

ONLINE APPENDIX TO

“THE USE AND USEFULNESS OF BIG DATA:
EVIDENCE FROM FINANCIAL ANALYSTS”

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

We compile a comprehensive 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 six 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)*

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

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

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

Table A1

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

In this table we present numbers of firms in our sample by Global Industry Classification Standard (GICS) sector, fractions of firms that are in the corresponding GICS sectors, numbers of firms in the CRSP/Compustat universe by GICS sector, fractions of firms that are in the corresponding GICS sectors, and the combined market value of the firms in our sample as a percentage of the combined market value of all firms in the CRSP/Compustat universe. 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	%	$\frac{\sum \text{Market Value}_{\text{Our Sample}}}{\sum \text{Market Value}_{\text{CRSP/Compustat}}}$
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

The Adoption of Alternative Data and Earnings Forecast Accuracy: Absolute Forecast Error

This table replicates Table 4 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). $I(\text{Alternative Data})$ is an indicator, which equals one if the corresponding analyst's forecast is supported by alternative data as described in Section 2 and zero otherwise. $I(\text{Alternative Data Category} = _)$ is an indicator variable, which equals one if the analyst uses alternative data from category $_$. $\sum \text{Alternative Data Categories}$ is the number of different alternative data categories the analyst uses. $I(\text{Alternative Data Source} = _)$ is an indicator variable, which equals one if the source of alternative data is $_$. All remaining variables are defined in Table 3. We report t -statistics in parentheses. Standard errors are double-clustered at the analyst- and the firm level. *, **, and *** denote significance at 10%, 5%, and 1% levels.

	(1)	(2)	(3)	(4)
$I(\text{Alternative Data})$	0.013** (2.08)			
$I(\text{Alternative Data Category} = \text{App Usage})$		0.020*** (5.19)		
$I(\text{Alternative Data Category} = \text{Consumer Sentiment})$		0.011 (1.40)		
$I(\text{Alternative Data Category} = \text{Employee})$		0.005 (0.67)		
$I(\text{Alternative Data Category} = \text{Geospatial})$		-0.011 (-1.51)		
$I(\text{Alternative Data Category} = \text{Point of Sale})$		0.004 (0.83)		
$I(\text{Alternative Data Category} = \text{Satellite Image})$		0.007 (1.25)		
$I(\text{Alternative Data Category} = \text{Web Search})$		0.014** (1.99)		
$I(\text{Alternative Data Category} = \text{Others})$		0.017*** (2.80)		
$\sum \text{Alternative Data Categories}$			0.011** (2.19)	
$I(\text{Alternative Data Source} = \text{In-House Data Science Team})$				0.021*** (2.81)
$I(\text{Alternative Data Source} = \text{Data Vendor})$				0.011* (1.97)
$I(\text{Alternative Data Source} = \text{Unknown})$				0.006 (1.04)

Table A2. Continued.

	(1)	(2)	(3)	(4)
<i>Forecast Age</i>	-0.022*** (-4.90)	-0.022*** (-4.90)	-0.022*** (-4.89)	-0.022*** (-4.90)
<i>Analyst/Firm Experience</i>	-0.003 (-0.73)	-0.003 (-0.74)	-0.003 (-0.78)	-0.003 (-0.77)
<i>Analyst Experience</i>	0.010* (2.02)	0.010* (2.00)	0.011** (2.05)	0.010* (2.00)
<i>#Firms Covered</i>	0.005 (1.58)	0.005 (1.59)	0.005 (1.52)	0.006 (1.68)
<i>Forecast Frequency</i>	0.004 (1.25)	0.004 (1.21)	0.004 (1.22)	0.004 (1.17)
<i>Broker Size</i>	-0.009 (-1.62)	-0.009 (-1.60)	-0.010 (-1.68)	-0.009 (-1.57)
Analyst-Firm Fixed Effects	Yes	Yes	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes	Yes	Yes
<i>N</i>	64,036	64,036	5,639	64,036
Adjusted <i>R</i> ²	0.822	0.822	0.822	0.822

Table A3
The Adoption of Alternative Data and Earnings Forecast Accuracy:
Do Analysts Not Discuss the Use of Alternative Data in Their Written Reports?

This table reports results from repeating the analysis tabulated in column (1) of Table 4, but either replaces $I(\text{Alternative Data})$ with $\text{Inferred } I(\text{Alternative Data})$ or adds $\text{Inferred } I(\text{Alternative Data})$ as another independent variable. $\text{Inferred } I(\text{Alternative Data})$ equals one if $I(\text{Alternative Data})$ equals zero but, at the same time, the corresponding analyst uses alternative data for another firm. We report t -statistics in parentheses. Standard errors are double-clustered at the analyst- and the firm level. *, **, and *** denote significance at 10%, 5%, and 1% levels.

	(1)	(2)
<i>I(Alternative Data)</i>	0.201*** (3.81)	
<i>Inferred I(Alternative Data)</i>	-0.106 (-1.54)	-0.154* (-1.81)
<i>Forecast Age</i>	-0.246*** (-10.75)	-0.248*** (-10.35)
<i>Analyst/Firm Experience</i>	0.058*** (4.64)	0.059*** (4.64)
<i>Analyst Experience</i>	0.063 (1.33)	0.067 (1.40)
<i>#Firms Covered</i>	0.038 (0.60)	0.042 (0.66)
<i>Forecast Frequency</i>	0.030 (1.00)	0.025 (0.80)
<i>Broker Size</i>	-0.081** (-2.13)	-0.080** (-2.06)
Analyst-Firm Fixed Effects	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes
<i>N</i>	64,036	64,036
Adjusted R ²	0.232	0.229

Table A4
The Adoption of Alternative Data and Earnings Forecast Accuracy:
Differences Across Industries

This table reports results from repeating the analysis tabulated in column (1) of Table 4, but estimates the regression separately for firms in Consumer Discretionary, Communication Services, Health Care, Consumer Staples, and Industrials (column (1)) and firms in Energy, Financials and Materials (column (2)). We report *t*-statistics in parentheses. Standard errors are double-clustered at the analyst- and the firm level. *, **, and *** denote significance at 10%, 5%, and 1% levels.

	Wide Use of Alternative Data (1)	Infrequent Use of Alternative Data (2)
<i>I(Alternative Data)</i>	0.208*** (3.24)	0.059 (0.59)
<i>Forecast Age</i>	-0.249*** (-7.72)	-0.211*** (-11.63)
<i>Analyst/Firm Experience</i>	0.122*** (2.64)	0.037** (2.44)
<i>Analyst Experience</i>	0.055 (0.94)	0.023 (0.48)
<i>#Firms Covered</i>	0.059 (0.74)	-0.026 (-0.27)
<i>Forecast Frequency</i>	0.035 (1.11)	-0.037 (-0.71)
<i>Broker Size</i>	-0.126** (-2.09)	-0.021 (-0.43)
Analyst-Firm Fixed Effects	Yes	Yes
Firm-Year Fixed Effects	Yes	Yes
<i>N</i>	50,656	13,380
Adjusted R ²	0.243	0.203