On the Capital Market Consequences of Big Data: Evidence from Outer Space

Zsolt Katona, Marcus O. Painter, Panos N. Patatoukas, and Jean Zeng

ONLINE SUPPLEMENT

Appendices

- **APPENDIX A:** Background on Remote Sensing Technology.

- **APPENDIX B:** Variable Definitions.

Supplemental Tables

- **TABLE A1:** Satellite Coverage by Company.

- **TABLE A2:** Forward-Looking Content of Satellite Data: Alternative Growth Rate.

- **TABLE A3:** Forward-Looking Content of Satellite Data: Alternative Return Window.

- **TABLE A4:** Comparison of Covered and Control Firms.

- **TABLE A5:** More Evidence on the Effect of Satellite Data on Individual Trading.

APPENDIX A
Background on Remote Sensing Technology

Mounting cameras to take pictures of the surface of the earth was the driving force behind early satellite launches. While the original purpose was oriented towards military applications and weather forecasting, it was not long before the first applications in economics research. Unlike most communication satellites that follow a geostationary orbit (at about 36,000km altitude) and remain in a fixed point above the equator relative to the surface of the earth, the satellites of interest to us orbit the earth at lower altitudes.

These remote sensing satellites typically provide full coverage of the earth's surface. The first publicly available data set originated from the U.S. Air Force's Defense Meteorological Satellite Program (DMSP) and NASA's Landsat system. The spacecraft in this program orbit the earth at altitudes around 700km and take advantage of the smaller distance and the different orbital characteristics to produce higher resolution images. Because of the lower orbit, these satellites move fast above the surface and orbit the earth about every 99 minutes or over fourteen times a day. The near polar orbits are set up such that they miss the poles only by few degrees and move mostly in a northerly/southerly direction taking images of the surface in "vertical" strips. Furthermore, the orbits are designed to be "sun-synchronous" such that the satellite passes a given latitude the same time of the day, every day. Since the earth rotates under the orbit, the cameras record a different strip of the surface on each revolution. Combining the different strips results in full coverage of the surface, where each point is covered at least once a day, at the same time of the day. With multiple satellites in a system the frequency can be increased.

RS Metrics sources raw satellite imagery from companies such as DigitalGlobe Inc., a division of Maxar Technologies and Airbus Defense and Space, formerly known as the European Aeronautic Defense and Space Company (EADS). These companies provide raw satellite images using their low orbit satellite constellations. For example, EADS launched the Pleiades 1A & 1B satellites as part of a new constellation in December 2011 and December 2012, respectively. Both satellites share the same sun-synchronous orbit, 180 degrees apart at an altitude of 694 kilometers with an orbital period of 99 minutes. The orbits are designed so that at least one of the satellites crosses over a given latitude/longitude at roughly the same local time every day. Each satellite photographs a north-south oriented swath of the surface of the Earth, with each swath shifting in the direction opposite to the rotation of the earth. Given the wide viewing angle and the resulting over 1 million square kilometers per day coverage capacity, the constellation provides daily revisit of each point at around the same local time. The satellites have very high-resolution cameras that provide a 0.5m resolution panchromatic and pan-sharpened multispectral images that capture a large part of the electromagnetic spectrum. This level of resolution makes it possible to measure parking lot traffic at the individual store-level.

APPENDIX B
Variable Definitions

Key Variables	
ΔSSS_{iq}	Year-over-year growth in domestic same-store sales. We obtained information on quarterly realizations of same-store sales from FactSet Fundamentals.
$\Delta FLRT_{iq}$	Year-over-year growth in parking lot fill rates. We obtained daily store-level information on parking lot traffic and capacity from RS Metrics.
$QRET_{iq}$	Buy-and-hold size and book-to-market adjusted stock return cumulated from the beginning to the end of quarter q. We obtained stock return data from CRSP.
$EARET_{iq}$	Buy-and-hold size and book-to-market adjusted stock return centered on the earnings announcement. We obtained stock return data from CRSP. We use the portfolio data from Kenneth French's website to calculate factor-adjusted returns.
$\Delta SHORT_{iq}^{Dem}$	The cumulative change in lender quantity on loan from the end of the quarter to two days before the earnings announcement divided by the number of shares outstanding in the company. We obtained daily data on lender quantity on loan from Markit.
$\Delta SHORT_{iq}^{Sup}$	The cumulative change in active supply of lendable shares from the end of the quarter to two days before the earnings announcement divided by the number of shares outstanding in the company. We obtained daily data on active lendable quantity from Markit.
$IndOIB_{iq}$	Individual order imbalance over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual order imbalance during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual order imbalance as buys minus sells divided by the total number of shares outstanding.
$IndBUYS_{iq}$	Individual buying activity over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual buying activity during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual buying activity as individual buys divided by the total number of shares outstanding.
$IndSELLS_{iq}$	Individual selling activity over the pre-earnings announcement window from the end of the quarter to two days before the earnings announcement adjusted for the average individual selling activity during the quarter. We use Boehmer's et al. (2021) method to identify individual trades and measure the daily individual selling activity as individual sells divided by the total number of shares outstanding.

$Spread_{iq}$	Effective spread defined as $2 \times (\|P_k - M_k\|)$, where P_k the price on a trade and M_k is the midpoint of the National Best Bid and Offer (NBBO) quotes for that trade. We measure the average effective spread from the end of the quarter to two days before the earnings announcement. Intraday trading data comes from the TAQ database.
$Price\ Impact_{iq}$	Price impact defined as $2 \times D_k(M_{k+5} - M_k)$, where M_k is the midpoint of the NBBO quotes on trade k, M_{k+5} is the midpoint five minutes after M_k, and D_k is equal to $+1$ if we identify the trade as buyer initiated and -1 if it is seller initiated. We follow Chakrabarty et al. (2006) to identify buyer versus seller-initiated trades. We measure the average price impact from the end of the quarter to two days before the earnings announcement. Intraday trading data comes from the TAQ database.
$Price\ Speed\ Ratio_{iq}$	Price speed ratio defined as the cumulative factor-adjusted return between days $t-10$ and $t-2$ divided by the cumulative factor-adjusted return between days $t-10$ and $t+1$, where day zero is the earnings announcement day.
$Price\ Jump\ Ratio_{iq}$	Price jump ratio defined as the cumulative factor-adjusted return between days $t-1$ and $t+1$ relative to day zero divided by the cumulative factor-adjusted return between days $t-10$ and $t+1$, where day zero is the earnings announcement day.

Other Characteristics	
$ln(Market\ Cap)$	Natural logarithm of total market capitalization.
$Tobin's\ Q$	Ratio of the sum of market value of equity plus the book value of long-term and short-term debt divided by the sum of book value of equity plus the book value of long-term and short-term debt
$Institutional\ Owners$	Percentage of shares outstanding held by institutions that manage over $100 million and report their quarterly holdings on SEC Form 13F and N-30Ds.
$Big\ 4\ Auditor$	Indicator variable that is equal to one if Deloitte, Ernest & Young, KPMG, or PwC audits the firm.
$Acquisition$	Indicator variable that is equal to one if the firm reported acquisition costs
$Restructuring$	Indicator variable that is equal to one if the firm reported restructuring costs.
Write Down	Indicator variable that is equal to one if the firm reported goodwill impairments or other asset write-offs.
$I(HVLT_{iq})$	Indicator variable equal to one if a firm's stock return volatility is above the median. We measure volatility as the standard deviation of stock returns for the past twelve months prior to the quarter.
$I(SMALL_{iq})$	Indicator variable equal to one if a firm's market cap is below the median market cap of all firms in that quarter.
$I(YOUNG_{iq})$	Indicator variable equal to one if a firm's age is below the median. We measure age relative to the firm's founding year using data available from Jay Ritter's website.

TABLE A1
Satellite Coverage by Company

This table reports information about the average store count and satellite coverage for each of the forty-four U.S. companies in our sample along with the starting date of RS Metrics coverage.

	Company Name	GICS Industry	Store Count	Coverage	Starting Date
1	Bed Bath & Beyond Inc. (BBBY)	Specialty Retail	1,468	46%	2011:Q3
2	Best Buy Co., Inc. (BBY)	Specialty Retail	2,403	38%	2011:Q3
3	Big 5 Sporting Goods Corporation (BGFV)	Specialty Retail	433	79%	2013:Q4
4	Big Lots, Inc. (BIG)	Multiline Retail	1,491	69%	2012:Q4
5	BJ's Restaurants, Inc. (BJRI)	Hotels, Restaurants & Leisure	170	77%	2013:Q4
6	Buffalo Wild Wings, Inc. (BWLD)	Hotels, Restaurants & Leisure	1,077	60%	2012:Q2
7	Burlington Stores, Inc. (BURL)	Specialty Retail	587	67%	2016:Q1
8	Cabela's Incorporated (CAB)	Specialty Retail	66	64%	2013:Q1
9	CarMax, Inc. (KMX)	Specialty Retail	177	81%	2016:Q4
10	Chipotle Mexican Grill, Inc. (CMG)	Hotels, Restaurants & Leisure	1,830	56%	2012:Q2
11	Conn's, Inc. (CONN)	Specialty Retail	108	79%	2015:Q2
12	Costco Wholesale Corporation (COST)	Food & Staples Retailing	741	40%	2017:Q4
13	Dick's Sporting Goods, Inc. (DKS)	Specialty Retail	769	61%	2015:Q2
14	Dillard's, Inc. (DDS)	Multiline Retail	293	66%	2016:Q4
15	Dollar General Corporation (DG)	Multiline Retail	12,246	33%	2013:Q2
16	Dollar Tree, Inc. (DLTR)	Multiline Retail	11,448	46%	2014:Q2
17	El Pollo Loco Holdings, Inc. (LOCO)	Hotels, Restaurants & Leisure	460	86%	2016:Q2
18	Home Depot, Inc. (HD)	Specialty Retail	2,264	61%	2011:Q1
19	J. C. Penney Company, Inc. (JCP)	Multiline Retail	1,062	66%	2011:Q4
20	Kohl's Corporation (KSS)	Multiline Retail	1,158	69%	2011:Q4
21	Kroger Co. (KR)	Food & Staples Retailing	3,892	51%	2016:Q1

22	Lowe's Companies, Inc. (LOW)	Specialty Retail	1,862	68%	2011:Q1
23	Lumber Liquidators Holdings, Inc. (LL)	Specialty Retail	361	72%	2013:Q3
24	Macy's Inc. (M)	Multiline Retail	855	30%	2013:Q1
25	Monro Inc. (MNRO)	Specialty Retail	1,085	42%	2012:Q3
26	Nordstrom, Inc. (JWN)	Multiline Retail	352	42%	2016:Q4
27	Panera Bread Company (PNRA)	Hotels, Restaurants & Leisure	1,821	59%	2011:Q4
28	Party City Holdco, Inc. (PRTY)	Specialty Retail	929	71%	2016:Q2
29	PetSmart, Inc. (PETM)	Specialty Retail	1,320	63%	2012:Q2
30	Pier 1 Imports, Inc. (PIR)	Specialty Retail	1,035	72%	2014:Q4
31	Ross Stores, Inc. (ROST)	Specialty Retail	1,491	63%	2015:Q1
32	Safeway Inc. (SWY)	Food & Staples Retailing	1,371	63%	2013:Q2
33	Sears Holdings Corporation (SHLDQ)	Multiline Retail	1,693	74%	2014:Q2
34	Sherwin-Williams Company (SHW)	Chemicals	4,202	48%	2012:Q3
35	Smart & Final Stores, Inc. (SFS)	Food & Staples Retailing	311	77%	2016:Q4
36	Staples, Inc. (SPLS)	Specialty Retail	2,067	43%	2011:Q4
37	Starbucks Corporation (SBUX)	Hotels, Restaurants & Leisure	21,942	8%	2012:Q2
40	TJX Companies Inc. (TJX)	Multiline Retail	3,785	38%	2016:Q1
38	Target Corporation (TGT)	Specialty Retail	1,819	71%	2011:Q3
39	The Container Store Group, Inc. (TCS)	Specialty Retail	80	66%	2014:Q4
41	Tractor Supply Company (TSCO)	Specialty Retail	1,412	46%	2012:Q1
42	Ulta Beauty Inc. (ULTA)	Specialty Retail	781	64%	2012:Q3
43	Walmart Inc. (WMT)	Food & Staples Retailing	10,957	18%	2011:Q1
44	Whole Foods Market, Inc. (WFM)	Food & Staples Retailing	444	61%	2015:Q2

TABLE A2
Forward-Looking Content of Satellite Data:
Additional Evidence Using Growth in Total Parking Lot Fill Rates

This table provides evidence that our results are unchanged when we replace growth in same-store parking lot fill rates (ΔSSS_{iq}) with overall growth without conditioning on same-store comparisons (ΔSSS^*_{iq}). Panel A reports results from the baseline linear regression model specification. Panel B reports results from the alternative specification that allows for a different coefficient on negative and positive values of growth in same-store parking lot fill rates. We report regression results using the standardized z-values of the continuous predictors. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests.

Panel A: Baseline Specification.

	Dependent Variable $= \Delta SSS^*_{iq}$		
	(1)	(2)	(3)
$\Delta FLRT_{iq}$	0.007**	0.007**	0.008***
	(3.35)	(3.59)	(4.82)
ΔSSS^*_{iq-1}	0.045***	0.044***	0.039***
	(16.46)	(18.23)	(13.23)
$QRET_{iq}$	.	0.007***	0.007***
		(3.77)	(8.47)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adjusted R²	68.7%	70.2%	71.7%
OBS.	650	650	650

Panel B: Alternative Specification.

	Dependent Variable $= \Delta SSS^*_{iq}$		
	(1)	(2)	(3)
$\Delta FLRT^-_{iq}$	0.012**	0.012**	0.012*
	(2.89)	(2.99)	(2.31)
$\Delta FLRT^+_{iq}$	0.006**	0.006**	0.006**
	(2.57)	(3.31)	(2.98)
ΔSSS^*_{iq-1}	0.044***	0.043***	0.038***
	(17.11)	(18.89)	(12.83)
$QRET_{iq}$	.	0.007**	0.007***
		(3.65)	(8.11)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adj. R²	69.2%	70.6%	71.8%
OBS.	650	650	650

TABLE A3
Forward-Looking Content of Satellite Data:
Additional Evidence Expanding the QTR Return Cumulation Window

This table provides evidence that growth in same-store parking lot fill rates predicts growth in same-store retailer sales. Panel A reports results from the baseline linear regression model in equation (1). Panel B reports results from the alternative specification that allows for a different coefficient on negative and positive values of growth in same-store parking lot fill rates. We report regression results using the standardized z-values of the continuous predictors. The sample includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, using two-tailed tests.

Panel A: Baseline Specification.

	Dependent Variable $= \Delta SSS_{iq}$		
	(1)	(2)	(3)
$\Delta FLRT_{iq}$	0.008**	0.008***	0.008***
	(3.46)	(3.84)	(4.29)
ΔSSS_{iq-1}	0.044***	0.043***	0.039***
	(15.39)	(17.04)	(13.09)
$QRET_{iq}^*$	.	0.008***	0.008***
		(4.28)	(6.91)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adjusted R²	69.1%	71.1%	72.2%
OBS.	650	650	650

Panel B: Alternative Specification.

	Dependent Variable $= \Delta SSS_{iq}$		
	(1)	(2)	(3)
$\Delta FLRT_{iq}^-$	0.017***	0.016***	0.014**
	(5.57)	(4.86)	(3.48)
$\Delta FLRT_{iq}^+$	0.006*	0.006**	0.006**
	(2.33)	(2.79)	(2.60)
ΔSSS_{iq-1}	0.043***	0.042***	0.038***
	(18.63)	(20.01)	(14.86)
$QRET_{iq}^*$	.	0.008***	0.007***
		(4.29)	(8.05)
Characteristic Controls	No	No	Yes
Firm Fixed Effects	No	No	Yes
Quarter Fixed Effects	No	No	Yes
Adj. R²	70.0%	71.7%	72.6%
OBS.	650	650	650

TABLE A4
Comparison of Covered and Control Firms

This table presents descriptive statistics, measured at the last quarter before the satellite coverage for the treated and control groups: ln(Market Cap) is the natural log of market cap; Tobin's Q is the ratio of the sum of market value of equity plus the book value of long-term and short-term debt divided by the sum of book value of equity plus the book value of long-term and short-term debt; Institutional Ownership is the percentage of stock owned by institutional investors; Big-4 Auditor is an indicator variable that is equal to one if Deloitte, Ernest & Young, KPMG, or PwC audits the firm; Acquisition is an indicator variable that is equal to one if the firm reported acquisition costs; Restructuring is an indicator variable that is equal to one if the firm reported restructuring costs; Write Down is an indicator variable that is equal to one if the firm reported goodwill impairments or other asset write-offs. The sample of treated and matched-control firms includes 4,420 observations with pre- and post-treatment coverage. ***, **, * denote significance at the 1%, 5%, and 10% level for a two-tailed test, respectively.

	Covered Firms		Control Firms		Dif. of Means	t-stat
	Mean	Median	Mean	Median		
ln(Market Cap)	8.82	8.90	7.23	7.36	1.60	5.53***
Tobin's Q	3.51	2.38	2.64	2.04	0.87	1.77*
Institutional Ownership	85.9%	87.8%	77.8%	79.6%	8.10%	2.41**
Big-4 Auditor	97.7%	100.0%	82.0%	100.0%	15.70%	3.13***
Acquisition	2.27%	0.00%	5.70%	0.00%	-3.43%	-1.01
Restructuring	0.00%	0.00%	0.88%	0.00%	-0.88%	-1.08
Write Down	2.27%	0.00%	6.14%	0.00%	-3.87%	-1.09

TABLE A5
Difference-in-Differences: More Evidence on the Effect of Satellite Data on Individual Trading

This table provides evidence that the informativeness of individual buying activity decreased after the introduction of satellite data. We report coefficient estimates based on the DID regression model in equation (3). Column 1 repeats our baseline results signing trades based on the sub-penny digit method of Boehmer et al. (2021), or BJZZ. Column 2 reports results signing trades based on the quote midpoint method of Lee and Ready (1991). Our current data access to TAQ is limited to the post-2009 period, thus the quote midpoint test does not include the earlier pre-treatment years. Columns 3 and 4 of Table R5 reports results focusing on firms with spreads below 10 cents and 5 cents, respectively. Column 5 reports results after eliminating stocks included in SEC's Tick Size Pilot (TSP). The treated group of retailers includes 650 firm-quarter observations from 2011:Q1 to 2017:Q4. We report regression results using the standardized z-values of the continuous predictors. For each retailer in the treated group, we use a symmetric event window before and after the initiation of satellite coverage. The matched control group includes an average of 2.4 size-matched competitors per retailer that operate in the same six-digit GICS industry and do not have satellite coverage. We report t-statistics in parentheses based on clustered standard errors by time. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively, based on two-tailed tests.

	Dependent Variable $=$ $EARET_{iq}$				
	BJZZ algorithm	Quote Midpoint	$< .1\ Spread$	$< .05\ Spread$	w/o TSP stocks
$IndBUYS_{iq}$	-0.013**	-0.003	-0.013*	-0.015**	-0.013**
	(-2.44)	(-0.45)	(-2.06)	(-2.61)	(-2.43)
$POST_{iq} \times IndBUYS_{iq}$	0.013*	0.011	0.014*	0.018**	0.012*
	(2.01)	(1.02)	(1.90)	(2.60)	(1.89)
$TREAT_{iq} \times IndBUYS_{iq}$	0.019**	0.019**	0.020***	0.023***	0.018***
	(3.51)	(2.56)	(3.65)	(3.68)	(3.36)
$POST_{iq} \times TREAT_{iq} \times IndBUYS_{iq}$	-0.024***	-0.029**	-0.027***	-0.032***	-0.022***
	(-3.76)	(-2.32)	(-3.97)	(-4.12)	(-3.40)
Characteristics (C_{iq})	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects (θ_i)	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects (δ_q)	Yes	Yes	Yes	Yes	Yes
Adj. R²	6.9%	6.3%	6.8%	7.4%	6.6%
OBS.	3,877	2,739	3,298	2,960	3,820