Is intangibles talk informative about future returns? Evidence from 10-K filings

Alexander David [a], Amir Hosseini [a, *], Anup Srivastava [a]

[a] University of Calgary, Haskayne School of Business, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

[*] Corresponding author. Tel.:+1 403-604-2149; E-mail addresses: alexander.david@haskayne.ucalgary.ca (A. David), seyedamirfarhang.hos@ucalgary.ca (A. Hosseini), anup.srivastava@ucalgary.ca (A. Srivastava)

June 2023

ABSTRACT

We construct a measure of intangible intensity — intangibles talk — based on textual analysis of discussions on intangibles in a firm's 10-K filings. Intangibles talk covers the main three categories of intangible value in the literature: innovation assets and information technology, brand and customer relations, and human resources. Our measure is correlated with prior accounting measures of intangibles in our panel of firms. We analyze the relationship between intangible value and stock returns by examining the informativeness of our measure about future returns. Returns from long and short value-weighted portfolios based on high and low values of intangibles talk, respectively, outperform traditional book-to-market value strategy and its intangible augmented versions. Our strategy delivers its strongest performance from 2008 to 2020 with an average annual returns of 7.9%, in contrast to the poor performance of value strategies for the same period. Our strategy generates an average annual alpha of 3.26% from 1995 to 2020 in the four-factor (three Fama and French factors plus momentum) model. Our alphas are higher than those generated from portfolios sorted on other indicators of intangible intensity shown in the literature. Positive alphas are concentrated in stocks with higher arbitrage risk, proxied by idiosyncratic volatility, suggesting that investors misprice stocks with higher intangible intensity.

JEL classification: E22; G14; O3
Keywords: 10-K filings, Intangibles talk, Intangible intensity, Portfolio returns, Idiosyncratic volatility

1. Introduction

The United States economy has experienced a dramatic shift towards intangible assets in recent decades. Investment in knowledge capital and organizational capacity among US firms has risen steadily (see Corrado, Hulten and Sichel, 2005; Eisfeldt and Papanikolaou, 2013; Enache and Srivastava, 2018; Ewens, Peters and Wang, 2019), allowing them to launch new products and services, or gain a competitive edge in existing marketplaces, through innovation, lower costs, and better customer relations. Due to current accounting rules about expensing of internally generated intangibles, they are largely missing from the balance sheet, and there exist no reliable measures of firms' total intangible capital. As a result, knowledge firms have more mispriced securities than do firms with physical assets (see Lev and Sougiannis, 1996; Daniel and Titman, 2006; Eisfeldt and Papanikolaou, 2013; Edmans, Li and Zhang, 2014). Meanwhile, the importance of intangibles in the U.S. economy keeps increasing, as each new cohort of public firms spends more on intangibles than its predecessor cohort (Corrado, Hulten and Sichel, 2005; Srivastava, 2014, 2023).

Numerous studies attempt to address investors' problem by estimating the value of internally generated capital, by capitalizing and amortizing research and development (R&D) and selling, general, and administrative (SG&A) expenses.[1] Edmans, Li and Zhang (2014) signifies the challenges of accurately defining and gauging intangible value. We extend this literature by proposing a new, text-based measure that relies on intangibles-related keywords appearing in firm's 10-K filings [following the glossary created by Filipovic and Wager (2019)]. We demonstrate the informativeness of our measure through alphas generated by long-short portfolios formed on our measure while controlling for momentum and Fama and French factors.

[1] Those studies use perpetual inventory model. See, Hulten and Hao (2008), Peters and Taylor (2017), Lev and Srivastava (2022), Eisfeldt, Kim and Papanikolaou (2020), Iqbal et al. (2022), and Falato et al. (2022)

US Generally Accepted Accounting Principles (GAAP) requires that expenditures on internally generated intangibles be immediately expensed. The same GAAP rules permit capitalization of expenditures on on property, plant, and equipment (PP&E) and acquired intangibles. As a result, information on in-house-developed intangible assets such as innovation, knowledge, and brand capital are not readily available in financial statements compared with data related to tangible capital (Belo et al., 2022). To address this accounting limitation, improving on measurements and proposing new methods to estimate intangible capital has been the focus of multiple studies in both finance and accounting literature, Peters and Taylor (2017), Enache and Srivastava (2018), Park (2019), Lev and Srivastava (2022), and Eisfeldt, Kim and Papanikolaou (2020). Most of those studies rely on perpetual inventory model, that is, capitalizing past R&D and SG&A expenditures. While the new methods yield improvements over models that ignore in-house intangibles, they suffer from two limitations. First, no consensus exists on what percentage of intangible expenditures should be capitalized. Capitalization percentages used by those studies range from 30% to 100%. Second, any capitalized intangible stock measure, based on past expenses, does not take into account the lottery type payoff that often comes from serendipitous investments. For example, the discovery of a search formula by Google founders led to a trillion-dollar valuation company, and no amount of capitalization of past expenses would yield a number close to the value of that discovery. Managers, however, are likely to describe the same successful discoveries and self-developed intangible assets in their communications with investors, if they are expected to create benefits for the company.

We contribute to the literature by identifying another measure of intangibles, which arguably tracks developed intangibles, but is not yet reflected in financial numbers presented in the balance sheet. We extract the informational content relating to intangible capital embedded in the text of a firm's 10-K filings. Financial Accounting Standards Board (FASB) Concepts Statement No. 5 prescribes strict criteria for recognition transaction in financial numbers, such as measurability, relevance, and reliability. Items that fail these criteria but are value-relevant nevertheless are often disclosed in footnotes and in the management discussion and analysis (MD&A) section of the 10-K. Managers describe their assessments of items that will impact future operations and whose discussion will enhance investors' understanding of firm operations. Any forward-looking information supplied in the MD&A section is expressly covered by the safe harbor rule, a legal provision that shields managers from liability if future projections go wrong. Textual portion, thus, is particularly useful for conveying "soft" information (Seamons and Rouse, 1997). Hence, the textual portion of 10-K filings could provide guidance to investors on the value of internally generated intangible capital, particularly the value that cannot be recognized in the balance sheet for lack of reliability and constraints imposed by accounting rules.

We conduct textual analysis to gauge the intensity of intangibles discussion in 10-K filings and consider it a proxy for intangible value. We focus on three distinct categories of intangible assets: innovation assets and information technology, brand and customer relations, and human resources. Our decision to concentrate on only these three categories is motivated by similar classifications of intangible investments used in the literature (Lev, 2000; Corrado, Hulten and Sichel, 2005). We construct our measure of intangibles talk using a glossary of intangible terms created by Filipovic and Wager (2019). A benefit of using a text-based measure, relative to capitalizing past expenses, is the possibility of mapping words to, and classifying them in, the aforesaid three categories, while separately analyzing their informativeness with respect to future returns.

Our textual measure assumes that the intensity of intangible-related discussion in 10-K filings represents the emphasis that management places on intangible development and its importance to firm operations. Our measure is the relative frequency of words on intangibles topics to total words in 10-K filings. Scaling the frequency of intangibles words by total words allows for comparison across firms with varying document sizes. We start by testing the validity of our measure. We find that intangibles talk positively correlates with conventional accounting measures of intangible investments such as R&D and SG&A expenditures (both scaled by total expense). Also, intangibles talk correlates positively with another indicator of intangible value called intangible capital advanced by Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020). Furthermore, our measure has a positive correlation with market-to-book ratio. Considering that intangibles talk is constructed based entirely on textual material, as opposed to market or book data, points to a strong link between intangibles discussions and underlying intangible capital. We view this link as preliminary evidence that intangibles talk tracks variations in intangible capital and intensity over time.

We analyze variations in our measure across firms in Fama and French twelve industries. As expected, industries such as health care and business equipment score the highest in intangibles talk, and the lowest intangibles talk belongs to the finance and energy industries. We also decompose intangibles talk into its three categories and investigate the highest-ranking industries under each of them. Results are consistent with intuition. For instance, while the healthcare industry scores the highest on intangibles talk focused on innovation assets and human resources, it scores among the lowest in the brand and customer relations category.

Our main tests of intangibles talk are informativeness with respect to future stock returns. Classic studies, as well as recent studies, investigate the mispricing of intangibles information. Lev and Sougiannis (1996) find a systematic mispricing of R&D-intensive stocks, and show that incorporating that information leads to an annual return of 4.57%. Chan, Lakonishok and Sougiannis (2001) find that stocks with high R&D relative to the market value of equity deliver an average of 6.12% annual returns. They show similar results for stocks with high advertising expenses. Other studies link excess returns to patent citations (Deng, Lev and Narin, 1999), software developments (Aboody and Lev, 1998), and employee satisfaction (Edmans, Li and Zhang, 2014). Our argument to support the hypothesis on the link between intangibles talk and future returns is similar to Edmans (2011), which points to the insufficient salience of intangibles information that could lead to its overlooking investors. Recent studies (Lev and Srivastava, 2022; Arnott et al., 2021; Choi, So and Wang, 2021) augment the book-to-market measure of value investing by intangibles estimates and show superior returns than non-augmented book-to-market based HML returns reported based on Fama and French (2015).

Prior studies on disclosure versus recognition argue that textual content on intangibles in documents such as 10-K filings are more likely to be ignored by investors than are reported numbers (Aboody, 1996; Davis-Friday et al., 1999; Ahmed, Kilic and Lobo, 2006). Based on this idea, we hypothesize that intangibles talk could convey informative signals that are likely missed by investors. Eisfeldt, Kim and Papanikolaou (2020) show that intangible augmented value factors outperform the traditional value factor and thus contribute to the literature by pushing the limits of the available accounting data to capture intangible value. We follow this idea to investigate our hypothesis. We first examine whether we can generate positive, risk-adjusted returns based on long-short portfolios formed on intangibles talk. We then compare these results with returns from other benchmark value signals, namely book-to-market and its intangible augmented versions from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020).

We begin our analysis by sorting portfolios based on intangibles talk, INT^{10K}. We follow the long-short sorting methodology presented in Fama and French (2015) and construct HML^{FF}, which captures the traditional value strategy based on book-to-market. Similarly, we construct HML^{PT} and HML^{EKP} using the intangible augmented book-to-market. The return analysis for INT^{10K} suggests positive and significant monthly returns between 1995 to 2020. To illustrate the relative magnitude of returns, we compare INT^{10K} returns against HML^{FF}, HML^{PT}, and HML^{EKP} portfolios (value strategies).

We find that INT^{10K} performance is strongest amongst the three value strategies for the sample period, covering July 1995 to June 2020, while excluding the years that represent the dot-com bubble burst (2000–2001).[2] We achieve an annualized average monthly return of 4.95%. The returns are particularly high at around 7.9% in the period between 2008 to 2020. The average returns are consistently positive for each year since 2007, with the exception of 2012 and 2016 (see Fig. 1).INT^{10K} performs most strongly during the period that records the worst performance of HML^{FF} as shown to be in the post–financial crisis era (2008–2020) (see Eisfeldt, Kim and Papanikolaou, 2020). Our graph of cumulative returns is consistent with the growing importance of intangible value (see Fig. 2). Our portfolio's return displays steady growth over time and reaches its highest levels by the end of our study period in 2020. INT^{10K}'s outperformance relative to HML^{FF}, HML^{PT}, and HML^{EKP}, especially in recent years, suggests that incorporating textual information holds great promise as a separate source of intangible value for investors besides accounting data. Furthermore, the fact that returns can be generated using textual data indicates that text-based intangibles information is not fully incorporated by the investors.

[Insert Figs.1 and 2 near here]

We conduct a more detailed examination on the relative informativeness of intangibles talk against other value indicators. We follow the strategy in Eisfeldt, Kim and Papanikolaou (2020); that is, we go long on INT^{10K} and short other value portfolios. This strategy enables us to show that our measure captures orthogonal, and perhaps superior, information compared with augmented value strategies. We find that this long-short portfolio generates significant positive returns over HML^{FF} except for the period after the dot-com bubble (2000–2007). We find results when the short leg of the portfolio is HML^{PT} and HML^{EKP}, providing

[2]We argue that the weak historical performance of INT^{10K} near the burst of the dot-com bubble results from the massive overvaluation of technology stocks in previous years. Technology stocks typically score high in intangibles talk and are thus mostly picked up by our sorting methodology, which relies on intangibles talk values.

strong evidence that our measure has additional, and arguably more unpriced, information than intangible augmented versions of value indicators in prior studies.

We next test whether the returns generated by intangibles talk represent premiums for risk or some other unpriced factor. We generate positive alphas while controlling for momentum (Carhart, 1997) and the three as well as the five Fama and French factors. The alpha averages 3.26% and 5.59% from 1995 to 2020 in the three- and five-factor settings, respectively. The alpha remains positive and significant when we replace HML^{FF} with portfolios HML^{PT} and HML^{EKP} as the value factor in the Fama and French regression models. This suggests that the INT^{10K} return cannot be explained by traditional risk measures. We also create five value-weighted portfolios ranked based on intangibles talk and show that the alpha of the highest-ranking portfolio minus the lowest-ranking portfolio is 6.46% and significant.

We test mispricing as an explanation for return informativeness of intangibles talk. We rely on the idea that higher idiosyncratic volatility (IVOL) should amplify returns that result from mispricing. Higher IVOL leads to greater arbitrage risk, which, in turn, limits the ability of rational investors to correct mispricing (Pontiff, 1996; Stambaugh, Yu and Yuan, 2015)[3]. In the case of intangible-intensive stocks, under higher IVOL market conditions, rational investors are less eager to bid aggressively against overlooking of intangibles information. Accordingly, we argue that, if intangible–intensive stocks are mispriced, then amplification of such an effect should be evident under market conditions with higher arbitrage risk. To test INT^{10K} abnormal returns in relation to IVOL, we first construct firm-level IVOL by estimating the volatility of the residuals from the three factors of Fama and French for daily returns (Ang et al., 2006). Our double-sorted portfolio returns based on IVOL and intangibles talk

[3]They show that among overpriced (underpriced) stocks the ones with the highest IVOL are the most overpriced (underpriced). Another recent study, Birru and Young (2020), utilizes IVOL to show stronger return predictions of investor irrationality. It measures, via IVOL, the firm-level uncertainty and shows that with higher uncertainty, limits to arbitrage allow for investor sentiment to be a stronger predictor of subsequent returns.

is consistent with the idea of mispricing. We find that alphas of portfolios sorted on intangibles talk increase in IVOL, are most significant for high IVOL stocks, and are insignificant at low levels of IVOL. We therefore conclude that intangible–intensive stocks experience greater mispricing under high arbitrage risk, as evidenced by the strongest INT 10K abnormal returns among high IVOL stocks. Our study mainly contributes to the emerging stream of literature, aiming to create alternative measures of intangible intensity, and using them to earn investment returns. In addition, our study is closely related to the literature on the use of textual analysis to predict returns (see Tetlock, 2007; Garcia, 2013; Jiang et al., 2019). We primarily rely on the bag of words approach in our textual analysis, which is the standard method used in the literature since Loughran and McDonald (2011). Most studies in the textual analysis literature measure tone sentiment or uncertainty, relying on various sources such as the Loughran & McDonald dictionary (see Loughran and McDonald, 2011). Our study in this regard deviates from such studies and is closer to those that use specialized glossaries to gauge the intensity of discussion surrounding particular topics such as climate change (see Engle et al. (2020)).

This paper proceeds as follows. In Section 2, we describe the methodology to construct intangibles talk and discuss and report its values across industries and firm characteristics. Section 3 presents the results from our portfolio analysis. Section 4 examines the common risk factors to explain the portfolio returns. In Section 5, we present and test the hypothesis that explains the abnormal returns of our portfolio. Section 6 concludes.

2. Data and methodology

2.1 SEC filings

The corpus or textual data for our analysis comes from is 10-K filings submitted to the Securities and Exchange Commission (SEC) by 12,184 public firms from January 1994 to December 2021. This adds up to a total of approximately 107,000 10-K filings in our analysis. The filings are collected using the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system[4]. In the parsing process, numbers, tables, and figures are removed for the raw text to be ready for textual analysis. To reduce the noise in the text, we remove all the stop words.[5]

To construct intangibles talk, we use the bag of words approach. A filing's vector of words is analyzed against a glossary of intangibles terms (see Table 1) based on the recently developed intangibles words list in Filipovic and Wager (2019). The list is derived from several studies on intangible assets such as Hall (2009) and Lev (2005, 2012). A more detailed description of how our intangibles word list was developed can be found in Filipovic and Wager (2019).

[Insert Table 1 near here]

We focus on three broad categories of intangible assets: innovation assets and information technology, brand and customer relations, and human resources. Lev (2005) claims that information technology is a component of this category that has grown in importance in recent years and is related to Internet platforms, software solutions, and so on. To account for this aspect of the organizational capital in intangible talk, we create a new combined category called innovation assets and information technology that is closer to what Wyatt (2008) calls

[4]We use the parsed documents publicly available on Loughran-McDonald webpage at: https://sraf.nd.edu/loughranmcdonald-master-dictionary/

[5]Examples of stop words are "a", "the", "are", "and", "could", "would."

technology resources. Table 2 contains the words from the intangible word list belonging to each category.

[Insert Table 2 near here]

Table 2 shows that words under the innovation assets and information technology category are mainly related to innovation along with online infrastructure such as databases, websites, and platforms. In the brand and customer relations category, we identify reputation, brands, and relations. In the last category, human resources, we emphasize skills, abilities, and competencies. Table 3 shows that most variance in intangibles words frequency across filings is related to the innovation assets and information technology category. This category contributes 47% to the total variance, and the words belonging to it account for 54% of the words in the glossary.

[Insert Table 3 near here]

We define intangibles talk for 10-K filing i with n intangibles words as the sum of the frequency across all intangibles words divided by the total number of words in the filing:

$$\mathit{Intangibles\ talk}_{\mathbf{i}} = \frac{\sum_{\mathbf{j=1}}^{\mathbf{n}} \mathit{Frequency\ of\ intangible\ word}_{\mathbf{j}}}{\mathit{Total\ words}_{\mathbf{i}}} \tag{1}$$

The ratio reflects the relative intensity of discussion surrounding intangibles topics in filings. As Filipovic and Wager (2019) assert, the intangibles word list is neither optimized nor reverse-engineered to fulfill return maximization objectives. In a similar fashion, we do not select targeted words from intangibles word list to maximize our portfolio returns using machine learning techniques.

Fig. 3 plots the top 30 most frequent intangibles words across firms in our sample. Some of the most common words, such as "employee", "customers", and "data," can be found under various discussion topics that are not related to intangible assets. Our glossary contains 128 words, of which the top three in Fig. 3 account for almost 20% of the total frequency while their share of the total frequency would be 2.3% under the equally distributed case. This illustrates the power law distribution of words frequency in natural languages, also known as Zipf's law.[6]

[Insert Fig. 3 near here]

2.2 Intangibles talk across industries

Fig. 4 depicts a summary of intangibles word frequency along with intangibles talk across the twelve Fama and French industries. Healthcare and business equipment rank the highest based on the median values of intangibles talk. The lowest-ranking industries are finance and energy. Intangible intensity in the healthcare and business equipment industries likely reflects their principal sources of competitive advantage, such as patents, data, technical knowhow, and information technology. In contrast, finance and energy rely heavily on tangible assets and financial capital. Later, we delve deeper into each industry's most frequent intangibles words to identify the main components of intangibles talk in that industry. Fig. 4, shows that the absolute frequency of intangibles words is high for some industries and the relative frequency (intangibles talk) can be low in some cases because of varying document sizes across industries.

[Insert Fig. 4 near here]

[6] According to Zipf's law the frequency of words is proportional to the inverse of their ranking: f(r) $\propto \frac{1}{r^{\alpha}}, with \alpha \approx 1$ (see Mandelbrot, 1961).

The portion of intangibles talk's three categories across the twelve industries reveal the concentration of each class of intangible assets in the economy. Fig. 5 shows that the consumer non-durables industry ranks high in the brand and customer relations category, indicating the importance of brands, sales and distribution network, and customer satisfaction in that industry. In contrast, the healthcare industry ranks low in the brand and customer relations category and ranks high in the human resources and innovation assets categories. This points to the central role that qualified professionals play in the healthcare industry. They are responsible for the quality and efficiency of services. In addition, patents for new drugs and medical devices are important sources of revenue in the pharmaceutical industry.

[Insert Fig. 5 near here]

Another important feature of our measure is that it uncovers the flavor of intangible value across firms and industries. Fig. 6 plots the top ten most frequent intangibles words by twelve industries. Words such as ”software”, ”data”, and ”technology” that fall under intangible value associated with information technology appear to be frequent and concentrated in the business equipment industry. Another word, ”employee,” appears as the most frequent word across most industries. As a caveat, the context in which common words such as ”employee” and ”data” are discussed is important in classifying them as intangibles words. One limitation of our measure is that it does not identify the context in which the words are presented. Nonetheless, when a group of words that refer to a particular category of intangible assets are concentrated in one industry, it indicates the importance of that category of intangible assets. For instance, the word ”customer” is relatively common across the majority of industries. Closely related words such as ”brand”, ”advertising”, ”franchise”, and ”marketing” are less common and are concentrated in industries such as consumer non-durables and wholesale and retail. This indicates the importance of brand and customer relations in certain industries.

[Insert Fig. 6 near here]

Overall, the textual analysis shows that emphasis on different categories of intangible assets varies across industries, as captured by our measure. The variance also aligns with the expected industry characteristics, consistent with intuition.

2.3 Validity of intangibles talk measure

In this section, we examine the relationship between intangibles talk and firm characteristics. Table 4 reports the time series average of median firm characteristics for firms in the high, middle, and low range of intangibles talk. Prior studies consider R&D to total expenses, SG&A to total expenses, and intangible capital as proxies for intangible intensity.[7, 8] We benchmark our measure against these proxies, by examining whether they vary in predicted directions across quantiles of firms formed based on our measure. All three proxies exhibit an increase by quantiles based on intangibles talk with notably high values amongst the firms in the third quantile. We also examine variation of the proxies across quantiles formed based on the three categories of intagibles talk.

R&D to total expenses shows the highest variation with our measure in the innovation assets and information technology category, with a jump from 0.01 to 0.13 between the second and third quantiles. The increase from the second to third quantile is sizable for all three

[7] R&D and SG&A are reported as expenses and occur over the normal course of business operations (i.e., flow variables). We divide them by total expenses instead of total assets to capture the variations in the flow of intangible investment over each year. This allows us to compare SG&A and R&D across firms with large and small asset bases.

[8] Intangible capital which is constructed based on the methods in Peters and Taylor (2017) (intangible capital^{PT}) and Eisfeldt, Kim and Papanikolaou (2020) (intangible capital^{EKP}), is a stock of knowledge and organizational capital that accumulates over time through expenses such as SG&A and R&D. We use both versions of intangible capital in our analysis. Intangible capital^{PT} and Intangible capital^{EKP} are available at the GitHub page affiliated with Eisfeldt, Kim and Papanikolaou (2020): `https://github.com/edwardtkim/intangiblevalue`

proxies of intangible intensity: R&D, SG&A, and intangible capital. This shows consistency between intangibles talk and proxies considered in the literature.

We also examine variations in firm characteristics, such as size, leverage, and profitability. No clear relationship is discernible between intangibles talk and sales-to-asset ratio. However, debt to total assets falls with intangibles talk, showing that firms in the high quantile are the least leveraged. The opposite holds for profitability-to-total assets ratio, with its highest in the third quantile of intangibles talk. This aligns with the research on the relation between R&D and profitability (Lev and Sougiannis, 1996, see). However, later studies such as Curtis, McVay and Toynbee (2020) suggest a decline in the relation between R&D and profitability since the 1980s.

We also examine the traditional book-to-market ratio, along with its intangible augmented versions put forth by Peters and Taylor (2017) (book-to-market$^{\text{PT}}$) and Eisfeldt, Kim and Papanikolaou (2020) (book-to-market$^{\text{EKP}}$).[9] We observe that book-to-market drops along the quantiles, with the lowest average book-to-market concentrated in the top quantile across all categories of intangibles talk. This shows consistency between our measure of intangible value and the one implied by book-to-market (which is the opposite of market-to-book ratio). With respect to intangible augmented book-to-market ratios, their highest values are in the lowest quantile of intangibles talk, with less variation across quantiles relative to book-to-market itself, perhaps indicating their partial capacity to capture intangible value.

We provide further evidence that intangibles talk is correlated with other indicators of intangible value in our panel regression analysis in Table 5. One standard deviation drop in book-to-market on average is associated with an approximately 0.18 standard deviation increase in intangibles talk. To capture the relation between SG&A and intangibles talk, we subtract R&D from SG&A because it is a constituent item of SG&A (Enache and Srivastava,

[9] The formula is book-to-market$^{\text{PT}}$ = (book value of equity + intangible capital$^{\text{PT}}$ + goodwill)/market value of equity. Similarly, book-to-market$^{\text{EKP}}$ uses intangible capital$^{\text{EKP}}$ in the formula.

2018). As expected, our results show that SG&A (without R&D) and R&D, both scaled by total expenses are positively related to intangibles talk. A standard deviation increase in SG&A and R&D to total expense ratios corresponds to about a 0.3 standard deviation rise in intangibles talk. A similar result holds for intangible capital based on both definitions of the variable from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020).

[Insert Table 5 near here]

We view the results from Table 5 as evidence that intangibles talk and conventional indicators of intangible value are strongly correlated. This strengthens our main assumption that intangibles talk is a proxy for intangible value and intensity across firms.

3. Returns analysis

This section examines the returns performance of portfolios sorted based on intangibles talk. We seek to determine whether intangibles talk is informative about future returns.

We use factor-mimicking portfolios to consider the informativeness of intangibles talk by analyzing portfolio returns over time while controlling for the common risk factors. We follow the long-short sorting method, described in Fama and French (2015), to construct a value-weighted portfolio. We use intangibles talk as an investment signal to identify the long and short portfolios every year in a sample of NYSE, Amex, and Nasdaq stocks, with data available from the Center for Research in Security Prices (CRSP). We analyze the period starting from July 1995 to June 2020. [10] In this section, we describe our sorting methodology and long-short portfolio performance. We then conduct numerous analyses using several subperiods, such as before and after the dot-com bubble burst and the 2008 financial crisis.

[10]Given data requirements from the previous year, our portfolio sorting method generates returns from July 1995 based on intangibles talk values available from 1994.

We then benchmark the returns of our portfolio against the value strategies based on book-to-market ratio, and its augmented versions with capitalized intangibles.

3.1 Intangibles talk and value strategies

We follow Fama and French in constructing long-short portfolios except that we use intangibles talk as the sorting variable instead of book-to-market ratio. We sort firms on June 30 of each year, based on intangibles talk calculated from the 10-K filings for the fiscal year that ended in the previous year. This method assumes that at least the December fiscal year-end firms have published their annual report by June 30 of the next year. The portfolio (INT^{10K}) is constructed based on last reported intangibles talk. We identify stocks above the 70th percentile of intangibles talk, and put them in the long portfolios while shorting those below the 30th percentile.[11] Portfolios are held constant from July 1, following the June 30 portfolio formation date, to June 30 of the next year, except for delisted stocks. Monthly returns are calculated for each long and short portfolio, by value weighting returns of their constituent securities using their share in total market cap at the end of December of the previous year. Annual returns are calculated using monthly returns from January to June, based on portfolios formed in June of the previous year, and from July to December, based on portfolios formed in June of this year.[12]

[11] A more detailed description of the sorting method can be found in Fama and French (2015). Using the NYSE median market cap as the breaking point, for the long leg of our portfolio, we average the returns of two portfolios, namely big and small stocks with high intangibles talk, and repeat the same procedure for the short leg of our portfolio.

[12] We test the validity of our sorting methodology, by constructing a portfolio based on book-to-market as defined in Fama and French (2015). We achieve a 95% correlation with the HML^{FF} returns reported in the Kenneth R. French data library: (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). We use the same sorting methodology to construct other portfolios described in this study based on various firm-level intangible value indicators including our measure.

We first compare the performance of INT^{10K} against portfolios formed based on book-to-market and its two intangible augmented versions (Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020)). In this analysis, HML^{FF} represents a value strategy that takes into account only reported value of assets, and HML^{PT} and HML^{EKP} incorporate non-reported intangible assets. That is, their book–to–market ratio, used for sorting portfolios, is constructed with equity book values enhanced by estimated intangible capital–augmented book values. Notably, the definition of intangible capital used to calculate the augmented versions differs between Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) and so do the returns generated by HML^{PT} and HML^{EKP}.

Performance statistics in Table 6 reports the average monthly returns of the four portfolios for the period between July 1995 and June 2020. None of the returns is statistically significant from zero on average. Nevertheless, the returns from INT^{10K} are the highest, and the Sharpe ratios suggest that HML^{EKP} and INT^{10K} provide the best-performing strategies. Fig. 2 shows that the returns from INT^{10K} drop around the dot-com bubble, arguably because many intangible intensive companies suffered large negative returns, and numerous others were delisted. We also report the portfolio performances excluding 2000 and 2001, the peak bubble years. INT^{10K} now outperforms all the value strategies with statistically significant average annualized returns of 4.95%.

[Insert Table 6 near here]

A more detailed examination of the subperiods is reported in the last three columns of Table 6. The performance of INT^{10K} is stronger compared with other value strategies from 1995 to 1999, but no portfolio shows any statistically significant returns. During the pre–financial crisis period from 2000 to 2007, INT^{10K} performs poorly and the returns of HML portfolios are positive with high Sharpe ratios but still are not statistically significant from zero. Examining the post–financial crisis period (after 2008) is important because of massive underperformance of HML^{FF} (Lev and Srivastava, 2022). INT^{10K} delivers the highest returns between 2008 to 2020 with a statistically significant return of 7.9% and a Sharpe ratio of 0.93, the highest amongst all four strategies. The results support two ideas: the growing importance of intangible value in the economy and the ex–post intangibles talk measure capturing intangible capital better than the ex-ante capitalization methods of Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020).

Similar to Eisfeldt, Kim and Papanikolaou (2020), we benchmark the performance of intangibles talk against other value investing by forming composite portfolios with INT^{10K} as the long leg and one of each value portfolios at a time as the short leg. This test aims to capture the effect of intangibles as measured in our textual measure while removing the effects of intangibles captured in the reported book values and their augmentation based on capitalization of intangibles. The composite portfolio essentially longs stocks with high intangibles talk and low book-to-market ratios (and their augmented versions) and shorts stocks with low intangibles value and high book-to-market (and their augmented versions).

Table 7 presents the returns of a composite portfolio across different periods. The first case is the portfolio that longs INT^{10K} and shorts HML^{FF}. This composite portfolio has a statistically significant positive return, on average, in the full sample period. When we exclude the dot-com bubble burst period, the performance of the portfolio improves to an average annualized monthly return of 7.13%. The portfolio's Sharpe ratio is 0.15 for the full sample and 0.41 when excluding the dot-com bubble. The returns increase for the subsample periods of 1995 to 1999 to 13.6% with a Sharpe ratio of 0.71. For the period from 2000 to 2007, the returns turn negative (-16.42%), reflecting the poor performance of INT^{10K} and the relatively better performance of HML^{FF}. The performance improves significantly between 2008 and 2020 and generates a 12.4% return per year on average, reflecting both outperformance of INT^{10K} and underperformance of HML^{FF}. When the short leg of the composite portfolio is HML^{PT} and HML^{EKP}, the results are similar. The returns of the composite portfolio are positive for the period between 1995 to 1999 and 2008 to 2020 and negative for the period from 2000 to 2007.

[Insert Table 7 near here]

Results follow the pattern presented in Eisfeldt, Kim and Papanikolaou (2020), showing that value investing based on intangible augmented book-to-market and intangibles talk both outperform HML^{FF}, particularly in the post–financial crisis era. Nevertheless, results demonstrate that our measure captures orthogonal, and arguably superior, information on intangibles compared with the ones based on capitalization of past expenses.

3.2 Other intangible intensity indicators

We turn to other indicators of intangible value from the literature and compare their portfolio performance against the performance of INT^{10K}. We create four portfolios based on four intangible proxies: SG&A expenses, and R&D expenses, and both versions of intangible capital from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020). We use intangible capital (scaled by total assets) to create the last two portfolios, not the book-to-market ratios augmented with intangibles. In constructing intangible capital, Peters and Taylor (2017) already account for R&D and SG&A expenses to a degree, and we include the two variables separately as well. We do this to account for any direct intangible value signal from these items that could be informative about future returns. We sort portfolios at the end of June every year based on the ratio of R&D and SG&A to total expenses reported at the end of the last fiscal year. Similarly, to incorporate intangible capital from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020), we sort portfolios based on intangible capital at the end of the fiscal year from the previous calendar year. We divide intangible capital by total assets at the beginning of that fiscal year. We name them INT^{PT}

and INT^{EKP} to distinguish them from HML^{PT} and HML^{EKP}. This exercise allows us to compare the returns of portfolios sorted on other measures of intangible intensity, which is a more comparable strategy to INT^{10K}.

The returns of these four portfolios, along with INT^{10K} are reported in Table 8. In the full sample between 1995 and 2020, none of the portfolios generate statistically significant returns. Similar to intangible augmented HML returns, when we exclude the years around the dot-com bubble, most strategies perform better, with INT^{10K} slightly outperforming the rest. Columns 3 to 5 do not report any significant outperformance by any of the portfolios, and the period from 2008 to 2020 presents significant positive returns for INT^{10K}, INT^{PT}, INT^{EKP}, and $R\&D$ portfolios. The returns of INT^{10K}, and INT^{PT} are essentially the same, with INT^{10K} having a slightly larger Sharpe ratio. Therefore, the results from Table 8 show that intangibles talk is at least as informative as the other measures of intangible intensity in the literature, if not superior.

[Insert Table 8 near here]

3.3 Returns across categories of intangibles talk

We break down intangibles talk into its three categories based on the words associated with each category (see Table 2). We repeat the sorting strategy based on each category of intangibles talk and report the returns. The returns of the portfolios based on these three categories are presented in Table 9. The results are similar to INT^{10K} over the full period and the subperiods. The highest returns are associated with the first category of intangibles talk related to innovation assets and information technology. However, the returns under all the categories are positive and significant when excluding the dot-com bubble years and especially in recent years. This indicates that the informativeness of intangibles talk is not limited to any single category and that the effect is present across all the categories.[13]

[Insert Table 9 near here]

4. Common risk factors and intangibles talk

We have established that INT^{10K} is informative about future returns and outperforms value strategies based on traditional and intangible augmented book-to-market, especially in recent years. We next examine whether the returns associated with intangibles talk are simply premiums for known risk factors. We first test the hypothesis that the return for INT^{10K} is compensation for systematic risk. We regress our portfolio's returns against the main systematic factors discussed in the literature, namely the three and five factors in Fama and French (2015) models plus momentum. We control for all the factors in the regression

$$\mathbf{R_t} = \boldsymbol{\alpha} + \boldsymbol{\beta_{MKT}}\mathbf{MKT_t} + \boldsymbol{\beta_{SMB}}\mathbf{SMB_t} + \boldsymbol{\beta_{HML}}\mathbf{HML_t} + \boldsymbol{\beta_{RMW}}\mathbf{RMW_t} + \boldsymbol{\beta_{CMA}} CMA_t + \boldsymbol{\beta_{MOM}} MOM_t + \epsilon_t \qquad (2)$$

where R_t is the return of INT^{10K} in month t, and α is the intercept that captures the abnormal returns after controlling for risk factors. MKT_t, SMB_t, HML_t, RMW_t, CMA_t, and MOM_t are, respectively, the returns of the market, size, value, profitability, investment, and momentum portfolios taken from Ken French's website. The standard errors are estimated

[13] We also sort stocks based on intangibles talk within industries, using industry-specific benchmarks. We use the twelve Fama and French industry classifications and form our portfolio using stocks from each industry. The returns are insignificant and mostly positive except in the finance industry, for which the returns are positive and statistically significant. This indicates that the informativeness of intangibles talk is not primarily a within-industry phenomenon, unlike value (see Asness, Porter and Stevens, 2000).

using Newey and West (1987), which allows for serially correlated and heteroskedastic error terms. The alphas are reported in Table 10 for the period between July 1995 and June 2020. The first three columns report the excess returns over the Fama and French three factors plus momentum. In the first column, as the baseline regression, we use HML^{FF} as the value factor. In the second and third columns, we use HML^{PT} and HML^{EKP} as the value factor, respectively, to account for intangible capital shown in prior studies. The alphas remain positive and statistically significant in the first three columns, with the highest excess return, on an annualized basis, reported at 4.48% and the lowest at 3.26%. The positive and significant alphas show that the informativeness of intangibles talk with respect to future returns is not fully explained by the common risk factors: market, size, value, and momentum. In Columns 4 to 6 in Table 10, we include RMW and CMA in our regressions. The excess returns remain positive and statistically significant, with the highest at 8.28% and the lowest at 5.59%. This result rules out the possibility that the returns associated with intangibles talk capture the effects of profitability (RMW). Profitability is positively associated with intangibles talk (see Table 4).

[Insert Table 10 near here]

The abnormal returns are similar for each category of intangibles talk as well. Table 11 reports the alphas over the Fama and French factors for the three portfolios corresponding to each category of intangibles talk. The highest alpha belongs to innovation assets and information technology with 2.92% and 5.09%, respectively, for the Fama and French three and five factors plus momentum. Therefore, while the magnitude of alpha varies across categories, the abnormal returns associated with intangibles talk are not limited to a particular class of intangible assets.

[Insert Table 11 near here]

The limited availability of 10-K filings restricts the exploration of excess returns from before 1995. However, the results here are comparable with the 3.48% excess returns (over the three Fama and French factors plus momentum) generated by the value-weighted portfolio that picks stocks based on employee satisfaction, reported between 1984 and 2009 by Edmans (2011). Similarly, the excess returns associated with R&D and advertising expenses reported in Lev and Sougiannis (1996) is around 4.57% in their sample, which goes back to 1975. Hence our results here point to to the link between intangible value and subsequent returns that seem to be persistent over the most recent years.

Table 10 better illustrates the relatively cheap (expensive) valuation of stocks with high (low) levels of intangibles talk in the market. We construct five portfolios of stocks ranked based on intangibles talk and estimate the alphas over the five-factor model. In addition, we investigate whether the observed excess returns in Table 10 are associated with intangible intensity in general. We repeat the exercise based on other intangible intensity indicators discussed so far and report their alphas over the five-factor model as well. These excess returns are presented in Table 12.

[Insert Table 12 near here]

The highest-ranking portfolio based on intangibles talk generates significant positive alpha, while a significant negative alpha is reported for the lowest-ranking portfolio, suggesting that intangible intensive stocks are underpriced. The results for the portfolio of stocks ranked based R&D to total expenses only partly confirm the previous findings by reporting a positive and significant alpha for the highest-ranking portfolio. When stocks are ranked based on intangible capital to total assets using Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) measures, the alphas are not significant for the highest-ranking portfolios, but the lowest-ranking portfolio in the case of intangible capital$^{\text{PT}}$ generates a significant negative alpha.

Regardless of the intangibles measured used to generate alphas, intangibles talk appears to capture the mispricing effect better than other indicators such as R&D and intangible capital. We view the results from this section and Section 3 as consistent with our hypothesis that the information on intangibles embedded in textual disclosures of 10-K filings is either ignored or not fully valued by investors.

5. Mispricing of intangibles talk with greater arbitrage risk

We next examine whether return informativeness of intangibles talk is related to mispricing associated with limits to arbitrage. We rely on IVOL as a measure for limits to arbitrage that amplifies mispricing. To the extent that the INT^{10K} alphas come from limits to arbitrage for mispriced stocks, they should increase in IVOL.

5.1 Mispricing and IVOL

Diversification of idiosyncratic risks is central to the capital asset pricing model (CAPM), yet its limiting effect on arbitrage is well proven in the literature. Stambaugh, Yu and Yuan (2015) claim that adverse price moves are more likely under higher IVOL and, therefore, it is a source of arbitrage risk. This is because a capital-constrained investor is forced to close her positions prematurely, under high IVOL conditions, before the subsequent price corrections. We follow Stambaugh, Yu and Yuan (2015), as well as subsequent studies on the effects of IVOL on mispricing (see Cao and Han, 2016; Birru and Young, 2020) to examine whether our results differ under different IVOL conditions.

Stambaugh, Yu and Yuan (2012) investigate the effects of IVOL on mispricing by constructing double-sorted portfolios based on IVOL and a mispricing measure. They examine 11 return anomalies identified in the literature that survive the three factors of Fama and

French. Their results indicate that the degree of mispricing is higher among high IVOL stocks in that the returns generated by going long on putative underpriced stocks and going short on putative overpriced stocks increases in IVOL. They also show that the negative IVOL-return relation (the so-called idiosyncratic volatility puzzle) exists among overpriced stocks and that the reverse is true for underpriced stocks (positive IVOL-return relation).

To the extent that returns from intangibles talk are because of mispricing of intangible-intensive stocks, we expect an amplification of such effect under high IVOL conditions. In addition, in line with the results in Stambaugh, Yu and Yuan (2012) on the IVOL-return relation, we expect IVOL-return relation to be negative among stocks that score the lowest in intangibles talk (that is, the putative overpriced stocks) and to be positive for the highest in intangibles talk stocks (that is, putative underpriced stocks).

We test our hypothesis by creating 25 double-sorted portfolios, five times five each, based on intangibles talk and IVOL. We then estimate the alphas of each portfolio over the five factors of Fama and French to examine the abnormal returns. Table 13 reports the alphas for the 25 portfolios. As we can see, the difference between abnormal returns of the highest and lowest-ranking portfolios (high minus low alpha) based on intangibles talk (rows) is not significantly different from zero among stocks with low IVOL. Meanwhile, the alpha difference for the bottom two rows with the highest IVOL is significant and positive. To account for size effects, we also report the alphas for the 25 portfolios separately for small and big firms. Table 14 shows that the results are similar for small and big firms separately. The results for the big firms imply the same, with the alpha difference being 19.53% among stocks with the highest IVOL.

[Insert Table 13 and 14 near here]

In Fig. 7, we plot the strong positive relation between IVOL and INT^{10K} abnormal returns over Fama and French's five factors. The highest alpha belongs to INT^{10K}, constructed using the stocks in the highest IVOL decile in our sample. More importantly, a positive, and increasing, trend appears in the magnitude of alpha across IVOL deciles with alphas going up to above 50% average annualized monthly returns in the highest decile.

[Insert Fig. 8 near here]

Another observation from Table 13, 14 comes from the sorting of stocks based on IVOL (columns), demonstrating the negative relation between IVOL and subsequent returns. This aligns with Ang et al. (2006) and the literature on the idiosyncratic volatility puzzle (negative IVOL-return relation). However, this relation disappears among stocks with the highest intangibles talk, for both small and big firms. This result aligns with Stambaugh, Yu and Yuan (2015), which shows that the negative relation of IVOL and returns is stronger for overpriced (low intangibles talk) stocks and is positive among underpriced (high intangibles talk) stocks.

Overall, the results suggest that the abnormal returns associated with intangibles talk is significant and growing in size with higher levels of IVOL and, thus, limits to arbitrage. The positive relation between alpha and IVOL is also in line with Birru and Young (2020), which finds that investor sentiment is a stronger predictor of subsequent returns (mispricing) in the presence of uncertainty (measured through IVOL).

The abnormal returns of INT^{10K} reported in Section 4 indicate that the informativeness of intangibles talk is not captured by the common risk factors. The sorting exercise provides strong evidence for the association between IVOL and INT^{10K} abnormal returns. This supports the mispricing hypothesis and therefore ties our findings to the literature that examines mispricing in relation to the limits to arbitrage.

6. Conclusion

We construct a textual measure of intangibles talk using 10-K filings and test its informativeness for future stock returns. We sort a long-short portfolio based on our measure and report significant and positive returns between 1995 and 2020, when we exclude the years 2000 and 2001, representing the dot-com bubble burst. We compare our portfolio's returns with value strategies based on traditional book-to-market and its intangible augmented versions in the literature and report a superior performance, especially in recent years. Our portfolio's historical returns suggest that intangibles talk picks up the intangible value associated with technology stocks, as evidenced by the portfolio's poor performance in the aftermath of the dot-com bubble. In addition, its outperformance after 2008 coincides with the poor performance of value strategies and supports the growing importance of intangible value in the US economy.

We test whether the common risk factors from Fama and French explain intangibles talk's informativeness. Our results indicate a positive and significant alpha over all the three and five factors from Fama and French plus momentum. We achieve the same results when we replace the value factor in the regressions with its intangible augmented versions from the literature. An implication of the abnormal returns of our portfolio is that investors do not price in intangibles information, perhaps due to the challenging nature of defining and valuing intangible assets. Our results align with similar findings on mispricing of R&D and employee satisfaction in stock valuations.

We test the mispricing hypothesis for intangibles talk's abnormal returns in the presence of limits to arbitrage. We show that the abnormal returns associated with intangibles talk is present only among stocks with higher levels of IVOL, which measure arbitrage risks. This confirms our mispricing hypothesis and shows that the limits to arbitrage exacerbate the mispricing of intangibles talk, similar to what is shown for mispricings of other origins in the literature.

Tables and figures

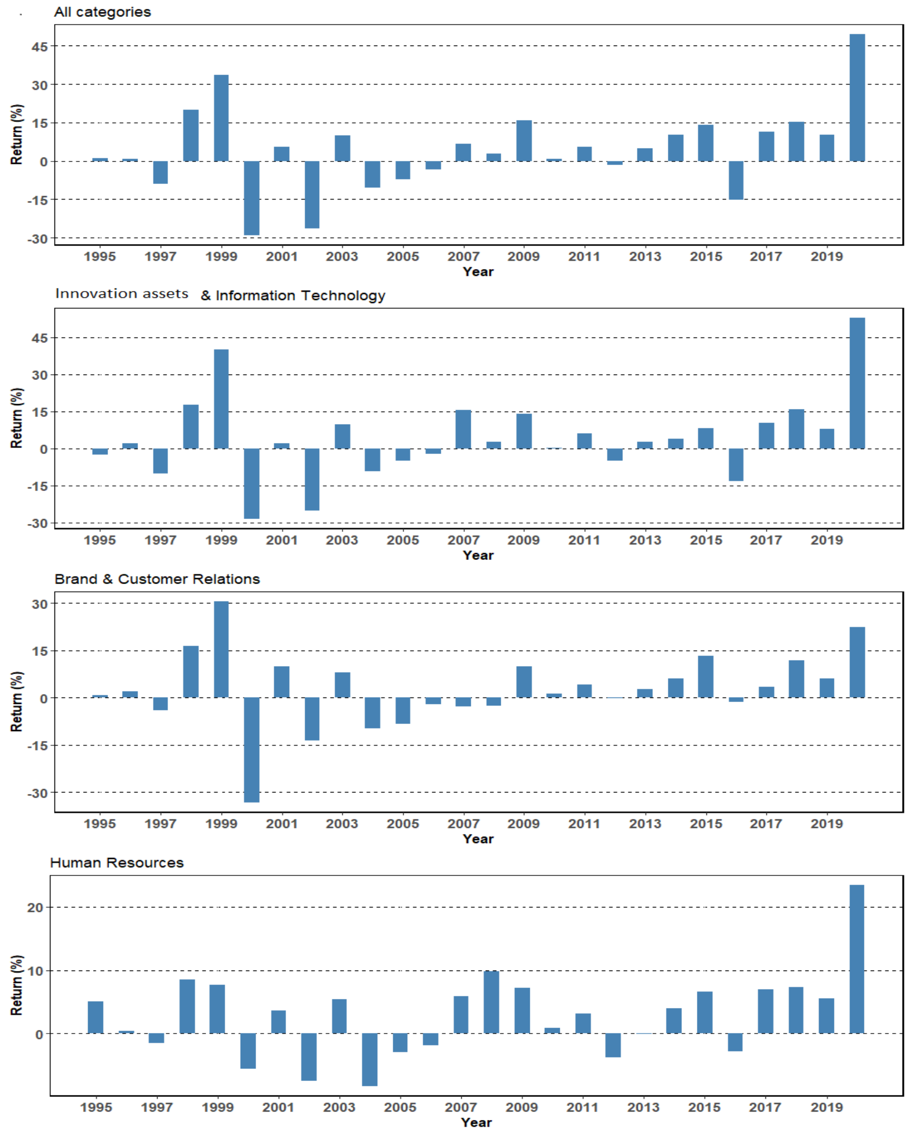


Figure 1: Average Monthly Return Performance

The figure above plots the average monthly returns of INT^{10K} for each year between July 1995 to June 2020. For each category stocks are sorted based on intangibles talk values in that category. The returns are monthly in percent per year (monthly return multiplied by twelve).

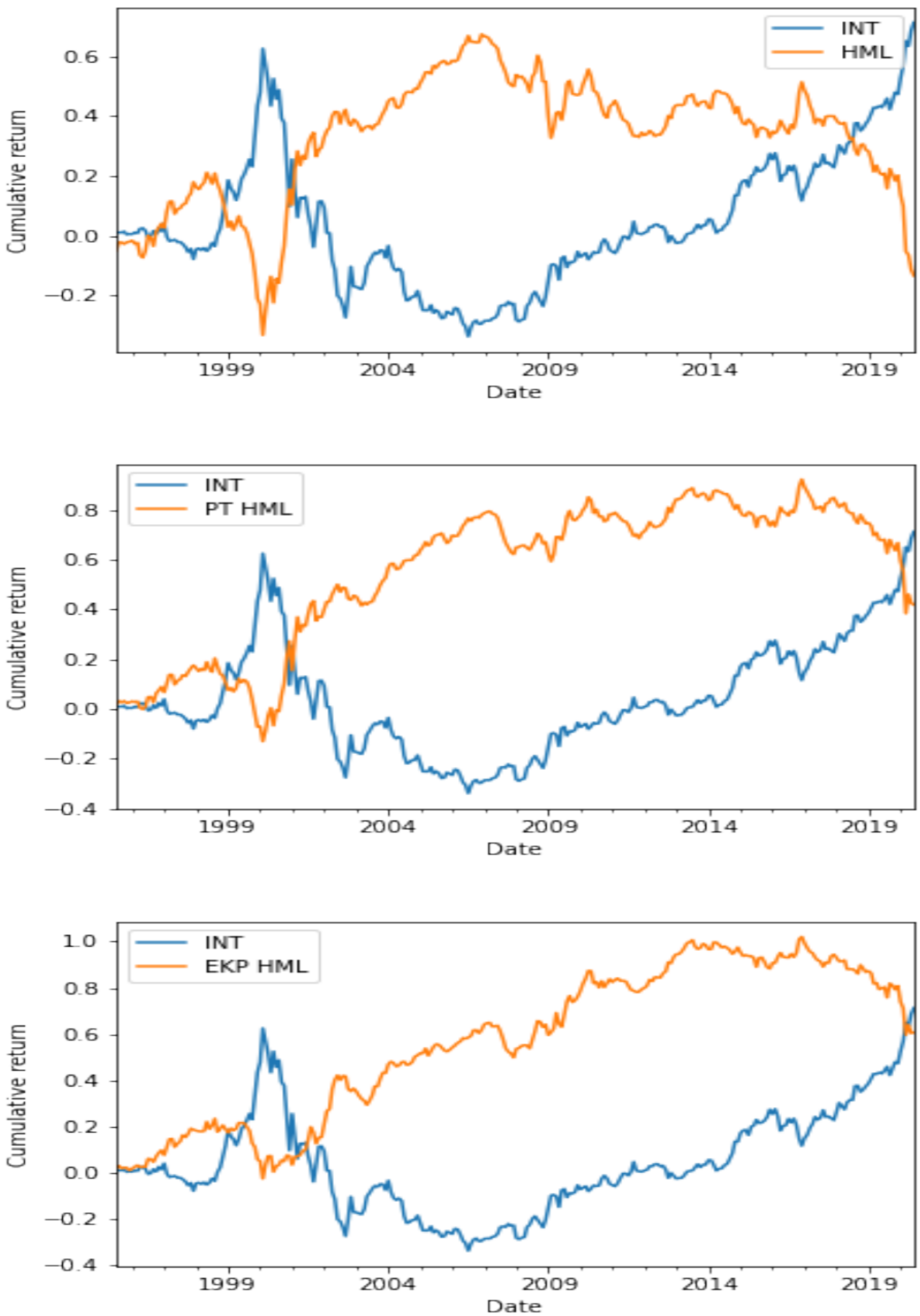


Figure 2: Cumulative Return Performance

The figure above plots the cumulative returns of one dollar invested in INT^{10K} in comparison to value strategies HML^{FF}, HML^{PT}, and HML^{EKP} for the period between July 1995 to June 2020.

ABILITIES	CUSTOMER RELATION	INNOVATION	NETWORKS	TALENT
ABILITY	CUSTOMERS	INNOVATIONS	PATENT	TALENTS
ADVERTISING	DATA	INNOVATOR	PATENTED	TEAM
ALGORITHM	DATABASE	INNOVATORS	PATENTS	TEAMS
AUTHORSHIP	DATABASES	INTELLECTUAL	PLATFORM	TEAMWORK
AUTHORSHIPS	DESIGN	intellectual property	PLATFORMS	TECHNOLOGIES
BRAND	DESIGNS	INTELLECTUAL PROPERTY	PRESENCE	TECHNOLOGY
BRANDING	DISCOVERIES	INTERNET	PRODUCTIVITY	TRADE MARK
BRANDS	DISCOVERY	INTERNET ACTIVITIES	PROTECTED DESIGN	TRADE MARKS
CLIENT	EMPLOYEE	INTERNET ACTIVITY	PROTECTED DESIGNS	TRADE NAME
CLIENT RELATIONS	EMPLOYEES	INVENT	REGISTERED DESIGN	TRADE SECRET
CLIENTS	EXPERIENCE	INVENTED	REGISTERED DESIGNS	TRADE SECRETS
COMPETENCE	EXPERT	INVENTING	RELATION	TRADEMARK
COMPETENCES	EXPERTISE	INVENTION	RELATIONS	TRADEMARKS
COMPETENCIES	EXPERTS	INVENTIONS	RELATIONSHIP	TRAINING
COMPETENCY	FORMULA	INVENTS	RELATIONSHIPS	USER
CONNECTIONS	FORMULAE	KNOWHOW	REPUTATION	USERS
CONNECTIVITY	FRANCHISE	KNOWLEDGE	RESEARCH	WEBSITE
CONSUMER	FRANCHISES	LABEL	RESEARCHES	WEBSITES
CONSUMERS	HUMAN	LABELS	SITE VISITS	WORKFORCE
COPYRIGHT	HUMAN CAPITAL	LICENCE	SKILL	
COPYRIGHTS	HUMAN RESOURCES	LICENCES	SKILLS	
CUSTOMER	INNOVATE	LOGO	SOFTWARE	
CUSTOMER BASE	INNOVATE PARTNERS	LOYALTY	SOLUTION	
CUSTOMER BASES	INNOVATED	MARKETING	SOLUTIONS	
CUSTOMER LIST	INNOVATES	NAMES	SYSTEM	
CUSTOMER LISTS	INNOVATING	NETWORK	SYSTEMS	

Table 1: intangibles words Glossary

This table shows the intangible word list with 128 words. We use this list of words to calculate intangibles talk for each 10-K filing. The words in the table are based on Filipovic and Wager (2019)).

Innovation assets & information technology		Brand & customer relations	Human resources
INNOVATE PARTNERS	INNOVATOR	CLIENT RELATIONS	HUMAN CAPITAL
INTELLECTUAL PROPERTIES	INNOVATORS	CUSTOMER BASE	HUMAN RESOURCES
INTELLECTUAL PROPERTY	INTELLECTUAL	CUSTOMER BASES	ABILITIES
INTERNET ACTIVITIES	INTERNET	CUSTOMER LIST	ABILITY
INTERNET ACTIVITY	INVENT	CUSTOMER LISTS	COMPETENCE
PROTECTED DESIGN	INVENTED	CUSTOMER RELATION	COMPETENCES
PROTECTED DESIGNS	INVENTING	ADVERTISING	COMPETENCIES
REGISTERED DESIGN	INVENTION	BRAND	COMPETENCY
REGISTERED DESIGNS	INVENTIONS	BRANDING	EMPLOYEE
SITE VISITS	INVENTS	BRANDS	EMPLOYEES
TRADE MARK	KNOWHOW	CLIENT	EXPERIENCE
TRADE MARKS	KNOWLEDGE	CLIENTS	EXPERT
TRADE NAME	LICENCE	CONNECTIONS	EXPERTISE
TRADE SECRET	LICENCES	CONNECTIVITY	EXPERTS
TRADE SECRETS	NETWORK	CONSUMER	HUMAN
ALGORITHM	NETWORKS	CONSUMERS	PRODUCTIVITY
AUTHORSHIP	PATENT	CUSTOMER	SKILL
AUTHORSHIPS	PATENTED	CUSTOMERS	SKILLS
COPYRIGHT	PATENTS	FRANCHISE	TALENT
COPYRIGHTS	PLATFORM	FRANCHISES	TALENTS
DATA	PLATFORMS	LABEL	TEAM
DATABASE	RESEARCH	LABELS	TEAMS
DATABASES	RESEARCHES	LOGO	TEAMWORK
DESIGN	SOFTWARE	LOYALTY	TRAINING
DESIGNS	SOLUTION	MARKETING	WORKFORCE
DISCOVERIES	SOLUTIONS	NAMES	
DISCOVERY	SYSTEM	PRESENCE	
FORMULA	SYSTEMS	RELATION	
FORMULAE	TECHNOLOGIES	RELATIONS	
INNOVATE	TECHNOLOGY	RELATIONSHIP	
INNOVATED	TRADEMARK	RELATIONSHIPS	
INNOVATES	TRADEMARKS	REPUTATION	
INNOVATING	WEBSITE	USER	
INNOVATION	WEBSITES	USERS	
INNOVATIONS			

Table 2: intangibles words by Each Category

Category	Portion of glossary	Variance share	Top words
Innovation assets & information technology	54%	46.74%	Data, System, Technology, Research, Intellectual Property
Brand & customer relations	27%	28.6%	Customers, Marketing, Consumer, Advertising, Relationship
Human resources	19%	24.66%	Employee, Ability, Experience Expertise, Talent

Table 3: Variance Shares of Intangibles Talk Categories

The table reports the share of the total variance coming from each category of intangibles words. The total variance is [$\Sigma_j \, v(j)$] summed across all the words in our glossary and $v(j)$ is the variance of frequency of word j across all the filings in our sample.

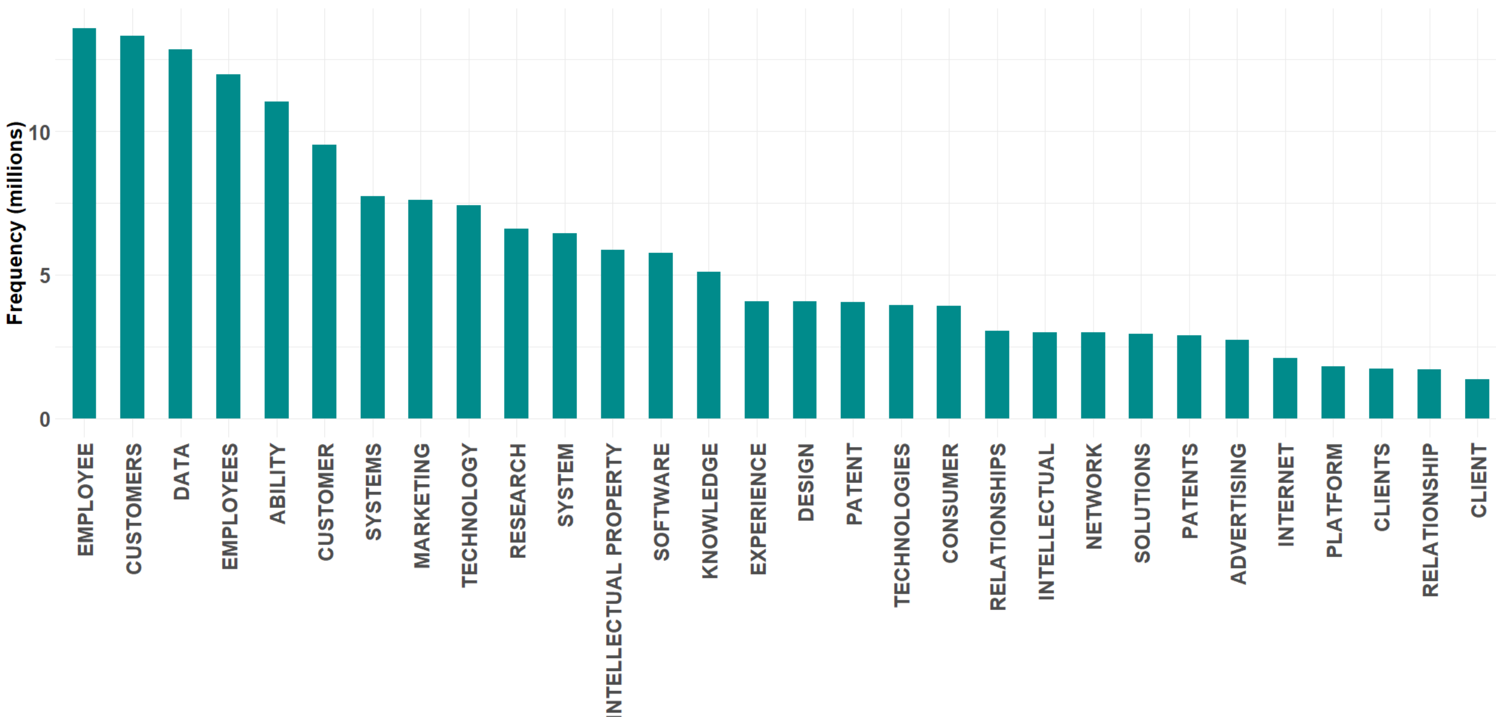


Figure 3: Most frequent intangible terms across all firms

.

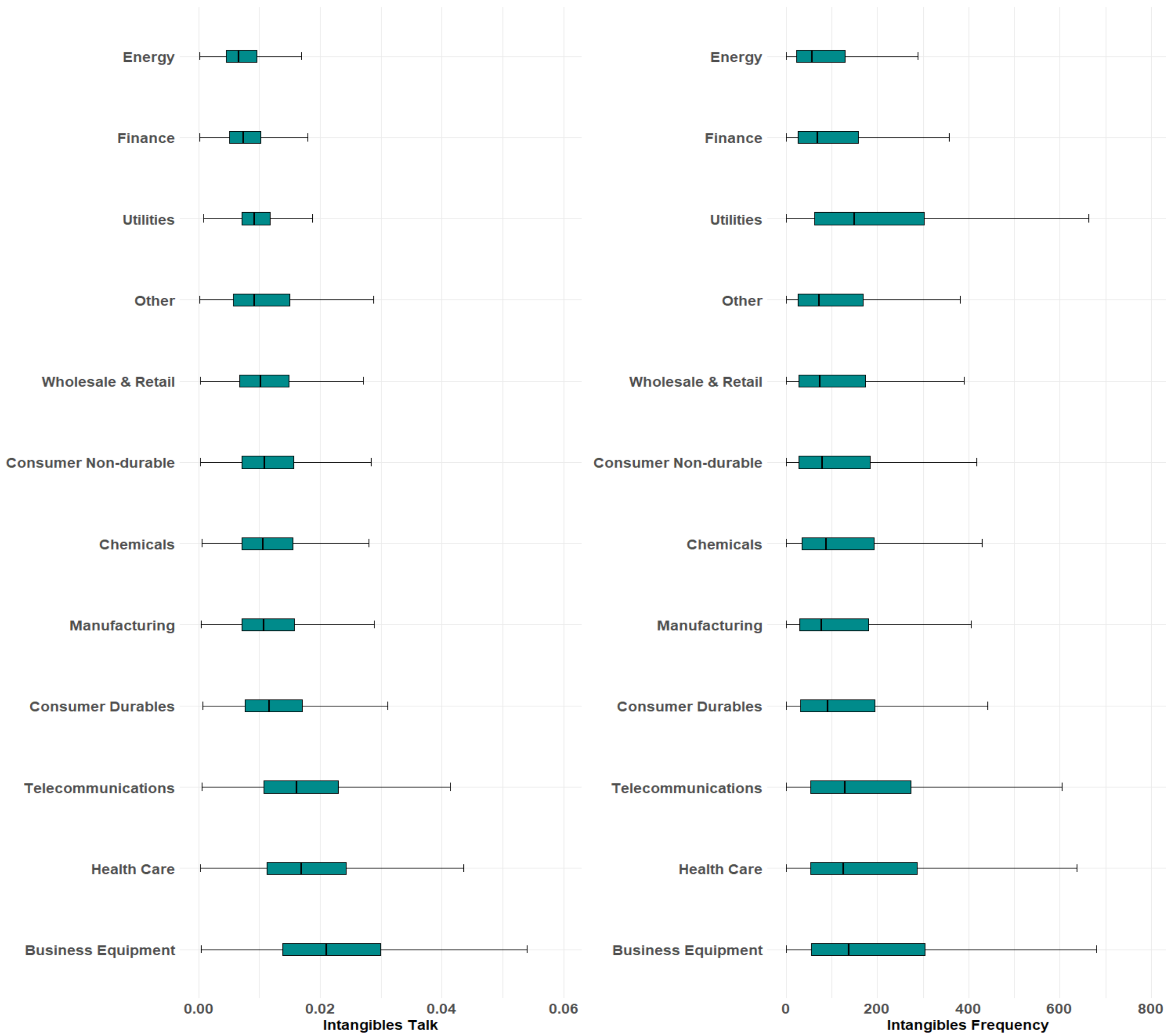


Figure 4: Intangibles Talk across Fama and French twelve industries

.

Figure 5: Intangibles by category in Fama and French twelve industries

.

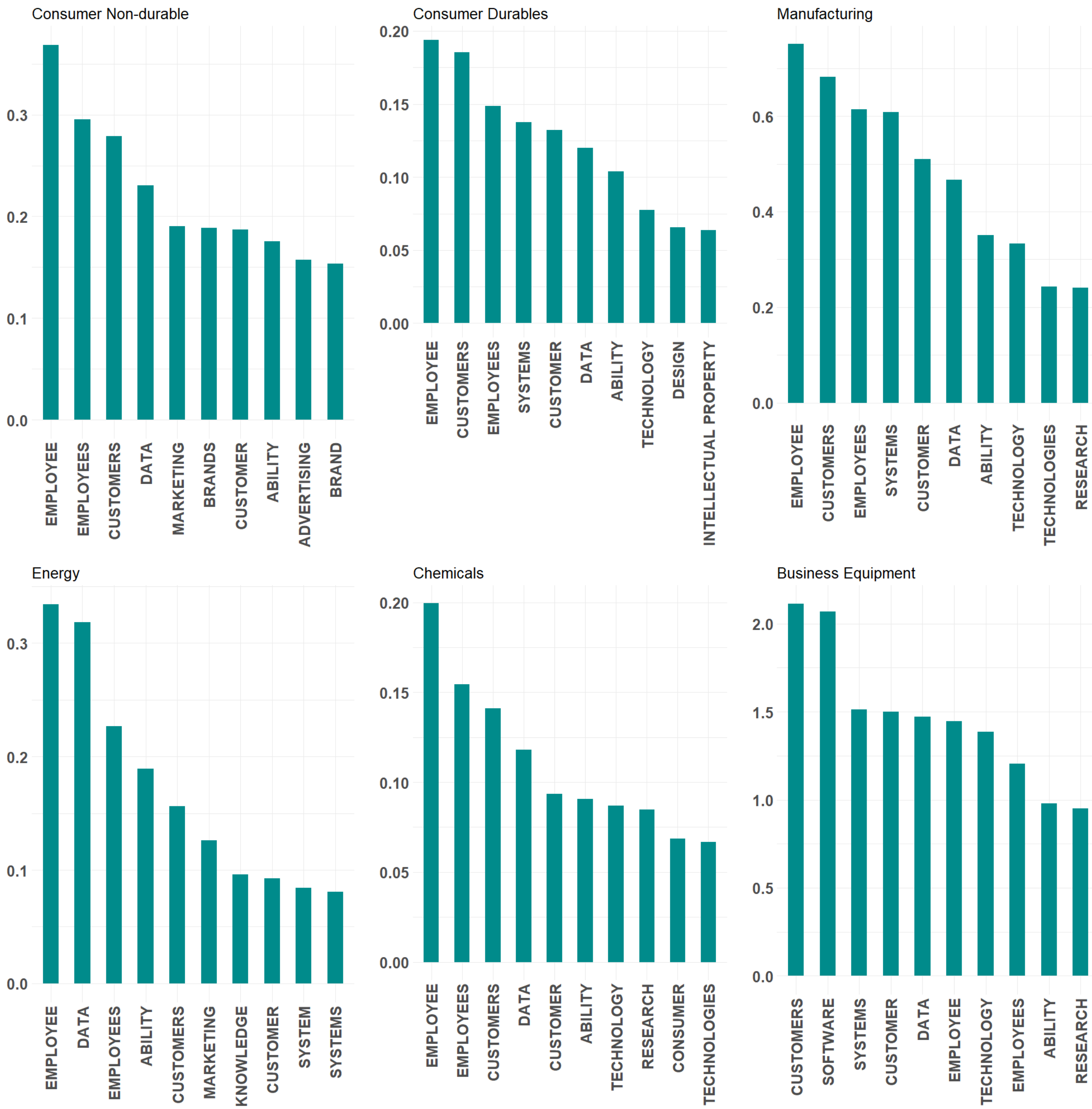
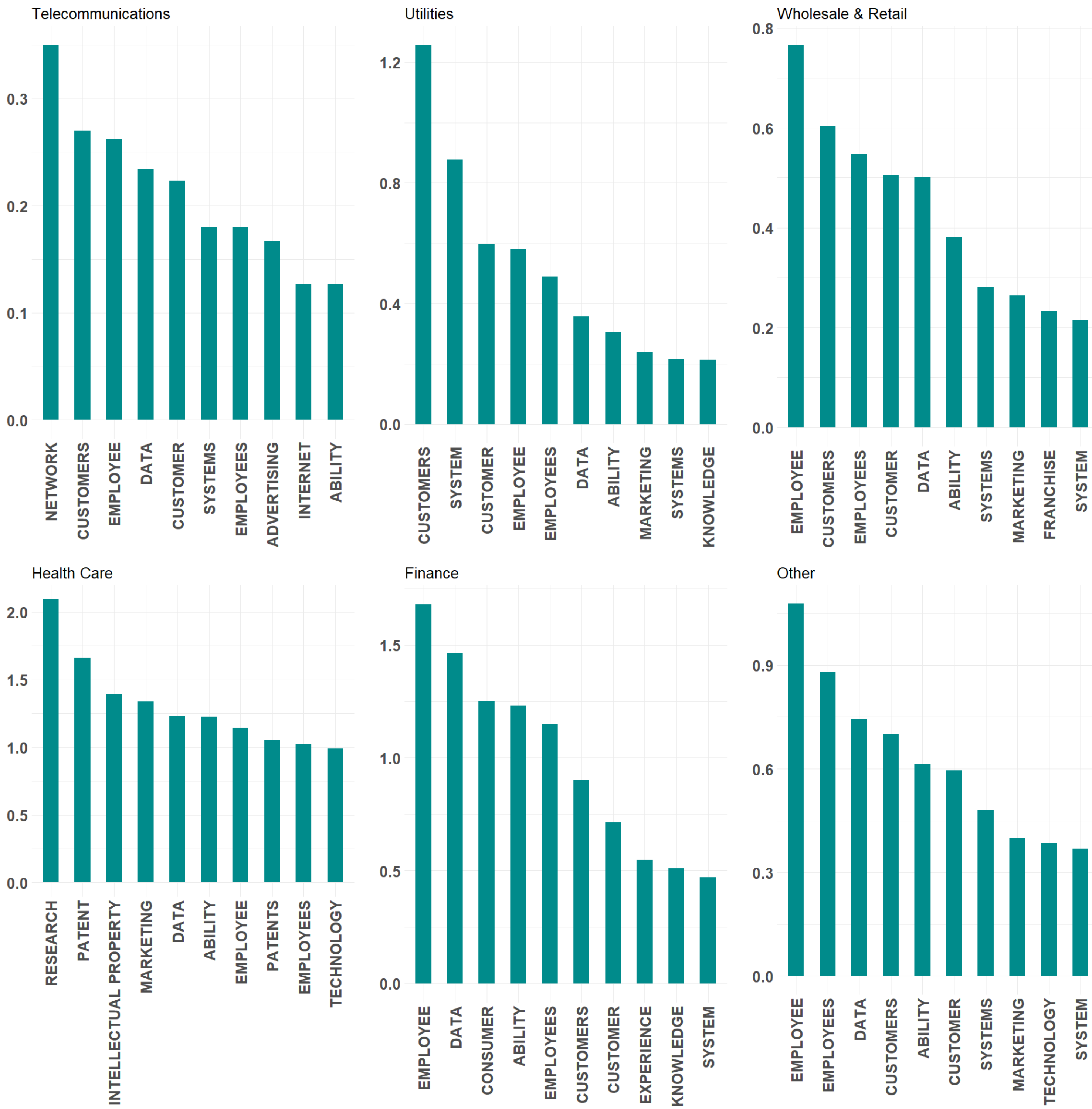


Figure 6: Most frequent intangible terms by Fama and French twelve industries

.


Telecommunications
0.0
0.1
0.2
0.3
NETWORK
CUSTOMERS
EMPLOYEE
DATA
CUSTOMER
SYSTEMS
EMPLOYEES
ADVERTISING
INTERNET
ABILITY
Utilities
0.0
0.4
0.8
1.2
CUSTOMERS
SYSTEM
CUSTOMER
EMPLOYEE
EMPLOYEES
DATA
ABILITY
MARKETING
SYSTEMS
KNOWLEDGE
Wholesale & Retail
0.0
0.2
0.4
0.6
0.8
EMPLOYEE
CUSTOMERS
EMPLOYEES
CUSTOMER
DATA
ABILITY
SYSTEMS
MARKETING
FRANCHISE
SYSTEM
Health Care
0.0
0.5
1.0
1.5
2.0
RESEARCH
PATENT
INTELLECTUAL PROPERTY
MARKETING
DATA
ABILITY
EMPLOYEE
PATENTS
EMPLOYEES
TECHNOLOGY
Finance
0.0
0.5
1.0
1.5
EMPLOYEE
DATA
CONSUMER
ABILITY
EMPLOYEES
CUSTOMERS
CUSTOMER
EXPERIENCE
KNOWLEDGE
SYSTEM
Other
0.0
0.3
0.6
0.9
EMPLOYEE
EMPLOYEES
DATA
CUSTOMERS
ABILITY
CUSTOMER
SYSTEMS
MARKETING
TECHNOLOGY
SYSTEM

	Intangibles Talk (All Categories)			Innovation assets & Information Technology		
	Low 30	Mid 40	High 30	Low 30	Mid 40	High 30
R&D to total expenses	0	0.02	0.12	0	0.01	0.13
SG&A to total expenses	0.24	0.23	0.38	0.27	0.21	0.4
Intangible capital$^{\text{PT}}$ to total assets	0.06	0.43	0.73	0.05	0.42	0.76
Intangible capital$^{\text{PT}}$ to sales	0.34	0.44	0.88	0.34	0.39	0.99
Sales to total assets	0.34	0.96	0.8	0.32	1.04	0.75
Sales to stockholder's equity	1.01	1.98	1.37	1.02	2.2	1.25
Price to sales	0.05	0.03	0.07	0.05	0.03	0.08
Debt to EBITDA	2.22	1.3	0.15	2.35	1.37	0.19
Debt to total assets	0.26	0.2	0.07	0.26	0.2	0.07
Profitability to total assets	0.09	0.28	0.35	0.08	0.29	0.33
Investment to physical capital	0.09	0.09	0.12	0.08	0.09	0.12
Market Cap (millions)	401.16	566.35	382.35	377.97	655.49	364.4
Book to market^{FF}	0.73	0.51	0.38	0.73	0.51	0.37
Book to market^{EKP}	1.2	1.21	1.12	1.23	1.22	1.07
Book to market^{PT}	0.85	0.73	0.63	0.87	0.73	0.6

	Brand & Customer Relations			Human Resources		
	Low 30	Mid 40	High 30	Low 30	Mid 40	High 30
R&D to total expenses	0.08	0.04	0.06	0.02	0.05	0.09
SG&A to total expenses	0.22	0.3	0.32	0.26	0.28	0.32
Intangible capital$^{\text{PT}}$ to total assets	0.14	0.23	0.6	0.12	0.37	0.58
Intangible capital$^{\text{PT}}$ to sales	0.41	0.53	0.62	0.43	0.53	0.64
Sales to total assets	0.45	0.68	0.95	0.53	0.73	0.85
Sales to stockholder's equity	1.1	1.3	1.75	1.21	1.43	1.57
Price to sales	0.05	0.06	0.05	0.05	0.05	0.05
Debt to EBITDA	1.46	1.05	0.58	1.64	1.11	0.22
Debt to total assets	0.21	0.18	0.11	0.23	0.18	0.09
Profitability to total assets	0.12	0.21	0.36	0.16	0.24	0.32
Investment to physical capital	0.09	0.09	0.11	0.09	0.1	0.11
Market Cap (millions)	447.57	387.42	480.18	425.67	401.57	473.28
Book to market^{FF}	0.58	0.61	0.46	0.66	0.56	0.44
Book to market^{EKP}	1.04	1.24	1.3	1.21	1.2	1.14
Book to market^{PT}	0.75	0.8	0.71	0.83	0.76	0.67

Table 4: Summary Statistics for Firm Characteristics

This table summarizes the characteristics of firms belonging to the above 70th percentile, between 30th to 70th percentile, and the bottom 30th percentile based on intangibles talk values. The values are the time-series average of the median firm characteristics within each percentile bucket. The sample period is from January 1994 to December 2019.

	Dependent variable:					
	Intangibles Talk$_t$					
	(1)	(2)	(3)	(4)	(5)	(6)
Book-to-Market$_t$	-0.176					0.006
	(-56.96)					(1.03)
$\frac{(SG\&A_t - R\&D_t)}{Total\, Expenses_t}$		0.299				0.255
		(88.05)				(40.63)
$\frac{R\&D_t}{Total\, Expenses_t}$			0.256			0.520
			(55.33)			(59.45)
$\frac{Intangible\, Capital_t^{PT}}{Total\, Assets_{t-1}}$				0.372		-0.512
				(112.43)		(-28.51)
$\frac{Intangible\, Capital_t^{EKP}}{Total\, Assets_{t-1}}$					0.370	0.525
					(108.74)	(29.98)
Observations	99,112	82,713	43,986	84,283	81,927	31,923
Adjusted R^2	0.032	0.086	0.065	0.130	0.126	0.218

Table 5: Intangibles Talk as a Proxy for Intangible Intensity

In this table, we report the pooling regression with firm-level intangibles talk as the dependent variable:

$$\text{Intangibles Talk}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta} X_{i,t} + \epsilon_{i,t} \qquad (1)$$

where $X_{i,t}$ are firm-level book-to-market, (SG&A - R&D)/Total Expense, R&D/Total Expense, intangible capital^{PT}. The panel covers the period between January 1994 to December 2021. To separately capture the effect of only SG&A we use (SG&A - R&D) since R&D is already recorded in SG&A values. All the variables are annual, with their negative values dropped, winsorized at 1% from above, and normalized by dividing by standard deviation. The error terms are clustered at the firm level and time fixed effects are accounted for.

		Full Sample (1995-2020)	Full Sample (Exc. Dot-com Bubble)	Dot-com Bubble (2000-2001)	1995-1999	2000-2007	2008-2020
		(1)	(2)	(3)	(4)	(5)	(6)
INT^{10K}	E[R]	3.61	**4.95**	-11.78	10.19	-6.78	**7.9**
		(1.26)	(2.19)	(-0.94)	(1.31)	(-1.32)	(3.24)
	σ	12.38	9.59	29.49	10.2	17.28	8.45
	Sharpe	0.29	0.52	-0.4	1	-0.39	0.93
HML^{FF}	E[R]	0.22	-2.18	**27.81**	-3.44	9.64	-4.51
		(-0.91)	(0.5)	(2.49)	(-0.44)	(1.66)	(-1.38)
	σ	11.13	9.44	21.75	10.12	12.33	10.4
	Sharpe	0.02	-0.238	1.28	-0.34	0.78	-0.43
HML^{PT}	E[R]	2.11	0.17	**24.4**	-1.79	9.41	-1.15
		(0.85)	(0.09)	(2.64)	(-0.36)	(1.85)	(-0.4)
	σ	10.32	8.61	21.27	8.34	12	9.63
	Sharpe	0.2	0.02	1.15	-0.21	0.78	-0.12
HML^{EKP}	E[R]	2.69	2.44	5.56	0.96	5.93	1.24
		(1.34)	(1.34)	(1.51)	(0.24)	(1.61)	(0.45)
	σ	8.57	8.21	12.12	7.43	9.02	8.65
	Sharpe	0.31	0.3	0.46	0.13	0.66	0.14

Table 6: INT^{10K} vs. Value Strategies

In this table, we summarize the risk and return associated with textual intangible value and other measures of value and intangible value in the literature. INT^{10K} is the portfolio sorted based on the intangible talk. HML^{FF} is sorted based on traditional book-to-market value. HML^{PT} and HML^{EKP} are sorted based on intangible augmented book-to-market calculated using Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) methods. The values in parentheses are Newey-West T-statistics that test the difference in the means to be zero. The full sample is from July 1995 to June 2020. We exclude the years 2000 and 2001 in the sample, excluding the dot-com bubble. The returns are monthly in percent per year (monthly return multiplied by twelve).

		Full Sample (1995-2020)	Full Sample (Exc. Dot-com Bubble)	1995-1999	2000-2007	2008-2020
		(1)	(2)	(3)	(4)	(5)
INT^{10K} - HML^{FF}	E[R]	**3.39**	**7.13**	**13.6**	**-16.42**	**12.4**
		(3.53)	(7.12)	(6)	(-9.66)	(9.13)
	σ	21.95	217.35	19.15	28.02	17.45
	Sharpe	0.15	0.41	0.71	-0.59	0.71
INT^{10K} - HML^{PT}	E[R]	1.5	**4.78**	**11.97**	**-16.18**	**9.05**
		(1.61)	(4.92)	(5.46)	(-9.84)	(6.88)
	σ	21.1	16.36	16.79	27.66	16.75
	Sharpe	0.07	0.29	0.71	-0.59	0.54
INT^{10K} - HML^{EKP}	E[R]	0.92	**2.51**	**9.23**	**-12.7**	**6.66**
		(1.06)	(2.77)	(4.51)	(-8.27)	(5.41)
	σ	18.21	15.31	15.12	23.4	14.99
	Sharpe	0.05	0.16	0.61	-0.55	0.44

Table 7: Performance Statistics of Composite Portfolios

In this table, we report $INT^{10K} - HML^{FF}$, which is the portfolio that longs INT^{10K} and shorts HML^{FF}. The same method holds for other rows. The values in parentheses are T-statistics that test the difference between the two means to be zero. The Sharpe ratio is $E[R_p - R_b]/\sigma(R_p - R_b)$. The full sample is from July 1995 to June 2020. The returns are monthly in percent per year (monthly return multiplied by twelve).

		Full Sample (1995-2020)	Full Sample (Exc. Dot-com Bubble)	Dot-com Bubble (2000-2001)	1995-1999	2000-2007	2008-2020
		(1)	(2)	(3)	(4)	(5)	(6)
	E[R]	3.61	**4.95**	-11.78	10.19	-6.78	**7.9**
		(1.26)	(2.19)	(-0.94)	(1.31)	(-1.32)	(3.24)
INT^{10K}	σ	12.38	9.59	29.49	10.2	17.28	8.45
	Sharpe	0.29	0.52	-0.4	1	-0.39	0.93
	E[R]	0.38	1.31	-10.32	9.47	**-7.82**	2.36
		(0.18)	(0.69)	(-1.18)	(1.8)	(-2.55)	(1.07)
SG&A$^{portf.}$	σ	10.35	9.15	19.51	9.48	12.07	9.15
	Sharpe	0.04	0.14	-0.53	1	-0.65	0.26
	E[R]	3.2	**4.84**	-15.66	13.69	-6.39	**5.56**
		(0.91)	(2.13)	(-0.93)	(1.64)	(-0.9)	(2.4)
R&D$^{portf.}$	σ	14.49	11.33	34.06	12.64	20.32	9.62
	Sharpe	0.22	0.43	-0.46	1.08	-0.31	0.58
	E[R]	2.98	**4.21**	-11.11	5.18	-5.31	**7.5**
		(1.14)	(1.91)	(-0.83)	(0.69)	(-1.23)	(2.67)
INT^{PT}	σ	11.81	9.51	26.73	11.45	14.56	9.63
	Sharpe	0.25	0.44	-0.42	0.458	-0.36	0.78
	E[R]	2.66	3.79	-10.3	3.79	-4.52	**6.85**
		(1.15)	(1.87)	(-0.82)	(0.59)	(-1.15)	(2.52)
INT^{EKP}	σ	10.6	8.71	10.43	7.22	12.69	8.94
	Sharpe	0.25	0.43	-0.44	0.36	-0.36	0.77

Table 8: INT^{10K} vs. Portfolios of Intangible Intensive Stocks

In this table, we summarize the risk and return associated with textual intangible value and other measures of value and intangible value in the literature. SG&A and R&D portfolios are sorted based on (SG&A - R&D)/Total Expense and R&D/Total Expense. INT^{PT} and INT^{EKP} are sorted based only on the intangible component of intangible augmented book-to-market calculated using Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020) methods. The values in parentheses are Newey-West T-statistics that test the difference in the means to be zero. The returns are monthly in percent per year (monthly return multiplied by twelve).

Category		Full Sample (1995-2020) (1)	Full Sample (Exc. Dot-com Bubble) (2)	Dot-com Bubble (2000-2001) (3)	1995-1999 (4)	2000-2007 (5)	2008-2020 (6)
Innovation assets & Information Technology	E[R]	3.44	**4.89**	-13.19	10.75	-5.32	**6.42**
		(1.14)	(2.18)	(-0.94)	(1.24)	(-0.92)	(2.54)
	σ	12.74	9.76	30.75	11.22	17.8	8.52
	Sharpe	0.27	0.5	-0.43	0.96	-0.3	0.75
Brand & Customer Relations	E[R]	2.37	**3.59**	-11.62	10.09	-6.49	**5.27**
		(1.1)	(2.04)	(-1.14)	(1.37)	(-1.71)	(3.37)
	σ	10.08	7.81	23.95	8	14.03	7
	Sharpe	0.24	0.46	-0.49	1.26	-0.46	0.75
Human Resources	E[R]	**2.49**	**2.79**	-1.01	3.9	-1.42	**4.49**
		(2.03)	(2.4)	(-0.2)	(1.77)	(-0.68)	(2.93)
	σ	6.54	5.85	12.06	5.08	8.51	5.41
	Sharpe	0.38	0.48	-0.08	0.77	-0.17	0.83

Table 9: INT^{10K} Categorical Returns

The values in parentheses are Newey-West T-statistics that test the difference in the means to be zero. The full sample is from July 1995 to June 2020. We exclude the years 2000 and 2001 in the sample, excluding the dot-com bubble. The returns are monthly in percent per year (monthly return multiplied by twelve).

	Dependent variable: INT^{10K}					
	(1)	(2)	(3)	(4)	(5)	(6)
$\alpha(\%)$	**3.26**	**4.48**	**4.32**	**5.59**	**6.23**	**8.28**
	(2.24)	(2.67)	(1.98)	(3.96)	(4.18)	(4.06)
β_{MktRF}	0.11	0.12	0.16	0.03	0.06	-0.03
	(2.29)	(2.47)	(2.20)	(0.92)	(1.41)	(-0.5)
β_{SMB}	0.16	0.24	0.23	0.02	0.10	-0.01
	(2)	(2.55)	(2.44)	(0.26)	(1.07)	(-0.19)
$\beta_{HML^{FF}}$	-0.85			-0.69		
	(-10.46)			(-8.83)		
$\beta_{HML^{PT}}$		-0.9			-0.75	
		(-13.42)			(-8.19)	
$\beta_{HML^{EKP}}$			-0.77			-0.25
			(-5.71)			(-2.37)
β_{RMW}				-0.41	-0.37	-0.65
				(-4.13)	(-3.22)	(-6.42)
β_{CMA}				-0.03	0.00	-0.49
				(-0.31)	(0.03)	(-2.79)
β_{UMD}	-0.11	-0.06	-0.05	-0.09	-0.05	-0.02
	(-1.63)	(-1.23)	(-0.58)	(-1.62)	(-1.26)	(-0.35)
Adj. $R^2(\%)$	62.72	62.41	35.22	68.06	66.5	52.17
obs	300	300	300	300	300	300

Table 10: Alphas - INT^{10K}

In this table, we report portfolio alphas and betas by regressing INT^{10K} against factor models. Columns (1) through (3) use the Fama and French (2015) three factors + momentum. HML portfolios is based on traditional book-to-market and its intangible augmented versions from Peters and Taylor (2017) and Eisfeldt, Kim and Papanikolaou (2020). Columns (4) through (6) use the Fama and French (2015) five-factor model + momentum. We include Newey-West T-statistics. The sample is monthly from July 1995 to June 2020. All coefficients are reported in percentage per year (monthly returns multiplied by twelve).

	Dependent variable: Categorical INT^{10K}					
Category:	Innovation assets &information technology	Band & customer relation	Human resources	Innovation assets & information technology	Band & customer relation	Human resources
$\alpha(\%)$	**2.92**	2.28	**2.38**	**5.09**	**4.6**	**2.66**
	(2.06)	(1.72)	(2.57)	(4.00)	(2.93)	(2.63)
β_{MktRF}	0.11	0.08	0.04	0.05	0.00	0.03
	(2.57)	(1.8)	(1.55)	(1.39)	(0.12)	(1.40)
β_{SMB}	0.19	0.04	0.07	0.05	-0.06	0.02
	(2.89)	(0.79)	(1.91)	(0.83)	(-1.03)	(0.5)
$\beta_{HML^{FF}}$	-0.92	-0.53	-0.37	-0.77	-0.35	-0.37
	(-13.55)	(-4.96)	(-11.5)	(-13.65)	(-3.95)	(-7.45)
β_{RMW}				-0.40	-0.34	-0.12
				(-4.59)	(-4.07)	(-1.51)
β_{CMA}				0.00	-0.16	0.12
				(-0.06)	(-1.25)	(1.15)
β_{UMD}	-0.09	-0.1	-0.04	-0.07	-0.08	-0.04
	(-1.41)	(-1.35)	(-1.07)	(-1.39)	(-1.28)	(-1.1)
Adj. $R^2(\%)$	70.78	37.94	41.75	75.78	43.68	44.74
obs	300	300	300	300	300	300

Table 11: Alphas - Categorical INT^{10K}

In this table, we report portfolio alphas and betas by regressing each category of INT^{10K} against factor models. Columns (1) through (3) use the Fama and French (2015) three factors + momentum. Columns (4) through (6) use the Fama and French (2015) five-factor model + momentum. We include Newey-West T-statistics. The sample is monthly from July 1995 to June 2020. All coefficients are reported in percentage per year (monthly returns multiplied by twelve).

Intangibles Talk →	Low	2	3	4	High	H-L
$\alpha(\%)$	**-3.56**	-0.47	-0.57	-0.58	**2.9**	**6.46**
	(-3.61)	(-0.45)	(-0.69)	(0.82)	(2.72)	(3.82)

R&D to Total Expenses →	Low	2	3	4	High	H-L
$\alpha(\%)$	-0.12	-3.12	-0.29	-1.23	**4.01**	**4.12**
	(-0.07)	(-1.99)	(-0.23)	(-0.87)	(3.02)	(2.37)

Intangible Capital$^{\text{PT}}$ to Total Assets →	Low	2	3	4	High	H-L
$\alpha(\%)$	**-2.11**	-2.14	1.17	2.11	1.27	3.38
	(-2.01)	(-1.82)	(1.17)	(1.52)	(1.1)	(1.91)

Intangible Capital$^{\text{EKP}}$ to Total Assets →	Low	2	3	4	High	H-L
$\alpha(\%)$	-2.02	**-2.33**	0.58	2.28	1.67	**3.69**
	(-1.86)	(-2.19)	(0.62)	(1.87)	(1.45)	(2.13)

Table 12: Alphas of Portfolios Sorted based on Intangible Intensity Indicators

The tables above report ten standard Fama and French regressions with 5 factors. In each regression, the dependent variable is the excess returns (over risk-free rate) of 5 value-weighted portfolios sorted based on intangibles talk, R&D to total expenses, Intangible Capital$^{\text{PT}}$ to total assets:

$$\text{R}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{MKT}} MKT_t + \boldsymbol{\beta_{SMB}} SMB_t + \boldsymbol{\beta_{HML}} HML_t + \boldsymbol{\beta_{RMW}} RMW_t + \boldsymbol{\beta_{CMA}} CMA_t + \epsilon_{i,t}$$

Where $R_{i,t}$ is the excess return (over risk-free rate) of a value-weighted portfolio at month t that is long in stocks belonging to the ith decile based on intangibles talk or R&D to total expense. The portfolio i is reshuffled yearly at the end of June based on the indicators values from the previous year. Only the alphas from the regressions are reported for expositional purposes. We include Newey-West T-statistics in parentheses. The sample is monthly from July 1995 to June 2020.

Intangibles $\rightarrow$	Low	2	3	4	High	$[H-L]$	
Talk							
Low IVOL	-1.94	1.18	0.43	1.43	1.17	3.11	(1.51)
2	-3.17	-0.26	-0.63	1.81	0.72	3.89	(1.55)
3	-3.02	-3.78	-0.89	-1.82	4.2	**7.22**	(2.72)
4	-9.21	-3.04	-3.97	-2.75	5.35	**14.56**	(3.67)
High IVOL	-12.69	-7.86	0.2	-0.04	2.7	**15.39**	(3.61)
$[H-L]$	**-10.76**	**-9.04**	-0.24	-1.47	1.53		
	(-2.97)	(-2.34)	(-0.07)	(-0.55)	(0.65)		

Table 13: Alphas of Portfolio Returns Sorted based on Idiosyncratic Volatility (IVOL), and Intangibles Talk

The tables above each report the alphas of 25 portfolios against the Fama-French 5-factor model:

$$R_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{MKT}} MKT_t + \boldsymbol{\beta_{SMB}} SMB_t + \boldsymbol{\beta_{HML}} HML_t + \boldsymbol{\beta_{RMW}} RMW_t + \boldsymbol{\beta_{CMA}} CMA_t + \epsilon_{i,t}$$

Where $R_{i,t}$ is the excess return (over risk-free rate) of value-weighted portfolio i. Element jk in each table is the alpha of a value-weighted portfolio that is long in stocks belonging to the jth quantile based on IVOL and the kth quantile based on intangibles talk. The portfolio is reshuffled yearly at the end of June based on idiosyncratic Volatility (IVOL) and intangibles talk from the previous year. Only the alphas from the regressions are reported for expositional purposes. We include Newey-West T-statistics in parentheses. The sample is monthly from July 1995 to June 2020.

Small Firms

Intangibles → Talk	Low	2	3	4	High	$[H-L]$	
Low IVOL	-0.95	2	-0.64	2.47	-0.36	0.59	(0.24)
2	0.46	0.67	-0.48	-2.14	1.68	1.22	(0.68)
3	1.91	-1.38	-0.43	-1.29	2.22	0.31	(0.11)
4	-4.65	-1.94	-1.4	-1.78	2.5	**7.15**	(2.72)
High IVOL	-9.95	-5.57	-6.19	-2.25	2.39	**12.34**	(3.24)
$[H-L]$	**-9**	**-7.57**	-5.55	-4.72	2.75		
	(-2.26)	(-2.31)	(-1.81)	(-1.77)	(0.83)		

Big Firms

Intangibles → Talk	Low	2	3	4	High	$[H-L]$	
Low IVOL	-2.41	0.73	0.51	1.33	-0.55	1.86	(0.83)
2	-4.17	-1.02	-1.78	1.31	3.75	**7.93**	(2.91)
3	-0.29	-3.03	0.48	-5.5	2.05	2.34	(0.58)
4	-10.3	-4.02	-4.78	-3.28	5.93	**16.23**	(3.71)
High IVOL	-14.84	-6.66	6.84	-0.4	4.69	**19.53**	(3.43)
$[H-L]$	**-12.43**	-7.39	6.33	-1.73	5.24		
	(-2.52)	(-1.47)	(1.25)	(-0.38)	(1.7)		

Table 14: Alphas of Portfolio Returns Sorted based on Size, Idiosyncratic Volatility (IVOL), and Intangibles Talk

The tables above each report the alphas of 25 portfolios against the Fama-French 5-factor model:

$$\mathrm{R}_{i,t} = \boldsymbol{\alpha_i} + \boldsymbol{\beta_{MKT}} MKT_t + \boldsymbol{\beta_{SMB}} SMB_t + \boldsymbol{\beta_{HML}} HML_t + \boldsymbol{\beta_{RMW}} RMW_t + \boldsymbol{\beta_{CMA}} CMA_t + \epsilon_{i,t}$$

Element ij in each table is the alpha of a value-weighted portfolio that is long in stocks belonging to the ith quantile based on IVOL and the jth quantile based on intangibles talk. The portfolio is reshuffled yearly at the end of June based on idiosyncratic Volatility (IVOL) and intangibles talk from the previous year. Only the alphas from the regressions are reported for expositional purposes. We include Newey-West T-statistics in parentheses. The sample is monthly from July 1995 to June 2020.

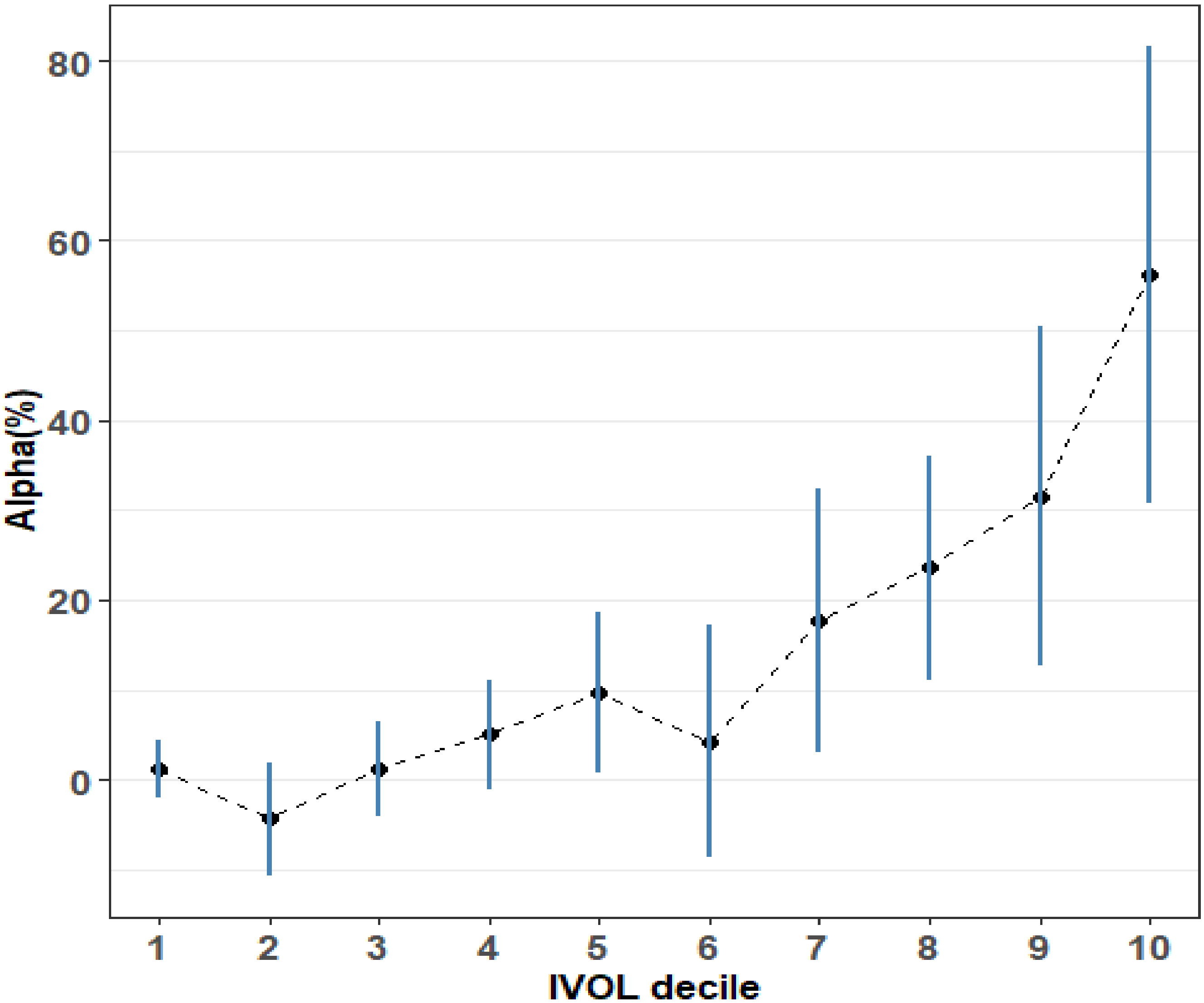


Figure 7: Alphas by IVOL Decile

The tables above plots the Fama-French 5-factor alpha of ten INT^{10K} by IVOL decile. Every year stocks are sorted into deciles based on their IVOL and INT^{10K} is constructed separately using the stocks in each decile. Alphas are estimated using annual returns (monthly returns multiplied by twelve). The blue line represents the ($\mu \pm 1.96 \times Newey-West\ standard\ errors$). The sample is monthly, covering the period between July 1995 to June 2020.

References

Aboody, David. 1996. "Recognition versus disclosure in the oil and gas industry." *Journal of Accounting Research* 34:21–32.

Aboody, David and Baruch Lev. 1998. "The value relevance of intangibles: The case of software capitalization." *Journal of Accounting Research* 36:161–191.

Ahmed, Anwer S, Emre Kilic and Gerald J Lobo. 2006. "Does recognition versus disclosure matter? Evidence from value-relevance of banks' recognized and disclosed derivative financial instruments." *Accounting Review* 81(3):567–588.

Ang, Andrew, Robert J Hodrick, Yuhang Xing and Xiaoyan Zhang. 2006. "The cross-section of volatility and expected returns." *Journal of Finance* 61(1):259–299.

Arnott, Robert D, Campbell R Harvey, Vitali Kalesnik and Juhani T Linnainmaa. 2021. "Reports of value's death may be greatly exaggerated." *Financial Analysts Journal* 77(1):44–67.

Asness, Clifford S, R Burt Porter and Ross L Stevens. 2000. "Predicting stock returns using industry-relative firm characteristics." *Social Science Research Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=213872".

Belo, F, V Gala, J Salomao and MA Vitorino. 2022. "Decomposing firm value (0898-2937)." *Journal of Fnancial Economics 143 (2), 619–639* .

Birru, Justin and Trevor Young. 2020. "Sentiment and uncertainty." *Fisher College of Business Working Paper* (2020-03):010.

Cao, Jie and Bing Han. 2016. "Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns." *Journal of Banking and Finance* 73:1–15.

Carhart, Mark M. 1997. "On persistence in mutual fund performance." *Journal of finance* 52(1):57–82.

Chan, Louis KC, Josef Lakonishok and Theodore Sougiannis. 2001. "The stock market valuation of research and development expenditures." *Journal of finance* 56(6):2431–2456.

Choi, Ki-Soon, Eric C So and Charles CY Wang. 2021. "Going by the Book: Valuation Ratios and Stock Returns." *Social Science Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3854022.".

Corrado, Carol, Charles Hulten and Daniel Sichel. 2005. Measuring capital and technology: An expanded framework. In *Measuring capital in the new economy*. University of Chicago Press pp. 11–46.

Curtis, Asher, Sarah McVay and Sara Toynbee. 2020. "The changing implications of research and development expenditures for future profitability." *Review of Accounting Studies* 25:405–437.

Daniel, Kent and Sheridan Titman. 2006. "Market reactions to tangible and intangible information." *Journal of Finance* 61(4):1605–1643.

Davis-Friday, Paquita Y, L Buky Folami, Chao-Shin Liu and H Fred Mittelstaedt. 1999. "The value relevance of financial statement recognition vs. disclosure: Evidence from SFAS No. 106." *Accounting Review* 74(4):403–423.

Deng, Zhen, Baruch Lev and Francis Narin. 1999. "Science and technology as predictors of stock performance." *Financial Analysts Journal* 55(3):20–32.

Edmans, Alex. 2011. "Does the stock market fully value intangibles? Employee satisfaction and equity prices." *Journal of Financial Economics* 101(3):621–640.

Edmans, Alex, Lucius Li and Chendi Zhang. 2014. Employee satisfaction, labor market flexibility, and stock returns around the world. Technical report National Bureau of Economic Research, Cambridge, MA.

Eisfeldt, Andrea L and Dimitris Papanikolaou. 2013. "Organization capital and the cross-section of expected returns." *Journal of Finance* 68(4):1365–1406.

Eisfeldt, Andrea L, Edward Kim and Dimitris Papanikolaou. 2020. Intangible value. Technical report National Bureau of Economic Research, Cambridge, MA.

Enache, Luminita and Anup Srivastava. 2018. "Should intangible investments be reported separately or commingled with operating expenses? New evidence." *Management Science* 64(7):3446–3468.

Engle, Robert F, Stefano Giglio, Bryan Kelly, Heebum Lee and Johannes Stroebel. 2020. "Hedging climate change news." *Review of Financial Studies* 33(3):1184–1216.

Ewens, Michael, Ryan H Peters and Sean Wang. 2019. Measuring intangible capital with market prices. Technical report National Bureau of Economic Research, Cambridge, MA.

Falato, Antonio, Dalida Kadyrzhanova, Jae Sim and Roberto Steri. 2022. "Rising intangible capital, shrinking debt capacity, and the US corporate savings glut." *Journal of Finance* 77(5):2799–2852.

Fama, Eugene F and Kenneth R French. 2015. "A five-factor asset pricing model." *Journal of Financial Economics* 116(1):1–22.

Filipovic, Zoran and Alexander F Wager. 2019. "The intangibles song in takeover announcements: Good tempo, Hollow Tune." *Social Science Research Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3330623. .

Garcia, Diego. 2013. "Sentiment during recessions." *Journal of Finance* 68(3):1267–1300.

Hall, Richard. 2009. The strategic analysis of intangible resources. In *Knowledge and strategy*. Routledge, London, England, pp. 181–195.

Hulten, Charles R and Xiaohui Hao. 2008. What is a Company Really Worth? Intangible Capital and the" Market to Book Value" Puzzle. Technical report National Bureau of Economic Research, Cambridge MA.

Iqbal, Aneel, Shivaram Rajgopal, Anup Srivastava and Rong Zhao. 2022. "Value of internally generated intangible capital." *Social Science Research Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3917998. .

Jiang, Fuwei, Joshua Lee, Xiumin Martin and Guofu Zhou. 2019. "Manager sentiment and stock returns." *Journal of Financial Economics* 132(1):126–149.

Lev, Baruch. 2000. *Intangibles: Management, Measurement, and Reporting*. Brookings Institution Press, Washington, DC.

Lev, Baruch. 2005. "Intangible assets: Concepts and measurements. In: Kempf-Leonard, K. (Ed.), Encyclopedia of Social Management." 2.

Lev, Baruch. 2012. *Winning Investors Over: Surprising Truths about Honesty, Earnings Guidance, and Other Ways to Boost Your Stock Price*. Harvard Business Review Press, Boston, MA.

Lev, Baruch and Anup Srivastava. 2022. "Explaining the recent failure of value investing." *Social Science Research Network.* "https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3442539".

Lev, Baruch and Theodore Sougiannis. 1996. "The capitalization, amortization, and value-relevance of R&D." *Journal of Accounting and Economics* 21(1):107–138.

Loughran, Tim and Bill McDonald. 2011. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *Journal of Finance* 66(1):35–65.

Mandelbrot, Benoit. 1961. "On the theory of word frequencies and on related Markovian models of discourse. In:Structure of language and its mathematical aspects." *American Mathematical Society, Providence* 12:190–219.

Newey, Whitney K and Kenneth D West. 1987. "Hypothesis testing with efficient method of moments estimation." *International Economic Review* pp. 777–787.

Park, Hyuna. 2019. "An intangible-adjusted book-to-market ratio still predicts stock returns." *Critical Finance Review* 25(1):207–236.

Peters, Ryan H and Lucian A Taylor. 2017. "Intangible capital and the investment-q relation." *Journal of Financial Economics* 123(2):251–272.

Pontiff, Jeffrey. 1996. "Costly arbitrage: Evidence from closed-end funds." *Quarterly Journal of Economics* 111(4):1135–1151.

Seamons, Quinton F and Robert W Rouse. 1997. "MD&A: How to use safe harbor to report "soft" information." *Journal of Corporate Accounting and Finance* 9(1):81–90.

Srivastava, Anup. 2014. "Why have measures of earnings quality changed over time?" *Journal of Accounting and Economics* 57(2-3):196–217.

Srivastava, Anup. 2023. “Trivialization of the bottom line and losing relevance of losses.” *Review of Accounting Studies (forthcoming)* .

Stambaugh, Robert F, Jianfeng Yu and Yu Yuan. 2012. “The short of it: Investor sentiment and anomalies.” *Journal of Financial Economics* 104(2):288–302.

Stambaugh, Robert F, Jianfeng Yu and Yu Yuan. 2015. “Arbitrage asymmetry and the idiosyncratic volatility puzzle.” *Journal of Finance* 70(5):1903–1948.

Tetlock, Paul C. 2007. “Giving content to investor sentiment: The role of media in the stock market.” *Journal of Finance* 62(3):1139–1168.

Wyatt, Anne. 2008. “What financial and non-financial information on intangibles is value-relevant? A review of the evidence.” *Accounting and Business Research* 38(3):217–256.