FROM TEXT TO INSIGHT - A NOVEL APPROACH TO MEASURING BUSINESS MODEL INNOVATION

A PREPRINT

Max Gabler
max.gabler@uni-ulm.de

Wanshu Jiang
wanshu.jiang@uni-ulm.de

Christoph Kiesl
christoph.kiesl@uni-ulm.de

Leonard Pöhls
leonard.poehls@uni-ulm.de

Alexander Rieber
alexander.rieber@uni-ulm.de

Ansgar Scherp
ansgar.scherp@uni-ulm.de

September 13, 2024

ABSTRACT

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 utilise business descriptions from 10-K filings between 2017 and 2023 to measure business model innovation. The methodology employed is as follows: Firstly, we employ Google's Gemini to summarize the core elements of the business model from the 10-K reports. Subsequently, we apply a BERT model (Bidirectional Encoder Representation from Transformers) and calculate the similarity between the summaries of a company over the observation period with the BERTScore. This approach enables us to identify changes in a company's business model over time. To evaluate the effectiveness of our measure, we regress firm performance on the distance between two summaries in consecutive years. Based on the existing literature, we hypothesise that there will be a positive relationship. We find that (. . .). The findings of this study offer insights into the extent to which textual similarities in regulatory reports can be employed as a reliable indicator for business model innovation. Thus, this method represent a novel approach to analyzing business model innovation over time.

Keywords 10-K • Business Model Innovation • BERT • Gemini

1 Introduction

Business model innovation (BMI) is a key activity to maintain competitiveness and even gain a competitive advantage (Pucihar et al. 2019; Teece 2018). It is therefore no surprise that the interest in BMI and methods of measuring it has grown rapidly over the last twenty years. Researchers have recently called for a BMI measurement instrument that is more comprehensive and advanced than already existing ones (Huang and Ichikohji 2023). The scale developed by Spieth & Schneider (2016) provides managers and practitioners with a measurement index for business model innovativeness. This measurement model only validates applicability of BMI theory (Huang and Ichikohji 2023) and is insufficient for longitudinal studies (Clauss 2017). Hence, this measure is not adequate for a time series analysis of BMI. Furthermore, it refers only to BMI as new-to-the-firm and is not able to grasp BMI in the sense of new-to-the-industry and new-to-market. Clauss (2017) developed a similar measure with comparable limitations.

This gap is addressed 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) use this filings to create word vectors about the companies products in order to cluster them into industries and thereby proposing a new industry classification. This study builds on their work and methods. We summarize these descriptions with Gemini, utilize the BERTScore as a similarity measure to calculate the similarities between companies and cluster them into industries. We compare our BERTScore clustering with the fixed industry classification (FIC) by Hoberg & Phillips (2016). Furthermore, we 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. There is evidence that an increase in BMI is associated with improved firm performance (Cucculelli and Bettinelli 2015; Latifi, Nikou, and Bouwman 2021; White et al. 2022). In order to test the validity of our measure, we regress financial numbers on our measure, hoping to find a positive relationship.

- Key findings (and Contribution, already down below)

Our contribution is made in two ways. Firstly, we build on the concept of alternative industry classification put forth by Hoberg & Phillips (2016) and propose an industry classification system based on a firm's BM. Secondly, we propose a novel measure for BMI that is sufficient for longitudinal studies.

The SEC mandates that the majority of public companies based in the United States submit specific documents in certain intervals. One such document is the annual 10-K filing. These filings follow a set order of topics and contain a range of information, including details about managerial discussions, risk factors for the company, legal proceedings and financial data. In the first section under the subtitle "Business," 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).

- paragraph 5 (robustness checks)

In spite of 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 a company's BM is comprised of: 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. A total of twelve indicators for measuring the innovativeness of the BM were identified through a comprehensive literature review and through engagement with industry practitioners. The external validity of the formative indicators was successfully validated through a survey of 200 experts in strategy and innovation management (Spieth and Schneider 2016). Clauss (2017) employs a very similar approach. After specifying the domain and dimensionality of BMI trough literature research, the author divides his scale into three hierarchical levels consisting of 41 reflective items, 10 subconstructs and three main dimensions, which are similar to the ones mentioned earlier. The scale was validated through two samples from the manufacturing industry and further demonstrated nomological validity (Clauss 2017). 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).

The process of text mining 10-K filings is not a novel concept. Hoberg & Phillips (2016) present a novel approach to defining industry boundaries. This is achieved through the parsing of the product descriptions provided by firm 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 propose two novel industry classification methods: the FIC and the text-based network industry classification (TNIC). Firstly, they cluster companies based on the similarity of the word vectors into fixed industries. Secondly, they define a minimum similarity threshold, above which firms are considered in the same industry. This relaxes their prior properties of binary membership transitivity and fixed industry location. This way 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. In their study, Lee & Hong (2014) examine the evolution of a firm's BM over time. The authors represent each document as a vector of keywords, which is similar to the approach utilized by Hoberg & Phillips (2016). After identifying the Item 1 part of the 10-K filings as the most crucial part for describing a firm's BM, Lee & Hong (2014) filter these for relevant sentences. Subsequently, the authors 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. Nevertheless, this approach is not without shortcomings. 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).

The rest of the paper proceeds as follows. Section 2 describes the preprocessing with Gemini, our data and our methodology. Section 3 compares the industry classification based on product descriptions by Hoberg & Phillips (2016) with our classification based on the BM. Section 4 lays out the BERT-model and our estimations. Section 5 discusses our results, and Section 6 concludes our study.

2 Data and Methodology

2.1 Preprocessing with Gemini

10-K filings are typically very large text documents, and Item 1 of these filings is no exception. Table 1 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 couple hundred words to tens of thousands. In order to utilise the entirety 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]. 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".[2] "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. Consequently, this prompt covers all aspects of the definition of BMI proposed by Spieth & Schneider (2016). 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. 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.

- result of this check

Table 1: Descriptive Statistics Original Filings

Year	Mean	Standard Deviation	Minimum	25th Percentile	Median	75th Percentile	Maximum
2016	7842	6104	155	3705	6026	10271	51227
2017	7542	6320	155	3522	5767	9700	70611
2018	7604	6272	180	3528	5771	9669	71700
2019	8009	6631	189	3669	5971	10410	78270
2020	8660	7195	171	3943	6449	10971	57980
2021	10324	8406	235	4670	7568	13563	78799
2022	9471	7997	171	4309	7042	11897	73937
2023	6646	4771	190	3660	5814	8401	43523

2.2 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 on the one hand has been made available to the SEC since 2001 in a publicly accessible list of 10,284 companies (Appendix), of which 7,590 are currently listed on NASDAQ, NYSE or over-the-counter. We extracted 10-K filings that were submitted between 2017 and 2023 based on underlying Central Index Keys (CIK). We exclude companies from the financial sector, namely companies with a SIC Code starting with six. Corresponding to Table 2, multiple steps of pre-processing were required to obtain the final amount of 21,683 observations for seven years. Financial key figures, including net income, total assets and others were originally extracted from the SEC, but also challenged with financial values from DataStream. A total of 4,225 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.

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

[2] 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 2: 10-K Sample Creation

Source/Filter	Sample Size	Observations Removed
1. Extracted 10-K filings from the SEC	35731	
2. Excluding filings from financial companies (SIC-code starting with ’6’)	27569	8162
3. Verify for Item 1 text availability (removed oberservations that are attributable to API quality)	23982	3587
4. Extracting dates for which the filings are reporting for and removing of duplicated filings	23971	11
5. Delete observations with incorrect date assignment	22161	1810
6. Merged Gemini processed Item 1 text to the underlying data set. We did not consider texts that were not processable	21697	464
7. Extract financials statements from SEC and merge them. Also remove observations for years prior to 2016	21686	11

Note:
Filings submitted between 2017 and 2023 are considered

TODO

- Descriptive Table3 for length of summary
- Description of Table2, Table3, and the final dataset

2.3 BERT and BERTScore

BERT is a pre-trained and transformer-based model for natural language processing (NPL) based on artificial neural networks. It works according to the so-called transformer architecture, which was first mentioned by Vaswani et al. (2017). According to these authors, this architecture consists of two main components, the encoder and the decoder. The encoder consists of several identical layers, which initially use the so-called self-attention mechanism to generate context-dependent representations of each word in the sentence. This mechanism can be parallelized and therefore enables different aspects of the context to be captured in the same way. The decoder, on the other hand, works in a similar way and is responsible for processing the information from the encoder and forming it into an output sequence. However, this is not relevant for BERT, as no sequence-to-sequence transformation is carried out in BERT. In contrast to Hoberg & Philips (2016) word-to-vec approach, BERT works bidirectional and takes into account the context from both sides of each word simultaneously. Therefore, BERT is able to capture deeper semantics in texts such as 10-K reports. The BERTScore now computes the cosine similarity between word or text meanings, that have been determined by representations (or embeddings) learned from BERT. The scale is from -1 to 1, where 1 describes a perfect similarity.

2.4 Methodology

After processing the data, we calculate the BERTScore between the summaries of different companies in one year and between the summaries for the same company over different years. The similarity between different companies in the same year is utilized in Section 3 to compare the FIC by Hoberg & Phillips (2016) with the industry classification by the BERTScore. Hoberg & Phillips (2016) calculate the cosine similarity between word vectors of product descriptions while we utilise the BERTScore to calculate the similarity between our BM summaries. The methodology and object of research differ between the two studys. But the product of a company is by the definition of Spieth & Schneider (2016) a part of its value offering and thereby a part 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 industry classification shows a similar distribution compared to the FIC.

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

As mentioned, our approach differs not marginally from the original paper by Hoberg & Phillips (2016). We fix the company and calculate the BERTScore between the summaries of different years. These summaries describe the BM of a company based on the 10-K filing of that company for that year. When a company innovates its BM over time, the 10-K filings change and thus the summaries of these filings. By taking the similarity between this summaries, we are able to calculate the distance between the summaries. This allows us to measure the change in a company's BM. Under the assumption that companies "don't run in circles", this measures BMI on the new-to-the-firm level. The assumption "don't run in circles" means, that a company that not marginally changes its BM, won't change it back to the same BM it had before. In the BMI literature, the positive relationship between BMI and firm performance is well explored by a wealth of studies examining this relationship (White et al. 2022). In the case that our measure indeed measures BMI, we expect to find a positive relationship between our BMI measure and firm performance. Therefore, we hypothesize that:

H3: Our measure for BMI shows a positive relationship with firm performance.

3 Estimation Strategy

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 provide support for the hypothesis that BMI has a positive effect on firm performance, with the effect increasing in line with the intensity of the innovation. Latifi et al. (2021) also find evidence that provides support for the hypothesis that a company engaging in BMI will improve its overall firm performance. In their study, the firm performance is evaluated subjectively. Zott & Amit (2008) analyse the effect of the BM and the product market strategy on firm performance. They measure the firm performance with the market value of equity as the stock price multiplied with the number of shares outstanding. 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.

Following paragraph is still preliminary: - Based on the literature (mostly on (White et al. 2022) and (Zott and Amit 2008)) we build our estimation strategy; use multiple multivariate regressions; - dependent variables: sales growth, market equity (growth), Tobins Q - independent variable: distance between summaries (makes the measure more intuitive; if distance grows/ similarity decreases, summaries are more dissimilar -> BMI) - controls: firm age and size, might need to consider how to measure size (employees, firm value), Tobins Q might also be used - fixed effects for year and industry might me useful. company FE?

4 Results and Discussion

4.1 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 similarity to each other into dynamic industries. Our approach differs in two ways: Firstly, in contrast to the TNIC and FIC, which employ word-to-vec, our approach utilises BERT to represent text. 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. In order to set a benchmark for our BERTScore industry classification, we have replicated the FIC approach as described by Hoberg & Phillips (2016). Subsequently, a comparison is made between the industry classification produced in this study and the aforementioned benchmark.

The data employed for the industry classification with the BERTScore is the same as described in Section 2. In order to establish a benchmark, 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 gvkeys, the matching of CIKs with gvkeys inevitably results in the loss of some observations. Ultimately, the clustering algorithm was applied to 1,958 firms for the year 2017. Hoberg & Phillips (2016) 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. They perform this step in increments of 50, but focus on 300 industries since this is very analogous to the number of industries of the SIC and the NAICS. Therefore, a comparison is made between our industry classification and the FIC based on 300 industries. 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. The FIC satisfies the binary membership transitivity property and the fixed location property. The binary membership transitivity property states that if two firms are in the same industry

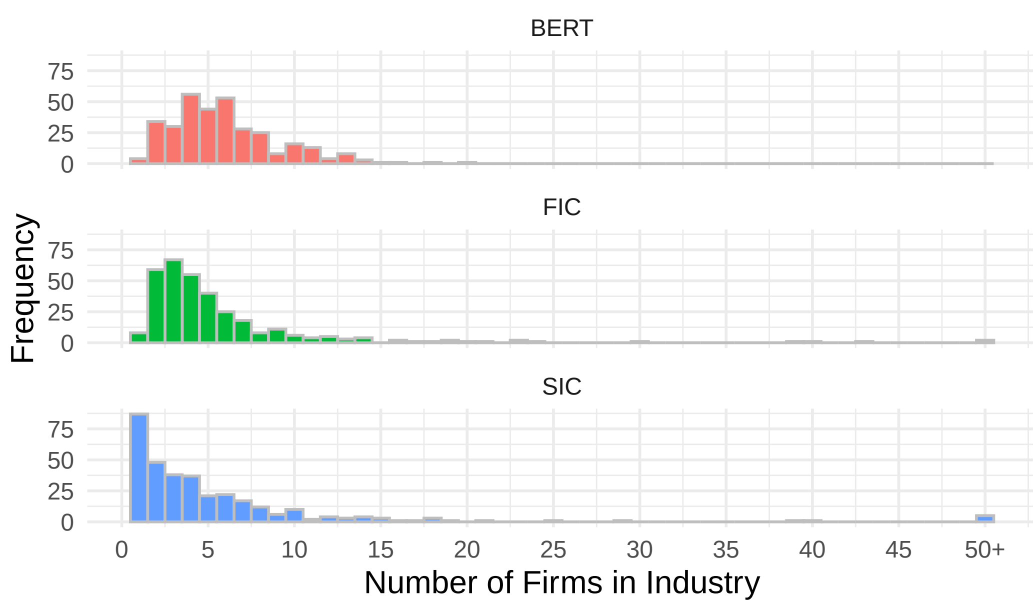
[3]For the database see: https://hobergphillips.tuck.dartmouth.edu.

and a third firm is in the same industry as one of the other two, then it is also in the same industry as the other. The fixed location property dictates that clusters are not recalculated on an annual basis, but rather remain fixed until the codes are updated (Hoberg and Phillips 2016).

Figure 1 presents a comparison of the distribution of industry size for the BERTScore classification and the FIC. 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. The distribution of the FIC is steeper than the on of the BERTScore classification. It is notable that the largest industry in the BERTScore classification comprises only 27 companies, whereas the FIC contains 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 is random. These findings do not provide support for H2.

The industries in the FIC are fixed, but by dropping the fixed location property, the industry classification are time-varying. Furthermore, if the binary membership transitivity property is relaxed, the result is the TNIC by Hoberg & Phillips (2016). The removal of the fixed location property encourages the notion of not fixing a specific year and calculating the similarities between different companies, but rather fixing a company and calculating the similarity between this firm and others over time.Such considerations result in a shift in focus away from industry classification towards an analysis of a company's evolution. As previously stated, 10-K filings are rich on information and contain a wealth of data beyond mere product descriptions. The additional information can be leveraged to investigate a company's BM. The combination of both rationales – fixing a company over time and analysing the BM – constitutes the primary concept of this study.

Figure 1: Comparison between BERT Classification, FIC and SIC



5 Conclusion

6 Acknowledgement

- Jonathan for IT Support
- Prof. Kranz for Repo

7 References

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.

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.

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 Kljaji Bortnar, 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.

8 Appendix

- distribution of BERT Scores