From Text to Insight - A Novel Approach to Measuring Business Model Innovation

Max Gabler Wanshu Jiang Christoph Kiesl Leonard Pöhls

The ability of a company to continuously innovate its business model is a pivotal determinant of long-term success in dynamic markets. It is therefore crucial to ensure the reliability of business model innovation measurement. In this study, we analyze business descriptions from 10-K filings from a sample of 4492 different companies between 2017 and 2023, which we summarized using a large language model (LLM). The innovation of business models is measured on the basis of the annual changes in the business descriptions, with BERTS scores serving as quantitative comparison indices for the business descriptions. Our findings indicate that above-market business model innovation positively influences a firm's sales growth but has a negative effect on Tobin's Q growth. Notably, small-cap companies experience significant sales growth linked to business model innovation. These insights highlight the potential of textual similarity in regulatory filings as a robust indicator of business model innovation, offering a novel approach to tracking such innovation over time.

Introduction

Business model innovation (BMI) is a key activity to maintain competitiveness and even gain a competitive advantage in todays fast paced markets (Pucihar et al. 2019; Teece 2018). It is therefore no surprise that the interest in BMI has grown rapidly over the last twenty years. In particular, research examining the impact of BMI on firm performance has been a prominent area of investigation, with numerous research papers published in this field (Cucculelli and Bettinelli 2015; Latifi, Nikou, and Bouwman 2021; Zott and Amit 2008; White et al. 2022). While the financial literature offers a wide range of established methods for measuring a company's performance, the BMI literature provides only a limited number of measures, all of which face similar challenges (White et al. 2022). Furthermore, these measures vary largely. In order to further validate and advance the BMI research field, more sophisticated and comprehensive measurement instruments are necessary (Huang and Ichikohji 2023).

Scales and measures used in the BMI literature (Clauss 2017; Spieth and Schneider 2016) provide managers and practitioners with a measurement index for BMI. But these measures only validate applicability of BMI theory (Huang and Ichikohji 2023) and are insufficient for longitudinal studies (Clauss 2017). Hence, these measures are not adequate for a time series analysis of BMI. We address this gap by proposing a novel approach to measuring BMI. US-based companies are obliged by the United States Security and Exchange Commission (SEC) to submit annual 10-K filings, wherein a detailed description of the company's business operations is required. Hoberg & Phillips (2016), on which we build this study, use these filings to cluster companies into industries. We, on the other hand, summarize these descriptions with Gemini and calculate the BERTScore between the summaries of different years for a single company. This approach enables the measurement of changes in the business model (BM) over time as the distance between the BM summary of one year to another. In order to test the validity of our measure, we regress sales growth and Tobin's Q growth on our measure. Additionally, we create our own industry classification based on BERTScores of business descriptions of same firms within the same year.

We conduct multivariate regressions and find that our measure is positively correlated with firm performance, showing its validity. Furthermore, we find that BMI has a greater impact on firm performance for smaller companies than medium and large companies. Consequently, we contribute to the literature by introducing a novel approach of measuring BMI, which overcomes drawbacks from existing approaches. Our measure and approach enable researchers to study BMI in longitudinal and large-scale studies. Lastly, we introduce our own industry classification based on a company's BM, which represents an alternative to existing ones.

The SEC mandates that the majority of public companies based in the United States submit an annual 10-K filing. In the first section a company presents its general business, encompassing information about its products and services. In some instances additional topics may be addressed, such as labor issues or competition (SEC 2024). In conclusion, this section contains the most useful information for describing a company's BM (Lee and Hong 2014). Furthermore, 10-K filings are a reliable source of information, given that US law prohibits false

or misleading statements in the filings. The SEC monitors the compliance of the companies with the requirements and comments where disclosure appears to be inconsistent (SEC 2024). Because of these guidelines, these descriptions are particularly suitable for a text analysis in order to quantify a BM.

Despite the growing interest in BMI and the increasing number of theoretical and empirical studies in this field, the research of BMI is still in a preliminary state (Huang and Ichikohji 2023). Consequently, there is considerable variation in the definitions of BMI, with some definitions being more similar to one another than others (Foss and Saebi 2017). Spieth & Schneider (2016) identify three core dimensions that comprise a company's BM: its value proposition, its value creation architecture and its revenue model logic. Based on this, BMI can be conceptualized as a change that is new-to-the-firm in at least one of these dimensions. Furthermore, Spieth and Schneider (2016) introduce a measurement model to evaluate these three dimensions of BMI. They develop an index by first specifying the contents, followed by a specification of the indicators and assessing their content validity, assessing the indicators collinearity and finally assessing the external validity. Clauss (2017) employs a very similar approach. After specifying the domain and dimensionality of BMI through literature research, the author divides his scale into three hierarchical levels, which are very similar to the ones of Spieth and Schneider (2016). We build on these conceptualizations to design our prompt we use in the pre-processing with Gemini. However, both measures are subject to three significant limitations. Firstly, both measures lack a temporal component. Consequently, they are inadequate for use in longitudinal studies or ex-post evaluations of BMI. Secondly, BMI is only measured at the new-to-the-firm level rather than at the new-to-the-industry or new-to-the-market level. Thirdly, both measures rely on interviews and questionnaires, which makes conducting large-scale studies time-consuming and reliant on the willingness of the companies to cooperate (Clauss 2017; Spieth and Schneider 2016). Our novel measurement approach tackles the first and third issue.

A number of studies have examined the relationship between BMI and the financial performance of a company. Cucculelli & Bettinelli (2015) investigate the effect of BMI on sales growth, return on sales (ROS) and total factor productivity (TFP). The results support the hypothesis that BMI has a positive effect on firm performance, with the effect increasing in line with the intensity of the innovation. Desyllas et al. (2022) find that BMI has a small effect on performance of incumbent firms. They measure firm performance by Tobin's Q growth. White et al. (2022) conducted a meta-analysis based on the extant BMI literature. They found a positive relationship between BMI and firm performance, and that this relationship is shaped by factors including the firm age, industry, the economic and political environment and BMI characteristics. Based on this, we derive the dependent and control variables in the estimation strategy.

Hoberg & Phillips (2016) present a novel approach to defining industry boundaries. They propose two novel industry classification methods: the fixed industry classification (FIC) and the text-based network industry classification (TNIC). Firstly, they cluster companies based on the similarity of word vectors into fixed industries. Secondly, they define a minimum

similarity threshold, above which firms are considered in the same industry. This second step relaxes their prior properties of binary membership transitivity and fixed industry location. The authors demonstrate shortcomings in the traditional industry classification systems such as the Standard Industry Classification (SIC) and the North American Industry Classification System (NAICS), which are not able to account for temporal changes. The new method is capable of capturing changes in industry boundaries and competitor sets over time, thereby providing a dynamic industry classification system. Based on the FIC we propose our own BERTScores industry classification and utilize it in our estimation. In their study, Lee & Hong (2014) examine the evolution of a firm's BM over time. After filtering the Item 1 parts of the 10-K filings for relevant sentences, Lee & Hong (2014) construct keyword vectors, which represent the concept of the BM. Therefore, the evolution of the BM is depicted as the change in the distribution of keywords over time. The authors advocate for a more robust methodology, such as incorporating multi-word phrases in the keyword vectors, to enhance the reliability of the approach (Lee and Hong 2014). Our study pursues a similar goal but with a novel methodology.

The remainder of the paper is organized as follows. Section 2 describes the origin and preparation of our data, the use and functioning of BERT, the preprocessing with Gemini and our methodology. Section 3 lays out our estimation strategy. Section 4 contains our results and discussion. Section 5 concludes our study.

Data and Methodology

The Dataset

We collect 10-K filings from the digital SEC Database, using the category "10-K" as extraction condition. Since the focus of our study lies on company's BM, we merely use the Item 1 part, since this is the most crucial part of the 10-K filings for describing the companies BM (Lee and Hong 2014). Our observations are limited to an intersection of such companies, which has been made available to the SEC since 2001 in a publicly accessible database. We extracted 10-K filings that were submitted between 2017 and 2023 based on underlying Central Index Keys (CIK). Occasionally, such filings are submitted retrospectively or are already submitted for the same year. We are therefore limiting the period for which we are reporting to 2016-2023, with fewer observations available for 2016 and 2023 as a result. We exclude companies from the financial sector, namely companies with a SIC Code starting with six. Corresponding to Table 2,

Table 1: 10-K Sample Creation

Source/Filter	Sample Size	Observations Removed
1. Original (exchanged listed) companies, whose 10-K filings are extracted from SEC	47,226	0
2. Removing observations from financial companies whose SIC-code start with '6'	37,750	9,476
3. Verify for Item 1 text availability (removed oberservations that are attributable to API quality)	32,611	5,139
4. Extracting dates for which the filings are reporting for and removing of duplicated filings	30,737	1,874
5. Delete observations with incorrect date assignment (some companies submitted two or more filings due to addendums or data quality)	30,131	606
6. Merged Gemini processed Item 1 text to the underlying data set. We did not consider texts that were not processable (e.g. due to recitation errors)	28,350	1,781
7. Extract financials statements from DataStream and merge them. Also remove observations for report-for-years prior to 2016	21,417	6,933

Note:
Filings submitted between 2017 and 2023 are considered.

multiple steps of pre-processing were required to obtain the final amount of 21,417 observations for seven years. Financial key figures, including sales, total assets, market values and Tobin's Q were extracted from DataStream. A total of 4,494 companies are included in the sample, although the availability of filings could not always be guaranteed for all years. This is due on the one hand to the quality of the API to the SEC and on the other hand to companies that did not file 10-K reports or were listed on the stock exchange for the entire period under review. Finally, we have access to the financial key figures of the companies for the respective year, the Item I text pre-processed with the help of Gemini, company-specific identification features and the conventional SIC industry classification. The information on SIC sector classification is limited to companies that are currently actively filing. Therefore, 609 firms are no longer actively filing, e.g. due to company bankruptcies or mergers.

BERT and BERTScore

BERT is a pre-trained and transformer-based model for natural language processing (NLP) based on artificial neural networks. It works according to the transformer architecture, first

mentioned by Vaswani et al. (2017). Unlike Hoberg & Philips' (2016) word-to-vec approach, BERT operates bidirectionally, considering the context from both sides of each word simultaneously. This bidirectional design helps clarify word meanings based on context, resulting in more accurate similarity calculations between texts that may use the same words but convey different meanings. BERT can also be fine-tuned for specific tasks, making it adaptable to different datasets, improving its performance even with limited labeled data. Thus, BERT effectively captures deeper semantics in texts such as 10-K reports. The BERTScore, a metric built on BERT embeddings, computes cosine similarity between word or text meanings learned by BERT, where a scale from -1 to 1 is used, with 1 representing perfect similarity.

In our study, we employ the 'bert-base-uncased' model in two distinct but complementary tasks. First, for inter-company comparisons, we use BERT to generate embeddings for product descriptions from different companies by processing each entry. The BERT tokenizer converts input text into tensors, and the model generates embeddings by averaging the token embeddings from the last hidden state. We finally compute pairwise cosine similarity scores between them to assess semantic similarity across companies' product descriptions. In parallel, we apply a similar approach to assess year-to-year evolution of product descriptions within individual companies. For each company (identified by its CIK), we retrieve product descriptions from two consecutive years and generate embeddings using the same BERT model. Here, we utilize the BERTScore metric to compare descriptions from adjacent years, such as 2017 and 2018. The similarity is calculated based on precision (P), recall (R), and F1-score, with the F1-score serving as the primary measure. This year-to-year comparison enables us to analyze how product descriptions evolve over time.

Preprocessing with Gemini

10-K filings are typically very large text documents, and Item 1 of these filings is no exception. Table 2 shows the descriptive measures of the length of the original Item 1 section in our final sample. The length of a document was measured by the word count without punctuation. The document length ranges from a minimum if 49 words up to 78,799 words. On average the documents are between 6,626 and 10,304 words long. In order to utilize the entirety

Table 2: Descriptive Statistics of Number of Words in Original Filings

Report-for Year	Average Word Count	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
2016	7,765	6,053	134	3,658	5,972	10,251	51,227
2017	7,426	6,162	134	3,474	5,689	9,538	70,611
2018	7,552	6,290	49	3,492	5,738	9,599	74,351
2019	7,943	6,562	59	3,646	5,961	10,241	78,270
2020	8,610	7,157	59	3,930	6,518	10,914	57,980
2021	10,304	8,347	235	4,671	7,614	13,637	78,799
2022	9,478	7,977	57	4,321	7,057	11,919	73,937
2023	6,626	4,750	190	3,667	5,807	8,366	43,523

Note:
All 21,417 original filings were considered.

of the information regarding the BM in the Item 1 section and pass the text to our BERT model, we decided to let Google's GenAI chatbot Gemini summarize them to a maximum length of 512 tokens. The summaries were created between 26 June 2024 and 6 August 2024. The model employed was Gemini Flash 1.5. The prompt was inserted at the beginning of each text file and it was passed via an API to Gemini [1]. Our prompt covers all aspects of the definition of BMI proposed by Spieth & Schneider (2016). For more details, see Appendix A. To assess the quality and accuracy of the summaries produced by Gemini, a random sample of 100 filings was selected for comparison with the original text. More precise,the original file was initially read with a focus on the points mentioned in the prompt. Subsequently, the summary was evaluated to ascertain whether it contained these same points. A list of the sample with the summaries is provided in the Appendix B.

Methodology

From the dataset containing all the summaries and financials per company per year we construct three datasets for our further analysis. The first dataset is used for the BERTScore industry classification. Therefore, we filter for all summaries of 2017 and for the first available summary of companies which do not appear first in 2017 but in a later year. We then calculate the BERTScores between all these summaries. The second and third dataset are used for the regression analysis of our measure. For the second dataset, we fix the company and calculate the BERTScore of the summaries of one year and its direct consecutive year as well as the sales growth rate. The third dataset contains the Tobin's Q growth calculated as the difference between the natural logarithm of Tobin's Q in the last and the first year of our observed period as well as the BERTScore between the summaries of the respective year of a company. The formulas of Sales Growth and Tobin's Q Growth are given by:

$$\text{Sales Growth} = (\frac{\text{Revenue in year } t}{\text{Revenue in year } t-1} - 1) \times 100$$

$$\text{Tobin's Q Growth} = (\frac{\text{Tobin's Q in year } t}{\text{Tobins's Q in year } t-1} - 1) \times 100$$

$$\text{for } t \in \{2017, \dots, 2023\}$$

The first dataset is utilized to compute our BERTScore industry classification. We firstly compare the BERTScore industry classification with the FIC by Hoberg & Phillips (2016) and the SIC. Furthermore, we use our BERTScore industry classification in our estimation. For the FIC, Hoberg & Phillips (2016) calculate the cosine similarity between word vectors of product descriptions, which they extracted from Item 1 of the 10-K filings. For our Industry classification we utilize the BERTScore to calculate the similarity between our BM summaries.

[1] We forked and used following Github repository: https://github.com/skranz/gemini_ex.

Based on these similarities we cluster the companies into industries via an agglomerative clustering algorithm. The methodology and object of research differ between the two studies. In accordance with the definition provided by Spieth & Schneider (2016), a company's product constitutes a component of its value proposition and, consequently, a constituent of the BM. Because the product is thereby entangled with the BM, companies that have similar products might have similar BMs. So despite the different methodology and object of research, we expect a similar distribution as Hoberg for the FIC, which is very granular and contains lot of small industries. Thus, we hypothesize:

H1: Our BERTScore industry classification shows a similar distribution compared to the FIC.

H2: Our BERTScore industry classification has a high overlap with the FIC.

As mentioned, our approach differs from the original paper by Hoberg & Phillips (2016). We fix the company and calculate the BERTScore of the summaries of one year and the following year. When a company innovates its BM over time, the 10-K filings change and thus the summaries of these filings. We subtract the BERTScores from one to get the distance instead of the similarity between summaries, because the distance yields a more intuitive interpretation: The higher the distance, the more do the BM summaries differ. Figure 1 shows the density function of the distance. The distance looks normally distributed and has a mean of 0.104 and a standard deviation of 0.027.

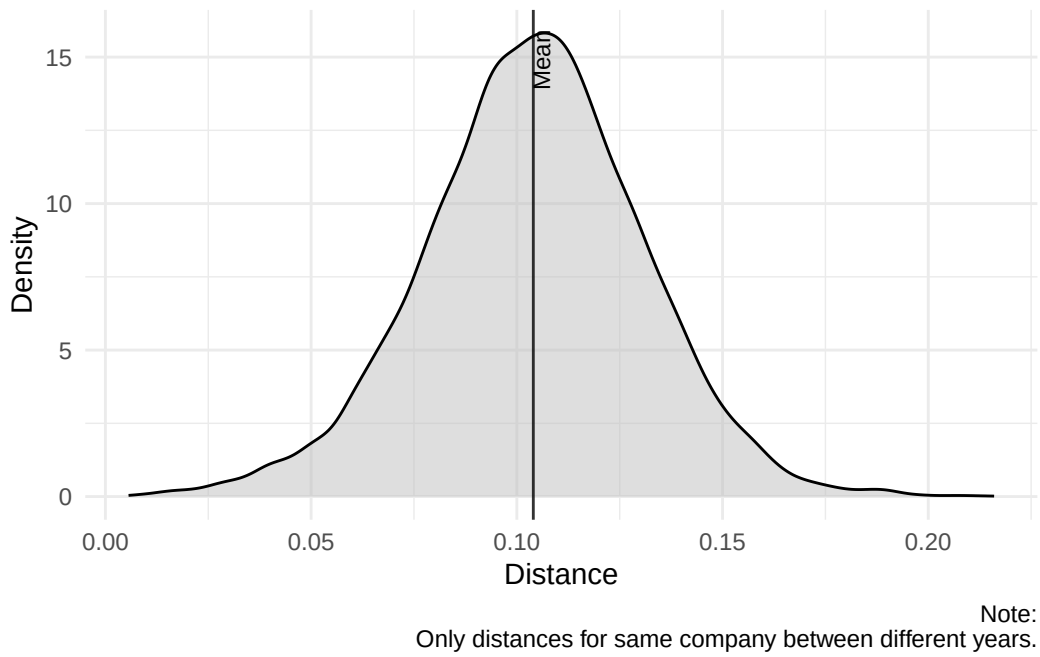

Figure 1: Density Function of the Distance

The values range from zero to 0.216. This means that on average the summaries of a company differ slightly. We attribute these small differences to our preprocessing rather than that companies on average slightly change their BM every year. Even if the summaries are very similar in terms of content, Gemini might use different phrases and wordings which result in different BERTScores and thereby in a higher distance. Hence, we normalize the distance by dividing by its mean. Furthermore, we subtract one and multiply it by one-hundred in order to further ease the interpretation of the coefficient. This results in the following definition of our measure:

$$\text{Normalized Distance} = \left(\frac{Distance_i}{\frac{1}{n}\sum_{i=1}^{n} Distance_i} - 1 \right) \cdot 100$$

where n denotes the number of observations in the dataset. The Normalized Distance measures the difference in the distance between two consecutive summaries of some company above the market average in percentage points. Under the assumption that the summaries reliably represent the firms' BM, the Normalized Distance measures the change of a company's BM above the market average in percentage points from one year to the next. We further assume that a company that changes its BM, will not return back to its original BM. Under this assumption the Normalized Distance measures BMI on the new-to-the-firm level. The literature posits that changes in the business model can be positively correlated with firm performance. In the case that our measure is able to measure BMI, we expect to find a positive relationship between our BMI measure and firm performance. Therefore, we hypothesize that:

H3: Our measure for BMI is positively correlated with firm performance.

When measuring BMI, Guldmann & Huulgaard (2020) assume several barriers to BMI in the circular economy, comparing differences between company sizes and equal company sizes. They found that even same-sized companies face different barriers to BMI, but also identify company size, industry and customer segments as drivers of differences in barriers to BMI. There is a gap in the literature when it comes to investigating the effect of BMI on the performance of different company sizes. We expect higher effects for companies with lower market capitalization (according to Guldmann & Huulgaard (2020), smaller companies are more flexible in changing established processes), while companies with medium or high market capitalization tend to show a less strong effect, which is why we put forward the following hypothesis:

H4: BMI has a stronger positive effect on the performance of small companies compared to medium and large companies.

Estimation Strategy

We test H3 using multivariate regression techniques. The independent variable is the Normalized Distance, as previously defined. Two different dependent variables are employed in order to measure a company's performance. In the initial specification, the sales growth is used as the dependent variable as by Cucculelli & Bettinelli (2015). For this specification, the second dataset is employed, and all observations with zero values for market value and total assets were removed in order to take the logarithms. We control for firm size by once using the logarithm of the market value and once the logarithm of total assets (Dang, Li, and Yang 2018). Moreover, outliners are winsorized only above the 99th percentile. For a more detailed discussion of this step, see Appendix C. The following regressions are used to estimate the normalized distance:

(1) $Y_{\text{Sales Growth}} = \beta_0 + \beta_1 X_{\text{Normalized Distance}} + \beta_2 \ln(X_{\text{Market Value}}) + \gamma_{\text{year}\times\text{industry}} + \epsilon_{it}$

(2) $Y_{\text{Sales Growth}} = \beta_0 + \beta_1 X_{\text{Normalized Distance}} + \beta_2 \ln(X_{\text{Total Assets}}) + \gamma_{\text{year}\times\text{industry}} + \epsilon_{it}$

(3) $Y_{\text{Tobin's Q Growth}} = \beta_0 + \beta_1 X_{\text{Normalized Distance}} + \beta_2 \ln(X_{\text{Market Value}}) + \gamma_{\text{year}\times\text{industry}} + \epsilon_{it}$

(4) $Y_{\text{Tobin's Q Growth}} = \beta_0 + \beta_1 X_{\text{Normalized Distance}} + \beta_2 \ln(X_{\text{Total Assets}}) + \gamma_{\text{year}\times\text{industry}} + \epsilon_{it}$

We use the interaction term of the year and our BERTScore industry classification as a fixed effect. The interaction term of the year and the industry is employed to capture potential heterogeneous effects of the industry classification across different years. The economic and political environment is not controlled for, as White et al. (2022) suggest, as all of the companies in question are based in the United States. In the second specification the growth of Tobin's Q is employed as described by Desyllas et al. (2022). To account for differences between companies, we us growth rates instead of absolute values. The third dataset is used for this specification. Since every company is one observation in the dataset, we control for the firm size once with the logarithm of the mean market value and once with the logarithm of the mean total assets of a company. The mean is calculated with all available years per company. Furthermore, the number of years between the initial and final years utilized for Tobin's Q growth is controlled for. The reason behind this is, that some companies in the dataset initially appear later than 2017 or lastly do not appear in 2023. We use fixed effects to control for this difference. Furthermore, we use industry fixed effects to control for heterogeneous effects between industries. As in the first specification, we do not control for the economic and political environment. We use a similar multivariate regression model to test H4:

(5) $Y_{\text{Sales Growth, k}} = \beta_{0,k} + \beta_{1,k} X_{\text{Normalized Distance}} + \beta_{2,k} \ln(X_{\text{Total Assets}}) + \gamma_{\text{year}\times\text{industry}} + \epsilon_{it,k}$

(6) $Y_{\text{Tobin's Q Growth, k}} = \beta_{0,k} + \beta_{1,k} X_{\text{Normalized Distance}} + \beta_{2,k} \ln(X_{\text{Total Assets}}) + \gamma_{\text{year}\times\text{industry}} + \epsilon_{it,k}$

whereby $k \in \{\text{Small-Cap}, \text{Mid-Cap}, \text{Large-Cap}\}$. In contrast to the previous specification, we only use sales growth as the dependent variable, as we would otherwise have too few observations available for our model due to the sample split. We divide the dataset into three different groups, using a flat split of market capitalization defined by NASDAQ as the thresholds. Accordingly, companies with a market capitalization of up to one billion US dollars are assigned to small-cap firms, companies up to five billion US dollars to mid-cap firms and companies larger than that to large-cap firms. As in the first specification, we use the interaction term of the year and our BERTScore industry classification as a fixed effect and control for firm size via the logarithm of the total assets.

Results and Discussion

Comparison

Our study builds on the idea of Hoberg & Phillips (2016) to utilize text data from 10-K filings to classify companies based on their product similarity into dynamic industries. They achieve this through the parsing of the product descriptions provided by Item 1 of firms 10-K filings and creating word vectors. Specifically, the authors identify and exclude proper nouns, which include common words and geographic locations. They then create word vectors for each firm and year, which enables the measurement of product similarity over time. They perform two steps to create the FIC. Firstly, a hierarchical agglomerative clustering algorithm is employed to cluster companies based on their similarity and maximize ex-post within cluster similarity. This enables a classification with any number of clusters. In the second step, the authors compute aggregated word vectors for each industry. These vectors now represent the industries. Subsequently, the similarity between industries and firms is calculated for each of the following years. From the second year onwards, firms are classified according to the industry with which they are most similar. Our approach differs in two ways: Firstly, in contrast to the TNIC and FIC, which employ word-to-vec, our approach utilizes BERT to represent text, which allows us to capture the context of words. Accordingly, the BERTScore is employed instead of the cosine similarity as our similarity measure. Secondly, our analysis is focused on the description of the BM rather than on the product descriptions. Nevertheless, in the following subsection, the BERTScore industry classification is compared with the FIC and the SIC. We employ the first dataset for the BERTScore industry classification and the comparison. The SIC codes come from the SEC website[2]. For the FIC we have utilized the similarity scores provided by Hoberg-Phillips Data Library[3]. The data consists of the gvkeys of two companies, the year and the cosine similarity between these two companies. In order to ensure comparability, only companies present in both the present study's dataset and that provided by the authors are included in the analysis. Because we use CIKs and accession numbers to identify firms and filings, and the fact that the data library employs Compustat's

[2]The list can be found here: https://www.sec.gov/search-filings/standard-industrial-classification-sic-code-list.
[3]For the database see: https://hobergphillips.tuck.dartmouth.edu.

gvkeys, the matching of CIKs with gvkeys inevitably results in the loss of some observations. Ultimately, for the comparison the clustering algorithm was applied to 1,958 of the 3,246 firms in our sample for the year 2017. In our dataset, companies are from 320 different SIC codes. Therefore, for the comparison the number of industries chosen for our industry classification and the FIC is 320.

Figure 2 compares the distribution of industry size for the BERTScore classification, the FIC and the SIC. Both the BERTScore classification and the FIC show a similar distribution, displaying a leftward skew with the majority of industries comprising fewer than ten firms.

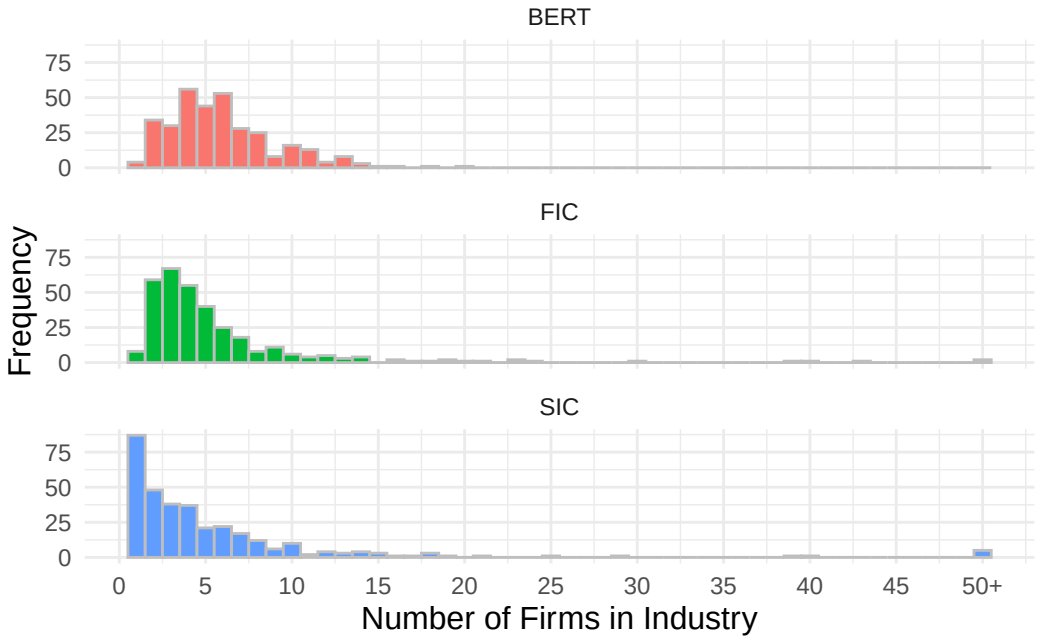

Figure 2: Distribution Comparison between BERT Classification, FIC and SIC

The SIC shows as well a left skewed distribution but with most industries only containing one company. The distribution of the FIC is steeper than the one of the BERTScore classification. It is notable that the largest industry in the BERTScore classification comprises only 20 companies, whereas the FIC and SIC contain industries with a greater number of firms, with some exceeding 50. This suggests that the BERTScore classification groups small to medium-sized industries, comprising between two and fourteen firms per industry, with fewer large industries. The FIC also comprises mostly of small to medium-sized industries, with a few larger ones. Despite these minor differences, this supports H1. The degree of homogeneity between the BERTScore classification and the FIC is 0.63, while the completeness is 0.6. This demonstrates only a medium degree of overlap between the two classifications. The Adjusted Rand Index (ARI) (Hubert and Arabie 1985) is situated at 0.0002, which is close to zero, indicating that the overlap might be random. These findings do not provide support for H2.

In order to use our BERTScore industry classification in our estimation, we classify all 3,246 companies from the year 2017 as described above. Since we use BERT and do not have word vectors for each industry, our methodology differs in the second step from Hoberg & Phillips (2016). We assign the remaining companies of which we do not have data for the year 2017 to the industry of the company which are already classified and which they are most similar to.

Results

Table 3 presents the four regressions related to hypothesis H3. As previously described, this relationship is examined using sales growth and Tobin's Q growth as firm performance indicators. Models 1 and 2 use sales growth as the dependent variable. Model 1 shows a significant positive effect for the Normalized Distance ($\hat{\beta}_1$ = 0.289, p < 0.01) on the sales growth. The results can be interpreted as that a company, that changes its BM in one year one percentage point more than the market average, experiences on average 0.289 percentage points more growth in sales. Model 2 as well shows this significant positive effect ($\hat{\beta}_1$ = 0.292, p < 0.01). Both regressions yield very similar results regardless of whether we include the logarithm of the market value ($\hat{\beta}_2$ = -1.014, p > 0.1) or the logarithm of the total assets ($\hat{\beta}_1$ = -2.155, p < 0.05) as a control variable for the firm size. These results provide support for H3.

Model 3 and 4 assess the effect of the Normalized Distance on Tobin's Q growth. Both Models show a negative significant effect ($\hat{\beta}_1$ = -0.007, p < 0.001 and $\hat{\beta}_1$ = -0.008, p < 0.001). This can be interpreted as indicating that a company that changes its BM by one percentage point more than the market average over the observation period will experience, on average, a reduction in Tobin's Q growth of 0.008, respectively 0.007, percentage points. As in Model 1 and 2, the coefficients for the Normalized Distance only change very little when controlling for the logarithm of the market value instead of the logarithm of the total assets. The results do not support H3, since we anticipated a positive effect. In summary, the findings indicate that BMI has a significant impact on firm performance. Specifically, BM changes have a positive effect on operational metrics such as sales growth, which can be viewed as a positive outcome. A differentiated view of market capitalization shows that the BMI of smaller companies predominantly has a positive impact on sales growth. In contrast, medium-sized companies have a negative effect, while large companies only show a marginal positive effect. However, the stock market does not respond as favorably to such changes. Tobin's Q growth decreases when companies adjust their business models, meaning the

Table 3: Regression Results I

	Sales Growth		Tobin's Q Growth	
	(1)	(2)	(3)	(4)
Normalized Distance	0.322***	0.326***	−0.098	−0.233
	(0.086)	(0.086)	(2.159)	(2.159)
log(Market Value)	−0.476		46.965*	
	(0.882)		(22.211)	
log(Assets)		−1.790*		61.665**
		(0.888)		(22.377)
Num.Obs.	12 014	12 014	12 014	12 014
Industry x Year Fixed Effects	Yes	Yes	Yes	Yes
R2	0.253	0.254	0.127	0.127

Note:
Industry classification by BERTScore.
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

ratio of market value to total capital declines. This effect is not necessarily problematic, as market value is influenced by stock trading, which incorporates subjective evaluations that are reflected in this metric.Table 4 illustrates how the previously demonstrated effects influences individual firm sizes. In the overall analysis, the effects are significant and support our hypothesis. When we differentiate the companies into Small-cap (6,573 observations),

Table 4: Regression Results II

	Sales Growth			Tobin's Q Growth		
	Small-Cap	Mid-Cap	Large-Cap	Small-Cap	Mid-Cap	Large-Cap
Normalized Distance	0.614***	0.010	0.121	−0.402	0.151	−0.057
	(0.150)	(0.203)	(0.097)	(4.568)	(0.144)	(0.076)
log(Assets)	−0.615	−5.552	−1.895	161.384**	−20.372***	−7.139***
	(1.853)	(5.947)	(2.133)	(56.499)	(4.206)	(1.667)
Industry x Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Num.Obs.	6459	2904	2651	6459	2904	2651
R2	0.328	0.419	0.566	0.147	0.390	0.576

Note:
Industry classification by BERTScore.
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Mid-cap (2,923 observations), and Large-cap (2,659 observations), the explanatory coefficients reveal substantial differences. The small firms exhibit by far the most significant effect on the

altered distance, which is also significant at the 1% level. In contrast, for large companies, we observe only approximately 20% of the effect of the business model change on sales figures. Furthermore, the coefficient does not even exceed one standard deviation, indicating a lack of significance. The situation is even less favorable for mid-sized companies, as evidenced by both the coefficient and standard deviation. The value of the distance effect accounts for merely 0,008% of that observed in large companies. Therefore, we can conclude that the previously observed effect across all firms can primarily be attributed to small firms. For medium to large firms, no discernible impact is evident.

Robustness Checks

To check the validity and reliability of this study, we conduct a series of robustness checks, illustrating alternative models. The revenue growth and the normalized distance are not logarithmized in this specification. Revenue growth can assume negative values and is a relative measure. The normalized distance is also a relative measure, and the distance itself is characterized by a norm distribution (according to Figure 1). Logarithmization is therefore not required. The use of market value and total assets as control variables makes it possible to control for the size of the company, which results on the one hand from supply and demand on the capital market, and on the other hand from the actual book value of the company. The correlation between these two variables is approximately 0.5. In Table 4, we perform a regression using SIC codes instead of the industry classification based on BERTScores as fixed effects.

Table 5: Robustness Check

	Sales Growth		Tobin's Q Growth	
	(1)	(2)	(3)	(4)
Normalized Distance	0.180	0.185	−0.009***	−0.009***
	(0.129)	(0.129)	(0.001)	(0.001)
log(Market Value)	−4.112***			
	(1.010)			
log(Assets)		−5.157***		
		(0.965)		
log(Mean Assets)			0.029**	
			(0.010)	
log(Mean Market Value)				0.049***
				(0.011)
Time Difference			0.098***	0.096***
			(0.018)	(0.018)
Industry x Year Fixed Effects	Yes	Yes	No	No
Industry Fixed Effects	No	No	Yes	Yes
Num.Obs.	11 187	11 187	2977	2977
R2	0.139	0.140	0.131	0.134
RMSE	219.86	219.71	1.33	1.33

Note:
Industry classification by SIC codes.
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

The number of observations decreases due to missing SIC codes for non-surviving companies. As a result Table 5 rather tend to contain a survivor ship bias (the number of observations decreases due to missing SIC codes for non-surviving companies). In addition, the coefficients for sales growth are no longer significant, the standard errors of the coefficients increase despite lower coefficient values. The value for R^2 also decreases for each model. This suggests that BERT-based industry classifications as fixed effects likely capture industry-specific endogenous effects, thereby enhancing both the model's validity and its informative value.

With regard to Table 4, other threshold values could also have been selected for a sample split. Each institution (index providers, stock exchanges, financial analysts, etc.) defines its own thresholds, although these do not usually vary greatly. In our study, we use a flat threshold for small-cap and large-cap firms for simplicity, although definitions for very small or very large firms would also require more classifications for a sample split. Due to the limited number of observations, we consider it sensible to dispense with further classifications.

Discussion

Interpretation

The findings of this study contribute to the expansion of the considerations put forth by Hoberg & Phillips (2016) Using BERT enables a semantic analysis of 10-K filings, allowing for a comprehensive company comparison. The categorization into industries can be updated annually, similar to Hoberg & Phillips (2016) approach, albeit with a somewhat greater effort involved. However, this additional effort is justifiable, as the preliminary work with Gemini facilitates a holistic examination of the 10-Ks. This aspect is a significant advantage of our approach, which we believe surpasses the Word-to-Vec method employed by Hoberg. We fully utilize the description of the BM, although certain limitations may exist. However, the loss associated with the previous method is substantially greater, as most words do not contribute to the analysis at all.

To broaden the scope of this study, we also address the financial aspects of the companies and how innovations impact these aspects. The previously applied model will be used to measure changes in the BM. This allows for the testing of new hypotheses and the generation of additional insights. In this study, the relationship with performance is measured using revenue growth and Tobin's Q. We find a significant dependence of innovation on sales and Tobin's Q, indicating that while revenue increases, the market value to asset ratio declines with increased changes in the BM. The result of revenue growth is not surprising; we anticipated this effect based on the premise that revenues would rise as companies aim to capture new markets and thereby generate more sales through changes in their business plans. Conversely, the decline in Tobin's Q is somewhat unusual. A possible explanation for this could be that investors trading stocks are often risk-averse, perceiving the changes negatively rather than viewing them as opportunities (time-varying risk aversion). Consequently, demand may decrease, and the stock price may also experience negative shifts. Furthermore, the profits generated by the company may temporarily decline due to these decisions. While the company's revenue is projected to increase on average, innovations may incur additional expenses that diminish profit and, consequently, dividends. (Business model innovation: it's not just about technology anymore) This could further reduce demand for the stock. Additionally, the new BM may create tensions with the existing one, leading to increased workload and other implications.

In our further analysis, we differentiate the companies by size to examine whether the effects of revenue growth vary accordingly. This assumption is confirmed, as significance is measured exclusively among smaller firms with a market capitalization of less than one billion. This implies that changes in the BM do not affect the performance of revenue growth in medium to large firms. A potential reason for this could be that smaller companies typically possess a less extensive product portfolio, making changes to a business segment or the addition of a new product more impactful.

In conclusion, we find that adjustments to the BM positively influence revenue growth for small firms. However, it is crucial not to misinterpret the results for larger firms by deeming

such adjustments unprofitable or nonsensical. This study focuses primarily on short-term effects, making it more challenging to observe the impact on the products of larger companies. Long-term profitability resulting from innovation is acknowledged but falls outside the scope of this analysis.

Limitation

This study represents an innovative extension of Hoberg & Phillips (2016) work regarding their approach. While using the BERTScore compels us to shorten and condense the descriptions of the BM, it integrates semantics into the similarity calculation. As a result, the entire text is considered, which appears to provide better informational value compared to solely relying on frequently occurring nouns. However, one potential critique of BERT is its limited ability to recognize implicit and subtle meanings (Acheampong, Nunoo-Mensah, and Chen 2021). The approach can only process the textual descriptions. In this work, we exclusively examine business models that provide little insight into current success in practice. Notably, there are no clear objective statements, as with many key financial indicators, that can be categorized as positive or negative (George and Bock 2011). A positive description of negative aspects is unlikely in this context, which justifies the use of the BERTScore.

The use of BERT, and the consequent limitation regarding text length, posed additional challenges in this research. In recent years, the possibilities for utilizing AI have grown immensely. With the help of GitHub Actions, we were able to obtain suitable access to Gemini. This brings us to the most critical point of this work. This approach raises transparency issues for the results. The system does not provide explainable intelligence, meaning we cannot fully verify how exactly the texts are generated. Our only option is to delegate the task and trust the results' accuracy.

Nevertheless, we consider the current approach as the best possible solution for condensing our BMs. On the one hand, the extracted texts do not follow the same structure across companies, and on the other, cutting the business models arbitrarily poses too great a risk. Losing up to 90 percent of the content would be unacceptable in this case, and due to structural differences, this becomes impossible beyond certain key points. Moreover, the summaries generated by GPT have sometimes been perceived as better than models specifically designed for this purpose (Goyal, Li, and Durrett 2023). A significant portion of this work involved data collection, which proved to be a major challenge. Due to the lack of access to Compustat, commonly used in financial studies and for SEC-related information, we had to rely on extensive research. Various approaches had to be abandoned as they failed to meet expectations. Nevertheless, we were able to complete our work, albeit with some compromises regarding the amount of data.

In conclusion, the findings of our study suggest that BMI is associated with improved performance, specifically in terms of operating revenue. However, we can only report a correlation rather than a causal effect, as the complexity of organizations is substantial, and our dataset

does not allow for definitive conclusions regarding causality. Nonetheless, this study can be viewed as a continuation of the approach by Hoberg & Phillips (2016), and the examination of BMI's impact on performance could be further explored in future studies, offering valuable extensions to this research.

Conclusion

The present study builds on the work of Hoberg & Phillips (2016) and extends their approach to improving traditional industry classifications by analyzing product descriptions in the context of BM. This research leverages modern methodologies to incorporate entire 10-K filings into the analysis. Unlike Hoberg & Phillips (2016), who focus on frequent nouns in a Word-to-Vec model, we employ BERT, which allows us not only to analyze the entirety of the 10-Ks but also to consider their semantic meaning. However, due to BERT's token length restrictions, we also use the AI tool Gemini to adjust the length of the 10-Ks. To remain consistent with the methodology of the referenced study, we group the BERTScores into industries through hierarchical clustering. We demonstrate that our classification exhibits a distribution similar to the FIC by Hoberg & Phillips (2016). Contrary to our expectations, there is no significant general overlap between the two classification methods.

Beyond extending the industry classification using 10-K filings, we also see further potential in the BERTScores. Therefore, we conducted additional analyses to examine the relationship between BMI and financial performance. Using regression analysis, we find evidence of a correlation: greater changes in the BM are associated with higher revenue growth and a lower Tobin's Q. These results, however, are based on the overall observations, without distinguishing between companies of different sizes. Upon further analysis using multiple regressions segmented by company size, we observe that the effect on revenue growth is statistically significant only for smaller firms.

In conclusion, we find that smaller firms benefit more from BMI, or in the short term, are the only ones that show a positive impact. Additionally, the insights and methodology presented in this study can be applied to discover new effects of BMI. For instance, an increase in R&D spending does not always result in innovation, a distinction that can be better determined using our new approach.

Acknowledgement

- Jonathan for IT Support
- Prof. Kranz for Repo
- bw unicluster

References

Acheampong, Francisca Adoma, Henry Nunoo-Mensah, and Wenyu Chen. 2021. "Transformer Models for Text-Based Emotion Detection: A Review of BERT-Based Approaches." *Artificial Intelligence Review* 54 (8): 5789–829. https://doi.org/10.1007/s10462-021-09958-2.

Clauss, Thomas. 2017. "Measuring Business Model Innovation: Conceptualization, Scale Development, and Proof of Performance." *R&D Management* 47 (3): 385–403. https://doi.org/10.1111/radm.12186.

Cucculelli, Marco, and Cristina Bettinelli. 2015. "Business Models, Intangibles and Firm Performance: Evidence on Corporate Entrepreneurship from Italian Manufacturing SMEs." *Small Business Economics* 45 (2): 329–50. https://doi.org/10.1007/s11187-015-9631-7.

Dang, Chongyu, Zhichuan Frank Li, and Chen Yang. 2018. "Measuring Firm Size in Empirical Corporate Finance." *Journal of Banking & Finance* 86: 159–76.

Desyllas, Panos, Ammon Salter, and Oliver Alexy. 2022. "The Breadth of Business Model Reconfiguration and Firm Performance." *Strategic Organization* 20 (2): 231–69. https://doi.org/10.1177/1476127020955138.

Foss, Nicolai J., and Tina Saebi. 2017. "Fifteen Years of Research on Business Model Innovation: How Far Have We Come, and Where Should We Go?" *Journal of Management* 43 (1): 200–227. https://doi.org/10.1177/0149206316675927.

George, Gerard, and Adam J. Bock. 2011. "The Business Model in Practice and Its Implications for Entrepreneurship Research." *Entrepreneurship Theory and Practice* 35 (1): 83–111. https://doi.org/10.1111/j.1540-6520.2010.00424.x.

Goyal, Tanya, Junyi Jessy Li, and Greg Durrett. 2023. "News Summarization and Evaluation in the Era of GPT-3." arXiv. http://arxiv.org/abs/2209.12356.

Guldmann, Eva, and Rikke Dorothea Huulgaard. 2020. "Barriers to Circular Business Model Innovation: A Multiple-Case Study." *Journal of Cleaner Production* 243 (January): 118160. https://doi.org/10.1016/j.jclepro.2019.118160.

Hoberg, Gerard, and Gordon Phillips. 2016. "Text-Based Network Industries and Endogenous Product Differentiation." *Journal of Political Economy* 124 (5): 1423–65. https://doi.org/10.1086/688176.

Huang, WenJun, and Takeyasu Ichikohji. 2023. "A Review and Analysis of the Business Model Innovation Literature." *Heliyon* 9 (7): e17895. https://doi.org/10.1016/j.heliyon.2023.e17895.

Hubert, Lawrence, and Phipps Arabie. 1985. "Comparing Partitions." *Journal of Classification* 2 (1): 193–218. https://doi.org/10.1007/BF01908075.

Latifi, Mohammad-Ali, Shahrokh Nikou, and Harry Bouwman. 2021. "Business Model Innovation and Firm Performance: Exploring Causal Mechanisms in SMEs." *Technovation* 107 (September): 102274. https://doi.org/10.1016/j.technovation.2021.102274.

Lee, Jihwan, and Yoo S. Hong. 2014. "Business Model Mining: Analyzing a Firm's Business Model with Text Mining of Annual Report." *Industrial Engineering and Management Systems* 13 (4): 432–41. https://doi.org/10.7232/iems.2014.13.4.432.

Pucihar, Andreja, Gregor Lenart, Mirjana Kljajić Borštnar, Doroteja Vidmar, and Marjeta Marolt. 2019. "Drivers and Outcomes of Business Model Innovation—Micro, Small and

Medium-Sized Enterprises Perspective." *Sustainability* 11 (2): 344. https://doi.org/10.3390/su11020344.

SEC. 2024. "Investor Bulletin: How to Read a 10-K." https://www.sec.gov/files/reada10k.pdf.

Spieth, Patrick, and Sabrina Schneider. 2016. "Business Model Innovativeness: Designing a Formative Measure for Business Model Innovation." *Journal of Business Economics* 86 (6): 671–96. https://doi.org/10.1007/s11573-015-0794-0.

Teece, David J. 2018. "Business Models and Dynamic Capabilities." *Long Range Planning* 51 (1): 40–49. https://doi.org/10.1016/j.lrp.2017.06.007.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need.(nips), 2017." *arXiv Preprint arXiv:1706.03762* 10: S0140525X16001837.

White, Joshua V., Erik Markin, David Marshall, and Vishal K. Gupta. 2022. "Exploring the Boundaries of Business Model Innovation and Firm Performance: A Meta-Analysis." *Long Range Planning* 55 (5): 102242. https://doi.org/10.1016/j.lrp.2022.102242.

Zott, Christoph, and Raphael Amit. 2008. "The Fit Between Product Market Strategy and Business Model: Implications for Firm Performance." *Strategic Management Journal* 29 (1): 1–26. https://doi.org/10.1002/smj.642.

Appendix

Appendix A

We used following prompt: "Summarize the business model from the following text. Answer with a continuous text and with five hundred twelve tokens at max. Set your focus on sources of revenue, the intended customer base, products, distribution channels and details of financing. Use only information from the following the text".[4] "Intended customer base" and "product" refer to the value offering, "distribution channels" refers to the value architecture, and "sources of revenue" and "details of financing" refer to the revenue model. The term 'tokens' was used deliberately in preference to 'words', given that the number of tokens and the number of words in a text may vary depending on the tokeniser. This way, we wanted to ensure that the whole summary is used by the BERT model. Table 6 presents the descriptive measures of the length of our summaries. On average our

Table 6: Descriptive Statistics of Number of Words of our Summaries

Report-for Year	Average Word Count	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
2016	355	156	116	273	331	409	5377
2017	356	154	117	274	331	407	5951
2018	350	109	31	275	330	404	1057
2019	358	186	56	276	333	407	6273
2020	367	214	56	276	337	422	6872
2021	370	159	76	281	344	432	5559
2022	371	161	138	283	345	425	6361
2023	366	207	150	275	333	419	5154

Note:
All 21,417 summaries were considered.

summaries are 350 till 371 words long with the 75th percentile ranging between 404 and 432 words. Looking at the maxima, some summaries are significantly longer than the average with a couple thousand words. Most of the summaries are short enough to be processed by BERT as a whole, while some summaries are too long for BERT. These outliners are only a few and therefore negligible.

Appendix B

To check the validity of the summaries generated by Gemini, we took a random sample of ten filings to evaluate them manually. The results are shown in Table 7.

[4]The spelling error in the last sentence of the prompt was found after processing Item 1. After evaluating the summaries, this error did not cause any issues.

Table 7: Subjective Evaluation of Random Item 1 Summaries

Accession ID	Company Name	Year
0000897101-17-000402	COMMUNICATIONS SYSTEMS INC	2016
0001615774-19-005785	NanoVibronix, Inc.	2018
0001326380-18-000033	GameStop Corp.	2018
0001683168-21-001026	FOCUS UNIVERSAL INC.	2020
0000768835-18-000030	BIG LOTS INC	2018
0001309108-23-000022	WEX Inc.	2022
0000900075-23-000034	COPART INC	2023
0001731122-20-000381	Data Storage Corp	2019
0000078749-17-000021	AGILYSYS INC	2017
0001509991-21-000031	Kosmos Energy Ltd.	2020

Note:
A random sample of ten filings was drawn

We find that the summaries are valid overall and consistently reflect the BM. Suitable representations of all Item 1 texts can be found in all corresponding summaries available in the sample. We therefore assume that the summaries are a valid representation of the original company descriptions. One conspicuous feature was that although strong diversification tends to be mentioned in the summaries, no possible interaction between these areas is discussed.

Appendix C

In our second dataset we winsorize only the top quantile. Figure 3 presents a boxplot of the distribution of the sales growth and also explains our reasoning behind this decision. The boxplot shows that the big majority of observations has a growth rate below 1,000 percent. Sales growth rates of multiple ten thousand percent or even of a couple million percent seem very unrealistic and are probably due to poor data quality. By winsorizing at the top one percent quantile, we get ride of these unrealistically high values. For the sake of completeness, Table 8 reports the regression results without winsorizing the data. All coefficients are not significant anymore while their values and their standard errors increase massively.

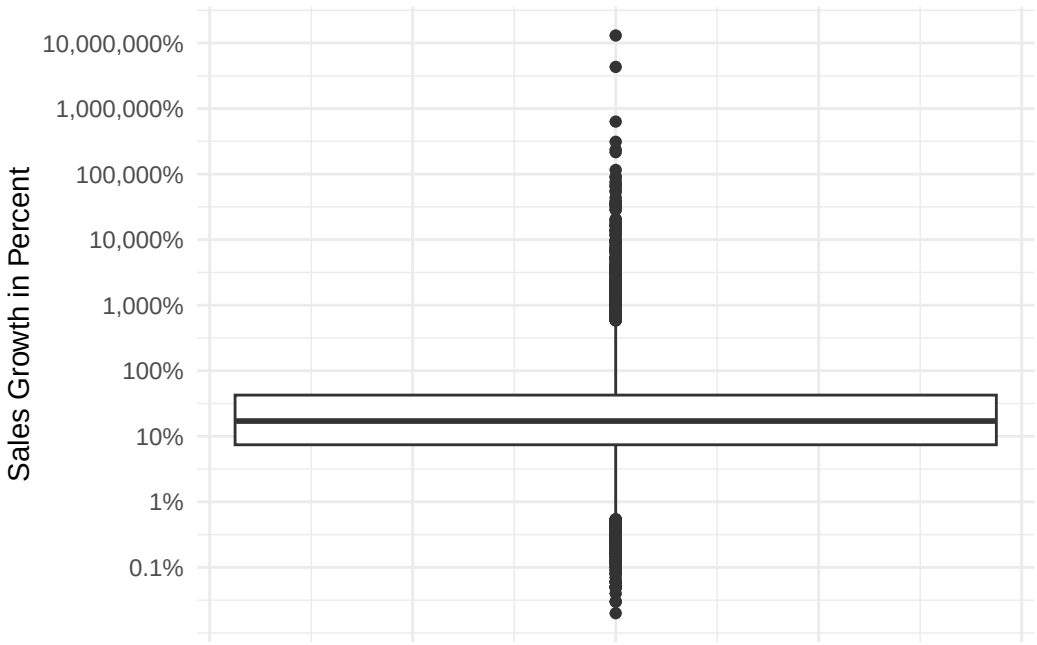

Figure 3: Distribution of the Sales Growth

Table 8: Regression Results (Not Winsorized)

	Sales Growth	
	(1)	(2)
Normalized Distance	41.999	41.800
	(50.080)	(50.087)
log(Market Value)	581.110	
	(420.543)	
log(Assets)		330.692
		(423.289)
Num.Obs.	12 155	12 155
Industry x Year Fixed Effects	Yes	Yes
R2	0.458	0.458
RMSE	91 778.53	91 784.54

Note:
This is the same estimation as in Table 3, but without winsorizing the data.
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001