# Measuring Knowledge Capital Risk

Pedro H. Braz Vallocci
University of California, Santa Cruz

June 18, 2023

## 1 Introduction

The transition towards a service- and knowledge-based economy has been accompanied by a sharp increase in intangible assets, most notably in research and development (R&D). Knowledge capital, defined as a firm's accumulated investments in R&D, accounts for an increasing share of public firms' valuation and has a different risk profile than physical capital. For example, knowledge capital-heavy firms are more exposed to the loss of key talent and changes in patent enforcement. However, to the best of my knowledge, there are no references in the literature that seek to understand if there are differences in risk premia among firms with different levels of knowledge capital and if agents are pricing in a sudden risk of economic disaster in knowledge capital. My paper seeks to fill this gap.

Spending on intangible assets qualifies as an investment since it reduces current consumption to increase future consumption (Corrado et al., 2009b,a). Moreover, endogenous growth models, such as models with expanding varieties (Romer, 1990), and with quality ladders (Grossman and Helpman, 1991; Atkeson and Burstein, 2019), imply that research expenses can spur growth.

The two latter models consider R&D as intermediate expenses. However, supply-side valuation models such as (Belo et al., 2013) argue that not only quasi-fixed physical capital and labor inputs but also knowledge and brand capital can determine a firm's market value. *Knowledge capital* is defined as a firm's accumulated investments in R&D. *Brand capital* is defined as a firm's accumulated expenses in advertising (Belo et al., 2019); *organization capital* is defined as the set of unique systems and processes employed in an enterprise, which increases with a firm's managerial quality scores (Eisfeldt and Papanikolaou, 2013; Bloom and Van Reenen, 2007). Knowledge capital, brand capital, and organization capital are components of a firm's intangible capital.

Hall (2001); McGrattan and Prescott (2001); Vitorino (2014); Eisfeldt et al. (2020); Li et al. (2014) confirm that intangible capital matters for aggregate stock market valuations and, more specifically, firm-level valuations. (Corrado et al., 2009b) estimated total intangible capital in 2003 to be 3.6 trillion, half of which as scientific and non-scientific R&D capital. (?) found that the omission of knowledge capital and its associated rents can explain up to 2/3 of the investment gap (the difference between marginal $q$ and average $Q$) in R&D intense sectors such as Healthcare and Chemicals.

However, equating the omission of intangible and knowledge capital to a mere mismeasurement of a firm's book value is misleading since intangible capital has unique characteristics and thus cannot be lumped together with physical capital. (Autor et al., 2020; Unger, 2019) shed light on a "scale-biased technological change" in R&D-heavy firms. Greater output scalability, the absence of geographical constraints for input sourcing or output distribution, and the greater importance of network effects lead to a "winner-takes-most" outcome in some industries (i.e., an uneven market power distribution) and a declining labor share, consistent with (Barkai, 2020).

Moreover, R&D-driven technological innovation, while generating research externalities in the form of knowledge available to the rest of society, may also be a force for creative destruction (Schumpeter, 1939) if it leads to mere resource reallocation across firms. Therefore, it is not evident that private-led innovation increases aggregate output. (Kogan et al., 2017; Garcia-Macia et al., 2019) show that innovation has a net effect of accelerating aggregate output and that own-product improvements by incumbents are more important than creative destruction.

R&D-intensive firms are also more likely to offshore profit shifting, i.e., to license their intellectual property rights to offshore subsidiaries in order to accrue derived profits at tax-havens. Compared to non-R&D-intensive firms (e.g., construction), it is easier to decouple the physical location of intellectual property production from its point of sale. Offshore profit shifting has grown since the 2000s, which coincides with the beginning of the productivity growth slowdown (Fernald, 2015). Adjusting for offshore profit shifting adds 2.1 log percentage points to R&D-heavy firms, which is ten times larger than for non-R&D. Labor productivity in R&D-intensive firms has also grown much faster than in non-R&D firms (Guvenen et al., 2021).

The riskiness of intangible assets differs systematically from tangible ones (Hansen et al., 2005). (Eisfeldt and Papanikolaou, 2013) points out that shareholders cannot entirely appropriate the cash flows from the key talent of the firm since the firm must always compensate key talent by its outside option. (Eisfeldt et al., 2018) shows that key talent partially owns the cash flow from intangible capital in the form of equity, finding that almost 40% of compensation to high-skilled labor happens as equity-based pay. Finally, (Ai et al., 2019) predicts that collateralizable assets, which do not include some categories of intangibles, provide insurance against aggregate shocks in the economy and should earn a lower expected return.

Specifically, knowledge capital is risky. Firms spend on R&D to increase their future profitability, which is not guaranteed. For example, research conducted by a pharmaceutical firm can lead to successful new drugs that lead to patent rents for several years or to no result at all. The riskiness of innovation firms, and the growing empirical dispersion of Tobin's q, also explains the relation between Tobin's q and aggregate investment has become tighter since the mid-1990s (Andrei et al., 2019). The riskiness of a financially constrained firm increases with its R&D intensity (Li, 2011). Besides the uncertainty of research investments, a firm is also susceptible to writing off part of its knowledge capital, e.g., when it narrowly loses a patent race (Peters and Taylor, 2017).

Knowledge capital heavy firms are especially susceptible to loss of key talent. (Eisfeldt and Papanikolaou, 2013) find that firms are more likely to list "loss of key talent" as a risk factor in their 10-K reports when they have high organization capital. It makes sense that the same pattern is valid for knowledge capital heavy firms, which are highly dependent on specialized and scarce workforce.

Firms vulnerable to loss of key talent are especially susceptible to immigration-related risks, e.g., the H-1B visa annual quota shortages. The H-1B visa, introduced with the Immigration and Nationality Act of 1990, established temporary renewable visas for college-educated specialty professional workers, most of whom work in STEM occupations. (Peri et al., 2015) shows that the growth in foreign-STEM workers may explain between 10 and 25% of the aggregate productivity growth and 10% of skill-bias growth between 1990 and 2010.

Following previous literature on time-varying disaster risks, such as (Gabaix, 2012; Barro, 2006; Rietz, 1988), (Gourio, 2012) shows that increases in the risk of an economic disaster can lead to an increase in risk premia and a decrease in unemployment and output. My work will develop upon (Gourio, 2012) to consider a knowledge capital heavy economy, and investigate if agents are pricing in the risk of sudden drops in the efficiency of knowledge investment (e.g., due to shortage of skilled workforce) or sudden firm-specific write-offs ("disasters") of knowledge capital (e.g., due to weakening in patent laws).

The current methods of identifying knowledge-heavy firms have proven to be insufficient due to a number of key issues. Firstly, there is a lack of consistency in the standards for R&D reporting across industries and firms. This discrepancy is further compounded by the fact that some firms do not disclose their R&D expenditure in their annual reports, making it challenging to accurately determine their level of knowledge intensity.

Secondly, the measures used to identify intangible assets, typically through indirect measures like Selling, General and Administrative expenses (SG&A), include components that are not related to knowledge, such as organizational capital. This can blur the distinction between knowledge-heavy and non-knowledge-heavy firms.

Lastly, relying on patents as a measure of a firm's knowledge intensity only tells part of the story. While patents reflect the final outcomes of R&D investments, they do not account for the internal learning processes that take place within firms, thus potentially undervaluing those that invest heavily in internal knowledge development, even if they do not have a high patent output.

Considering these shortcomings, several research questions emerge: Firstly, is it possible to better identify knowledge-heavy firms by performing text analysis on the risk factors reported in their annual reports? By analyzing the language and terminology used, can we glean more information about the firm's knowledge intensity?

Secondly, if we can indeed identify knowledge-heavy firms in this manner, does this information influence how agents price these firms? In other words, do market participants factor in different risks for knowledge-heavy firms compared to non-knowledge-heavy firms?

Finally, we must question if using topic modeling to categorize the firms' self-declared risk factors can provide a more sensible and accurate categorization of firms by risk. By examining the topics that firms themselves identify as risks, can we create a more nuanced classification system for firms based on the nature and magnitude of their risk exposure?

# 2 Methodology and Data

## 2.1 Latent Dirichlet Allocation (LDA)

In this study, I use Latent Dirichlet Allocation (LDA), a topic modeling technique, to identify latent topics within a comprehensive corpus of 121,839 firm annual reports spanning the years 2006 to 2022. This approach allows me to uncover patterns and themes in the data that may not be immediately apparent.

Latent Dirichlet Allocation (LDA) is a generative statistical model that is widely employed for topic modelling within the field of natural language processing (NLP). Its effectiveness hinges on the fundamental assumption that each document in a given corpus can be seen as a mixture of a certain number of latent topics, denoted as $k \in 1, ..., K$, each of which carries a particular weight, $\omega_{i1}, ..., \omega_{iK}$. Each of these topics is assigned a word probability vector, $\theta_k$, defining the likelihood of each word appearing under this topic Blei et al. (2003).

Under this model, if we denote $X_i$ as the vector of word counts with length $n_i$ in the $i$th document, the word distribution in a document is modeled as a multinomial distribution. The probability of $X_i$ can be written as:

$$X_i \sim Multinomial\left(n_i, \omega_{i1}\theta_1 + \cdots + \omega_{iK}\theta_K\right) \tag{1}$$

This equation represents the fact that the observed words in the document are generated by a mixture of topics, with each topic contributing to the document with a certain weight.

The output of an LDA operation is twofold: firstly, it generates a list of topics, with each topic represented as a collection of words. Secondly, it offers a weight distribution across these topics for each document, indicating the degree to which each topic is present in a given document.

It is important to note that LDA is an unsupervised learning method. This means that it operates without any predefined labels, instead learning and inferring patterns directly from the data. This characteristic makes LDA a versatile tool, able to extract valuable insights from large and complex corpora of text data.

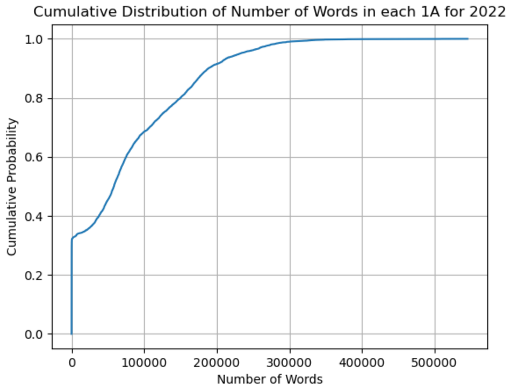
## 2.2 Data

I retrieve the annual reports (10-Ks) for all publicly listed firms since 2013 from the Securities and Exchange Commission's EDGAR database, using their API. A 10-K is a comprehensive document that provides an overview of the company's financial performance and operations over a year, offering a detailed picture of a company's business. To ensure transparency and accuracy, laws and regulations strictly prohibit companies from making false or misleading statements in their 10-Ks. Additionally, under the Sarbanes-Oxley Act, a company's Chief Financial Officer (CFO) and Chief Executive Officer (CEO) are required to certify the accuracy of the 10-K (SEC: Office of Investor Education and Advocacy (2011)).

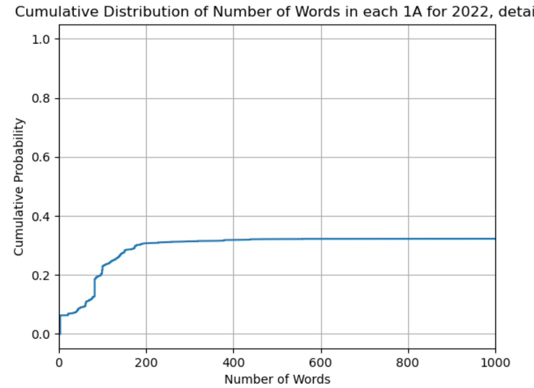
Item 1A of a 10-K ("Risk Factors") includes information about the most significant risks that apply to the company or its securities. I extract the item 1A information from each 10-K using XML parsing and `BeautifulSoup`, and removed supposedly less meaningful characters such as punctuation and numbers.

## 2.3 Filtering firms

Following Golubov and Konstantinidi (2019); Stambaugh and Yuan (2016), I filter firms by considering only ordinary common shares, traded on NYSE, AMEX, and NASDAQ exchanges; and following Stambaugh and Yuan (2016) I exclude those whose prices are less than $5 in 2016 dollars.




(a) Cumulative Distribution of Number of Words




(b) Cumulative Distribution of Number of Words, Zoom

Figure 1: Cumulative Distribution of Number of Words in 2022

Reporting risk factors is mandatory for most firms; however, there are exceptions for asset-backed issuers and smaller reporting companies. Asset-backed issuers are defined as issuers whose reporting obligation arises from either the registration of an offering of asset-backed securities under the Securities Act or the registration of a class of asset-backed securities. Firms that are not required to disclose risk factors either leave Item 1A empty or write a placeholder text specifying that, due to their nature, they are not required to disclose risk factors. Consequently, there is an abnormal frequency of 10-K filings with a significantly low number of words, as depicted in Figure 1. In subsequent stages, 1A texts with an insufficient word count are discarded. The threshold adopted was 200 words.

The total count of filtered firms by year is shown in Table 1.

## 2.4 Text conversion to bag-of-words

After filtering firms, I utilize the `spacy` Python library to perform lemmatization on all filtered texts. Lemmatization is a technique used to convert words to their root form, thus ensuring semantic consistency. For example, variations like "take", "took", and "taken" are simplified to "take". `spacy` leverages WordNet, an extensive lexical database of English maintained by Princeton University.

In order to discern common collocations—like "patent application"—which impart more semantic depth compared to individual words, this study utilizes collocation detection as detailed in Mikolov et al. (2013). This approach generates meaningful bigrams and trigrams. A minimum count threshold of 5 is established for collocations, ensuring that only statistically significant combinations are included in the dictionary.

Table 1: File Counts by Year

| Year | Total_1As | Filtered |
|---|---|---|
| 2006 | 5685 | 2466 |
| 2007 | 6445 | 2714 |
| 2008 | 6931 | 2305 |
| 2009 | 8244 | 2190 |
| 2010 | 8122 | 2290 |
| 2011 | 8019 | 2356 |
| 2012 | 7797 | 2316 |
| 2013 | 7560 | 2401 |
| 2014 | 7560 | 2518 |
| 2015 | 7531 | 2528 |
| 2016 | 7196 | 2431 |
| 2017 | 6896 | 2394 |
| 2018 | 6804 | 2418 |
| 2019 | 6683 | 2404 |
| 2020 | 6531 | 2332 |
| 2021 | 6936 | 2308 |
| 2022 | 6899 | 1885 |

The complete subset of words, bigrams, and trigrams found is used to create a dictionary, which

Finally, I converted all texts to a bag-of-words format using the dictionary and the n-gramized texts. This format retains the count of appearances for each word in a document, $c_{ij}$, but disregards word order, resulting in the final representation of the corpus.

To further augment my analysis, I match firms based on their Central Index Key (CIK) and Permanent Company Number (PERMNO), linking their annual reports to several other data sources. This includes daily stock data (which I aggregate on a weekly basis), Compustat data, and measurements of the firms' knowledge capital, their accumulated patent value, and the level of skill in their respective industries as indicated by existing literature.

## 2.5 Topic modeling

Having the corpus and the dictionary, I apply unsupervised topic modeling, specifically using the Latent Dirichlet Allocation technique, to the entire corpus of documents. These documents contain risk factors for different firms over various years. As part of the model's setup, I set a parameter for the number of topics, denoted as $k$, and feed the model with the dictionary I previously constructed. The choice of $k$ is often done *ad hoc* and is primarily driven by interpretability, as noted by Gentzkow (2019).

An example of output from such a model is shown in Figure 2.

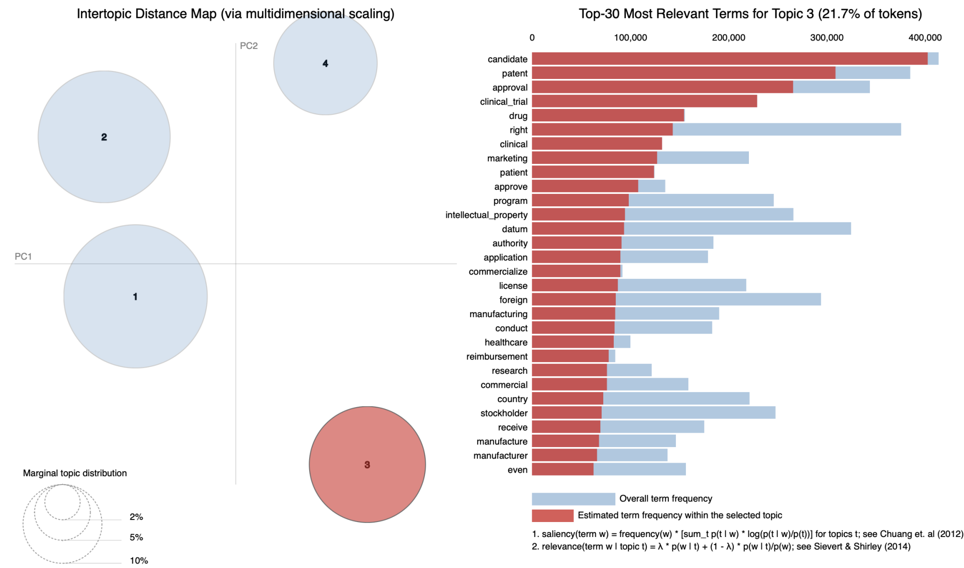



Figure 2: A graphic representation of a four-topic model on firms' risk factors since 2006.

After merging a four-topic map between firms and topic intensities to firms' identifying data, I obtain a topic map as shown in Table 2.

Table 2: Sample of topic map

| conm | year | CIK | topic_0 | topic_1 | topic_2 | topic_3 |
|---|---|---|---|---|---|---|
| BOEING CO | 2015 | 12927 | 0.875 | 0.109 | 0.016 | 0 |
| UNIFI INC | 2022 | 100726 | 0.893 | 0.107 | 0 | 0 |
| UTAH MEDICAL PRODUCTS INC | 2007 | 706698 | 0.021 | 0.457 | 0 | 0.519 |
| SPOK HOLDINGS INC | 2010 | 1289945 | 0.356 | 0.643 | 0 | 0 |
| APTARGROUP INC | 2015 | 896622 | 0.791 | 0.146 | 0 | 0.062 |
| OASIS PETROLEUM INC | 2018 | 1486159 | 1 | 0 | 0 | 0 |
| PROGRESSIVE CORP-OHIO | 2011 | 80661 | 0.051 | 0.238 | 0.711 | 0 |
| RENTRAK CORP | 2009 | 800458 | 0.017 | 0.982 | 0 | 0 |
| UNVL STAINLESS ALLOY PRODS | 2015 | 931584 | 0.957 | 0.042 | 0 | 0 |
| QUALCOMM INC | 2015 | 804328 | 0 | 0.998 | 0 | 0 |

# 3 Results

For every topic map, I define ”topic_kk” as the topic that has the highest loading within high-tech sectors in the economy, defined as SIC codes 283, 357, 466, 367, 382, 384, 737 (Brown et al. (2009))

Table 3: Topic averages by hi-tech status

| hi_tech | topic_0 | topic_kk | topic_2 | topic_3 |
|---|---|---|---|---|
| 0 | 0.49 | 0.22 | 0.27 | 0.03 |
| 1 | 0.1 | 0.58 | 0.02 | 0.3 |

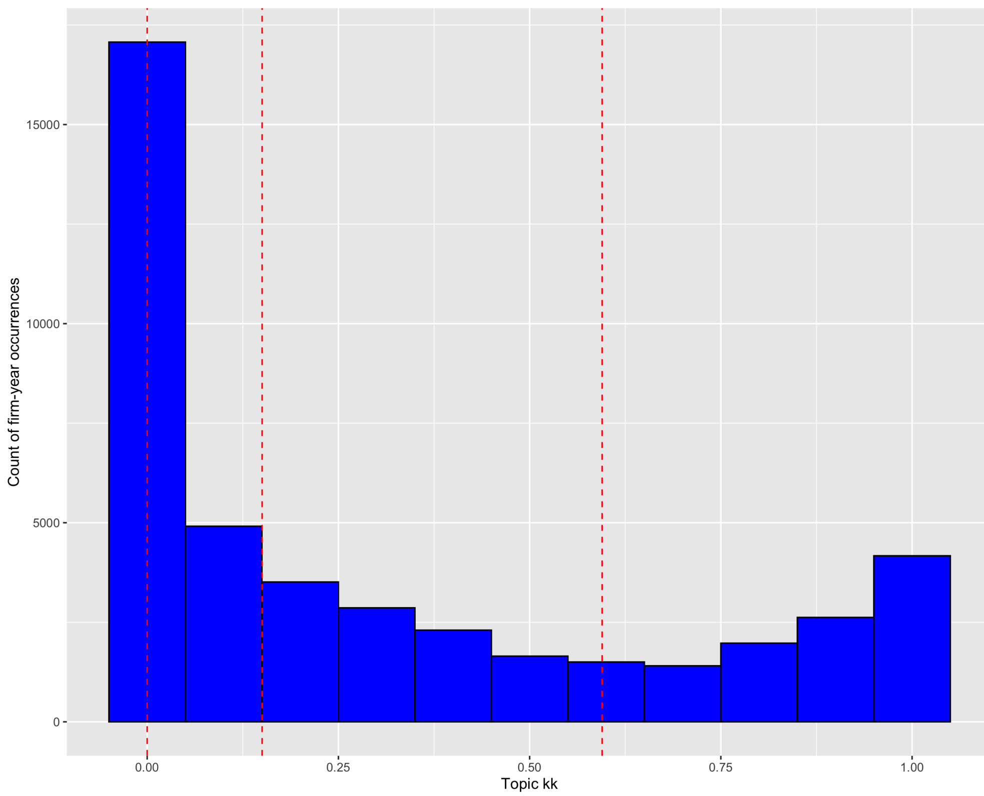



Figure 3: The histogram bars represent the frequency of topic_kk values, while the red dashed lines indicate the quartile dividers.

The mean topic intensity by year is shown in Figure 4. The figure illustrates the consistent upward trend in the mean intensity of knowledge-capital related language among a firm’s risk factors, starting from 2009.

I also analyze the performance of firms between differnet levels of knowledge-capital language. I create a varaible `kk_bin`, which assumes values from 1 to 5, according to the

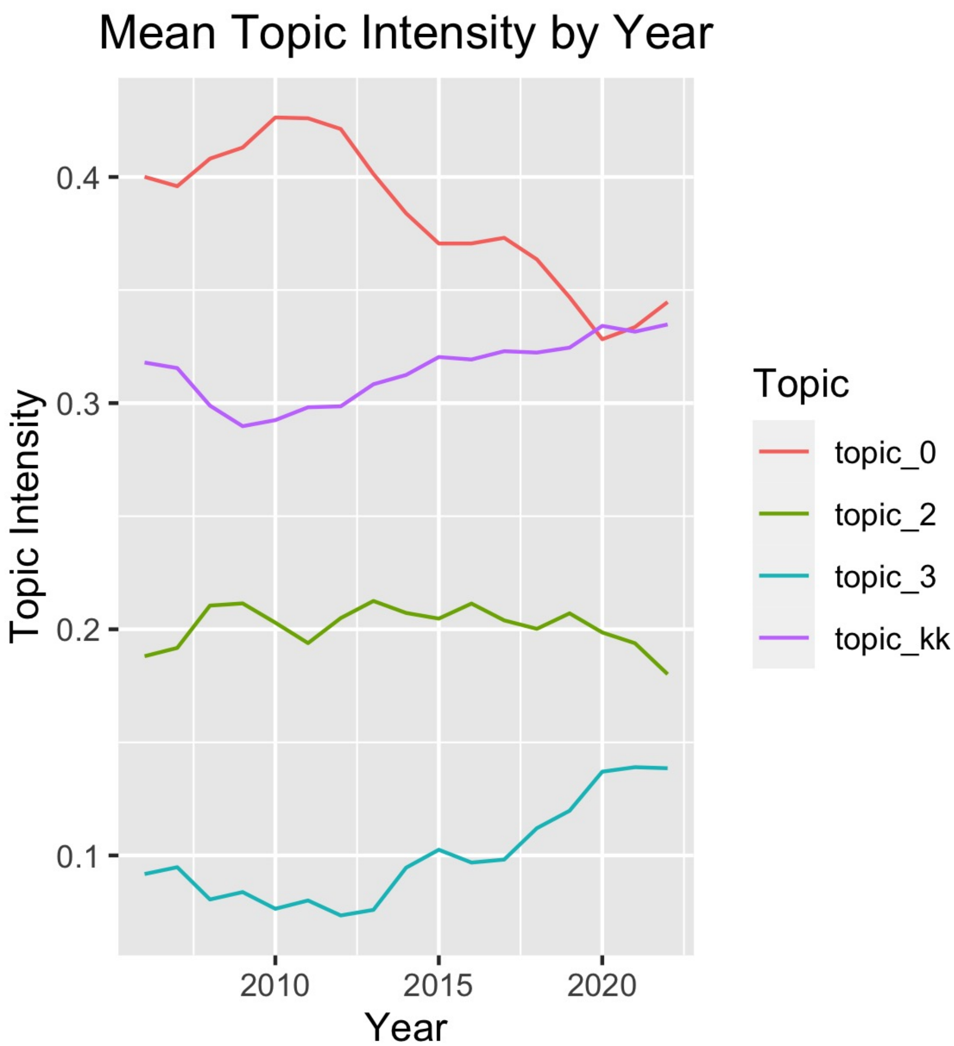



Figure 4: Mean topic intensity by year

topic's intensity (1 if between 0 and 0.2, 2 if between 0.2 and 0.4, up until 5, if between 0.8 and 1).

Figure **??** shows the monthly standard deviation of returns for the highest and lowest bins of the knowledge capital topic. Figure **??** shows the accumulated returns for each bin of the high knowledge capital topic.

Table **??** shows the average topic intensity for low- and high-tech firms, as described by.

Figure 7 creates the correlation between the intensity of each of the topics and the industry's Belo et al. (2017) average labor-skill level of employees in a given narrowly-defined industry. Figure 8 creates the correlation matrix between topic intensities and a firms' patent intensities, defined as the ratio between the firms' cumulated value of their patents as defined by Kogan and Papanikolaou (2019) and their total assets.

Table 4: Regression Summary

| | *Dependent variable:* | |
|---|---|---|
| | formula3ff | eretw |
| | (1) | (2) |
| kkhml | 0.0002 | 0.002 |
| | (0.003) | (0.003) |
| HML | $-0.006^{***}$ | $-0.006^{***}$ |
| | (0.001) | (0.001) |
| SMB | $-0.001^{***}$ | $-0.002^{***}$ |
| | (0.0005) | (0.001) |
| Mkt.RF | $0.007^{***}$ | $0.007^{***}$ |
| | (0.002) | (0.002) |
| Constant | $-0.004^{**}$ | $-0.003$ |
| | (0.002) | (0.002) |
| Observations | 25 | 25 |
| $R^2$ | 0.822 | 0.818 |
| Adjusted $R^2$ | 0.786 | 0.782 |
| Residual Std. Error (df = 20) | 0.001 | 0.00001 |
| F Statistic (df = 4; 20) | $23.072^{***}$ | $22.532^{***}$ |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

Table 5: Topic averages by hi-tech status

| hi_tech | topic_0 | topic_kk | topic_2 | topic_3 |
|---|---|---|---|---|
| 0 | 0.49 | 0.22 | 0.27 | 0.03 |
| 1 | 0.1 | 0.58 | 0.02 | 0.3 |

Table 6: Regression Summary

| | *Dependent variable:* | |
|---|---|---|
| | formula3ff | eretw |
| | (1) | (2) |
| kkhml | 0.0002 | 0.002 |
| | (0.003) | (0.003) |
| HML | $-0.006^{***}$ | $-0.006^{***}$ |
| | (0.001) | (0.001) |
| SMB | $-0.001^{***}$ | $-0.002^{***}$ |
| | (0.0005) | (0.001) |
| Mkt.RF | $0.007^{***}$ | $0.007^{***}$ |
| | (0.002) | (0.002) |
| Constant | $-0.004^{**}$ | $-0.003$ |
| | (0.002) | (0.002) |
| Observations | 25 | 25 |
| $R^2$ | 0.822 | 0.818 |
| Adjusted $R^2$ | 0.786 | 0.782 |
| Residual Std. Error (df = 20) | 0.001 | 0.00001 |
| F Statistic (df = 4; 20) | $23.072^{***}$ | $22.532^{***}$ |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

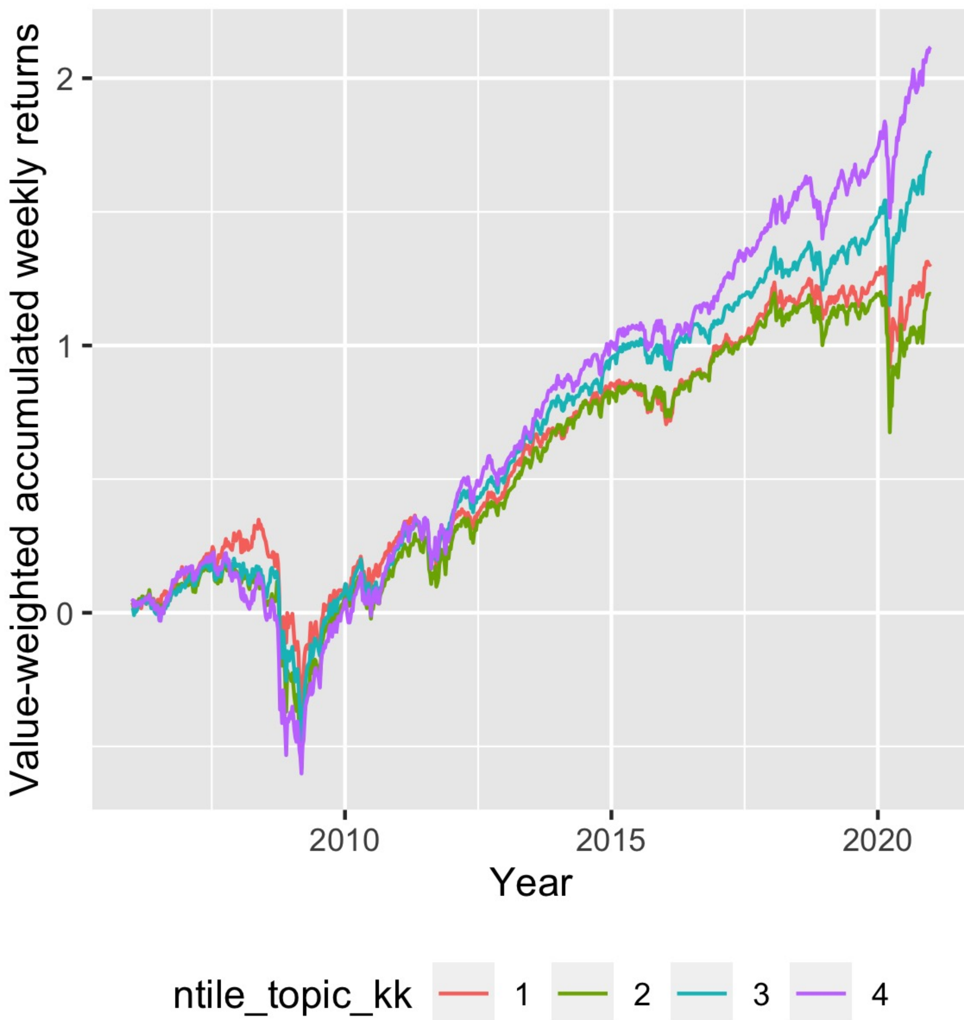



Figure 5: Asset-weighted accumulated weekly returns

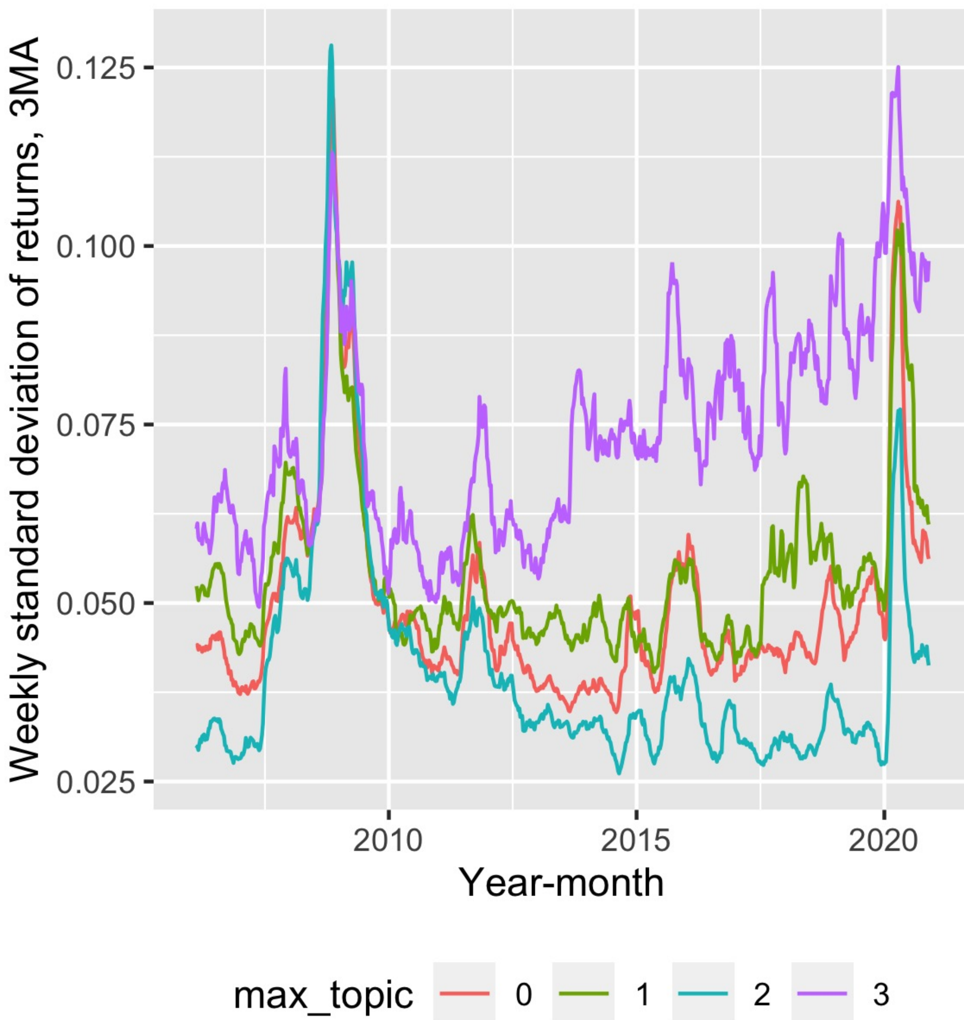



Figure 6: Weekly standard deviation of returns by group

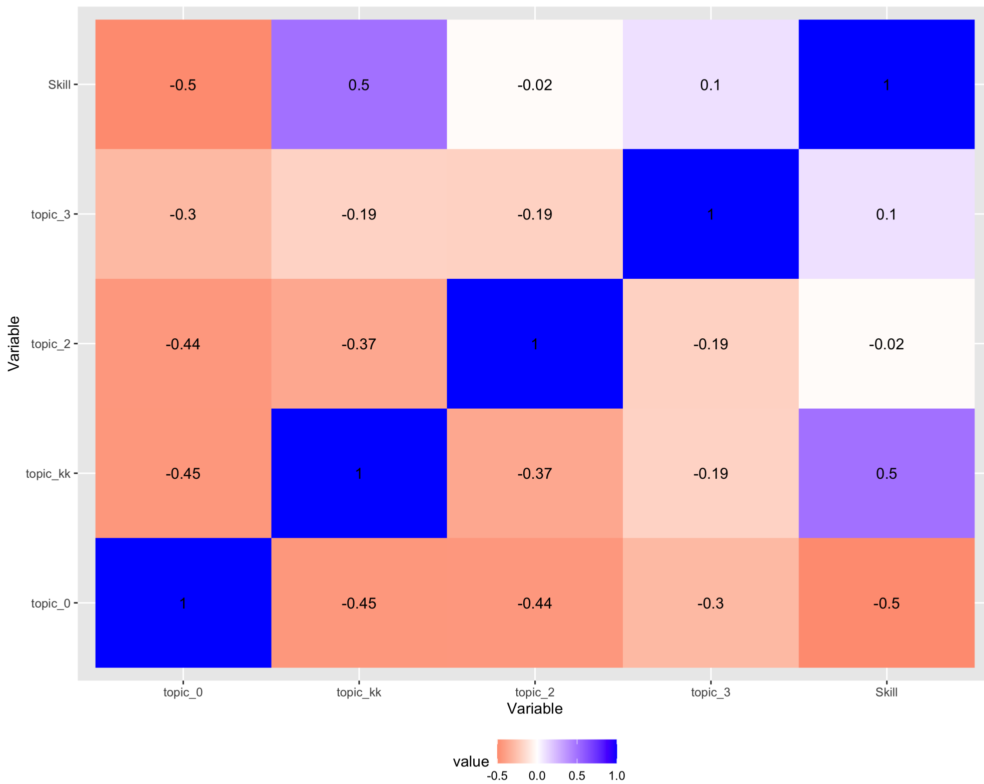



Figure 7: Correlation matrix between topics and skill

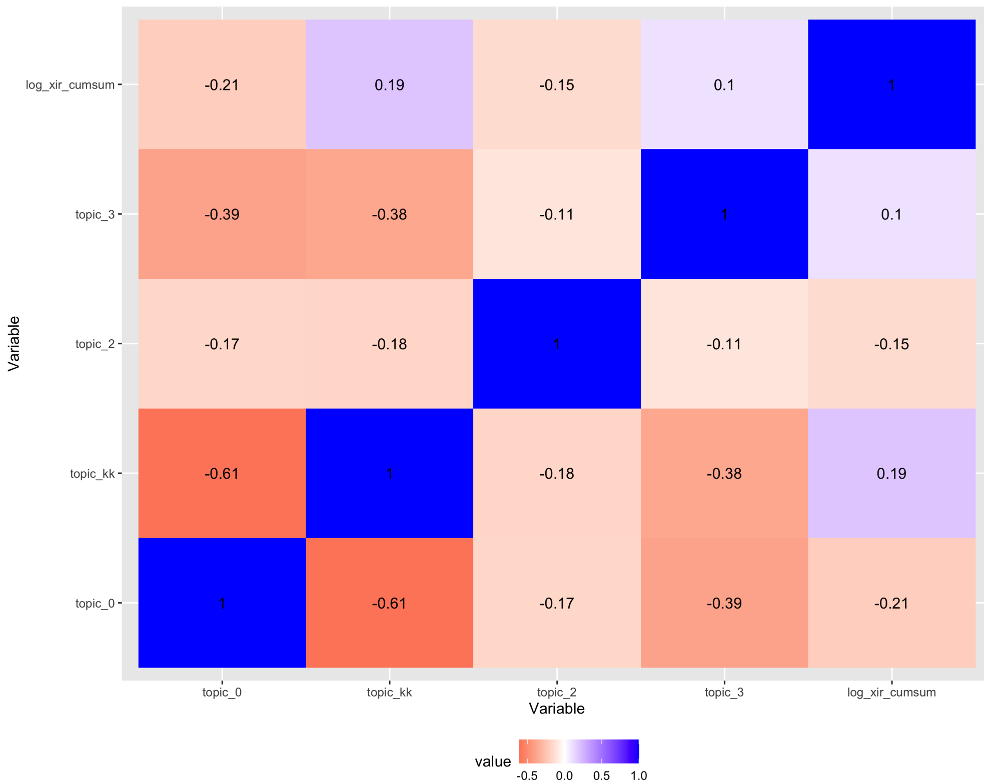



Figure 8: Correlation matrix between topics and patent intensity

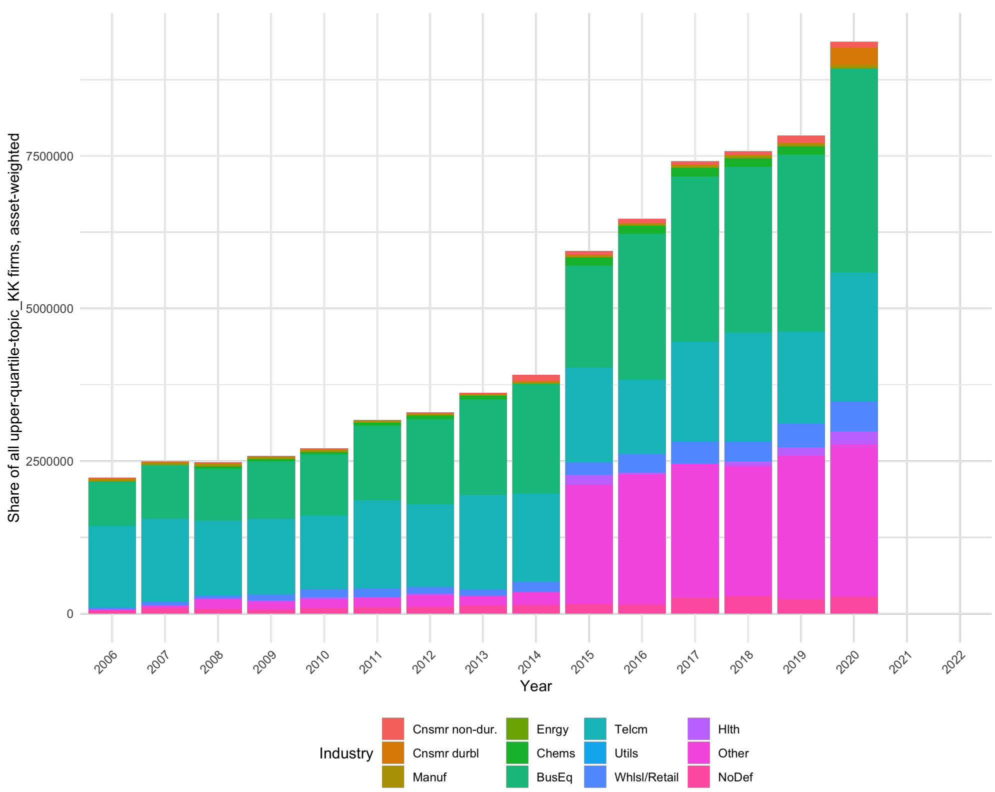




Figure 9: Stacked plot of firms' total assets in the upper quartile of KK topic intensity

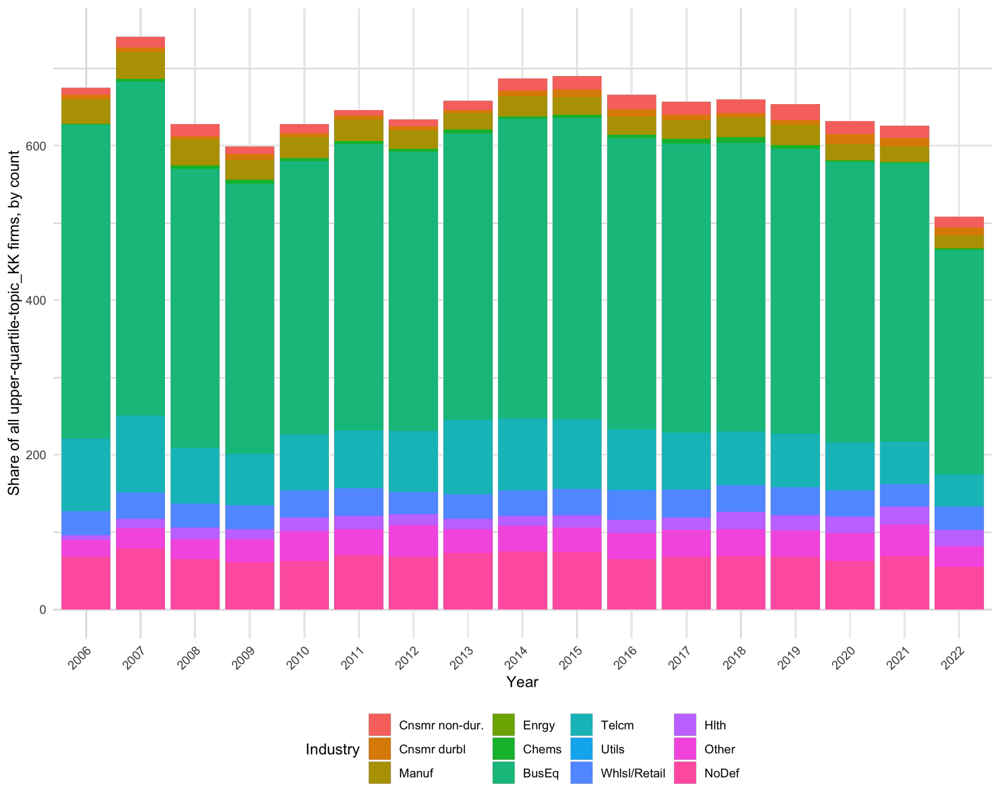



Figure 10: -

# References


Ai, Hengjie, Jun Li, Kai Li, and Christian Schlag (2019) "The Collateralizability Premium."

Andrei, Daniel, William Mann, and Nathalie Moyen (2019) "Why did the q theory of investment start working?."

Atkeson, Andrew and Ariel Burstein (2019) "Aggregate Implications of Innovation Policy," *J. Polit. Econ.*, 127 (6), 2625–2683.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen (2020) "The Fall of the Labor Share and the Rise of Superstar Firms*," *Q. J. Econ.*, 135 (2), 645–709.

Barkai, Simcha (2020) "Declining labor and capital shares," *J. Finance*, 75 (5), 2421–2463.

Barro, Robert J (2006) "Rare disasters and asset markets in the twentieth century," *Q. J. Econ.*, 121 (3), 823–866.

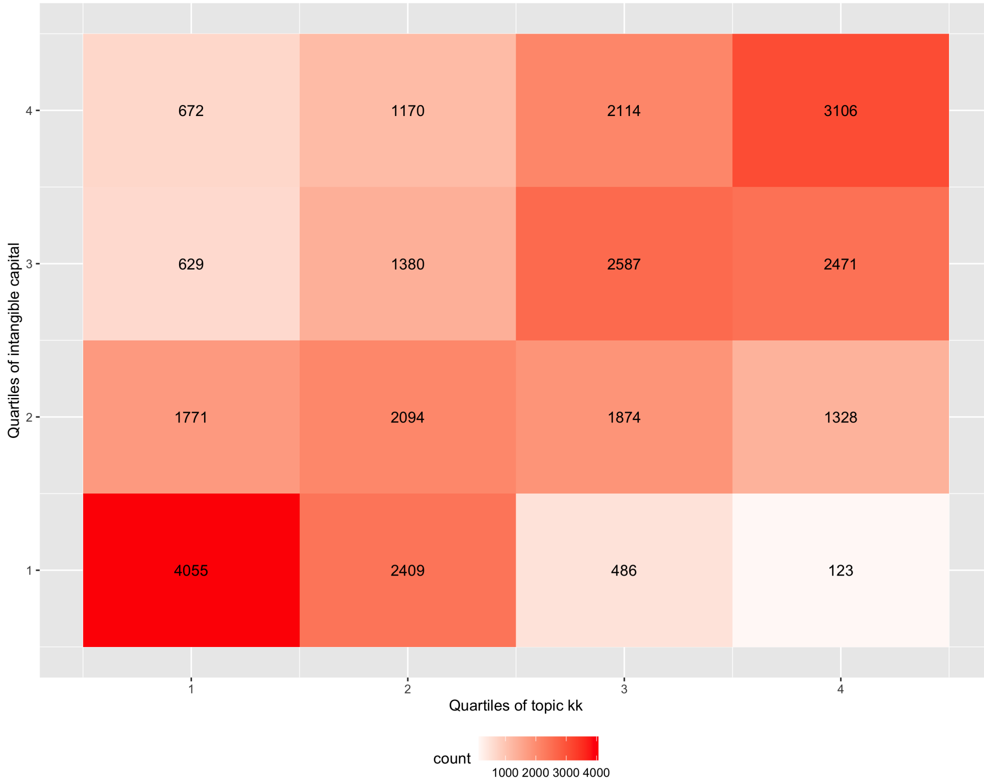



Figure 11: -

Belo, Frederico, Vito D Gala, Juliana Salomao, and Maria Ana Vitorino (2019) "Decomposing Firm Value," *J. financ. econ.*, 143 (2), 619–639.

Belo, Frederico, Jun Li, Xiaoji Lin, and Xiaofei Zhao (2017) "Labor-force heterogeneity and asset prices: The importance of skilled labor," *Rev. Financ. Stud.*, 30 (10), 3669–3709.

Belo, Frederico, Chen Xue, and Lu Zhang (2013) "A supply approach to valuation," *Rev. Financ. Stud.*, 26 (12), 3029–3067.

Blei, David M, Andrei Ng, and Michael Jordan (2003) "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*

Bloom, N and J Van Reenen (2007) "Measuring and explaining management practices across firms and countries," *Q. J. Econ.*, 122 (4), 1351–1408.

Brown, James R, Steven M Fazzari, and Bruce C Petersen (2009) "Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom," *J. Finance*, 64 (1), 151–185.


Figure 12: -


Corrado, Carol, John Haltiwanger, and Daniel Sichel (2009a) *Measuring Capital in the New Economy*: University of Chicago Press.

Corrado, Carol, Charles Hulten, and Daniel Sichel (2009b) "Intangible capital and u.S. Economic growth," *Rev. Income Wealth*, 55 (3), 661–685.

Eisfeldt, Andrea L, Antonio Falato, and Mindy Z Xiaolan (2018) "Human Capitalists," April.

Eisfeldt, Andrea L, Edward Kim, and Dimitris Papanikolaou (2020) "Intangible Value,"Technical Report w28056, National Bureau of Economic Research.

Eisfeldt, Andrea L and Dimitris Papanikolaou (2013) "Organization Capital and the Cross-Section of Expected Returns."

Fernald, J G (2015) "Productivity and Potential Output before, during, and after the Great Recession," *NBER macroeconomics annual*.

Gabaix, Xavier (2012) "Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance," *Q. J. Econ.*, 127 (2), 645–700.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J Klenow (2019) "How destructive is innovation?" *Econometrica*, 87 (5), 1507–1541.

Golubov, Andrey and Theodosia Konstantinidi (2019) "Where is the risk in value? Evidence from a market-to-book decomposition," *J. Finance*, 74 (6), 3135–3186.

Gourio, François (2012) "Disaster risk and business cycles," *Am. Econ. Rev.*, 102 (6), 2734–2766.

Grossman, Gene M and Elhanan Helpman (1991) "Quality ladders in the theory of growth," *Rev. Econ. Stud.*, 58 (1), 43.

Guvenen, Fatih, Raymond J Mataloni, Jr, Dylan G Rassier, and Kim J Ruhl (2021) "Offshore profit shifting and aggregate measurement: Balance of payments, foreign investment, productivity, and the labor share,"Technical report, Updated from NBER Working Paper.

Hall, Robert E (2001) "The stock market and capital accumulation," *Am. Econ. Rev.*, 91 (5), 1185–1202.

Hansen, Lars Peter, John C Heaton, and Nan Li (2005) "Intangible risk," in *Measuring Capital in the New Economy*, 111–152: University of Chicago Press.

Kogan, L, D Papanikolaou, A Seru, and Noah Stoffman (2017) "Technological innovation, resource allocation, and growth," *The Quarterly Journal*.

Kogan, Leonid and Dimitris Papanikolaou (2019) "Technological Innovation, Intangible Capital, and Asset Prices," *Annu. Rev. Financ. Econ.*, 11 (1), 221–242.

Li, Dongmei (2011) "Financial Constraints, R&D Investment, and Stock Returns."

Li, Erica X N, Laura Xiaolei Liu, and Chen Xue (2014) "Intangible Assets and Cross-Sectional Stock Returns: Evidence from Structural Estimation," May.

McGrattan, Ellen R and Edward C Prescott (2001) "Is the Stock Market Overvalued?" January.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (2013) "Distributed Representations of Words and Phrases and their Compositionality."

Peri, G, K Shih, and C Sparber (2015) "STEM workers, H-1B visas, and productivity in US cities," *J. Labor Econ.*

Peters, Ryan H and Lucian A Taylor (2017) "Intangible capital and the investment-q relation."

Rietz, Thomas A (1988) "The equity risk premium a solution," *J. Monet. Econ.*, 22 (1), 117–131.

Romer, Paul M (1990) "Endogenous Technological Change," *J. Polit. Econ.*, 98 (5, Part 2), S71–S102.

Schumpeter, Joseph A (1939) *Business cycles*, 1: Mcgraw-hill New York.

SEC: Office of Investor Education and Advocacy (2011) "Investor Bulletin: How to Read a 10-K."

Stambaugh, Robert F and Yu Yuan (2016) "Mispricing Factors," *Rev. Financ. Stud.*, 30 (4), 1270–1315.

Unger, Gabriel (2019) "Scale-Biased Technological Change,"Technical report, Harvard Working Paper.

Vitorino, Maria Ana (2014) "Understanding the Effect of Advertising on Stock Returns and Firm Value: Theory and Evidence from a Structural Model," *Manage. Sci.*, 60 (1), 227–245.