"Risk Factors that Matter Textual Analysis of Risk Disclosures for the Cross-Section of Returns"

Alejandro Lopez-Lira, Wharton University of Pennsylvania

Swedish House of Finance Conference on
Financial Markets and Corporate Decisions
August 19-20, 2019

Risk Factors that Matter

Textual Analysis of Risk Disclosures for the Cross-Section of Returns

Alejandro Lopez-Lira[1]

[1]Phd Candidate
The Wharton School
University of Pennsylvania

August 2019

I am grateful for the financial support provided by The Mack Institute for Innovation Management and The Rodney L. White Center for Financial Research

Why do we want to model returns with risk factors?

- Explain simply why different companies earn different returns
  - Investing in companies more exposed to systematic risks earns compensation for taking those risks

- $r^e_{i,t+1} = \alpha_{i,t} + \beta'_{i,t} R^e_{t+1} + \epsilon_{i,t+1}$, $\alpha_{i,t} = 0$
  - 'Free' by theory: Factor structure of the SDF

- Which risks? Economic content comes in picking the factors

Overview of usual Factor Models

Empirical Models

- Perform great on very puzzling portfolios
  - Value premium
  - Momentum
- Extremely good for detecting if a new anomaly is in the span of the previous ones

Statistical Factor Models

- Amazing statistical performance
- Really good for covariance estimation
- No need to manually select the factors

What more do we want from risk factors?

- Interpretable
  - Explain returns of strategies
  - Cost of capital
  - What does exposure to the second PCA means?
  - Which risk does the profitability factor proxy for?

- Represent economic risk
  - Are these risks? Anomalies? Proxies for firms first order conditions?
    - Kozak, Nagel, and Santosh (2018)

Big picture

"The best hope for finding pricing factors ... is to try to understand the fundamental macroeconomic sources of risk" (Cochrane 2005)

- Collect all of the risks in the economy $\Rightarrow$ end the anomaly literature?
- Firms probably understand the risks they face better than we do

What to do?

- It would be great to get a list of all of the risks in the economy
- $SDF_t = m(c_t, r_t, c_{t+1}, r_{t+1}...)$
- In equilibrium $y_t = c_t$
- So if $y_t = h_t(f_t)$
- $SDF_t = g_t(f)$

Road map

1. Use machine learning to extract all of the risks perceived by the firms
2. Get risk exposures for each firm
3. Explain differences in expected returns and covariances using revealed risks
4. Bonus: Risk Factors are described by words

Related Literature

- Text Analysis searching for specific risks
  - Hassan, Hollander, van Lent and Tahoun (2017), Grotteria (2019)
    - Political risk
  - Loughran, McDonald, and Pragidis (2019)
    - Oil news
- Topic Modelling
  - Israelsen (2014)
    - Disclosed risks associated with commonly used asset pricing risk factors
  - Hanley, Hoberg (2018)
    - Dynamic interpretation of emerging risks in the Financial Sector
  - Cong, Liang, Zhang (2019)
    - Clustering word embeddings
- Text Analysis
  - Cohen, Malloy, Nguyen (2018)
- Machine Learning
- Factor Models

How to get a list of all of the (real) risk factors in the economy?

- Firms are required to disclose all of the risk that they face

- Use machine learning to extract all of these risks!

Machine Learning, but not the usual type!

Unsupervised Machine Learning

1. Helps us understand the data
2. No information from the returns is used
3. Designed to get meaningful risks

Supervised Machine Learning and PCA

1. Fits the data
2. Use realized returns to find the factors
3. Designed to get the best statistical performance

Advantages of using text analysis to get risk factors

- No subjectivity in choosing the risk factors
  - Take them directly from the firms
  - They are the ones that best understand the risks they face
- The factors unambiguously represent economic risk
  - Which risks are priced?
  - What assets can we price?
- No information on past returns
  - Effectively out-of-sample
  - No data mining
  - No p-hacking
- Interpretable risk factors
  - Risk Factors are described in plain words

Data

- Monthly returns, annual disclosures
- Firms disclose in a specific section "Risk Factors" the risks they face
  - Legally required since 2006
- Can we trust the risk disclosures? Yes!
  - Face legal action if they fail to obey the regulation
  - Managers provide risk factor disclosures that meaningfully reflect the risks they face Campbell et al. (2014)
  - The type of risk the firm faces determines whether it devotes a greater portion of its disclosures towards describing that risk type (e.g. Gaulin (2017) and Campbell et al. (2014))

Extract: Apple 10-K 2010 Section 1A (10 pages)

- "Demand ... could differ ... [because] of the strengthening of the U.S.dollar"
- "The Company uses some custom components..."
- "Due to the highly volatile and competitive nature of the [industry], the Company must continually introduce new products"

Extract: Apple 10-K 2010 Section 1A International Risk

The Company's business is subject to the risks of international operations.

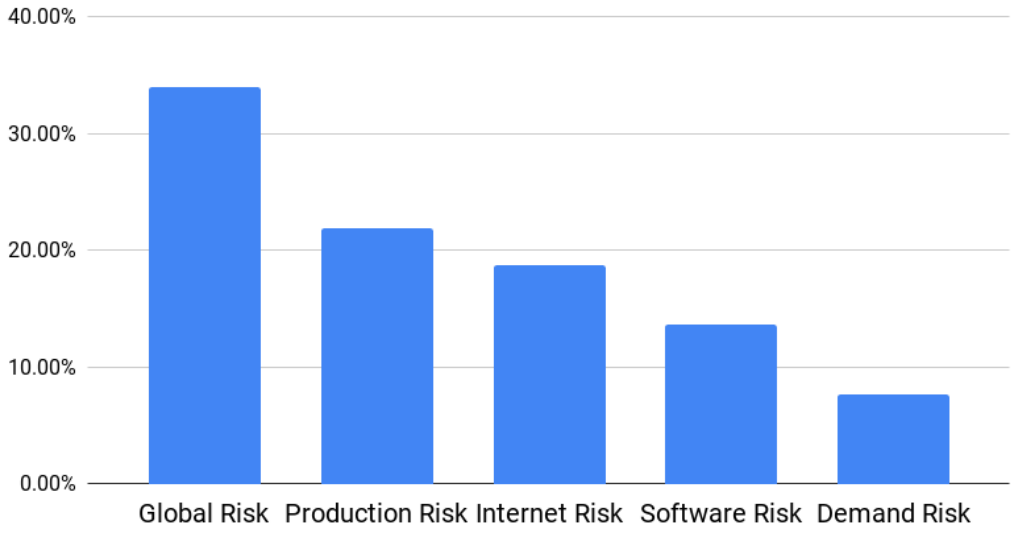
The Company derives a significant portion of its revenue and earnings from its international operations. Compliance with U.S. and foreign laws and regulations that apply to the Company's international operations, including without limitation import and export requirements, anti-corruption laws, tax laws (including U.S. taxes on foreign subsidiaries), foreign exchange controls and cash repatriation restrictions, data privacy requirements, labor laws, and anti-competition regulations, increases the costs of doing business in foreign jurisdictions, and any such costs, which may rise in the future as a result of changes in these laws and regulations or in their interpretation. Furthermore, the Company has implemented policies and procedures designed to ensure compliance with these laws and regulations, but there can be no assurance that the Company's employees, contractors, or agents will not violate such laws and regulations or the Company's policies. Any such violations could individually or in the aggregate materially adversely affect the Company's financial condition or operating results.

The Company's financial condition and operating results also could be significantly affected by other risks associated with international activities, including but not limited to, economic and labor conditions, increased duties, taxes and other costs, political instability, and changes in the value of the U.S. dollar versus local currencies. Margins on sales of the Company's products in foreign countries, and on sales of products that include components obtained from foreign suppliers, could be materially adversely affected by foreign currency exchange rate fluctuations and by international trade regulations, including duties, tariffs and antidumping penalties. Additionally, the Company is exposed to credit and collectability risk on its trade receivables with customers in certain international markets. There can be no assurance it can effectively limit its credit risk and avoid losses, which could materially adversely affect the Company's financial condition and operating results.

The Company's primary exposure to movements in foreign currency exchange rates relate to non-U.S. dollar denominated sales in Europe, Japan, Australia, Canada and certain parts of Asia, as well as non-U.S. dollar denominated operating expenses incurred throughout the world. Weakening of foreign currencies relative to the U.S. dollar will adversely affect the U.S. dollar value of the Company's foreign currency-denominated sales and earnings, and generally will lead the Company to raise international pricing, potentially reducing demand for the Company's products. In some circumstances, due to competition or other reasons, the Company may decide not to raise local prices to the full extent of the dollar's strengthening, or at all, which would adversely affect the U.S. dollar value of the Company's foreign currency denominated sales and earnings. Conversely, a strengthening of foreign currencies, while generally beneficial to the Company's foreign currency-denominated sales and earnings, could cause the Company to reduce international pricing and incur losses on its foreign currency derivative instruments, thereby limiting the benefit. Additionally, strengthening of foreign currencies may also increase the Company's cost of product components denominated in those currencies, thus adversely affecting gross margins.

The Company has used derivative instruments, such as foreign currency forward and option contracts, to hedge certain exposures to fluctuations in foreign currency exchange rates. The use of such hedging activities may not offset any or more than a portion of the adverse financial effects of unfavorable movements in foreign exchange rates over the limited time the hedges are in place.

Apple's Risk Exposure 2016



Explaining the black box: LDA

- Statistical model
- Each document can be described by a distribution over topics
- Each topic can be described by a distribution over words
- LDA adds priors and makes it formal

Document-Term Matrix

2016	Forecasts	IMF	WBG	and	as	cut	discuss	economy	growth	issues	meet	to	warning
0	1	1	0	0	1	1	0	0	0	1	0	0	1
0	0	1	1	1	0	0	1	1	0	0	1	1	0
2	0	0	1	0	0	0	0	0	1	1	0	0	1

3 Documents x 14 terms

Why not pick the risks with dictionary methods?

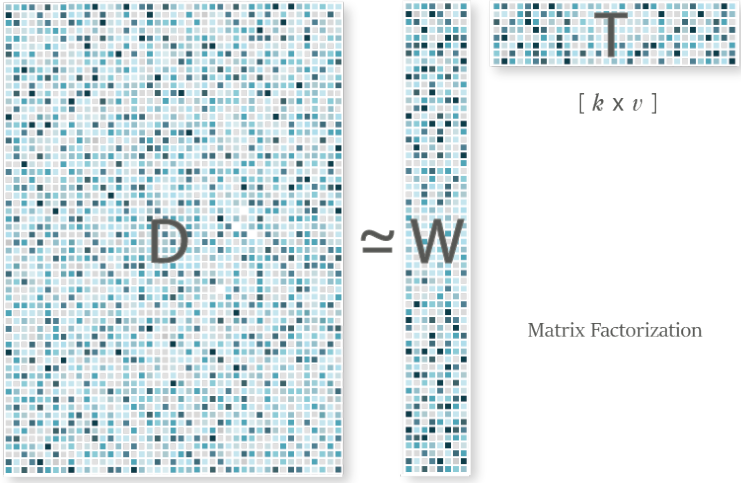
- Dictionary Methods
  - Define set of words of interest
  - Count frequencies across documents
- Subjectivity in picking the risks
  - Effectively imposing which risk matter
- We want the risks to arise naturally

Document-Term Matrix

2016	Forecasts	IMF	WBG	and	as	cut	discuss	economy	growth	issues	meet	to	warning
0	1	1	0	0	1	1	0	0	0	1	0	0	1
0	0	1	1	1	0	0	1	1	0	0	1	1	0
2	0	0	1	0	0	0	0	0	1	1	0	0	1

3 Documents x 14 terms

LDA

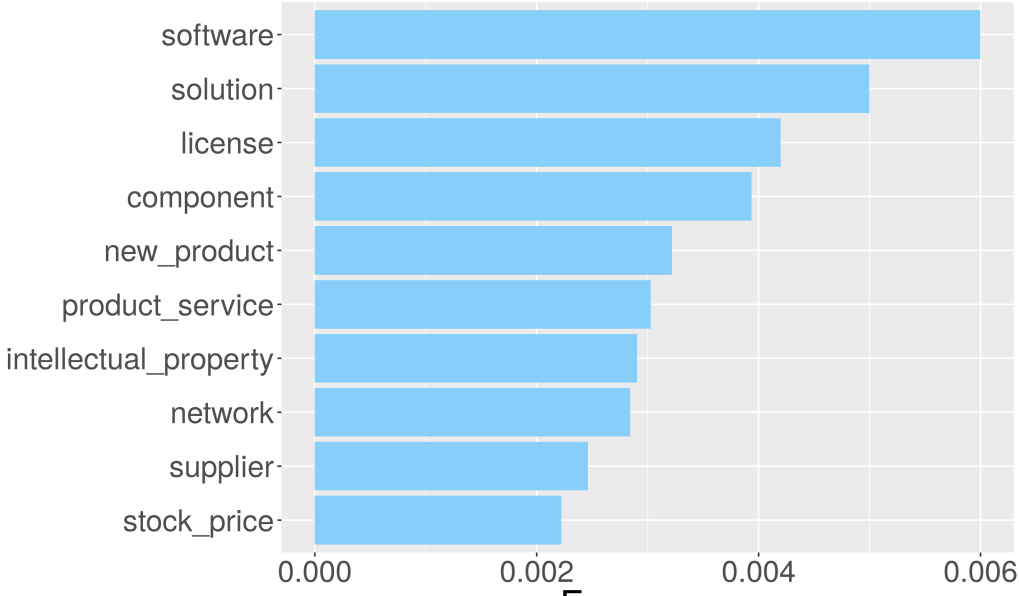


Matrix Factorization

Risk Topics

china chinese rmb registration currency foreign_exchange economy dollar entity approval enterprise procedure transfer legal authority subsidiary tax foreign foreign_currency offshore chinese_government shareholder relevant court safe
client government_agency u__government contractor professional agency procurement funding subcontractor government_contract bid budget work defense award audit perform engagement spending federal_government equipment terminate termination backlog profit
solution client datum software security_breach datum_center organization offering subscription user harm grow internet breach error privacy card platform content member infrastructure implementation electronic product_service provider
solution software license component network stock_price vendor country distributor supplier product_service release partner harm new_product
collaborator patent research approval trademark study drug fda license collaboration grant commercial partner trial
institution school education educational employment participation default_rate agency grant enrollment private online loan funding federal university fiscal_year eligibility participate course authorization commission lose title proprietary
advertising content advertiser television medium satellite internet channel mobile video radio user cable online network station digital license entertainment fcc
test laboratory testing production manufacturing supplier instrument patent perform cell clinical approval human animal country distributor license research commercial research_development intellectual_property process manufacture diagnostic fda
network internet website user provider mobile subscriber product_service wireless traffic fcc brand datum content partner carrier online offering consumer privacy
home construction land housing subcontractor joint_venture real_estate electricity inventory california lender buyer mortgage finance warranty utility build loan community shortage mortgage_loan installation residential building solar
equipment manufacturer transportation fuel lease crop dealer rental labor credit rail hotel pricing volume vehicle
warrant holder board_director redemption share_common amend pursuant exercise isec least approve vote nasdaq target shareholder liquidation entity combination market_price_common_stock preferred_stock officer director
supplier country product_service currency global tax international segment exposure dollar foreign_currency
production reserve drilling oil_gas oil_natural_gas_price hydraulic_fracturing mine oil lease water mining produce drill gas natural_gas pipeline commodity_price formation quantity exploration
vessel charter marine travel maintenance shareholder freight lease trade country air fuel fleet oil safety engine ship international port equipment capacity shipping carrier offshore transportation
unitholder unit crude_oil taxable_income fiduciary_duty limited_partner person volume affiliate energy tax irs partner treat pipeline refinery partnership_agreement natural_gas production cash_distribution transportation ferc terminal
restaurant store merchandise franchise import manufacturer food retail fiscal net_sale brand open labor lease good supplier consumer trend trademark location locate inventory retailer vendor distribution_center
coal shareholder tax spin_off manager board_director taxable court loan trust separation note holder owner affiliate inc irs income mine group opinion transfer common_share subsidiary corporation
vehicle production raw_material supplier subsidiary industrial energy emission site plant steel metal fuel environmental chemical automotive produce
patient physician hospital medicare health medicaid payor healthcare hipaa medical care health_care federal drug pharmacy center reimbursement billing cm provider
device fda country distributor reimbursement approval patient clearance supplier recall physician hospital treatment healthcare clearance_approval medical procedure
supplier semiconductor manufacturer component inventory manufacture equipment oem china net_sale test country capacity production
consumer brand distributor trademark labeling brand_product packaging health retailer fda food retail marketing direct manufacturer raw_material supplier net_sale ingredient
clinical_trial drug patient manufacturer manufacture collaboration collaborator treatment license patent datum approve trial candidate indication approval clinical regulatory_approval pharmaceutical commercialize fda study regulatory_authority
debt note credit_facility credit_agreement outstanding license covenant lease stockholder lender capacity affiliate default liquidity subsidiary equipment merger

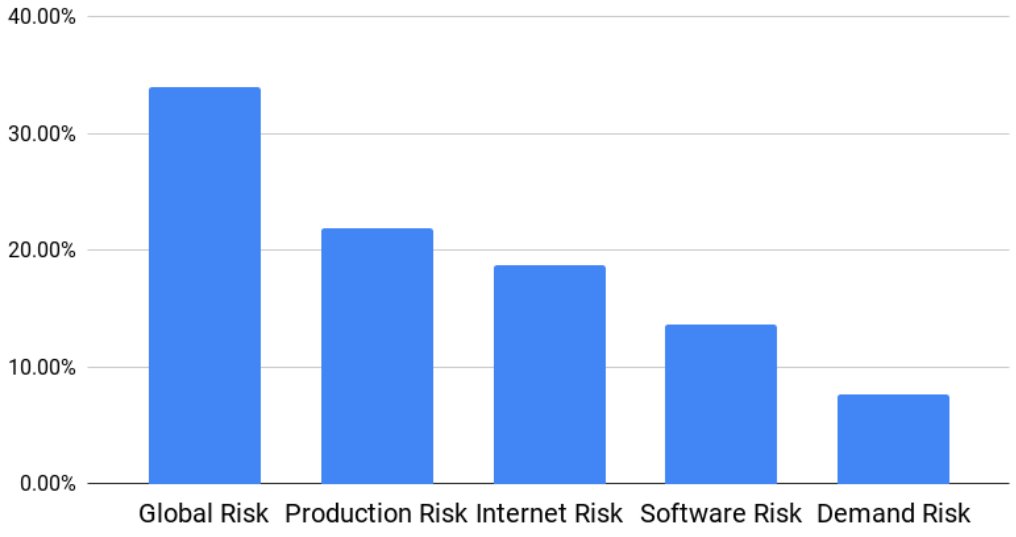
Technology and Innovation Risk



Systematic and Idiosyncratic Risks

- Firms more similar in the risk topic space are more correlated
- Not always induced by β exposure $\Rightarrow$ naturally generates α
- For the systematic risks more exposure implies higher correlation
- Systematic risk exposure is important for prediction
- How far can we get with using only systematic risks?

Apple's Risk Exposure 2016



Systematic and Idiosyncratic

Table: Descriptive statistics

Statistic	N	Mean	St. Dev.	Pctl(25)	Pctl(75)
Pairwise Correlation	3,347,132	0.20	0.15	0.10	0.30
Risk Simmilarity	3,347,132	0.14	0.14	0.03	0.20
Beta Exposure	3,347,132	1.25	0.41	0.97	1.50
Book-to-Market Distance	3,347,132	1.05	3.20	0.17	0.92
Size Distance	3,347,132	2.23	1.69	0.89	3.21

Systematic and Idiosyncratic

Table: Correlation Matrix of Distances and Exposures

	Pairwise Correlation	Risk Similarity	Beta Exposure	Book-to-Market Distance	Size Distance
Pairwise Correlation	1	0.19	0.35	0.06	0.12
Risk Similarity	0.19	1	0.03	0.03	0.06
Beta Exposure	0.35	0.03	1	0.06	0.13
Book-to-Market Distance	0.06	0.03	0.06	1	0.16
Size Distance	0.12	0.06	0.13	0.16	1

Systematic and Idiosyncratic

Table 14: Impact of risk similarity on correlation

	Dependent variable:		
	Pairwise Correlation		
	(1)	(2)	(3)
Risk Simmilarity	0.202***	0.190***	0.184***
	(0.001)	(0.001)	(0.001)
	$t = 368.520$	$t = 362.560$	$t = 353.547$
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Beta Exposure		0.130***	0.129***
		(0.0002)	(0.0002)
		$t = 695.019$	$t = 682.636$
		$p = 0.000$	$p = 0.000$
Book-to-Market Distance			−0.003***
			(0.00002)
			$t = -122.887$
			$p = 0.000$
Size Distance			−0.005***
			(0.00005)
			$t = -117.119$
			$p = 0.000$
Constant	0.170***	0.014***	0.031***
	(0.0001)	(0.0003)	(0.0003)
	$t = 1,554.111$	$t = 53.403$	$t = 108.795$
	$p = 0.000$	$p = 0.000$	$p = 0.000$
Observations	3,619,868	3,347,132	3,347,132
R^2	0.036	0.159	0.168
Adjusted R^2	0.036	0.159	0.168
Residual Std. Error	0.150 (df = 3619866)	0.139 (df = 3347129)	0.138 (df = 3347127)
F Statistic	135,807.300*** (df = 1; 3619866)	316,361.400*** (df = 2; 3347129)	168,520.600*** (df = 4; 3347127)
Note:			*p<0.1; **p<0.05; ***p<0.01

Systematic and Idiosyncratic

- Systematic risk exposure is important for prediction
- More systematic risks have more predicting power
- How far can we get with using only systematic risks?

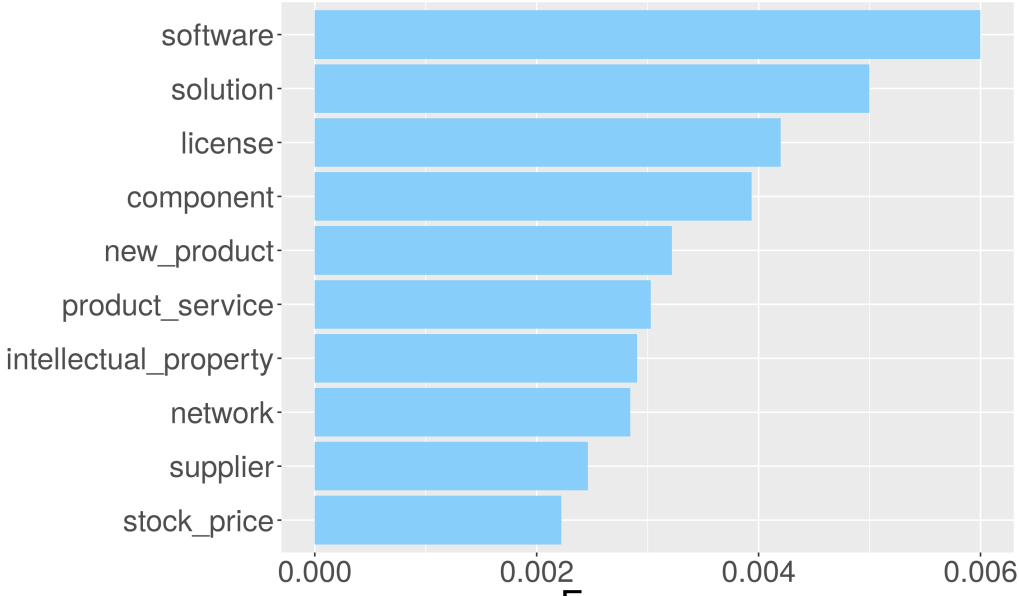
Average proportion of the risk disclosures allocated to each risk for the most discussed risks in the year 2006

Technology Risk	Production Risk	International Risk	Demand Risk	Total
0.11	0.09	0.08	0.08	0.36

Number of firms that spend more than 25% of the time discussing each topic

Year	Technology Risk	Production Risk	International Risk	Demand Risk	Percentage of Total Firms
2006	413	364	343	264	0.54
2007	442	354	324	245	0.52
2008	388	270	356	223	0.50
2009	355	305	412	215	0.51
2010	300	275	387	211	0.48
2011	285	266	422	221	0.49
2012	258	252	468	202	0.50
2013	261	237	452	205	0.49
2014	248	215	479	203	0.48
2015	230	197	493	196	0.46
2016	213	171	505	205	0.44

Technology and Innovation Risk



Technology and Innovation Risk

Company Name	Market Value (Millions)
MICROSOFT CORP	354392
ORACLE CORP	166066
CISCO SYSTEMS INC	144516
QUALCOMM INC	81885
EMC CORP/MA	49896
HP INC	48628
ADOBE SYSTEMS INC	45530
ILLUMINA INC	28136
VMWARE INC -CL A	23870
ELECTRONIC ARTS INC	19873

Table: Biggest 10 Companies that are exposed more than 25% to the Technology and Innovation Risk

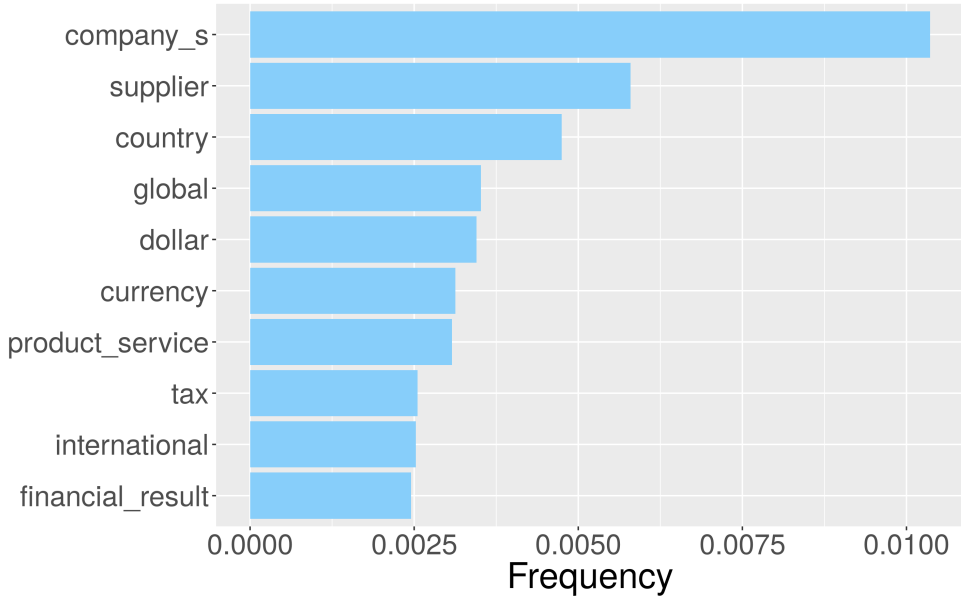
Are we picking up industries?

Industries are not flexible enough to capture common risks

Table: Number of firms by SIC code for firms that are exposed to the Technology Risk Factor

2-Digit SIC Code	Industry	Division	Number of firms
35	Manufacturing	Industrial and Commercial Machinery and Computer Equipment	43
36	Manufacturing	Electronic and other Electrical Equipment and Components, except Computer Equipment	58
38	Manufacturing	Measuring, Analyzing, and Controlling Instruments; Photographic, Medical and Optical Goods; Watches and Clocks	18
73	Services	Business Services	82

International Risk

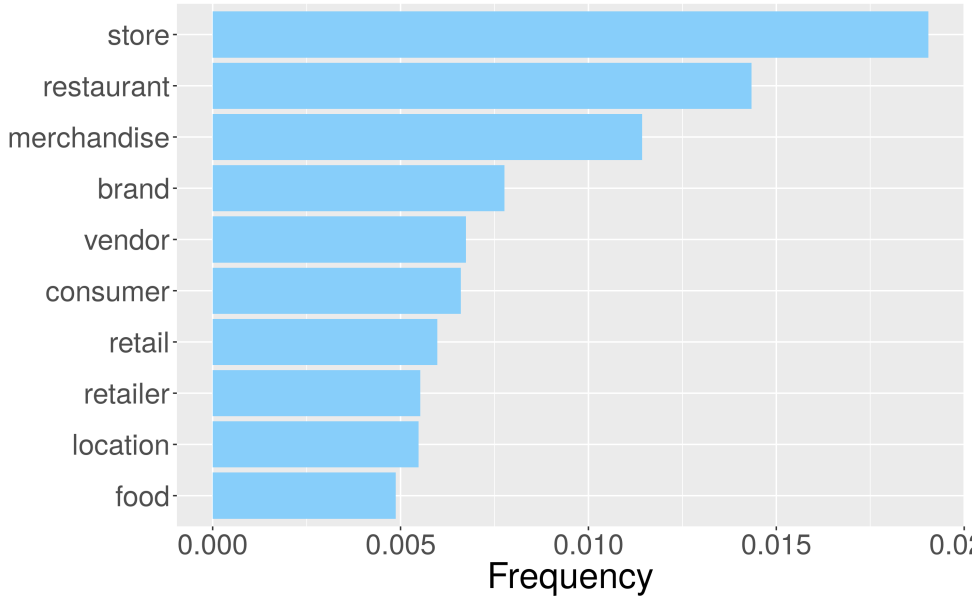


International Risk

Company Name	Market Value (Millions)
APPLE INC	615336
EXXON MOBIL CORP	323960
PROCTER & GAMBLE CO	212388
AT&T INC	211447
PFIZER INC	199329
COCA-COLA CO	185759
CHEVRON CORP	169378
ORACLE CORP	166066
INTEL CORP	162776
MERCK & CO	146899

Table: Biggest 10 Companies that are exposed more than 25% to International Risk

Demand Risk


store
restaurant
merchandise
brand
vendor
consumer
retail
retailer
location
food
0.000
0.005
0.010
0.015
0.0
Frequency

Demand Risk

Company Name	Market Value (Millions)
WALMART INC	209830
HOME DEPOT INC	157452
MCDONALD'S CORP	107129
NIKE INC	92880
STARBUCKS CORP	84413
LOWE'S COMPANIES INC	65211
COSTCO WHOLESALE CORP	61335
TJX COMPANIES INC	47267
TARGET CORP	43613
YUM BRANDS INC	30681

Table: Biggest 10 Companies that are exposed more than 25% to Demand Risk

From Risk Topics to Risk Factors

- How much each company allocates to discuss each risk?
  - θ_i, document proportions
- Consider firms part of each portfolio if they allocate more than 25% of their disclosure in that risk
- Value weight the firms in the portfolio
- Subtract risk-free rate

Selected Statistics

	Technology Risk	Production Risk	International Risk	Demand Risk	Market Portfolio
Mean	0.96	1.13	0.73	0.86	0.70
Sd.	5.58	6.27	4.12	4.29	4.41
Annualized Sharpe Ratio	0.59	0.62	0.62	0.69	0.55
BM	0.52	0.67	0.54	0.62	0.61
Size	6.09	6.18	7.69	6.81	6.71

Is it a reasonable description of expected returns?

- Industry portfolios
- Book to Market portfolios
- Anomalies
- $GRS \propto \frac{\alpha' \Sigma^{-1} \alpha}{1+\mu' \Sigma^{-1} \mu}$
- Low GRS $\Rightarrow$ low evidence of mispricing $\Rightarrow$ high p-value
  - Null is that the model is correct, $\alpha = 0$
- $r^e_{i,t+1} = \alpha_{i,t} + \beta'_{i,t} R^e_{t+1} + \epsilon_{i,t+1}$, $\alpha_{i,t} = 0$

GRS Test

	49 Industry + 25 B-to-M			49 Industry + 25 B-to-M + 15 α		
	GRS	p-value	R^2	GRS	p-value	R^2
Text-based 4 Factor Model	1.52	0.061	0.69	2.09	0.018	0.64
Fama-French 5 Factor Model	1.85	0.012	0.76	3.05	0.001	0.72
Mispricing Factors	1.67	0.044	0.76	2.47	0.006	0.73
q-factor Model	1.81	0.024	0.75	2.48	0.005	0.71

GRS Test

	49 Industry Portfolios			25 Book-to-Market Portfolios			15 Anomaly Portfolios		
	GRS	p-value	R^2	GRS	p-value	R^2	GRS	p-value	R^2
Text-based 4 Factor Model	0.88	0.679	0.63	1.83	0.019	0.8	1.34	0.21	0.21
Fama-French 5 Factor Model	1.55	0.045	0.68	1.91	0.013	0.94	1.12	0.35	0.43
Mispricing Factors	1.22	0.223	0.68	1.70	0.037	0.92	0.68	0.75	0.52
q-factor Model	1.47	0.073	0.67	1.88	0.017	0.92	1.13	0.35	0.43

In the next version of the paper :)

- Adding the information from conference calls
- Systematic vs idiosyncratic
- Prediction
- Beta
- Interaction with the anomalies
- Combine with supervised machine learning

Summary

- Firms have a significant understanding of the risk they are facing
- Information revealed by the firms can provide guidance on how to improve our theoretical asset pricing models
- Interpretable Risk Factors
- Represent economic risk for the firms
- Comparable statistical power
  - We can prize many assets using the firms' revealed risks