

The Name of the Title Is Hope

MAX GABLER, WANSHU JIANG, CHRISTOPH KIESL, LEONARD PÖHLS, ALEXANDER RIEBER,
and ANSGAR SCHERP

A clear and well-documented L^AT_EX 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.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: 10-K, Business Model Innovation, BERT, Gemini

ACM Reference Format:

Max Gabler, Wanshu Jiang, Christoph Kiesel, Leonard Pöhls, Alexander Rieber, and Ansgar Scherp. 2018. The Name of the Title Is Hope. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

Business model innovation (BMI) is a key activity to maintain competitiveness and even gain a competitive advantage [? ?]. 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 [?]. The scale developed by Spieth & Schneider [?] provides managers and practitioners with a measurement index for business model innovativeness. This measurement model only validates applicability of BMI theory [?] and is insufficient for longitudinal studies [?]. 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 [?]: 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 [?] 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.

Authors’ Contact Information: Max Gabler; Wanshu Jiang; Christoph Kiesel; Leonard Pöhls; Alexander Rieber; Ansgar Scherp.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

2 Related Work

Our study borders several research fields and thus we have a variety of literature to our disposal. 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 [?]. Consequently, there is considerable variation in the definitions of BMI, with some definitions being more similar to one another than others [?]. Spieth & Schneider [?] 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 [?] 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 [?]. Clauss [?] 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 [?]. 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 [? ?].

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 [? ?]. Cucculelli & Bettinelli [?] 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. [?] 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 [?]. White et al. [?] 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 [?] 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 [?] examine the evolution of a firm's BM over time. The authors thus utilize the procedure proposed by Miner et al. [?] to represent each document as a vector of keywords, which is similar to the approach utilized by Hoberg & Phillips [?]. After identifying the Item 1 part of the 10-K filings as the most crucial part for describing a firm's BM, Lee & Hong [?] 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 [?].

2.4 Summary

BMI is complex concepts which makes it hard to measure. Existing scale-measures fulfill their purpose but are not sufficient to be used in large-scale, longitudinal studies. Furthermore, researchers have to rely on interviews and questionnaires, thus

- BMI is a complex concept which is hard to measure; Measure of Spieth fulfills its purpose
- Use Spieth definition for BM and BMI
- There have been attempts to utilize text data from 10-Ks in different research fields
- BMI is related to firm performance -> test our measure against this assumptions

3 Methods

3.1 Assumptions

3.2 Gemini

3.3 BERT

3.4 BERTScore

3.5 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 [?].

//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 7590 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 - Descriptive Table1 for document length of original filings

- Description of Table1, i.e. The intersection varies from year to year between 3138 and 4929
- Descriptive Table2 for document length of processed filings
- Description of Table2 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