Predicting Performance Using Consumer Big Data

Kenneth Froot, Namho Kang, Gideon Ozik, and Ronnie Sadka

Kenneth Froot
is the André R. Jakurski Professor of Business Administration, Emeritus, at Harvard University Graduate School of Business in Cambridge, MA.
kfroot@hbs.edu

Namho Kang
is an assistant professor of finance at Bentley University in Waltham, MA.
nkang@bentley.edu

Gideon Ozik
is an affiliate professor at EDHEC Business School in Cambridge, MA.
gideon.ozik@edhec.com

Ronnie Sadka
is the Haub Family Professor, chairperson of the Department of Finance, and senior associate dean of faculty at the Carroll School of Management, Boston College, in Chestnut Hill, MA.
sadka@bc.edu

KEY FINDINGS

- Consumer big data contain information on various types of consumer activity, including store visits (IN-STORE), reactions to corporate brands (BRAND), and web searches (WEB). Real-time sales proxies constructed based on these activity types predict firms' revenue, earnings, and returns.
- The speed of information dissemination varies depending on the types of consumer activity. WEB information is incorporated in stock prices more quickly than the information in IN-STORE and BRAND.
- The increase in WEB and decrease in IN-STORE during the pandemic suggest that firms' online sales have increased and offline sales have decreased. In addition, the return predictability of sales proxies has increased during the pandemic.

ABSTRACT

To predict firms' fundamentals, the authors construct three proxies for real-time corporate sales from fully distinct information sources: in-store foot traffic (IN-STORE), web traffic to companies' websites (WEB), and consumers' interest level in corporate brands and products (BRAND). The authors demonstrate that trading using these proxies, estimated for a sample of 330 firms over 2009–2020, results in significant net-of-transaction-costs profitability. During the pandemic, WEB activity increased significantly whereas IN-STORE experienced a remarkable decrease, reflecting the migration of consumers from physical stores toward online retailers. The results suggest that the information contained in IN-STORE and BRAND is not immediately available to investors, whereas the WEB information diffuses more quickly, and overall information diffusion worsened during the pandemic.

Big data sources that describe aspects of consumer/investor behavior have become one of the fastest-growing themes in many disciplines, including empirical finance. In what already sounds quaint today, a few years ago UBS analysts 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 (Zhu 2019). In addition to the publicly realized benefit of more informative stock prices, there is also evidence suggesting that the information in these data sources is only partially—not yet fully—incorporated into market prices. For example, Froot et al. (2017) used consumer big data to proxy for firms' consumer-store visits and showed that these proxies help predict earnings announcement returns.

In this article, we study the ability of big data to predict firms' fundamentals and stock returns. In particular, we contribute to the literature by investigating various sources of consumer data for a broader cross-section of firms. Whereas Froot et al. (2017) used in-store sales information for roughly 60 firms, we use two additional types of information—web traffic and brand awareness, estimated from various big data sources—for an expanded universe of 330 firms. We ask whether all three types of information are efficiently incorporated into asset prices, which is an important economic question for academics and practitioners alike, especially as they try to understand which type of data should be paid more attention, given the large amounts of alternative data available these days. We also consider the costs of trading before reaching final conclusions. Thus, this article also offers practical applications for investment professionals.

DATA AND VARIABLES

Institutional Background: Alternative Data and Investment

Adoption of alternative data in asset management has experienced rapid growth in the last few years. EY's Global Hedge Funds and Investor Survey (Lee 2017) indicated that 46% of hedge funds used or expected to use nontraditional data in their investment processes. Furthermore, it is estimated that 60% of managers have their front-office team experimenting with nontraditional data. AlternativeData.org, a directory of data vendors, lists more than 400 alternative data providers and estimates that the alternative data industry grew four-fold from 2017 to 2020. As alternative data become available to a wider audience, it is important to enhance the knowledge of the investment management community regarding the associated challenges and implications.

Alternative data market and providers. Kolanovic and Krishnamachari (2017) classified alternative datasets into three categories: sources related to individuals, business processes, and sensors. Individual-type sources include datasets generated from individual activities in social media and specialized websites, as well as consumers' web searches and personal data (e.g., email receipts). Business-process datasets are generated from commercial transactions (e.g., credit card transaction logs) or provided by private and public agencies, and sensor types includes datasets generated from satellites or other sensory devices. The datasets examined in this article are related to the individual type because they proxy for individual consumer activities, both offline and online, and for consumer interest in brands. About a third of the alternative data vendors listed on AlternativeData.org are categorized as web data, web traffic, or social/sentiment, categories relevant to the proxies examined in this article.

Investment strategies and academic research. As researchers have studied the usefulness of alternative datasets, various investing strategies have been developed as a result. Sentiment/natural language processing analysis of the individual data sources, such as social media, is a popular approach. Many academic studies have provided evidence of the profitability of such strategies. Chen et al. (2014) studied Seeking Alpha, a popular financial blog, and found that positive sentiment measures predict earnings announcements and future stock returns. Bartov, Faurel, and Mohanram (2015) used the Twitter feed to extract aggregate sentiment before earnings announcements. Da, Engleberg, and Gao (2011) showed that Google search volume is a momentum forecaster for near-term stock returns and, further out, leads to subsequent reversals. Froot et al. (2018) showed that sentiment measured from professional media sources reinforces and positively predicts recent returns.

Huang (2018) examined customer product reviews on Amazon.com and found that customer ratings positively predict stock returns.

Studies also show that fundamental analysis based on big data sources can create indicators that are more granular in time and location. Sales forecasts using web searches, credit card transactions, and satellite images are well-known applications. Froot et al. (2017) created from these information sources sales proxies for consumer-store visits and showed that these proxies help predict earnings announcement returns. Zhu (2019) demonstrated that availability of satellite data and consumer transaction information influences the efficiency of stock prices, which are better indicators of future earnings and investment opportunities. Katona et al. (2020) showed that availability of satellite image data creates opportunities for sophisticated investors to formulate trading strategies at the expense of individual investors.

Although most alternative datasets are focused on equity or specific sectors, the application of such datasets to macro variables, such as inflation and consumer credit, have also been studied. Cavallo and Rigobon (2016) and Cavallo (2017) used the vast number of online prices on the web and created more granular and precise price indexes, which is of use to macro investors.

Data Sources

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

IN-STORE measures consumer activity at retail stores. Specifically, it estimates consumers' intention to visit or shop at a retail store by using activities such as searches for driving directions to a store location, queries concerning store hours, or downloads of discount coupons. These consumer activities at retail stores are collected from various sources, including millions of consumer devices. 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 covers only 66 retail firms.

WEB is an estimate of consumer visits to consumer-firm websites. In particular, it is estimated by observing the 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 of 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. Unlike WEB, which estimates consumer activities using firm-site visits, BRAND 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, we derive proxies to cover two distinct periods, one to measure quarterly sales activity and the other to measure monthly sales activity. The quarterly 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). Monthly sales proxies are estimated from the growth rate of consumer activities during the most recent 3-month period over the average for the past 12-month period. The growth rates are calculated using the log differences.

Sample and Summary Statistics

The full sample includes 331 US public companies in various industries during the period 2009–2020. It comprises 178 companies in consumer sectors (consumer discretionary and consumer staples; 136 and 42 companies for each sector, respectively);[1] 76 companies in financials and information technology sectors (47 and 29 companies, respectively); and 77 companies spanning across the health care, real estate, communications, and industrial and materials sectors. We use CRSP to obtain stock market variables, including stock returns and prices, and Compustat to obtain information on financial statements. Analyst forecasts are obtained from IBES.

Exhibit 1 shows the summary statistics of the monthly sales proxies and other main variables. SIZE is the natural logarithm of market capitalization at the end of the fiscal quarter. BE/ME is the natural logarithm of the book-to-market ratio, measured as of the most recent fiscal year ending at least three months before the fiscal quarter end. MOM is a buy-and-hold return during the past 12-month period.

Panel A provides a snapshot of our sample coverage with respect to the CRSP universe as of 2017, for which our sample size is the largest. The sample size and the sector coverage are stable over time. Compared to the entire CRSP universe, the sample size is about 7% of the total number of firms in CRSP. This proportion increases to 28% when the sample is restricted to the consumer sectors. However, the sample coverage is much more representative than the sample size indicates: The sample covers 33% of total sales and 46% of total market capitalization of CRSP, as well as 59% of total sales and 73% of market capitalization of firms in the consumer sectors.

Panel B shows that IN-STORE has a higher mean (6.7%) and standard deviation (28%) than other proxies. The mean of WEB is 3.5%, and its standard deviation is 23.3%, whereas BRAND has a mean of 0.4% and a standard deviation of 11.1%. IN-STORE firms tend to have low book-to-market ratios and low past returns compared to firms that BRAND and WEB cover, indicating recent struggles of big-box retailers.

Panel C provides the correlations among the sales proxies and firm characteristics. The upper right corner reports Pearson correlations, and the lower left corner provides Spearman rank correlations. The panel shows that the correlations among the three sales proxies are all positive and statistically significant. For example, IN-STORE has a correlation of 0.56 with BRAND and 0.48 with WEB. Panel C also shows that all three proxies are highly correlated with MOM, indicating that firms with a high level of consumer activities have recently experienced high stock returns.

[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.

EXHIBIT 1

Summary Statistics

Panel A: Sample Coverage

	CRSP Universe ($ billion)			Our Sample ($ billion)			Coverage		
Sectors	**N**	**Sales**	**Market Cap**	**N**	**Sales**	**Market Cap**	**N**	**Sales**	**Market Cap**
Consumer Discretionary	488	3,082	3,393	136	1,566	2,282	27.9%	50.8%	67.3%
Consumer Staples	158	1,985	2,389	42	1,409	1,915	26.6%	71.0%	80.2%
Consumer Total	646	5,067	5,782	178	2,975	4,197	27.6%	58.7%	72.6%
Health Care	778	2,474	4,044	22	930	1,882	2.8%	37.6%	46.5%
Financials	772	3,049	4,838	47	1,076	2,548	6.1%	35.3%	52.7%
Information Technology	615	1,929	5,229	29	714	2,930	4.7%	37.0%	56.0%
Others	1700	8,635	11,441	55	1,317	2,828	3.2%	15.2%	24.7%
Non-Consumer Total	3865	16,087	25,552	153	4,037	10,188	4.0%	25.1%	39.9%
Grand Total	4511	21,153	31,334	331	7,012	14,385	7.3%	33.1%	45.9%

Panel B: Descriptive Statistics

	Sales Proxies				Firm Characteristics (mean)		
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Mkt Cap (mil)**	**BE/ME**	**Mom**
In-Store	6564	0.067	0.278	0.011	20,836.89	0.670	0.099
Brand	39356	0.004	0.111	0.000	39,794.80	1.417	0.154
Web	33123	0.035	0.233	0.008	27,060.05	1.557	0.157

Panel C: Correlations

	In-Store	Brand	Web	Size	BE/ME	Mom
In-Store		0.556	0.481	–0.019	–0.006	0.102
		[0.000]	[0.000]	[0.130]	[0.673]	[0.000]
Brand	0.602		0.261	–0.003	–0.011	0.037
	[0.000]		[0.000]	[0.523]	[0.030]	[0.000]
Web	0.527	0.264		–0.013	0.040	0.075
	[0.000]	[0.000]		[0.022]	[0.000]	[0.000]
Size	0.036	0.001	0.003		–0.417	0.137
	[0.004]	[0.912]	[0.640]		[0.000]	[0.000]
BE/ME	–0.018	–0.030	0.031	–0.462		–0.059
	[0.162]	[0.000]	[0.000]	[0.000]		[0.000]
Mom	0.114	0.041	0.063	0.181	–0.199	
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	

NOTES: This exhibit provides the summary statistics of sales proxies calculated from various big-data sources. In-Store estimates consumer intention to visit or shop at a retail store. Brand measures the level of consumer interest from activities such as searches for sample-company products and brand names over various platforms. Web is an estimate of consumer visits to firm websites. The monthly sales proxies are calculated from the growth rate of consumer activities during the most recent 3-month period from the average of consumer activities over the past 12-month period. Size is the natural logarithm of market capitalization at the end of the fiscal quarter. BE/ME is the ratio of the book value of equity to the market value of equity. MOM is the cumulative returns of the past 12 months. Panel A shows the overall sample coverage by comparing the total sales and the market capitalization of the sample firms with those of the entire CRSP universe as of 2017. Panel B provides the descriptive statistics of the monthly sales proxies as well as firm characteristics. The upper right corner of Panel C reports Pearson correlations, and the lower left corner of the panel provides Spearman rank correlations. The *p*-values of correlations are reported in brackets. The sample period is 2009–2020.

FIRM FUNDAMENTALS AND SALES PROXIES

Exhibit 2 examines the predictive power of the sales proxies for firms' fundamentals. Specifically, Panels A, B, C, and D show results from the regressions of revenue growth, standardized unexpected revenue (SUR), standardized unexpected earnings (SUE), and analyst forecast errors (AFE), respectively, on the sales proxies.

EXHIBIT 2

Predicting Firm Fundamentals Using Sales Proxies

	In-Store		Brand				Web			
Sales Proxy	All		All		Consumer		All		Consumer	
Sector	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Revenue Growth										
Coefficient	0.475	0.356	0.529	0.589	1.195	1.290	0.128	0.185	0.195	0.286
t-Value	[26.61]	[16.94]	[24.53]	[27.33]	[27.99]	[30.90]	[13.37]	[17.90]	[13.17]	[18.20]
N	1,725	1,725	8,805	8,805	3,511	3,511	9,291	9,291	5,492	5,492
R^2	29.1%	46.0%	6.4%	22.8%	18.3%	37.5%	1.9%	15.7%	3.1%	19.7%
Fixed Effects	N	Y	N	Y	N	Y	N	Y	N	Y
Panel B: SUR										
Coefficient	0.536	0.495	0.271	0.403	0.527	0.685	0.461	0.338	0.525	0.371
t-Value	[3.80]	[3.55]	[2.20]	[3.29]	[2.70]	[3.59]	[8.96]	[6.17]	[7.41]	[5.00]
N	1,658	1,658	8,406	8,406	3,356	3,356	8,881	8,881	5,251	5,251
R^2	0.9%	12.9%	0.1%	11.0%	0.2%	13.4%	0.9%	12.5%	1.0%	13.0%
Fixed Effects	N	Y	N	Y	N	Y	N	Y	N	Y
Panel C: SUE										
Coefficient	0.289	0.369	0.050	0.165	0.385	0.517	0.294	0.246	0.239	0.178
t-Value	[1.99]	[2.50]	[0.42]	[1.33]	[1.98]	[2.60]	[6.17]	[4.67]	[3.57]	[2.45]
N	1,658	1,658	8,406	8,406	3,356	3,356	8,881	8,881	5,251	5,251
R^2	0.2%	7.5%	0.0%	3.8%	0.1%	4.8%	0.4%	5.0%	0.2%	5.7%
Fixed Effects	N	Y	N	Y	N	Y	N	Y	N	Y
Panel D: AFE										
Coefficient	0.071	0.152	0.059	0.121	0.266	0.391	0.193	0.155	0.239	0.255
t-Value	[0.78]	[1.70]	[0.85]	[1.76]	[2.37]	[3.53]	[5.55]	[4.11]	[5.16]	[5.10]
N	1,943	1,943	10,048	10,048	4,027	4,027	9,123	9,123	5,306	5,306
R^2	0.0%	12.3%	0.0%	9.1%	0.1%	11.0%	0.3%	12.1%	0.5%	11.5%
Fixed Effects	N	Y	N	Y	N	Y	N	Y	N	Y

NOTES: The exhibit shows the regressions of the quarterly revenue growth, standardized unexpected revenue (SUR), standardized unexpected earnings (SUE), and analyst earnings forecast errors (AFE) on the sales proxies. Revenue growth is the quarterly revenue growth for firm *i* as of fiscal quarter *t*. The SUR for stock *i* in quarter *t* is calculated as $[(S_{i,t} - S_{i,t-4}) - r_{i,t}]/\sigma_{i,t}$, where $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and average, respectively, of $(S_{i,t} - S_{i,t-4})$ over the preceding eight quarters. The SUE 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 per share announced for fiscal quarter *t* for firm *i*, and $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and average, respectively, of $(E_{i,t} - E_{i,t-4})$ over the preceding eight quarters. AFE is estimated as $(E_{i,t} - F_{i,q})/P_{i,t}$, where $F_{i,t}$ is the mean analyst forecasted EPS for quarter *t*, and $P_{i,t}$ is quarter-end price. All variables are standardized to a mean of zero and unit standard deviation. The sales proxy for the fiscal quarter *t* 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 $t - 1$). Firm and time fixed effects are included for the even-numbered columns. The *t* statistics are reported in brackets. The sample period is 2009–2020.

The revenue growth is the percentage increase in the quarterly revenue for firm *i* as of fiscal quarter *t*. The SUR 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*, and $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and the average, respectively, of $(S_{i,t} - S_{i,t-4})$ over the preceding eight quarters. SUE 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 per share announced for fiscal quarter *t* for firm *i*, and $\sigma_{i,t}$ and $r_{i,t}$ are the standard deviation and the average, respectively, of $(E_{i,t} - E_{i,t-4})$ over the preceding eight quarters. AFE is estimated as $(E_{i,t} - F_{i,t})/P_{i,t}$, where $F_{i,t}$ is mean analysts' forecasted EPS for quarter *t*, and $P_{i,t}$ is quarter-end price. For Exhibit 2, all variables are standardized to have a mean of zero and unit standard deviation for ease of interpretation. Except for IN-STORE, which covers only consumer-sector firms, we report results for the consumer-sector subsample separately.

Panel A shows that all three proxies are significantly related to firms' revenue growth. The coefficients on all three proxies are positive and highly significant. For BRAND and WEB, the association between the sales proxies and revenue growth is much stronger for consumer-sector firms. For example, the R^2 of BRAND is 6.4% for the entire sample (Model 3), whereas it is 18.3% for consumer sector firms (Model 5). However, the R^2 of WEB is much smaller than those of other proxies, indicating that the relation between WEB and revenue growth is much weaker.

Panels B and C examine the relations of sales proxies with SUR and SUE, the unexpected portion of revenue and earnings growths. The panels report that although all three proxies are significantly associated with the SUR and SUE, the coefficients and R^2 of WEB tend to be more significant than those of other proxies. For example, in Panel B, the regression of the SUR on WEB (Model 7) yields a coefficient of 0.461 (*t*-value of 8.96) and an R^2 of 0.9%, indicating that a one-standard-deviation change in WEB leads to a change in SUR corresponding to 46% of its standard deviation. The *t*-value of the coefficient and the R^2 for the model are larger than the values of the corresponding models for the other proxies. The results also show that the information contained in BRAND and WEB is more relevant for consumer firms. For example, in Panel C, although BRAND is not significantly related to the SUE when the entire sample is used (Models 3 and 4), the consumer-sector subsample results in Models 5 and 6 indicate that BRAND also contains significant information regarding the unexpected earnings growth of consumer companies.

Panel D shows that the sales proxies are also significantly associated with AFE. Similar to the SUE results, the prediction using BRAND and WEB is stronger for consumer-sector firms. For example, a one-standard-deviation change in BRAND is associated with a change in AFE of about 6% of its standard deviation (Model 3). For consumer-sector firms, the magnitude increases to a change of about 27% of the standard deviation (Model 5).

STOCK RETURNS AND SALES PROXIES

In this section, we first study whether the availability of alternative data sources helps to predict the returns of calendar-time portfolios. We then examine whether the information in the sales proxies is impounded in prices in a timely manner. Finally, we examine whether the predictability leads to profitable trades after considering transaction costs.

Portfolio Returns

In Exhibit 3, we analyze portfolio strategies constructed based on the sales proxies. For brevity, we focus on the consumer-sector subsample because the previous analyses show that the information in the sales proxies is more pertinent for consumer-sector firms.

We construct portfolios in each calendar month, based on monthly calendar-time sales proxies. Specifically, the sales proxies in month *t* are estimated from the growth rate of consumer activities during the most recent 3-month period over the past 12-month period. Quintile portfolios are formed based on the sales proxies in month *t*. Average returns of the portfolios, in excess of the market, are calculated for months $t + 1$, $t + 2$, and $t + 3$. Alphas of high-minus-low portfolios are calculated using the Fama–French six factors (Fama–French three, RMW, CMA, and Momentum).[2] Panel A shows the results of equal-weighted portfolios, and Panel B reports the value-weighted portfolio results.

[2] Our results are similar when Carhart's four factors (Fama–French three + Momentum) are used.

EXHIBIT 3

Calendar Time Portfolio Returns

Sales Proxies	In-Store			Brand			Web		
Quintiles\Month	*t* + 1	*t* + 2	*t* + 3	*t* + 1	*t* + 2	*t* + 3	*t* + 1	*t* + 2	*t* + 3
Panel A: Equal-Weighted Portfolios									
Low	–0.68%	–0.90%	–0.83%	0.20%	–0.06%	0.22%	–0.08%	0.13%	0.23%
	[–1.22]	[–1.88]	[–1.58]	[0.73]	[–0.18]	[0.67]	[–0.19]	[0.37]	[0.57]
2	0.06%	–0.36%	0.00%	0.06%	0.10%	0.11%	0.39%	–0.19%	0.14%
	[0.13]	[–0.72]	[0.00]	[0.27]	[0.53]	[0.49]	[1.08]	[–0.50]	[0.38]
3	0.03%	0.16%	–0.18%	0.20%	0.29%	0.24%	0.44%	0.53%	0.07%
	[0.06]	[0.39]	[–0.39]	[0.87]	[1.51]	[1.04]	[1.31]	[1.47]	[0.21]
4	0.10%	0.49%	0.20%	0.31%	0.35%	0.24%	0.38%	0.41%	0.47%
	[0.22]	[0.96]	[0.40]	[1.42]	[1.58]	[1.06]	[1.03]	[1.01]	[1.24]
High	1.21%	1.02%	0.68%	0.50%	0.41%	0.29%	0.48%	0.63%	0.44%
	[1.98]	[1.69]	[1.39]	[1.76]	[1.23]	[1.06]	[1.19]	[1.68]	[1.13]
High–Low	1.89%	1.93%	1.52%	0.31%	0.46%	0.07%	0.56%	0.50%	0.22%
	[3.20]	[3.35]	[2.87]	[1.09]	[1.58]	[0.22]	[1.48]	[1.52]	[0.60]
Alpha (H–L)	1.93%	1.80%	1.38%	0.32%	0.55%	0.26%	0.37%	0.41%	0.25%
	[3.08]	[3.06]	[2.39]	[1.11]	[1.77]	[0.78]	[0.88]	[1.13]	[0.65]
Panel B: Value-Weighted Portfolios									
Low	–0.41%	–0.58%	–0.45%	–0.04%	–0.46%	–0.11%	–0.50%	–0.21%	0.04%
	[–0.94]	[–1.33]	[–1.14]	[–0.17]	[–2.00]	[–0.46]	[–2.06]	[–0.82]	[0.15]
2	–0.04%	–0.18%	0.05%	–0.01%	–0.19%	0.12%	0.43%	–0.06%	–0.02%
	[–0.12]	[–0.50]	[0.17]	[–0.07]	[–0.92]	[0.58]	[1.97]	[–0.26]	[–0.08]
3	–0.12%	–0.13%	–0.42%	–0.26%	0.10%	–0.25%	0.24%	0.13%	0.18%
	[–0.31]	[–0.37]	[–1.28]	[–1.13]	[0.50]	[–1.10]	[1.01]	[0.57]	[0.70]
4	0.16%	0.13%	–0.03%	0.05%	0.25%	0.21%	0.12%	0.54%	0.36%
	[0.45]	[0.36]	[–0.07]	[0.24]	[1.05]	[0.95]	[0.43]	[2.10]	[1.38]
High	0.66%	0.56%	0.56%	0.50%	0.37%	0.24%	0.12%	0.29%	0.34%
	[1.77]	[1.44]	[1.35]	[2.08]	[1.48]	[0.95]	[0.46]	[1.09]	[1.29]
High–Low	1.07%	1.13%	1.01%	0.54%	0.83%	0.34%	0.62%	0.50%	0.30%
	[2.00]	[2.18]	[2.00]	[1.88]	[2.60]	[1.00]	[2.05]	[1.52]	[0.85]
Alpha (H–L)	1.48%	1.13%	0.70%	0.46%	0.95%	0.46%	0.50%	0.48%	0.33%
	[2.60]	[2.02]	[1.32]	[1.49]	[2.87]	[1.33]	[1.58]	[1.40]	[0.95]

NOTES: This exhibit presents, for the subsample of firms in the consumer sectors, the average excess returns of quintile portfolios formed based on different types of sales proxies. Monthly sales proxies are estimated from the growth rate of consumer activities during the most recent 3-month period over the past 12-month period. Quintile portfolios are formed based on the sales proxies at month *t*. Average returns of the portfolios, in excess of the market, then are calculated for months *t* + 1, *t* + 2, and *t* + 3. Alphas of high-minus-low portfolios are calculated using the Fama–French six factors. Panel A shows the results of equal-weighted portfolios, and Panel B reports the value-weighted portfolio results. The *t* statistics are reported in brackets. The sample period is 2009–2020.

Panel A shows that IN-STORE strongly predicts equal-weighted portfolio returns up to three months after the formation period. The high-minus-low portfolio returns and alphas are not only statistically significant but also economically sizable. For example, the high-minus-low portfolio of IN-STORE provides average returns of about 1.9% in months *t* + 1 and *t* + 2 (22.7% and 23.2% per annum, respectively). On the contrary, portfolios formed on BRAND and WEB do not provide strong returns in month *t* + 1. However, the high-minus-low of BRAND provides a significant alpha for a month after formation (*t* + 2). The alpha of the portfolio is 0.55% (6.6% per annum) with a *t*-statistic of 1.78, which is significant at the 10% level. No WEB portfolios display significance at any conventional level. These results imply that the information in WEB is likely to be quickly disseminated, whereas the BRAND information is incorporated into prices with a delay.

Panel B shows that the portfolio results using BRAND and WEB are generally stronger when the value-weighted method is used. Similar to equally weighted portfolio results, the value-weighted portfolios of IN-STORE provide significantly positive returns for up to three months. However, unlike Panel A, BRAND strongly predicts portfolio returns in months $t + 1$ and $t + 2$. For example, the high-minus-low portfolio of BRAND provides a significantly positive return of about 0.5% in month $t + 1$ (6.5% per annum) and 0.8% in month $t + 2$ (10.0% per annum). The value-weighted portfolio formed on WEB also generates a significant return for month $t + 1$. These results suggest that the information in BRAND and WEB may be more useful for predicting returns of large firms, indicating that consumer online activities, whether visiting websites or expressing interest in companies' brands, are more relevant to the sales of large firms.

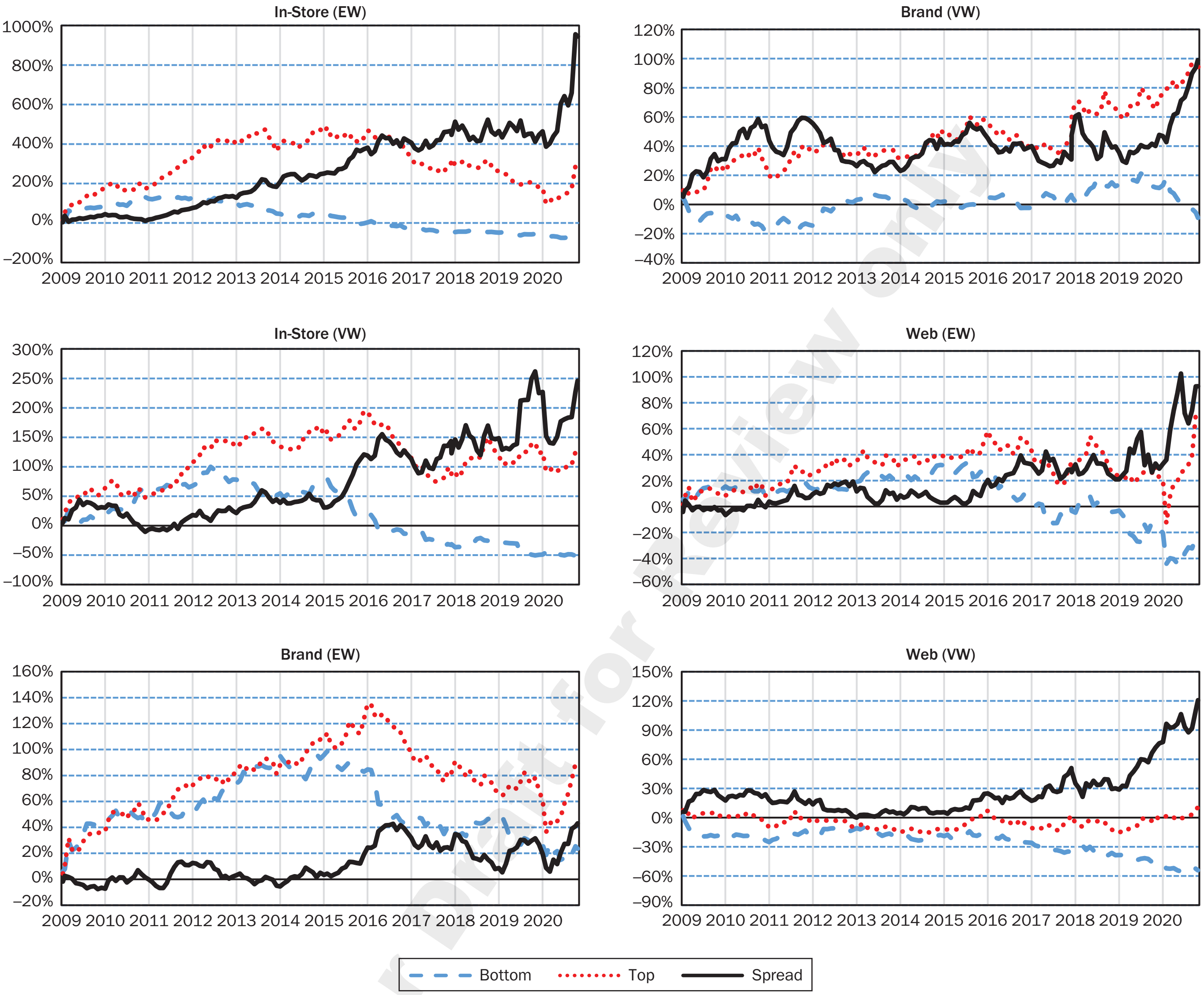
To visualize the overall economic magnitude of the portfolio strategies, we plot the cumulative excess returns of quintile portfolios for the consumer-sector subsamples. Specifically, Exhibit 4 shows the cumulative excess return of the highest (top) and the lowest (bottom) quintile portfolios, as well as the high-minus-low (spread) portfolios of the sales proxies. The first column shows the cumulative returns of equally weighted portfolios, and the second column plots the cumulative returns of value-weighted portfolios. The exhibit shows that the cumulative returns of the high-minus-low portfolio of IN-STORE are of significant magnitude. In addition, consistent with Exhibit 3, the cumulative return of the value-weighted high-minus-low portfolio formed on BRAND, about 100% over the sample period, is much larger than that of the equally weighted portfolio. Overall, the exhibit demonstrates sizable economic magnitudes of return for the portfolios formed based on the sales proxies.

The results in Exhibits 3 and 4 call into question whether the information contained in the sales proxies is disseminated in a timely manner. Specifically, Exhibit 3 shows that the portfolio strategy based on WEB tends to generate weaker results than portfolio strategies based on other proxies. In addition, although IN-STORE and BRAND significantly predict returns of month $t + 2$, WEB does not have any predictive power beyond month $t + 1$. Therefore, we conjecture that although the information contained in WEB is readily available to investors, the information in other proxies is disseminated with a delay. To examine this, we employ an event study approach and calculate the average cumulative returns for the 65-trading-day period starting the next trading day of portfolio formation each month. The event period coincides approximately with a three-month period after portfolio formation.

Exhibit 5 plots the daily cumulative returns of high-minus-low value-weighted portfolios formed based on the sales proxies. The *x*-axis indicates the number of trading days from portfolio formation, and the *y*-axis shows the daily cumulative returns. We calculate the daily cumulative returns up to 65 trading days, which roughly correspond to three months after formation. The exhibit shows that there are steady increases in cumulative returns for the portfolios formed based on IN-STORE and BRAND at up to three months, indicating that the information contained in those sales proxies is incorporated in stock prices with a delay. However, the WEB portfolio displays a distinct pattern compared with the other proxies. The increase in the cumulative return occurs mostly before day 30, and the cumulative return stays flat after that, indicating that the information in WEB is reflected in the price more quickly compared to other proxies. Overall, these results suggest that the information in WEB is accessible to investors relatively quickly, whereas the BRAND and IN-STORE information is disseminated to investors with a delay.

EXHIBIT 4

Cumulative Excess Returns of Portfolios Sorted by Sales Proxies



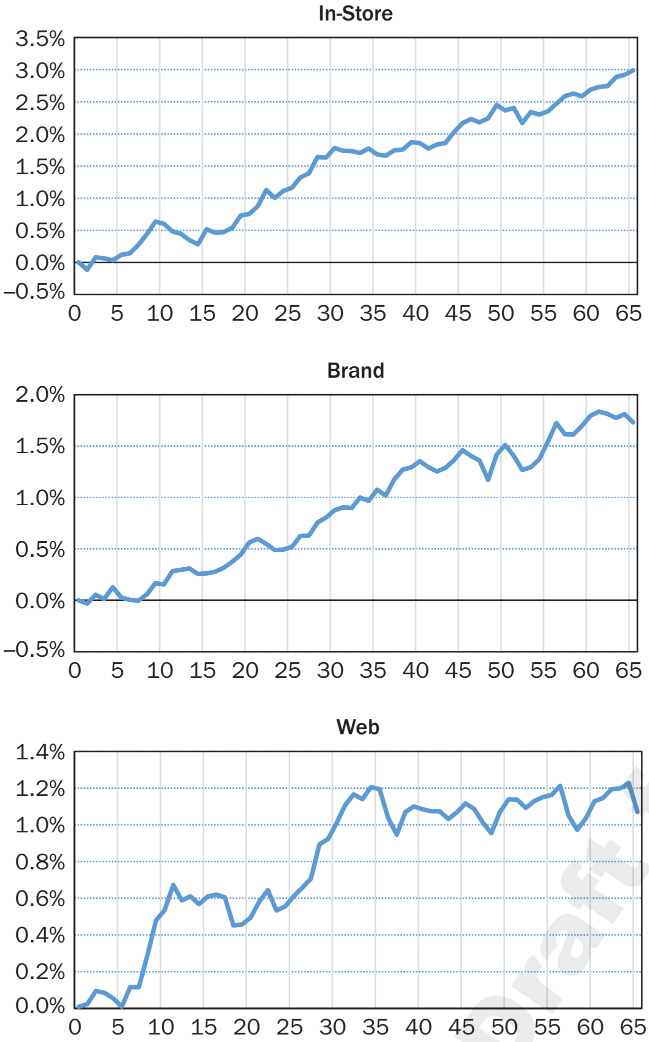
NOTES: This exhibit plots the cumulative excess returns of quintile portfolios formed based on the sales proxies for the subsample of firms in the consumer sectors. The blue dashed line plots the cumulative returns, in excess of the market, of the lowest quintile portfolio. The red dotted line represents the cumulative excess returns of the highest quintile portfolio. The black solid line indicates the cumulative returns from taking a long position on the highest quintile portfolios and a short position on the lowest quintile portfolios. Monthly sales proxies are estimated from the growth rate of consumer activities during the most recent 3-month period over the past 12-month period. Quintile portfolios are formed based on the sales proxies in month *t*. The portfolios then are held for *t* + 1. The first chart in each grouping shows the results of equally weighted (EW) portfolios, and the second reports the value-weighted (VW) portfolio results. The sample period is 2009–2020.

Profitability Net of Transaction Costs

The results in Exhibit 3 show that, in general, the portfolio strategy based on sales proxies provides a positive alpha. However, the results do not show whether actual trades net of transaction costs are profitable. Although investors have access to the information and resources necessary to analyze this, they may be discouraged from doing so if transaction costs are large enough to offset the potential trade profits.

EXHIBIT 5

Event Time—Daily Cumulative Returns



NOTES: This exhibit plots the daily cumulative returns of high-minus-low quintile portfolios formed based on the sales proxies for the subsample of firms in the consumer sectors. Monthly sales proxies are estimated from the growth rate of consumer activities during the most recent 3-month period over the past 12-month period. Value-weighted quintile portfolios are formed based on the sales proxies in month *t*. The daily cumulative returns for the high-minus-low quintile portfolios then are calculated from day one of month *t* + 1. The *x*-axis indicates the trading days since portfolio formation, and the *y*-axis shows the daily cumulative returns. The daily cumulative returns are calculated over 65 trading days, which roughly correspond to three months after portfolio formation. The sample period is 2009–2020.

We use the transaction-cost estimates provided by Frazzini, Israel, and Moskowitz (2015), who calculated actual trade execution costs separately along various dimensions, including firm size, transaction types (i.e., long and short positions), and firms' stock exchange listings. The estimated trade-execution costs are based on live trading data of a large institutional money manager over the period 1998 through 2013. Specifically, they estimated the average execution costs to be 11.21 bps for large-cap stocks, 21.27 bps for small-cap stocks, 14.79 for US-based long positions, 10.02 for US-based short positions, 8.81 for stocks listed on NYSE-Amex, and 11.44 for stocks listed on NASDAQ. Although Frazzini, Israel, and Moskowitz (2015) defined stocks that are included in the Russell 1000 Index as large-cap stocks, we use the S&P 500 Index as the threshold for large-cap stocks, thereby making our cost estimates more conservative.

Exhibit 6 estimates the average returns and Fama–French six-factor alphas of high-minus-low portfolio returns, net of transaction costs, for months *t* + 1, *t* + 2, and *t* + 3. As in Exhibit 3, we focus on the consumer-sector subsample. Panels A, B, and C report the results using the three different transaction-cost estimates (i.e., size, type, and exchange, respectively).

The exhibit shows that the results reported in Exhibit 3 are generally robust, albeit weaker for some specifications, to the transaction costs. The average returns and alphas of portfolios based on IN-STORE, net of transaction costs, are significantly positive for months *t* + 1 and *t* + 2 and are robust to different estimates of transaction costs. For example, Panel A shows that the equal-weighted IN-STORE portfolio for month *t* + 1 generates a 1.58% alpha (*t*-value of 2.51) when execution costs are estimated based on firm size. In addition, the value-weighted portfolios of BRAND provide significantly positive returns and alphas for month *t* + 2, robust to different transaction cost estimates. Overall, the results show that return predictability of sales proxies leads to profitable trades.

CONSUMER ACTIVITIES DURING THE PANDEMIC PERIOD

How has the COVID-19 pandemic affected consumer activities? The pandemic has claimed the lives of more than 600,000 people in the United States alone at the time of writing and has had significant impacts on economic activities and the everyday life of people worldwide. Businesses and schools are closed, and many people work from home or have lost their jobs. As people stay home, except for essential activities such as grocery shopping, the pandemic has had a significant

EXHIBIT 6

Performance Net of Transaction Costs

Sales Proxies		In-Store			Brand			Web		
Portfolios		*t* + 1	*t* + 2	*t* + 3	*t* + 1	*t* + 2	*t* + 3	*t* + 1	*t* + 2	*t* + 3
Panel A: Size										
EW	Return	1.53%	1.57%	1.16%	0.04%	0.20%	–0.19%	0.26%	0.19%	–0.09%
		[2.59]	[2.73]	[2.20]	[0.16]	[0.69]	[–0.58]	[0.67]	[0.58]	[–0.25]
	Alpha	1.58%	1.44%	1.02%	0.06%	0.29%	0.00%	0.06%	0.10%	–0.06%
		[2.51]	[2.46]	[1.77]	[0.20]	[0.93]	[–0.01]	[0.15]	[0.28]	[–0.16]
VW	Return	0.80%	0.86%	0.74%	0.31%	0.61%	0.12%	0.39%	0.26%	0.06%
		[1.49]	[1.66]	[1.46]	[1.10]	[1.90]	[0.35]	[1.28]	[0.81]	[0.19]
	Alpha	1.21%	0.85%	0.43%	0.23%	0.72%	0.23%	0.27%	0.25%	0.10%
		[2.12]	[1.53]	[0.80]	[0.76]	[2.19]	[0.68]	[0.85]	[0.72]	[0.29]
Panel B: Transaction Type										
EW	Return	1.64%	1.68%	1.27%	0.06%	0.21%	–0.18%	0.31%	0.25%	–0.03%
		[2.78]	[2.92]	[2.40]	[0.21]	[0.73]	[–0.54]	[0.83]	[0.76]	[–0.09]
	Alpha	1.68%	1.55%	1.13%	0.07%	0.30%	0.01%	0.12%	0.16%	0.00%
		[2.68]	[2.64]	[1.96]	[0.25]	[0.97]	[0.03]	[0.28]	[0.44]	[–0.01]
VW	Return	0.82%	0.89%	0.76%	0.29%	0.58%	0.10%	0.37%	0.25%	0.05%
		[1.54]	[1.70]	[1.51]	[1.02]	[1.82]	[0.28]	[1.23]	[0.76]	[0.14]
	Alpha	1.23%	0.88%	0.46%	0.21%	0.70%	0.21%	0.25%	0.23%	0.09%
		[2.17]	[1.57]	[0.85]	[0.68]	[2.12]	[0.61]	[0.80]	[0.68]	[0.24]
Panel C: Stock Exchanges										
EW	Return	1.67%	1.71%	1.30%	0.11%	0.26%	–0.13%	0.36%	0.29%	0.01%
		[2.83]	[2.97]	[2.46]	[0.38]	[0.90]	[–0.38]	[0.94]	[0.89]	[0.03]
	Alpha	1.72%	1.58%	1.16%	0.12%	0.35%	0.06%	0.16%	0.20%	0.04%
		[2.73]	[2.69]	[2.01]	[0.42]	[1.13]	[0.18]	[0.39]	[0.56]	[0.11]
VW	Return	0.87%	0.93%	0.81%	0.35%	0.64%	0.15%	0.43%	0.31%	0.11%
		[1.63]	[1.79]	[1.60]	[1.20]	[1.99]	[0.44]	[1.42]	[0.94]	[0.31]
	Alpha	1.28%	0.93%	0.50%	0.26%	0.75%	0.26%	0.31%	0.29%	0.14%
		[2.25]	[1.66]	[0.94]	[0.86]	[2.28]	[0.77]	[0.98]	[0.85]	[0.41]

NOTES: The exhibit estimates, for the consumer-sector subsample, the net-of-transaction-costs alpha of the high-minus-low portfolio strategy formed based on the sales proxies. We use the transaction cost estimates from Frazzini, Israel, and Moskowitz (2015). Specifically, transaction costs are estimated based on firm size, transaction type (long or short), and stock exchange. High-minus-low portfolios are constructed by taking a long position on the highest quintile portfolio and a short position on the low quintile portfolio, net of transaction costs. Alphas of high-minus-low portfolios are calculated using the Fama–French six factors. Monthly sales proxies are estimated from the growth rate of consumer activities during the most recent 3-month period over the past 12-month period. Quintile portfolios are formed based on the sales proxies at month *t*. The average returns of the portfolios then are calculated for months *t* + 1, *t* + 2, and *t* + 3. Panel A shows the portfolio alpha using the transaction cost estimate based on firm size, and Panels B and C use the transaction cost estimates based on transaction type and stock exchanges. The sample period is 2009–2020.

impact on consumer behavior as well. Therefore, in Exhibit 7, we investigate how the pandemic has affected consumer activities.

First, in Panel A, we compare various statistics of sales proxies for the pre-pandemic period with those for the pandemic period. The pandemic period begins in March 2020 and ends in December 2020. The pre-pandemic period is 2009 through February 2020. The exhibit reports significant differences in sales proxies between the pre-pandemic and pandemic periods. Whereas there is a significant decrease in IN-STORE, WEB displays a sharp increase during the pandemic. The decrease of 6.3% in IN-STORE is remarkable considering the pre-pandemic average is 7.2%. The increase in WEB by 1.3% is also economically sizable compared to the pre-pandemic average of 2.9%.

EXHIBIT 7
Pandemic Period

Panel A: Sales Proxies

Sample		Consumer Firms			Firms with All Three Signals		
Sales Proxy		**In-Store**	**Brand**	**Web**	**In-Store**	**Brand**	**Web**
Average	Pre (a)	7.16%	1.12%	2.89%	7.70%	2.08%	2.02%
	During (b)	0.85%	0.21%	4.22%	1.59%	0.50%	7.27%
	Diff (b – a)	–6.31%	–0.90%	1.33%	–6.11%	–1.58%	5.25%
	t-Value	[–4.87]	[–2.49]	[1.84]	[–3.47]	[–1.54]	[3.76]
Std Dev	Pre (a)	27.63%	10.48%	23.43%	26.38%	14.86%	20.82%
	During (b)	28.65%	17.00%	25.43%	30.33%	21.79%	24.74%
	Diff (b – a)	1.02%	6.52%	2.00%	3.95%	6.93%	3.92%
	F-Value	[1.08]	[2.63]	[1.18]	[1.32]	[2.15]	[1.41]
	P-Value	(0.261)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)

Panel B: Portfolio Characteristics

		In-Store		Brand		Web	
Sales Proxy		**EW**	**VW**	**EW**	**VW**	**EW**	**VW**
Returns	Pre (a)	1.48%	1.07%	0.18%	0.35%	0.29%	0.49%
	During (b)	7.26%	1.07%	1.97%	3.12%	4.15%	2.32%
	Diff (b – a)	5.77%	0.00%	1.79%	2.77%	3.86%	1.83%
	t-Value	[1.20]	[–0.00]	[0.95]	[2.81]	[1.35]	[1.06]
Std Dev	Pre (a)	5.98%	6.08%	3.07%	3.39%	3.96%	3.45%
	During (b)	15.09%	10.16%	5.92%	2.98%	8.99%	5.39%
	Diff (b – a)	9.11%	4.07%	2.85%	–0.41%	5.03%	1.95%
	F-Value	[6.36]	[2.79]	[3.71]	[0.77]	[5.16]	[2.45]
	P-Value	(0.000)	(0.004)	(0.000)	(0.656)	(0.000)	(0.010)

NOTES: This exhibit investigates how the pandemic affects the sales proxies and portfolio characteristics. Panel A compares various statistics of sales proxies for the pre-pandemic period with those for the pandemic period. Panel B examines the average returns and the standard deviations of high-minus-low portfolios formed based on the sales proxies for the pre-pandemic period and for the pandemic period. The pandemic period begins in March 2020 and ends in December 2020. The pre-pandemic period is 2009 to February 2020.

These differences between the two periods are even more striking when we restrict the sample to firms that have all three proxies. This restriction makes the changes in the different proxies caused by the pandemic more comparable. The result shows that the decrease in IN-STORE and the spike in WEB are of similar magnitude in absolute value, at 6.1% and 5.3%, respectively. This opposite pattern in IN-STORE and WEB clearly reflects the shift in consumer activity from physical stores to online sites during the pandemic. Interestingly, there is also a decrease in BRAND activities, albeit much smaller in magnitude. The decrease in BRAND accompanied by the increase in WEB is likely due to increased consumer focus on essential items and less focus on non-necessity brands during the pandemic.

Panel B examines the portfolio characteristics of the pre-pandemic period and those of the pandemic period. The average returns of high-minus-low portfolios tend to be larger during the pandemic, indicating that the return predictability of sales proxies is stronger during this period. Although the differences between the pre-pandemic and pandemic period are not statistically significant (with the exception of the value-weighted portfolio constructed based on BRAND), the economic magnitude of the differences is large. For example, the value-weighted high-minus-low portfolios constructed on WEB (BRAND) have an average monthly return that is 1.8% (2.8%)

larger than that during the pre-pandemic period.[3] The large economic magnitude of the differences is driven by firms in both the top and bottom quintiles, indicating that the information contained in the sales proxies is disseminated more slowly during the pandemic compared to the pre-pandemic period.

Overall, the results in Exhibit 7 show that there is a significant shift in consumer activities due to the pandemic, from physical stores to online venues. In addition, the predictability of the portfolio strategy using the sales proxies increases, indicating that the information in the sales proxies is incorporated in stock prices more slowly during the pandemic.

CONCLUSION

To study the power of big data in predicting firms' fundamentals and stock returns, we examine data sources that contain information on consumer activities, such as visits to retail stores, visits to firms' websites, and consumer interest in products and brand names. For the sample of roughly 330 firms and the time period of 2009–2020, we use these data sources to develop three sales proxies: IN-STORE, WEB, and BRAND.

Our analyses show that all three proxies predict firms' (unanticipated) revenue growth and earnings surprises, which suggests that the proxies contain value-relevant information about firms' fundamentals. However, the information content in each sales proxy reflects different aspects of fundamentals. IN-STORE and BRAND are significant predictors of revenue growth, whereas WEB provides relatively strong forecasts of earnings surprises. In addition, predictions are significantly stronger for companies in consumer sectors.

The sales proxies are also useful in predicting portfolio returns. However, when it comes to predicting company returns, WEB is considerably weaker than IN-STORE and BRAND. This seems consistent with the speed and ease with which WEB can be constructed in real time compared with IN-STORE and BRAND. We infer from this that WEB information is impounded more quickly into prices than IN-STORE and BRAND information. The profitability of the portfolio strategies based on the sales proxies is generally robust after transaction costs are considered.

Finally, we examine the sales proxies and portfolio results during the COVID-19 pandemic period. Although there is a significant decrease in IN-STORE during the pandemic period, WEB activities increase sharply, reflecting increased consumer online activity while staying home. In addition, the predictability of portfolio strategies based on the sales proxies increases, indicating that information dissemination during the pandemic is slower than in the pre-pandemic period.

REFERENCES

Bartov, E., L. Faurel, and P. Mohanram. 2015. "Can Twitter Help Predict Firm-Level Earnings and Stock Returns?" *The Accounting Review* 93: 25–57.

Cavallo, A. 2017. "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers." *American Economic Review* 107: 283–303.

Cavallo, A., and R. Rigobon. 2016. "The Billion Prices Project: Using Online Prices for Measurement and Research." *Journal of Economic Perspectives* 30: 151–178.

[3] The lack of statistical power is likely due to the short time period and the heightened market volatility during the pandemic period.

Chen, H., P. De, Y. Hu, and B. H. Hwang. 2014. “Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media.” *Review of Financial Studies* 27: 1367–1403.

Da, Z., J. Engleberg, and P. Gao. 2011. “In Search of Attention.” *The Journal of Finance* 66: 1461–1499.

Frazzini, A., R. Israel, and T. J. Moskowitz. “Trading Costs of Asset Pricing Anomalies.” Working paper, 2015.

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: 143–162.

Froot, K., X. Lou, G. Ozik, R. Sadka, and S. Shen. “Media Reinforcement in International Financial Markets.” Working paper, 2018.

Huang, J. 2018. “The Customer Knows Best: The Investment Value of Consumer Opinions.” *Journal of Financial Economics* 128: 164–182.

Katona, Z., M. O. Painter, P. N. Patatoukas, and J. Zeng. “On the Capital Market Consequences of Alternative Data: Evidence from Outer Space.” Working paper, 2020.

Kolanovic, M., and R. T. Krishnamachari. “Big Data and AI Strategies: Machine Learning and Alternative Data Approach to Investing.” J.P. Morgan, 2017.

Lee, M. “Hedge Funds Now Embracing Emergent Technology.” 2017, https://www.ey.com/en_gl/wealth-asset-management/how-will-you-use-innovation-to-illuminate-competitive-advantages/.

Ozik, G., and R. Sadka. 2013. “Big Data and Information Edge.” *Hedge Funds Review* (December 2013/January 2014): 32–34.

Zhu, C. 2019. “Big Data as a Governance Mechanism.” *Review of Financial Studies* 32: 2021–2061.