
THE NAME OF THE TITLE IS HOPE

A PREPRINT

Max Gabler
max.gabler@uni-ulm.de

Wanshu Jiang
wanshu.jiang@uni-ulm.de

Christoph Kiesel
christoph.kiesel@uni-ulm.de

Leonard Pöhls
leonard.pohls@uni-ulm.de

Alexander Rieber
alexander.rieber@uni-ulm.de

Ansgar Scherp
ansgar.scherp@uni-ulm.de

August 16, 2024

ABSTRACT

A clear and well-documented \LaTeX document is presented as an article formatted for publication by ACM in a conference proceedings or journal publication. Based on the “acmart” document class, this article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work.

Table of contents

1	Introduction	2
2	Related Work	2
2.1	Measuring Business Model Innovation	2
2.2	BMI and Firm Performance	3
2.3	Text Mining 10-K Filings	3
2.4	Summary	3
3	Methods	3
3.1	Assumptions	3
3.2	Gemini	3
3.3	BERT and BERTScore	4
3.4	Hierarchical Agglomerative Clustering	4
4	Experimental Apparatus	4

4.1	Datasets	4
4.2	Pre-Processing	4
4.3	Procedure	4
5	Results	4
6	Discussion	4
7	Conclusion	4
8	Limitations	4
9	Acknowledgement	4

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.

- Solution approach (both after we actually are final with approach and have (some) results)
- Results

Our contribution is made in a number of ways. Firstly, we tackle two issues raised in the study by Lee & Hong (2014): we employ a more reliable and contemporary methodology for extracting the business model (BM) from 10-K filings and we are able to extend the scope of their study. Secondly, 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. Thirdly, we propose a novel measure for BMI that is sufficient for longitudinal studies.

2 Related Work

Our study borders several research fields and thus we have a variety of literature to cover. Section 2.1 presents state of the art measures of BMI while Section 2.2 examines the relationship between BMI and the financial performance of a firm. Section 2.3 deals with previous approaches to text mine 10-K filings. Finally, in Section 2.4 a summary of the findings from the related work is presented.

2.1 Measuring Business Model Innovation

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 through 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).

2.2 BMI and Firm Performance

A number of studies have examined the relationship between BMI and the financial performance of a company. These studies differ in terms of sample size and characteristics, methodology and choice of measures (Foss and Saebi 2017; White et al. 2022). Cucculelli & Bettinelli (2015) for example investigate the effect of BMI on sales growth, return on sales (ROS) and total factor productivity (TFP) in a sample of 376 Italian SMEs in the years 2000-2010. 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) use a cross-industry sample of 563 European SMEs and provide 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 in accordance with the model proposed by Venkatraman & Ramanujam (1986). White et al. (2022) conducted a meta-analysis based on the extant BMI literature. After an extensive search of the literature, they identified 77 studies comprising 26,050 firms from 26 countries and six continents. 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.

2.3 Text Mining 10-K Filings

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. In 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. Furthermore, the authors' findings indicate a correlation between firms' R&D and advertising expenditures and an increase in product differentiation. This underscores the significance of product differentiation. In their study, Lee & Hong (2014) examine the evolution of a firm's BM over time. The authors thus utilize the procedure proposed by Miner et al. (2012) to 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).

2.4 Summary

The concept of BMI is complex, with different definitions and dimensions, which makes it challenging to measure. The existing scale-based measures fulfill their purpose, but are insufficient for use in large-scale, longitudinal studies. Furthermore, researchers must rely on interviews and questionnaires, which makes them reliant on the willingness of firms to cooperate. Despite these difficulties, researchers have identified ways to measure a positive relationship between BMI and a firm's performance. This relationship is shaped not only by the choice of measurement of BMI and firm performance but also by factors such as firm age, industry, and economic and political environment.

- What do we learn from Text Mining part

3 Methods

3.1 Assumptions

3.2 Gemini

10-K filings are in general very large text documents, where Item 1 of these filings is no exception. Table 1 shows the average, minimum and maximum length of the original Item 1 part in our sample. In order to utilize all information regarding the BM in the Item 1 part 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. We inserted our prompt at the beginning of each text file and passed it via an API to Gemini¹. We used following prompt: "Summarize the business model from the following

¹We forked and used following Github Repository: https://github.com/skranz/gemini_ex.

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 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. Thus, every dimension of the definition of BMI by Spieth & Schneider (2016) is covered by this prompt. To check quality and accuracy of the summaries by Gemini, we draw a random sample of 100 filings and compare the original text with the summary. More precise, we first read the original file,

TODO - Descriptive Table1 for document length of original filings

3.3 BERT and BERTScore

3.4 Hierarchical Agglomerative Clustering

4 Experimental Apparatus

4.1 Datasets

We collect 10-Ks 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 only use the Item 1 part, since this is the most crucial part of the 10-K filings for describing the company’s 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 listed (on stock exchange). On the other hand, we consider companies that filed 10-K reports with the SEC between 2017 and 2023
 //→ rewrite as step by step, how we got to the final list of companies

We exclude companies from the financial sector, namely companies with a SIC Code starting with six. We consider the filings from 2017 to 2023.

TODO

- Table2 like Alex suggested
- Descriptive Table3 for document length of processed filings
- Description of Table3 and the final Dataset

4.2 Pre-Processing

- everything regarding Gemini in Pre-Processing

Since the parts of section I are generally not well structured for further processing, we use Gemini to summarize the texts with emphasis on the business description. Based on this data, we create similarity scores between different companies and the same companies from different years. Finally, we match these firms with several financial numbers that we extracted from Eikon/Datastream.

4.3 Procedure

Before generating the final datasets, we had to iteratively coordinate various steps over several phases in order to generate the data consistently. The first step was to extract the required company identifiers (CIKs) from the SEC database.

5 Results

6 Discussion

7 Conclusion

8 Limitations

9 Acknowledgement

- Jonathan for IT Support
- Prof. Kranz for Repo

²The spelling error in the last sentence of the prompt was found after processing the Item 1. After evaluating the Summaries, this error did not cause any issues.

- 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>.
- 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>.
- Miner, Gary, Dursun Delen, John F. Elder, Andrew Fast, Thomas Hill, Robert Nisbet, Jennifer Thompson, et al. 2012. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*. Waltham, Mass., Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo: Elsevier, Academic Press.
- 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>.
- 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>.
- Venkatraman, N., and Vasudevan Ramanujam. 1986. "Measurement of Business Performance in Strategy Research: A Comparison of Approaches." *The Academy of Management Review* 11 (4): 801. <https://doi.org/10.2307/258398>.
- 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>.