_{i,t}$ is the quarterly earnings as of fiscal quarter *t* for firm *i*, $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and average, respectively, of $(E_{i,t} - E_{i,t-4})$ over the preceding eight quarters. Analyst FE is estimated as $(A_{i,q} - F_{i,q})/P_{i,q}$, where $A_{i,q}$ is quarterly earnings per share announced for quarter *q* of stock *i*, $F_{i,q}$ is mean analysts' forecasted EPS, and $P_{i,q}$ is quarter-end price. Analyst FE is standardized to mean zero and unit standard deviation. WQS is the sales proxy for the fiscal quarter *t*, which is obtained from the growth rate of consumer activities during Fiscal Quarter *t* from the quarterly average of consumer activities over the past four quarters (*t*-4 to *t*-1). *t* statistics are reported in the bracket. The sample period is 2009–2017.

**Panel A: SCE - All Firms**

| Sales Proxy | In-Store (Consumer Firms) | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Coefficient | 0.323 | 0.305 | 0.288 | 0.129 | 0.175 | 0.186 | 0.441 | 0.487 | 0.380 |
| t value | [2.11] | [1.98] | [1.88] | [0.81] | [1.07] | [1.15] | [7.52] | [7.88] | [5.82] |
| N | 1283 | 1283 | 1283 | 7065 | 7065 | 7065 | 7898 | 7898 | 7898 |
| (Avg) $R^2$ | 0.3% | 4.8% | 9.5% | 0.0% | 2.6% | 4.2% | 0.7% | 3.8% | 4.8% |
| Fixed Effects | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time |

**Panel B: SCE - Consumer Firms**

| Sales Proxy | In-Store | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Coefficient | 0.323 | 0.305 | 0.288 | 0.454 | 0.472 | 0.447 | 0.504 | 0.531 | 0.426 |
| t value | [2.11] | [1.98] | [1.88] | [1.87] | [1.90] | [1.81] | [5.64] | [5.68] | [4.32] |
| N | 1283 | 1283 | 1283 | 2871 | 2871 | 2871 | 4441 | 4441 | 4441 |
| (Avg) $R^2$ | 0.3% | 4.8% | 9.5% | 0.1% | 2.6% | 4.2% | 0.7% | 4.1% | 5.4% |
| Fixed Effects | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time |

**Panel C: Analyst FE - All Firms**

| Sales Proxy | In-Store (Consumer Firms) | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Coefficient | 0.203 | 0.253 | 0.215 | 0.095 | 0.195 | 0.186 | 0.171 | 0.146 | 0.058 |
| t value | [1.98] | [2.54] | [2.16] | [1.00] | [2.08] | [1.99] | [4.30] | [3.65] | [1.36] |
| N | 1383 | 1383 | 1383 | 7506 | 7506 | 7506 | 8210 | 8210 | 8210 |
| (Avg) $R^2$ | 0.3% | 11.5% | 14.3% | 0.0% | 10.9% | 11.7% | 0.2% | 12.4% | 13.5% |
| Fixed Effects | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time |

**Panel D: Analyst FE - Consumer Firms**

| Sales Proxy | In-Store | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Coefficient | 0.203 | 0.253 | 0.215 | 0.266 | 0.356 | 0.302 | 0.298 | 0.256 | 0.146 |
| t value | [1.98] | [2.54] | [2.16] | [1.83] | [2.56] | [2.19] | [5.22] | [4.49] | [2.43] |
| N | 1383 | 1383 | 1383 | 3028 | 3028 | 3028 | 4351 | 4351 | 4351 |
| (Avg) $R^2$ | 0.3% | 11.5% | 14.3% | 0.1% | 14.1% | 16.6% | 0.6% | 14.0% | 15.9% |
| Fixed Effects | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time | N/A | Firm | Firm+Time |

**Table 4: Event Studies: Quarterly Earnings Announcement Returns**

The table shows the average returns around earning announcement dates by quintiles of each sales proxy. The event window is the period between one day prior to the earnings announcement date and three day afterward. Returns are calculated in excess of the market returns of the corresponding periods. Quintiles of sales proxies are calculated using the following process. In month *t*, we pool firms that have fiscal quarter ending during the three-month rolling period of *t-2* to *t*, and rank the firms based on the proxies to obtain quintile cutoff values. Then, we use the quintile cutoff values to assign quintile ranks for the firms that have fiscal quarter ending in month *t*. The last row of each panel reports the results of the hypothesis testing for the mean difference between the highest and the lowest quintiles. Panel A reports the results of the entire sample, while Panel B uses the subsample of the consumer sectors. The sample period is 2009–2017.

**Panel A: All Firms**

| Sales Proxy | In-store (Consumer Firms) | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| Quintile | N | Mean | t Value | N | Mean | t Value | N | Mean | t Value |
| 1 | 279 | -1.34% | -1.95 | 1514 | -0.02% | -0.16 | 1711 | -0.07% | -0.35 |
| 2 | 303 | -0.53% | -0.93 | 1595 | 0.18% | 1.15 | 1772 | 0.69% | 3.62 |
| 3 | 315 | 0.67% | 1.33 | 1621 | 0.33% | 2.36 | 1776 | 0.20% | 1.11 |
| 4 | 318 | 0.94% | 1.78 | 1601 | 0.28% | 2.01 | 1752 | 0.22% | 1.17 |
| 5 | 295 | 1.89% | 3.47 | 1508 | 0.60% | 3.81 | 1697 | 0.69% | 3.61 |
| 5-1 | | 3.23% | 3.70 | | 0.62% | 2.87 | | 0.76% | 2.83 |

**Panel B: Consumer Sectors**

| Sales Proxy | In-store | | | Brand | | | Web | | |
|---|---|---|---|---|---|---|---|---|---|
| Quintile | N | Mean | t Value | N | Mean | t Value | N | Mean | t Value |
| 1 | 279 | -1.34% | -1.95 | 590 | -0.30% | -1.07 | 892 | -0.13% | -0.43 |
| 2 | 303 | -0.53% | -0.93 | 639 | -0.11% | -0.42 | 972 | 0.56% | 1.85 |
| 3 | 315 | 0.67% | 1.33 | 667 | 0.62% | 2.4 | 997 | 0.51% | 1.85 |
| 4 | 318 | 0.94% | 1.78 | 638 | 0.52% | 2.23 | 971 | 0.28% | 0.95 |
| 5 | 295 | 1.89% | 3.47 | 566 | 0.99% | 3.24 | 927 | 1.06% | 3.51 |
| 5-1 | | 3.23% | 3.70 | | 1.29% | 3.12 | | 1.19% | 2.79 |

**Table 5: Regressions of Earnings-Announcement Returns**

This table reports the regression results of the earnings-announcement returns on the sales proxies and control variables. Panel A shows the results of the entire sample, while Panel B reports the analyses using the subsample of the consumer sectors. The dependent variable is the returns around earnings-announcement dates for fiscal quarter-*q* earnings. The announcement return is calculated as the return in excess over the market during the period of one day before the earnings announcement date and three day after. WQS is the sales proxy for the fiscal quarter *q*, which is obtained from the growth rate of consumer activities during Fiscal Quarter *q* from the quarterly average of consumer activities over the past four quarters (*q*-4 to *q*-1). PQS is the corporate sales proxy for the period beginning after the fiscal-quarter-*q* end and ending prior to the announcement date for quarter-*q* earnings. Loss is an indicator variable that equals one if the announced earnings is negative, and zero otherwise. Past Return is the cumulative return in excess over the market for the period of thirteen to three days prior to the earnings announcement. The sample period is 2009–2017.

**Panel A: All Firms**

| Variables\Sales Proxy | N/A | In-Store (Consumer Firms) | | Brand | | Web | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WQS | | 0.065 | 0.053 | 0.017 | 0.015 | 0.009 | 0.007 |
| | | [4.51] | [3.74] | [3.05] | [2.50] | [2.21] | [1.72] |
| PQS | | -0.027 | -0.025 | -0.004 | 0.002 | 0.002 | 0.004 |
| | | [-2.15] | [-2.11] | [-0.81] | [0.36] | [0.98] | [1.57] |
| Analyst FE | 1.886 | | 4.799 | | 1.173 | | 1.800 |
| | [16.21] | | [12.19] | | [9.49] | | [14.49] |
| SIZE | -0.002 | | -0.003 | | -0.003 | | -0.003 |
| | [-4.61] | | [-1.32] | | [-4.53] | | [-4.29] |
| LogBM | 0.00 | | 0.01 | | 0.00 | | 0.00 |
| | [2.09] | | [1.60] | | [0.19] | | [1.63] |
| MOM | -0.006 | | 0.000 | | -0.018 | | -0.005 |
| | [-1.81] | | [0.03] | | [-4.92] | | [-1.45] |
| Loss | -0.013 | | -0.039 | | 0.000 | | -0.012 |
| | [-3.51] | | [-3.52] | | [0.02] | | [-2.83] |
| PastRet | -0.02 | | -0.16 | | 0.03 | | -0.02 |
| | [-1.00] | | [-3.49] | | [1.83] | | [-0.98] |
| Intercept | 0.045 | 0.002 | 0.047 | 0.003 | 0.055 | 0.004 | 0.048 |
| | [5.21] | [0.78] | [1.53] | [4.03] | [4.92] | [4.39] | [4.89] |
| N | 8937 | 1482 | 1349 | 7647 | 6683 | 8347 | 7152 |
| Adj $R^2$ | 3.5% | 1.2% | 13.6% | 0.1% | 2.2% | 0.1% | 3.6% |

**Panel B: Consumer Sectors**

| Variables\Sales Proxy | N/A | In-Store | | Brand | | Web | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WQS | | 0.065 | 0.053 | 0.045 | 0.042 | 0.017 | 0.011 |
| | | [4.51] | [3.74] | [4.75] | [3.99] | [2.61] | [1.82] |
| PQS | | -0.027 | -0.025 | -0.022 | -0.018 | 0.003 | 0.003 |
| | | [-2.15] | [-2.11] | [-2.69] | [-1.97] | [1.00] | [0.88] |
| Analyst FE | 2.149 | | 4.799 | | 1.115 | | 2.007 |
| | [13.61] | | [12.19] | | [6.24] | | [12.10] |
| SIZE | -0.002 | | -0.003 | | -0.003 | | -0.002 |
| | [-2.86] | | [-1.32] | | [-2.26] | | [-2.26] |
| LogBM | 0.00 | | 0.01 | | 0.00 | | 0.00 |
| | [3.13] | | [1.60] | | [1.49] | | [2.68] |
| MOM | -0.002 | | 0.000 | | -0.021 | | -0.002 |
| | [-0.44] | | [0.03] | | [-3.42] | | [-0.36] |
| Loss | -0.025 | | -0.039 | | -0.006 | | -0.021 |
| | [-4.59] | | [-3.52] | | [-0.85] | | [-3.50] |
| PastRet | -0.05 | | -0.16 | | 0.03 | | -0.04 |
| | [-2.28] | | [-3.49] | | [1.03] | | [-1.71] |
| Intercept | 0.047 | 0.002 | 0.047 | 0.003 | 0.054 | 0.007 | 0.044 |
| | [3.60] | [0.78] | [1.53] | [2.82] | [2.64] | [4.13] | [3.00] |
| N | 4835 | 1482 | 1349 | 3001 | 2786 | 4556 | 4020 |
| Adj $R^2$ | 4.8% | 1.2% | 13.6% | 0.7% | 2.7% | 0.2% | 4.7% |

**Table 6: Pre-Earnings-Announcement Returns**

This table reports the regression results of the pre-earnings-announcement returns on the sales proxies, and other control variables. Panel A shows the results of the entire sample, while Panel B reports the analyses using the consumer-sector subsample. The dependent variable is the pre-earnings-announcement return. The pre-earnings-announcement return is calculated as the return in excess over the market for the period of 25 to two days prior to the earnings announcement date. WQS is the sales proxy for the fiscal quarter *q*. PQS is the corporate sales proxy for the period beginning after the fiscal-quarter-*q* end and ending prior to the announcement date for quarter-*q* earnings. The sample period is 2009–2017.

**Panel A: All Firms**

| Variables\Sales Proxy | N/A | In-Store (Consumer Firms) | | Brand | | Web | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WQS | | 0.041 | 0.038 | 0.001 | 0.013 | 0.006 | 0.008 |
| | | [3.24] | [2.86] | [0.19] | [1.87] | [1.54] | [1.78] |
| PQS | | -0.001 | 0.002 | 0.002 | 0.000 | 0.003 | 0.005 |
| | | [-0.06] | [0.17] | [0.36] | [0.05] | [1.53] | [2.28] |
| Analyst FE | 1.325 | | 0.207 | | 0.644 | | 1.499 |
| | [10.66] | | [0.56] | | [4.61] | | [11.45] |
| SIZE | -0.001 | | 0.001 | | -0.005 | | -0.001 |
| | [-2.19] | | [0.42] | | [-6.03] | | [-1.48] |
| LogBM | 0.00 | | 0.01 | | 0.00 | | 0.00 |
| | [1.81] | | [1.96] | | [1.26] | | [0.97] |
| MOM | -0.020 | | 0.013 | | -0.029 | | -0.023 |
| | [-6.04] | | [1.67] | | [-7.21] | | [-6.40] |
| Loss | 0.011 | | 0.005 | | 0.026 | | 0.014 |
| | [2.65] | | [0.50] | | [5.55] | | [3.09] |
| Intercept | 0.028 | -0.001 | -0.008 | 0.007 | 0.087 | 0.006 | 0.023 |
| | [2.97] | [-0.37] | [-0.29] | [7.91] | [6.83] | [6.32] | [2.17] |
| N | 8937 | 1482 | 1349 | 7646 | 6683 | 8346 | 7152 |
| Adj $R^2$ | 1.8% | 0.7% | 1.1% | 0.0% | 2.3% | 0.1% | 2.5% |

**Panel B: Consumer Sectors**

| Variables\Sales Proxy | N/A | In-Store | | Brand | | Web | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WQS | | 0.041 | 0.038 | 0.008 | 0.022 | 0.015 | 0.013 |
| | | [3.24] | [2.86] | [0.89] | [2.15] | [2.61] | [2.10] |
| PQS | | -0.001 | 0.002 | -0.004 | -0.006 | 0.005 | 0.004 |
| | | [-0.06] | [0.17] | [-0.49] | [-0.66] | [1.66] | [1.53] |
| Analyst FE | 1.700 | | 0.207 | | 0.903 | | 1.721 |
| | [10.98] | | [0.56] | | [5.30] | | [10.42] |
| SIZE | -0.001 | | 0.001 | | -0.005 | | 0.000 |
| | [-0.81] | | [0.42] | | [-4.20] | | [-0.20] |
| LogBM | 0.00 | | 0.01 | | 0.00 | | 0.00 |
| | [1.34] | | [1.96] | | [0.57] | | [0.25] |
| MOM | -0.008 | | 0.013 | | 0.001 | | -0.010 |
| | [-2.01] | | [1.67] | | [0.16] | | [-2.16] |
| Loss | 0.003 | | 0.005 | | 0.031 | | 0.010 |
| | [0.64] | | [0.50] | | [4.34] | | [1.58] |
| Intercept | 0.016 | -0.001 | -0.008 | 0.005 | 0.088 | 0.004 | 0.007 |
| | [1.23] | [-0.37] | [-0.29] | [4.23] | [4.53] | [3.15] | [0.48] |
| N | 4835 | 1482 | 1349 | 3001 | 2786 | 4556 | 4020 |
| Adj $R^2$ | 2.5% | 0.7% | 1.1% | 0.0% | 2.3% | 0.3% | 2.9% |

## Table 7: Post-Earnings-Announcement Returns

This table reports the regression results of the post-earnings-announcement returns on the sales proxies and other control variables. Panel A shows the results of the entire sample, while Panel B reports the results of the consumer-sector subsample. The dependent variable is the return of each firm in excess over the market for the period beginning on 4 days after the quarter-*q* earnings announcement date and ending on 60 days after the announcement date. WQS is the sales proxy for the fiscal quarter *q*. PQS is the corporate sales proxy for the period beginning after the fiscal-quarter-*q* end and ending prior to the announcement date for quarter-*q* earnings. The sample period is 2009–2017.

**Panel A: All Firms**

| Variables\Sales Proxy | N/A | In-Store (Consumer Firms) | | Brand | | Web | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WQS | | -0.017 | -0.021 | -0.015 | -0.014 | -0.015 | -0.013 |
| | | [-0.78] | [-0.95] | [-1.72] | [-1.41] | [-2.57] | [-1.98] |
| PQS | | 0.059 | 0.054 | 0.017 | 0.019 | 0.022 | 0.021 |
| | | [3.19] | [2.91] | [2.42] | [2.53] | [6.91] | [6.15] |
| Analyst FE | 0.439 | | -0.034 | | -0.117 | | 0.462 |
| | [2.45] | | [-0.06] | | [-0.59] | | [2.41] |
| SIZE | 0.000 | | 0.005 | | -0.004 | | 0.000 |
| | [-0.14] | | [1.53] | | [-3.38] | | [-0.40] |
| LogBM | 0.00 | | 0.02 | | 0.00 | | 0.00 |
| | [1.72] | | [3.02] | | [2.28] | | [0.30] |
| MOM | 0.018 | | 0.014 | | 0.016 | | 0.016 |
| | [3.93] | | [1.06] | | [2.86] | | [3.02] |
| Loss | 0.025 | | 0.021 | | 0.016 | | 0.034 |
| | [4.22] | | [1.21] | | [2.37] | | [5.14] |
| PastRet | 0.06 | | -0.22 | | 0.01 | | 0.03 |
| | [2.37] | | [-3.12] | | [0.39] | | [1.26] |
| Intercept | 0.005 | -0.011 | -0.071 | 0.007 | 0.069 | 0.008 | 0.010 |
| | [0.36] | [-2.73] | [-1.50] | [5.71] | [3.83] | [5.59] | [0.67] |
| N | 8924 | 1481 | 1348 | 7642 | 6679 | 8343 | 7148 |
| Adj $R^2$ | 0.5% | 0.6% | 1.9% | 0.1% | 0.6% | 0.6% | 1.1% |

**Panel B: Consumer Sectors**

| Variables\Sales Proxy | N/A | In-Store | | Brand | | Web | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| WQS | | -0.017 | -0.021 | -0.014 | -0.006 | -0.016 | -0.013 |
| | | [-0.78] | [-0.95] | [-0.92] | [-0.37] | [-1.86] | [-1.37] |
| PQS | | 0.059 | 0.054 | 0.020 | 0.011 | 0.026 | 0.022 |
| | | [3.19] | [2.91] | [1.61] | [0.84] | [6.30] | [5.10] |
| Analyst FE | 0.819 | | -0.034 | | 0.246 | | 0.774 |
| | [3.47] | | [-0.06] | | [0.91] | | [3.10] |
| SIZE | 0.000 | | 0.005 | | -0.001 | | 0.001 |
| | [0.19] | | [1.53] | | [-0.78] | | [0.36] |
| LogBM | 0.00 | | 0.02 | | 0.01 | | 0.00 |
| | [1.35] | | [3.02] | | [2.78] | | [-0.13] |
| MOM | 0.020 | | 0.014 | | 0.014 | | 0.019 |
| | [3.10] | | [1.06] | | [1.58] | | [2.66] |
| Loss | 0.022 | | 0.021 | | -0.005 | | 0.034 |
| | [2.71] | | [1.21] | | [-0.43] | | [3.71] |
| PastRet | 0.04 | | -0.22 | | -0.08 | | 0.03 |
| | [1.25] | | [-3.12] | | [-1.72] | | [0.88] |
| Intercept | -0.005 | -0.011 | -0.071 | 0.001 | 0.033 | 0.006 | -0.009 |
| | [-0.25] | [-2.73] | [-1.50] | [0.57] | [1.05] | [2.58] | [-0.40] |
| N | 4827 | 1481 | 1348 | 3000 | 2786 | 4556 | 4020 |
| Adj $R^2$ | 0.6% | 0.6% | 1.9% | 0.0% | 0.3% | 0.9% | 1.3% |