ONLINE APPENDIX TO

"THE USE AND USEFULNESS OF BIG DATA IN FINANCE"

Figure A1
List of Alternative Data Vendors and In-house Data Science Teams

We compile a list of data-science teams and alternative-data vendors by combining the vendor list of AlternativeData.org, a platform that connects users to providers of alternative data, with that of J. P. Morgan's 2019 Alternative Data Handbook. The figure below lists all the seven in-house data-science teams and all the 513 alternative-data vendors. *denote in-house data-science teams.

AlphaWise (Morgan Stanley)*
Barclays Investment Sciences and Data Science Team (Barclays)*
Piper Jaffray Web Analytics (PiperJaffray, now Piper Sandler Companies)*
RBC Elements (Royal Bank of Canada)*
UBS Evidence Lab (UBS)*
Wolfe quant team (Wolfe Research)*
Kyber Data Science (Cowen)*

1010Data
7Park
Aberdeen
Accern
Accrete
Aclima
Acuris
AddThis
Advan
Affinity Solutions
AggData
Agribotix
Agricultural Research Federation
Airports Council International
AirSage
ALASA
Alexandria
AllTheRooms
Almax Information Systems
Alpha Hat
AlphaFlow
AlphaLetters
Alphamatician
Alphasense
Alt Hub
Alternate DNS
Amareos
Amass Insights
Amenity Analytics
American Trucking Association
Ampere Analysis
Anonymous Provider
AnthemData
Apertio Technologies
ApexData
AppAnnie
Applaudience
Apptopia
Arab Air Carrier Organization
Arabesque S Ray
ARC
Arch Metrics
AreaMetrics
ARM Insight
Ascend Worldwide Limited
Astutex
Audit Analytics
aWhere
Barchart
BayStreet Research
Beijing Chuang Yi Fang Technology
Beijing UC Science & Technology
Benzinga
Big Byte Insights
Bird.i
Bitly
Bitvore
BizQualify
Black Box (TDn2k)
Black Sky
Bloomberg Tesla Tracker
BMLL Technology
Bombora
Borrell
Boxoffice Media
Brain Company
BrandLoyalties
BrandWatch
Brave New Coin
Brickstream
Bridg
Broughton Capital
Buddy
BuildFax
BuiltWith
Business Intelligence Advisors
Business Monitor International
Capella Space
CB Richard Ellis Inc.
CDU-TEK: Central Dispatching Department of Fuel Energy Complex of Russia
Chain Store Guide Information Services
ChemOrbis
China National Chemical Information Center
China Real Estate Information Corporation
Civic Science
ClipperData
CogniSent
Comlinkdata
CompStak
ComScore
Consumer Edge
Cooltrader
CQG
Crain Communications Inc.
CreditRiskMonitor
Crimson Hexagon
Cropnosis
CropProphet
CrowdThnk
Cruise Analytics
Cuebiq
Cuemacro
CyberStream
Data Guru Limited
Data Simply
Datalogix
Dataminr
Datamyne
Dataprovider.com
DataPulse
Datarama
DataSift
Datastoxx
DataStreamx
DataTrek
DataWeave
DataYes
Dawex
DecaData
DeepAffects
Del Mar Networks
Delphia
DemystData
Descartes Labs
Digital Globe
DigitalMR
Doane Advisory Service
Dodge
Drawbridge
Drewry Shipping Consultants Ltd
Drillinginfo
DroneDeploy
Dun & Bradstreet
EagleAlpha
Earnest Research
Earthcube
EcommerceDB
Edison
Edmunds
EEDAR
Eilers & Krejcik Gaming
Emolument
Endor
EnerKnol
ENGAGE Research
Enigma
Entgroup
EntSight
EODData
EPFR
Epsilon
eSignal
Estimize
Eurekahedge
Euromonitor International
Event Registry
EventVestor
Everest Group
Exante Data
Exerica
Experian Footfall
ExtractAlpha
FactSet Revere
FactSquared
Fashionbi
FastBooking
FeatureX
FHS - Swiss Watch Data
Finweavers
First Data Merchant Services Corporation
First Data SpendTrend
First to Invest
Flexport
FN Arena
FNGO
Foursquare
Fraud Factors
Freestyle Media
FreightWaves
FTR Freight Transport Research Associates
Fysical
GDELT
Genscape
Geocento
GeoQuant
GeoSpark Analytics, Inc
Geospatial Insight
Geotab
GeoWiki
GfK Boutique Research

Global Tone Communication (GTCOM)
GNIP
Good Judgment
GovSpend
Grandata
Granular.ai
Grapedata
Greenwich.HR
Gro Intelligence
GroundTruth (xAd)
GS Dataworks
Guidepoint
Gyana
h2o
Headset
Health Forum
HealthVerity
Heckyl
HFR
Hillside Partners
humanpredictions
Huq Industries
HySpecIQ
ICEYE
ICI
IFI CLAIMS Patent Services
iiMedia Research
IMS Quintiles
Index Marketing Solutions Limited
IndexMath
Inferess
InformaFinancialIntelligence
InfoTEK Publishing House
InfoTrie
Innovata
Inovayt
Insights Data Solutions
InSpectrum
Intelius
Interconnect Analytics
Intermodal Association of North America
International Data Corporation Inc.
Internet Truckstop
Intrinio
Investing.com
IPqwery
iResearch
Irisys
iSentia
iSentium
iSpot
ISS Analytics
ISSB Ltd
Jettrack.io
Jiguang
Jumpshot
JustData
JWN Energy
Kayrros
KD Interactive
Knowsis
Kpler
ktMINE
Kyber Data Science
Landsat on AWS:
Legal Shield
Legis
Lexalytics
LikeFolio
LIMRA
LinkUp
LISTedTECH
ListenFirst
Lota Data
Lucena Research
Lyra Insight
M Science
Magna Global Research
Manfredi & Associates
Manheim
MariData
MarineTraffic
Marinexplore
MarketCheck
MarketPsych
Marketscout Corporation
MASSIVE Data Heights
MasterCard Advisors
MatterMark
Mavrx
Measurable AI
MedMine
Meltwater
Metricle
MIDiA Research
Millennium Research Group Inc.
MixRank
MKT Mediastat
Mobiquity Networks
Money Dashboard
MoneySuperMarket
NAIP
Narrative.io
New Generation Research
Newscred
Newswhip
Nexant Inc.
NEXRAD on AWS
NIC
Nikkei
Nowcast
NPD
Off-Highway Research Limited
Omega Point: a PM platform with AI intelligence
Omney Data
One Click Retail
OpenCorporates
OpenSignal
OpenstreetMap
Optimum Complexity
Orb Intelligence
Orbital Insight
OTAS
Ovum Ltd Us Branch
Owl Analytics
Pacific Epoch (China)
Panjiva
Panvista Analytics
Parsely
PatentSight
PatSnap
Paynxt360
Percolata
PipeCandy
Pitchbook
PlaceIQ
Placemeter
Placer.ai
Planet Labs
Pluribus Labs
Prattle
Predata
Predict HQ
Premonition
PriceStats
PROME
Prosper Insights & Analytics
PsychSignal
QL2
Quad Analytix
Quandl
Quantcube
Quantxt
Quest Offshore
QuestMobile
Quexopa
Rakuten Intelligence
RandomWalk
RavenPack
Real Capital Analytics
Real Estate Data
Realrents
Realyse
Re-analytics
Redbook Research Inc.
RedTech
REIS
RelateTheNews
RelationshipScience
RepRisk
Repustate
RetailNext
Return Path
Reveal Mobile
Revelio Labs
Reviewshake
Rezatec
Rigdata
RigLogix
Rigup
Rook Research
RootMetrics
RS Metrics
RunningAlpha
RVIA
RxData.net
Rystad Energy
Safegraph
Sandalwood
Satellite Imaging Corporation
SatScout
Savvr
SciDex Alpha
Scoop Analytics
Scrapehero
Scutify
Second Measure
Seer Aerospace
Selerity

Semiconductor Equipment & Materials International
Semlab
Sense360
Sensor Tower
Sentifi
Sentiment Trader
Sequentum
SESAMm
Sg2 (MarketPulse)
Sharablee
ShareIQ
ShareThis
ShareThis, Inc.
ShopperTrak
Shoppertrak Rct Corporation
Sigmai
Signal.co
SimilarWeb
SJ Consulting Group Inc.
Sky Watch
Skydeo
Slice Intelligence
Slingshot Aerospace
SmarterWorks
SMB Intelligence
Smith Travel
SNL Kagan
Social Alpha
Social Market Analytics
Space Know
SpaceKnow
Spacelist
SpaceNet on AWS
Spire Global
Spring Pond Partners
Standard Media Index
Statistical Survey
Statlas
Stax
Steel Orbis
StockTwits
STR
StreetLight Data
Suburbia
SumZero
SuperData
SuperFly
Superfly insights
Sustainalytics
Suzy
T.H. Capital
Tailwind Imaging
Tala
Talismatic
TalkingData
Tecnon Orbichem
Tegus
TellusLabs
Teragence
Terra Bella
Terrain Tiles
TerraQuanta
Thasos
The Climate Corporation
The Fertilizer Institute
TheySay
Thinknum
ThinkTopic
TickerTags
Tipigo
Tipranks
TMT Analysis
Towergate Informatics
Trackur
Tradesparq
TransCore
Transport Topics Publishing Group
Trendeo
Tribe Dynamics
Triton Research
TrustData
TrustedInsight
TruValue Labs
Tussell
TVeyes
TXN
TYR Data
Uber Media
Umbra Lab
Unacast
Understory
Unmetric
Upswell Group
Ursa
Urthecast
Venpath
Verbatim Advisory Group
Veronis Suhler Stevenson
Vertical Knowledge
Verto Analytics
Vessel Finder
VesselsValue
Vestdata
VidaMinds
Vigilant
Vortexa
Wall Street Horizon
Wards Automotive Group
Waste Analytics
WDZJ.com
Webhose.io
Wikimapia
Windward
Woodseer
World View
WXshift
Xebral
X-mode
Yewno
YipitData
Yodlee / Envestnet
Zaoshu.io
Zephyr
Zhiwei Data

Table A1

Number and Fraction of Firms by Industry: Our Sample versus the CRSP/Compustat Universe

In this table we present the numbers of firms in our sample by Global Industry Classification Standard (GICS) industry sector, the fractions of firms that are in the corresponding GICS industry sectors, the numbers of firms in the CRSP/Compustat universe by GICS industry sector, the fractions of firms that are in the corresponding GICS industry sectors, and the combined market values of the firms in our sample as a percentage of the combined market values of all firms in the CRSP/Compust universe by GICS industry sector. Our sample contains all firms in the Dow Jones Industrial Average Index from June 1 2009 through May 31 2019.

	Our Sample	%	CRSP/Compustat Universe	%	$\sum$ Market Value $_{\text{Our Sample}}$ / $\sum$ Market Value $_{\text{Crsp/Compustat}}$
Energy	2	6%	362	8%	17%
Materials	2	6%	261	5%	9%
Industrials	5	14%	577	12%	17%
Consumer Discretionary	3	9%	519	11%	11%
Consumer Staples	5	14%	166	3%	31%
Health Care	4	11%	882	18%	22%
Financials	5	14%	816	17%	13%
Information Technology	6	17%	632	13%	40%
Communication Services	3	9%	220	5%	16%
Utilities	0	0%	107	2%	0%
Real Estate	0	0%	234	5%	0%

Table A2

How Much Incremental Insight Is There in Alternative Data? Using Absolute Forecast Error

This table replicates Table 3, but the dependent variable is now the absolute forecast error of analyst *i* predicting earnings of firm *j*, scaled by the absolute value of the actual earnings, multiplied by (-1). We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and year-month levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
I(Alternative Data)	0.013**	
	(3.98)	
I(Category = App Usage)		0.020***
		(2.45)
I(Category = Sentiment)		0.011*
		(1.75)
I(Category = Employee)		0.005
		(0.81)
I(Category = Geospatial)		-0.011**
		(-2.50)
I(Category = Point of Sale)		0.004
		(1.48)
I(Category = Satellite Image)		0.008
		(0.73)
I(Category = Web Traffic)		0.014**
		(2.06)
I(Category = Others)		0.016***
		(3.00)
Forecast Age	-0.022***	-0.022***
	(-9.73)	(-9.71)
Analyst/Firm Experience	-0.003	-0.003
	(-0.62)	(-0.65)
Analyst Experience	0.010*	0.010**
	(2.08)	(2.08)
#Firms Covered	0.005	0.005
	(1.18)	(1.15)
Forecast Frequency	0.004	0.003
	(1.55)	(1.51)
Broker Size	-0.000	-0.000*
	(-1.62)	(-1.70)
Analyst-Firm Fixed Effects	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes
N	64,036	64,036
Adjusted R^2	0.822	0.822

Table A3
Summary Statistics

This table reports summary statistics for all variables in our main tests. Appendix 2 defines all variables. All continuous variables are winsorized at the 1% and 99% levels.

Variables	Mean (1)	SD (2)	P25 (3)	P50 (4)	P75 (5)	# of Obs. (6)
Acc	0.013	0.751	-0.393	0.152	0.601	64,036
I(Alternative Data)	0.088	0.283	0	0	0	64,036
Forecast Age	4.913	1.120	4.575	5.236	5.631	64,036
Analyst/Firm Experience	6.691	6.828	1.781	4.510	9.189	64,036
Analyst Experience	13.873	9.541	5.732	11.934	21.904	64,036
#Firms Covered	2.907	0.370	2.708	2.944	3.135	64,036
Forecast Frequency	6.362	0.679	6.038	6.450	6.819	64,036
Broker Size	87.102	50.213	47	84	116	64,036
Number of 8-Ks	15.704	7.310	10	14	21	64,036
Return Volatility	0.012	0.005	0.009	0.011	0.013	64,036
Earnings Surprise	0.001	0.016	-0.002	0.001	0.004	64,036
I(Earnings Restatement)	0.320	0.466	0	0	1	64,036
Discretionary Accruals	0.111	0.151	0.018	0.063	0.141	64,036
I(Lack of Preferential Access to Management)	0.752	0.432	1	1	1	64,036
I(In-House Data Science Team)	0.158	0.365	0	0	0	64,036
$\sum Colleagues_{Alternative\ Data}$	2.807	2.667	1	2	4	64,036
Size	11.769	0.779	11.217	11.854	12.263	64,036
M/B	4.256	5.503	1.859	2.921	4.484	64,036
Momentum	0.083	0.154	-0.013	0.076	0.177	64,036
Cumulative Abnormal Returns	-0.033	2.686	-1.245	-0.030	1.217	64,007
Earnings Forecast Change	0.001	0.033	0	0	0	57,761
Target Price Change	0.008	0.047	0	0	0	55,432
Recommendation Change	0.000	0.156	0	0	0	60,363

Description of Analysis Tabulated in Online Appendix Table A4

An analyst's decision to adopt alternative data may coincide with an analyst's decision to exert greater effort covering the corresponding firm. To assess the relevance of this possibility, we construct measures of analyst effort that have been used in prior literature (Merkley, Michaely, and Pacelli [2017], Hwang, Liberti, and Sturgess [2019], Grennan and Michaely [2020]). We then test whether the adoption of alternative data comes with greater effort.

Our regression equation is similar to regression equation (6):

$$Effort_{i,f,t} = \eta_{i,f} + \theta_{f,t} + \beta\, I(Alternative\ Data_{i,f,t}) + \gamma`\, Controls + \varepsilon_{i,f,t} \qquad (9)$$

First, for each analyst/firm/year, we compute the number of days between the earnings announcement and the analyst's most recent forecast prior to the corresponding earnings announcement, multiplied by (-1). We also compute the number of forecast revisions made by the corresponding analyst for the corresponding firm's earnings. Analysts who exert greater effort should issue earnings forecasts that are less stale (Merkley, Michaely, and Pacelli [2017]) and, in general, update their earnings forecasts more frequently (Hwang, Liberti, and Sturgess [2019]).

Motivated by Grennan and Michaely [2020], we also construct the following measures based on analysts' earnings conference call behavior. First, we construct an indicator, which equals one if the analyst participated in the earning conference call discussing the corresponding firm's annual earnings and zero otherwise. Within the subset of analysts who participate in an earnings conference call, we also construct: (a) the total number of questions posed by the analyst, (b) the total number of words spoken by the analyst, (c) *Easy-to-measure Earnings Topics*, which, following Grennan and Michaely [2020] equals one if an analyst's questions contain the words "*sale*," "*margin*," "*price*," or "*capital*," and (d) *Hard-to-measure Earnings Topics*, which, following Grennan and Michaely equals one if an analyst's questions contain the words "*adapt*," "*brand*," "*engage*," or "*technology*." We obtain our earnings conference call data through Refinitiv.

We report our findings in Table A4. For our regressions based on analysts' forecasts, we find that the estimates of *I(Alternative Data)* are small in magnitude and not statistically significant. That is, we find that the adoption of alternative data changes neither the timeliness of forecasts nor the number of forecast revisions.

Similarly, for our regressions based on analysts' conference call behavior, we find that the adoption of alternative data changes neither the number of questions asked, nor the number of words spoken, nor the types of questions asked. We do find that adopting alternative data marginally increases the likelihood of attending a conference call; the corresponding estimate of *I(Alternative Data)* is 0.040 (*t*-statistic = 1.67).

Table A4
Alternative Data Adoption and Analyst Effort

This table reports coefficient estimates from regressions of various measures of analyst effort on whether an analyst explicitly references the use of alternative data in her written report. The observations are at the analyst/firm/year level. The regressions are identical to that in column (1) of Table 3, except that the dependent variables are proxies for analyst effort. In column (1), analyst effort is measured by the number of forecast revisions made by the corresponding analyst for the corresponding firm's earnings. In column (2), analyst effort is measured by the number of days between the date of the analyst's last forecast prior to the earnings announcement date and the earnings announcement date, multiplied by (-1). The dependent variables in columns (3) through columns (7) are an indicator if the analyst participated in the earning conference call discussing the corresponding firm's annual earnings, the total number of questions posed by the analyst, the total number of words spoken by the analyst, and whether the analyst's questions pertained to "easy-to-measure earnings topics," or "hard-to-measure earnings topics." We no longer include *Forecast Age* and *Forecast Frequency* as controls. We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and year-month levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Analyst Forecasts and Reports		Conference Call Behavior				
	Number of Forecast Revisions	Timeliness of Forecast	Attendance	Number of Questions Asked	Number of Words Spoken	Easy-to-Measure Topic	Hard-to-Measure Topic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I(Alternative Data)	0.059	0.919	0.040*	-0.048	-2.281	-0.074	-0.027
	(0.46)	(0.26)	(1.67)	(-0.31)	(-0.47)	(-1.26)	(-0.99)
Analyst/Firm Experience	0.023	3.296	0.004	0.017	0.743	-0.026	-0.009
	(0.40)	(1.22)	(0.55)	(0.46)	(0.37)	(-1.04)	(-0.73)
Analyst Experience	0.168**	14.549**	0.012*	0.009	6.579**	0.058**	0.007
	(2.60)	(2.19)	(1.67)	(0.13)	(2.60)	(2.25)	(0.40)
#Firms Covered	0.551***	15.555**	0.057	0.317	20.217*	0.024	0.046
	(3.75)	(2.28)	(1.41)	(1.47)	(1.97)	(0.24)	(0.96)
Broker Size	-0.003**	-0.063	0.000	0.003**	0.082	-0.001*	0.001**
	(-1.98)	(-0.81)	(0.40)	(2.22)	(1.22)	(-1.96)	(2.01)
Analyst-Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5,831	5,831	5,831	2,007	2,007	2,007	2,007
Adjusted R^2	0.418	0.521	0.475	0.644	0.539	0.095	0.172

Table A5
How Much Incremental Insight Is There in Alternative Data? Results by Alternative Data Category

This table reports results from repeating the analysis tabulated in column (2) of Table 3, but we now estimate separate regressions for each indicator variable, *I(Category = __)*. We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and year-month levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I(Category = App Usage)	0.384*** (4.52)							
I(Category = Sentiment)		0.216*** (3.47)						
I(Category = Employee)			0.225*** (3.60)					
I(Category = Geospatial)				0.066 (0.63)				
I(Category = Point of Sale)					0.208*** (4.31)			
I(Category = Satellite Image)						0.135 (1.23)		
I(Category = Web Traffic)							0.183*** (3.08)	
I(Category = Others)								0.200*** (4.34)

Table A5. Continued.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Forecast Age	-0.247***	-0.248***	-0.248***	-0.248***	-0.248***	-0.248***	-0.247***	-0.248***
	(-12.22)	(-12.26)	(-12.21)	(-12.18)	(-12.18)	(-12.19)	(-12.24)	(-12.17)
Analyst/Firm Experience	0.058***	0.059***	0.061***	0.059***	0.059***	0.059***	0.058***	0.058***
	(2.70)	(2.79)	(3.00)	(2.81)	(2.81)	(2.81)	(2.74)	(2.76)
Analyst Experience	0.065	0.064	0.062	0.065	0.063	0.065	0.065	0.065
	(1.17)	(1.15)	(1.13)	(1.17)	(1.14)	(1.17)	(1.18)	(1.17)
#Firms Covered	0.048	0.041	0.046	0.043	0.040	0.042	0.043	0.038
	(0.97)	(0.80)	(0.91)	(0.86)	(0.79)	(0.85)	(0.86)	(0.59)
Forecast Frequency	0.026	0.027	0.027	0.027	0.029	0.027	0.026	0.030
	(0.84)	(0.89)	(0.86)	(0.87)	(0.95)	(0.87)	(0.85)	(0.76)
Broker Size	-0.001	-0.001	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(-1.07)	(-1.05)	(-0.94)	(-1.00)	(-0.99)	(-1.00)	(-1.05)	(-1.01)
Analyst-Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	64,036	64,036	64,036	64,036	64,036	64,036	64,036	64,036
Adjusted R^2	0.229	0.228	0.228	0.228	0.228	0.228	0.229	0.229

Table A6
How Much Incremental Insight Is There in Alternative Data?
Simultaneously Drawing From Multiple Categories and Differences in Data Source

This table reports results from repeating the analysis tabulated in column (1) of Table 3, but we now replace *I(Alternative Data)* with ∑ *Categories* in column (1), which is the number of different alternative data categories the corresponding analyst explicitly relies on. In column (2), we replace *I(Alternative Data)* with *I(Source = In-House Data Science Team)* and *I(Source = Data Vendor)*, which equal one if the corresponding analyst explicitly references the use of alternative data and the referenced source is an in-house data science team and an external vendor, respectively. *I(Source = Unknown)* equals one if the corresponding analyst explicitly references the use of alternative data but does not state the source. The results in column (1) are based the subset of forecasts explicitly supported by alternative data. We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and year-month levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
∑ *Categories*	0.176*	
	(1.77)	
I(Source = In-House Data Science Team)		0.260***
		(3.52)
I(Source = Data Vendor)		0.179***
		(3.34)
I(Source = Unknown)		0.184***
		(5.06)
Forecast Age	-0.180***	-0.246***
	(-3.70)	(-12.36)
Analyst/Firm Experience	-0.017	0.058***
	(-1.25)	(2.79)
Analyst Experience	0.386	0.059
	(1.54)	(1.08)
#Firms Covered	-0.277	0.041
	(-1.05)	(0.81)
Forecast Frequency	-0.090	0.027
	(-0.87)	(0.97)
Broker Size	0.004	-0.000
	(1.51)	(-0.95)
Analyst-Firm Fixed Effects	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes
N	5,639	64,036
Adjusted R^2	0.371	0.232

Table A7

How Much Incremental Insight Is There in Alternative Data? The Role of Learning

This table reports coefficient estimates from regression of forecast accuracy on variables reflecting an analyst's level of expertise working with alternative data. The observations are at the analyst/firm/report-date level. The regressions are identical to that in column (1) of Table 3 except for that we replace *I(Alternative Data)* with $\sum Revisions_{\text{Alternative Data}}$, which is the number of previous forecasts revisions supported by alternative data. We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and year-month levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Learning from Past Experience
$\sum Revisions_{\text{Alternative Data}}$	0.053***
	(4.07)
Forecast Age	-0.147***
	(-3.28)
Analyst/Firm Experience	-0.002
	(-0.14)
Analyst Experience	0.310
	(1.32)
#Firms Covered	-0.236
	(-0.79)
Forecast Frequency	-0.052
	(-0.50)
Broker Size	0.004
	(1.55)
Analyst-Firm Fixed Effects	Yes
Firm-Year Fixed Effects	Yes
N	5,639
Adjusted R^2	0.393

Table A8
How Much Incremental Insight Is There in Alternative Data? The Role of Learning by Alternative Data Category

This table reports results from repeating the analysis tabulated in Online Appendix Table A7, but we now compute the number of forecasts revisions explicitly supported by alternative data from a particular category, $\sum$ *Revisions* $_{Category}$. We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and firm levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\sum$ *Revisions* $_{App\ Usage}$	0.121							
	(1.29)							
$\sum$ *Revisions* $_{Sentiment}$		0.041						
		(0.91)						
$\sum$ *Revisions* $_{Employee}$			0.111*					
			(1.78)					
$\sum$ *Revisions* $_{Geospatial}$				-0.123				
				(-1.39)				
$\sum$ *Revisions* $_{Point\ of\ Sale}$					0.051			
					(0.74)			
$\sum$ *Revisions* $_{Satellite\ Image}$						0.257***		
						(6.51)		
$\sum$ *Revisions* $_{Web\ Traffic}$							0.033**	
							(2.31)	
$\sum$ *Revisions* $_{Others}$								0.044*
								(1.69)

Table A8. Continued.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Forecast Age	0.148***	0.124**	0.122**	0.017	0.025	0.043	0.200***	0.063*
	(3.09)	(2.22)	(2.32)	(0.36)	(0.63)	(1.04)	(3.80)	(1.72)
Analyst/Firm Experience	2.415**	8.036***	0.859***	1.009	-1.204	0.037	1.881	-0.039
	(2.52)	(5.56)	(6.31)	(0.84)	(-1.19)	(0.27)	(0.92)	(-0.06)
Analyst Experience	-1.238	-6.412***	0.132	-0.088	2.342***	0.597	0.490	1.192
	(-1.62)	(-5.64)	(0.39)	(-0.10)	(2.78)	(0.84)	(0.25)	(1.60)
#Firms Covered	0.411**	-0.577	-4.907***	1.400*	-0.502	-1.791	-0.220	3.141**
	(2.06)	(-1.65)	(-8.54)	(2.05)	(-0.50)	(-1.41)	(-0.32)	(2.47)
Forecast Frequency	-0.808*	0.351*	0.819***	-0.252**	0.071	0.005	0.052	-0.601**
	(-1.90)	(1.80)	(8.71)	(-2.67)	(0.22)	(0.04)	(0.29)	(-2.12)
Broker Size	-0.015	-0.002	-0.016**	-0.018**	0.001	-0.019	0.012*	-0.010
	(-1.62)	(-0.78)	(-2.17)	(-2.58)	(0.30)	(-0.87)	(1.87)	(-1.16)
Analyst-Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	476	1,062	543	257	1,080	171	1,944	1,322
Adjusted R^2	0.523	0.548	0.683	0.695	0.512	0.795	0.531	0.489

Table A9 Alternative Data and Stock Market Reactions in the Medium- and Long-run

This table repeats the analysis in Table 6 over medium- and long horizons. [2, 5] reflects the abnormal return summed from the second trading day after the revision to five trading days after the revision, that is, roughly one calendar week after the revision. [2, 21] reflects the abnormal return summed from the second trading day after the revision to 21 trading days after the revision, that is, roughly one calendar month after the revision. [2, 63] reflects the abnormal return summed from the second trading day after the revision to 63 trading days after the revision, that is, roughly three calendar months after the revision. [2, 252] reflects the abnormal return summed from the second trading day after the revision to 252 trading days after the revision, that is, roughly one calendar year after the revision. We report *t*-statistics in parentheses. We double-cluster our standard errors at the analyst- and year-month levels. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	[2, 5]		[2, 21]		[2, 63]		[2, 252]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Earnings Forecast Change								
I(Alternative Data) × *Δ*	-7.318**	-6.537**	-16.598**	-14.554***	-30.037**	-16.969**	-58.857**	-6.775
	(-2.48)	(-2.34)	(-2.34)	(-2.63)	(-2.22)	(-2.00)	(-2.02)	(-0.70)
Δ	-0.101	-0.487	0.411	-1.864	1.510	-9.028***	5.641	-16.737***
	(-0.14)	(-0.73)	(0.19)	(-1.12)	(0.35)	(-3.06)	(0.53)	(-4.04)
I(Alternative Data)	0.069	0.075	0.303**	0.355**	0.898***	0.182	2.610**	-0.183
	(1.45)	(1.47)	(2.25)	(2.42)	(2.66)	(0.50)	(2.01)	(-0.38)
Firm Characteristics Controls	Yes	No	Yes	No	Yes	No	Yes	No
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Analyst-Firm Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Firm-Year Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	37,955	37,955	37,955	37,955	37,955	37,955	37,955	37,955
Adjusted R^2	0.022	0.047	0.037	0.122	0.029	0.271	0.012	0.708

Table A9. Continued.

	[2, 5]		[2, 21]		[2, 63]		[2, 252]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel B: Target Price Change								
I(Alternative Data) × Δ	0.358	0.147	1.836	1.966	-3.487	-1.122	-10.763	-0.772
	(0.28)	(0.13)	(0.80)	(0.95)	(-0.81)	(-0.33)	(-0.89)	(-0.13)
Δ	-1.959***	-0.847*	-4.449***	-3.829***	-7.528***	-13.045***	-7.307*	-23.171***
	(-4.29)	(-1.90)	(-3.98)	(-4.21)	(-3.92)	(-8.57)	(-1.79)	(-10.04)
I(Alternative Data)	0.059	0.059	0.265*	0.315*	0.789**	0.119	2.057	-0.339
	(1.15)	(1.08)	(1.93)	(1.97)	(2.30)	(0.31)	(1.57)	(-0.64)
Firm Characteristics Controls	Yes	No	Yes	No	Yes	No	Yes	No
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Analyst-Firm Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Firm-Year Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	34,697	34,697	34,697	34,697	34,697	34,697	34,697	34,697
Adjusted R^2	0.022	0.046	0.036	0.120	0.028	0.272	0.011	0.704

Table A9. Continued.

	[2, 5]		[2, 21]		[2, 63]		[2, 252]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel C: Recommendation Change								
I(Alternative Data) × *Δ*	0.170	0.148	0.095	-0.079	-1.243	-1.799*	-0.303	-2.043
	(0.55)	(0.54)	(0.18)	(-0.17)	(-1.10)	(-1.72)	(-0.14)	(-1.49)
Δ	0.001	0.013	-0.214	-0.225	-0.657**	-0.669**	-0.389	0.201
	(0.02)	(0.17)	(-1.25)	(-1.44)	(-2.19)	(-2.28)	(-0.64)	(0.55)
I(Alternative Data)	0.077	0.063	0.303**	0.337**	0.914***	0.256	2.416*	-0.101
	(1.54)	(1.14)	(2.31)	(2.14)	(2.66)	(0.66)	(1.80)	(-0.20)
Firm Characteristics Controls	Yes	No	Yes	No	Yes	No	Yes	No
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Analyst-Firm Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Firm-Year Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
N	37,848	37,848	37,848	37,848	37,848	37,848	37,848	37,848
Adjusted R^2	0.020	0.044	0.035	0.116	0.029	0.265	0.012	0.697