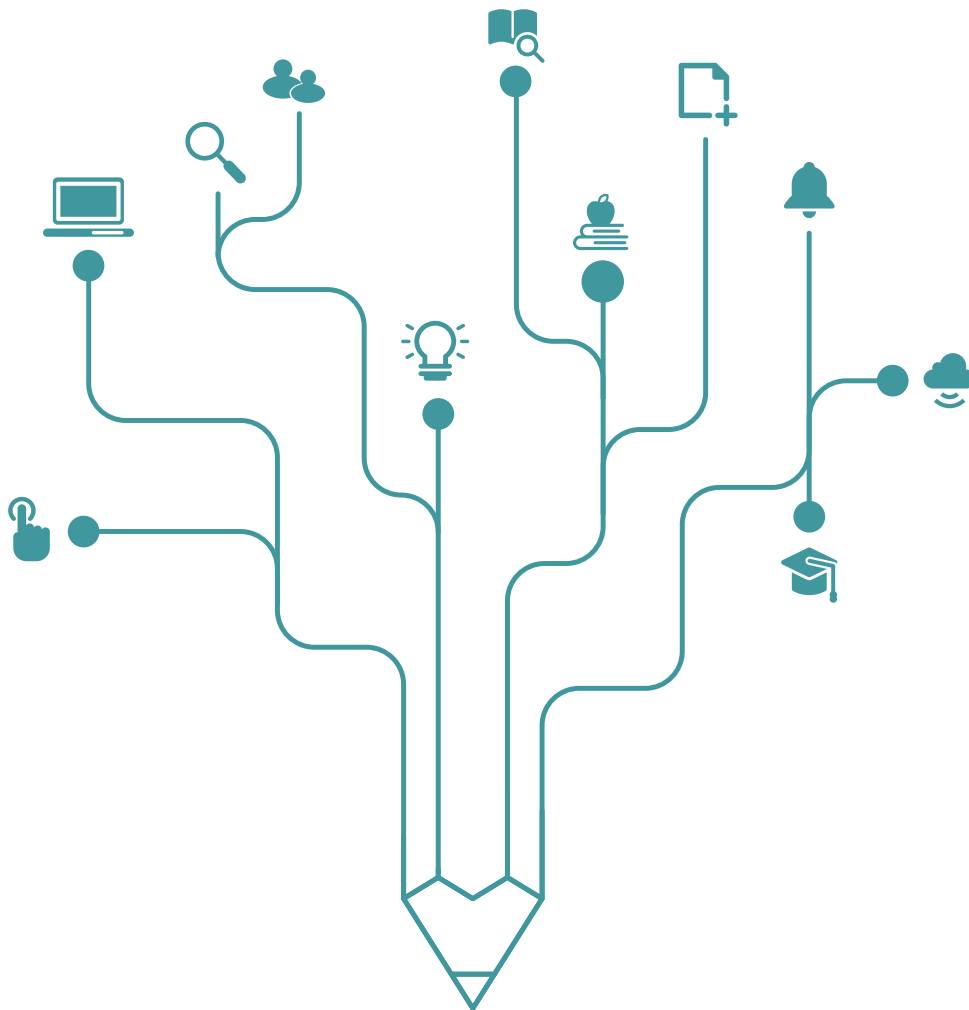


Topic Modeling in Fintech Regulations

Yunjung Yang (KDI School of Public Policy and Management)

Seohyun Lee (KDI School of Public Policy and Management)



Topic Modeling in Fintech Regulations *

YunJung Yang¹ Seohyun Lee²

June 18, 2024

Abstract

This paper analyzes South Korea's fintech policy landscape from 2000 to 2023, employing topic modeling techniques such as Non-negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA). We identify key topics and investigate the trends and shifts of recurring topics by examining 67 fintech-related policy documents issued by the national government. Three dominant topics are institutional support for the fintech sector, strengthening regulation for safe and secure provision of fintech services, and fintech regulatory reform for economic growth. Keywords and their weight distribution provides useful insights into the core elements of each theme, illuminating policy priorities and objectives. Our findings suggest that the evolution of key fintech policy themes are connected to political shifts and significant trigger events. By leveraging Intertopic Distance Maps, we examine the relative dominance and commonality of topics. The results show the evidence of the interplay of policy goals to support financial innovation and to ensure financial stability.

JEL Classification: G18, G28

Keywords: Fintech, Topic Modeling, Financial regulations

*We thank seminar participants at the Korean Association for Policy Studies for useful suggestions.

¹First author, KDI School of Public Policy and Management, E-mail: yyjsarah@gmail.com

²Corresponding author, KDI School of Public Policy and Management, E-mail: slee6@kdis.ac.kr.

1 Introduction

In recent years, fintech sectors has experienced remarkable growth, driven by technological advancements and innovations. In turn, policy responses to fintech industry has created and reshaped the infrastructure for financial services. Identifying patterns, trends and relationships within fintech policies serves as a foundational step to understand the intricate dynamics underpinning the fintech policy landscape. This understanding will be crucial for anticipating and navigating significant challenges that manifest in response to pivotal incidents, contributing to more informed and adaptive regulatory environment.

This research embarks on a comprehensive exploration of the multifaceted fintech policies, delving into South Korea's two-decade-long repository of fintech policy documents. South Korea, with its distinctive blend of cutting-edge technological advancements and a dominant traditional banking sector, emerges as a compelling focal point for this study. South Korea is one of the most developed countries in information and communications technologies. Despite its technological prowess, South Korea's fintech landscape is profoundly influenced by the presence of its traditional banking industry. Unraveling the practical implications of this coexistence within the fintech ecosystem, particularly its impact on regulatory decision-making, promises interesting insights into the factors that mold fintech policies in an ever-evolving global landscape.

This research aims to provide a thorough overview of South Korea's fintech policies through the lens of topic modeling analysis. The analysis centers on 67 fintech-related policy documents issued by the national government from 2000 to 2023. Employing topic modeling techniques such as Non-negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA), the study categorizes these policies based on frequently used words, referred to as "topics".

By uncovering underlying themes, relationships between policies, and identifying shifts in policy objectives, the study offers a multifaceted depiction of the fintech regulatory landscape. In doing so, this paper is to furnish valuable insights for policy makers and industry stakeholders, empowering them to formulate effective regulations and foster sustainable growth within the dynamic fintech sector. Given that policies have lasting effects on politics and society (Pierson, 1993), an

exploration of the evolutionary trajectory of policies provide insights that can significantly influence future policy decisions.

The paper seeks to address several key questions. First, it looks at how fintech policies evolved in South Korea from 2000 to 2023, and investigate key trends, challenges, and opportunities that have emerged in this dynamic regulatory landscape. Second, we identify the thematic areas and topics that dominate fintech policy documents in South Korea, and examine how these themes have evolved over time. By investigating two key shifts in regulatory statements, we find that these shifts correlate with changes in political regimes and significant events in South Korea's fintech landscape.

Our research intersects with two key strands of literature. The first strand of literature focuses on the need for effective fintech policymaking to accommodate technological advancements while ensuring a fair regulatory environment. Scholars focus on understanding coalition formation, regulation governance, and regulation strategy's complexity in shaping the regulatory landscape (Whitford and Anderson, 2021; Ehrentraud et al., 2020; Batunanggar, 2019). These studies observe conflicts between traditional financial institutions and ICT-based newcomers, including disputes over regulatory frameworks, data privacy, and consumer protection. Literature also advocates for a comprehensive approach, combining theoretical models and case studies to analyze fintech policy development. The Narrative Policy Framework (Im and Park, 2022) is proposed to examine policy narratives and advocate groups and to offer insights into advocacy coalition formation and their influence on the legislative process. While the framework attempts to improve our understanding of the intricate dynamics at play in the policy process, it relies solely on theoretical models. Recognizing the importance of quantitative analysis, existing studies utilize methodologies such as regression analysis (Haddad and Hornuf, 2019; Holtfort et al., 2021; Siek and Sutanto, 2019; Muganyi et al., 2022), bibliometric analysis (Liu et al., 2020) and topic modeling analysis (Li et al., 2017).

The second strand of literature that is related to our research emphasizes text analyses for an empirical tool of policy science. Analyzing text as data has been a significant component of policy

science, while the computational method of topic modeling has gained traction since 2015 (Isoaho et al., 2021). Interpreting policy text in social scientific research has been experimented with various topic modeling methods (Müller-Hansen et al., 2021 (environmental policy) and Saheb and Saheb, 2023 (artificial intelligence policy), among many).

The existing literature highlights the significance of political narrative and stakeholder dynamics to understand the complexity of policy development, which aligns with the perspective of this research. While previous studies have mainly focused on factors affecting the growth of fintech industry or its impact on the financial sector, our research offers a unique and comprehensive exploration of fintech policy itself. Notably, the paper pioneers advanced text analysis methods, contributing to the area in quantitative analysis. Moreover, by focusing on South Korea, a nation balancing cutting-edge technology and a dominant traditional banking sector, this study fills a critical gap in current research landscape by providing a data-driven examination of fintech policies.

The rest of the paper is organized as follows. Section 2 describes the data and Section 3 explains the empirical methodologies employed in the study. Section 4 presents the main findings and Section 5 concludes.

2 Data

The primary dataset for this research consists of 67 fintech policy documents issued by the South Korean government. This dataset spans the years from 2000 to 2023, providing a comprehensive overview of the evolutionary trajectory of fintech policies in South Korea. The data is retrieved from the KDI Economic Policy Time Series service¹, a reliable source that systematically curates government announcements on economic policies, media reports, research institute findings, and public opinion poll results dating back to the 1990s. The service operates under the Economic Policy Information Database (EPIC) of the KDI Economic Information and Education Center (EIEC), aggregating data from 15 economic-related ministries and affiliated organizations.

Examining this dataset reveals notable fluctuations in fintech policy activities over the years

¹<https://epts.kdi.re.kr/>

(Figure 1). In particular, 2015 emerges as the most active year marked by 13 policy document releases. This suggests potential correlation with specific trigger events, media attention, or legislative initiatives during the period. The trend continues in the year 2019 witnessed by the issuance of 10 policies, while 2018 and 2016 also show notable activity with 8 and 7 policies issued, respectively. In contrast, eleven years experienced minimal fintech policy development, signaling a dynamic regulatory environment characterized by periods of intense policymaking and relative stagnation.

A plausible explanation for this pattern is linked to the cycle of presidential regimes. Over the period from 2000 to 2023, South Korea witnessed five presidential elections², resulting in the reigns of six presidents. The cycle of presidential terms appears to influence the fintech policy landscape, with peaks in policy activity occurring in the middle of each regime. The observation aligns with historical trends, as the initial stages of a regime often prioritize foundational economic aspects like GDP and infrastructure, while fintech, perceived as sophisticated industrial advancement, takes a backseat.

The evolution of fintech policies in South Korea are influenced not only by political agendas but also by other factors. For instance, the stagnation in fintech policies during Lee Myung-bak's presidency (2008-2013) could be attributed to factors like the Free Trade Agreement with the United States or policy priorities towards other industries. The national initiative for Lee administration emphasized tasks such as resource diplomacy (overseas resource development) and infrastructural development, including the Grand Canal for the Korean peninsula. In addition, although the term 'fintech' was introduced in South Korea in the early 2000s, South Korean fintech industry did not achieve noticeable interests or growth due to regulatory barriers and public certificate requirements until 2015. Such temporal shifts of policy focus warrant further exploration to understand the nuanced interplay between policy evolution, economic shifts, and political landscapes.

The boxplot in Figure 2 illustrates the distribution of word counts in 67 fintech policy documents. The policy document with maximum word count, "Comprehensive Measures to Prevent

²February 24, 2003, February 24, 2008, February 24, 2013, March 10, 2017, March 9, 2022

Recurrence of Personal Financial Information Leaks”, issued in March 2013, is closely followed by the second longest document, “IT/financial Convergence Support Plan”, released in Jan 2015. Lengthier policy documents imply either detailed government measures or the complexity of the addressed issue. However, concise policies can be equally impactful, exemplified by the one with the minimum word count, “2023 Financial Services Commission Work Plan” issued in Jan 2023. This document outlines major initiatives for the year, promising comprehensive support for consulting, policy funding, and datasets for fintech companies. Emphasis is also placed on regulatory policies such as the establishment of a virtual asset discipline system and the issuance discipline system for piece investment and securities-type digital assets. The significance of a document lies not only in its length but also in its ability to address issues and resonate with stakeholders. Most of the documents cluster around 200 to 350 words.

Figures 3 and 4 provide an insight into the policy distribution among government agencies, highlighting major players such as the Financial Services Commission (FSC), the Financial Supervisory Service (FSS), the Ministry of Economy and Finance (MOEF), the Ministry of Science and ICT (MSI), and the Ministry of SMEs and Startups (MSS). The graph reveals that nearly 80% of policies are issued by a single government agency, with the Financial Services Commission (40.5%), Ministry of Economy and Finance (26.9%), and Financial Supervisory Service (7.5%) taking the lead. Collaboration between multiple government agencies is evident in about one-fifth of fintech policies.

The Financial Services Commission (FSC) holds a pivotal role in overseeing financial policy, promoting policies for domestic financial companies, supervising the Financial Supervisory Service (FSS) to ensure fair financial transactions. Operating as a consensus-based administrative agency, with members such as the chairman, standing and non-standing committee members, it makes final decisions on major financial service matters. Additionally, input from ex officio members, such as the Vice Minister of MOEF, the Governor of the FSS, the President of the Korea Deposit Insurance Corporation, and the Vice Governor of the Bank of Korea, further contribute to its diverse expertise. The Ministry of Economy and Finance (MOEF) manages fiscal and eco-

conomic policies, with a focus on distributing the national budget, exercising jurisdiction over the national treasury, and overseeing state-owned property. The Financial Supervisory Service (FSS), a specialized corporation under FSC, independently inspects and supervises financial entities. This collaborative yet separate structure maintains a balanced regulatory approach, preventing excessive government control.

3 Methodology

To derive meaningful insights from the dataset, we employ various text analysis techniques with a primary focus on topic modeling. Topic modeling is a collective term for a family of computational algorithms that are used to model text in a collection of documents as arising from a much smaller set of topics. These algorithms relate to latent variable discovery, dimensionality reduction, as well as clustering (Isoaho et al., 2021). To offer unique perspectives, we utilize two distinct topic modeling methods: Non-negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA). The topic modeling process of the paper is summarized in the following (see Figure 5).

3.1 Data pre-processing

To refine the text data for effective topic modeling analyses, the initial step involves tokenization, an essential natural language processing (NLP) technique used for pre-processing and cleaning text data. Tokenization breaks down sentences into individual words (*tokens*). We use the Okt tokenizer from the KoNLPy library, a robust NLP tool designed for handling Korean text in Python.

To gauge the importance of words within the text data, we leverage the Term Frequency-Inverse Document Frequency (TF-IDF) technique. This widely-used method in information retrieval and text mining evaluates the significance of each word in the document collection (Kim and Gil, 2019). TF-IDF quantifies both term frequency (TF) in a document and inverse document frequency (IDF), indicating how unique the term is across the entire corpus. Consequently, a numerical value is assigned to each word, reflecting its importance within a specific document relative to the entire collection. Words that are more frequent in a specific document but relatively rare across the entire

corpus receive higher TF-IDF scores.³ Figure 6 shows the word cloud after pre-processing of policy documents.

3.2 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF), introduced by Paatero and Tapper (1994), serves as an algorithm to identify key topics and frequently used words within the 67 fintech policy documents. The initial work on NMF (Lee and Seung, 1999; Lee and Seung, 2000) emphasizes that NMF components capture coherent and relevant information about the underlying themes in the documents. The goal of NMF is to find a non-negative matrix providing a compact, interpretable representation of documents and terms. In the NMF algorithm, the input matrix V , a TF-IDF matrix, contains original text data vectored in documents and terms. The cells of matrix V represent the TF-IDF scores. Through the factorization process, NMF decomposes the TF-IDF matrix into two matrices, W (the basis matrix representing topics in the documents) and H (the coefficient matrix representing the importance of topics).

After constructing the TD-IDF matrix, NMF is applied with three parameters. First, the number of topics is set to 5. It specifies the number of distinct themes the algorithm aims to uncover within documents. Second, the number of top words to display for each topic is set to 10. This parameter sets the quantity of the most representative words capturing the essence of each topic. Third, the number of top documents to display for each topic is set to xx. This identifies key documents closely associated with each topic.

The optimal number of topics for the NMF model is determined by the topic coherence score method. Topic coherence quantifies how interpretable a topic is based on the co-occurrence patterns of words in a topic. The idea is that if certain words co-occur frequently within the same documents, they are likely to form coherent topics (Röder et al., 2015). Lower C_V coherence scores typically indicate more coherent topics. Using the Gensim library to compute C_V coherence, the optimal number of topics that provides the best performance is three (Figure 7).

³TF-IDF vectorization effectively reduces the importance of irrelevant words such as stop words.

3.3 Latent Dirichlet Allocation (LDA) method

In addition to NMF, we utilize Latent Dirichlet Allocation (LDA) as a topic modeling method. Developed by Blei et al. (2003), LDA discovers hidden themes within documents using probabilistic models based on word distributions. The LDA model involves a complex interplay of distributions and parameters to capture the underlying structure of the documents in the corpus. While LDA traditionally operates on raw word frequency counts without any preprocessing technique, we opt for TF-IDF as input for a more accurate representation of word importance. TF-IDF provides more accurate keywords by weighing their importance in specific documents against their rarity in the whole corpus.⁴

The most important tuning parameter for performing LDA model is the number of topics. The optimal number of topics for LDA is determined using perplexity scores.⁵ Perplexity score measures how well a model predicts words within a dataset, with lower scores indicating better prediction (Cha and Cho, 2012). The study employs the elbow method to identify the point where perplexity scores start to level off in the curve. Figure 8 illustrates how perplexity scores change with different topic numbers. After experimenting with a range of topics, we find that 10 is the optimal number of topics providing the best performance. Adding more topics shows diminishing returns in predictive capabilities.

We use Scikit-Learn (sklearn), a Python machine learning library to implement the LDA model. T-SNE, a dimensionality reduction technique can be applied to visualize clustering determined by LDA in a 2-D embedding.

⁴The TF-IDF-LDA combination has been employed in various papers, emphasizing its advantages in enhancing topic modeling results (Kim and Gil, 2019).

⁵We employ pyLDAvis to cluster optimal topics into 10, resulting in a visually informative intertopic distance map in Section 4.

4 Results

4.1 The NMF topic modeling results

Understanding the themes of fintech policies

We first explore the thematic areas that have consistently dominated fintech policy documents in South Korea. The NMF topic modeling results have successfully identified three key recurring themes: (a) institutional support for fintech business, (b) strengthening regulation for safe and secure fintech service, and (c) regulatory innovation for economic growth. Table 1 presents quantitative metrics of top ten topics, showcasing the keywords and their weights associated with each theme. Each keyword within a topic is assigned a weight, which is a value between 0 and 1. Weight represents keyword's relative importance within the topic. This means a higher the weight, the more central that keyword is to the topic.

Topic 1 (*institutional support for fintech business*) revolves around the policy support extended to businesses in the fintech sector. The discussion includes terms such as "finance", "support", "business", "innovation", and "overseas expansion". A deep dive into documents indicates their focus on supporting fintech companies' innovation and global expansion, considering financial technology and innovation, institutional support, regulatory aspects, and services promotion for overseas markets. Key documents contributing to this theme include the Fintech Industry Revitalization Plan (May 4, 2015), Basic Plan for Financial Center Creation and Development (October 12, 2017; June 21, 2023), Fintech Scale-up Promotion Strategy (December 5, 2019), and Plan to Revitalize Fintech Innovation (March 21, 2018).

Topic 2 (*strengthening regulation for safe and secure fintech Service*) focuses on implementing security measures for a trustworthy fintech landscape. Terms like "electronic", "transactions", "security", "safety" and "consumer" feature prominently in this topic. Governmental actions, as reflected in key documents like Measures to Improve the Safety of IT and Electronic Financial Transactions (February 13, 2003), Comprehensive Measures to Strengthen the Safety of Electronic Financial Transactions (September 20, 2005.9.20), and others, aim to ensure the reliability

of digital financial services and protect consumers. While the central theme revolves around promoting secure electronic financial transactions, documents vary in their target audiences – some targeting financial industry directly, others broader financial consumers. The areas of government intervention also differ, ranging from safeguarding electronic financial transactions through company-focused regulation, to using customized financial education and measures to prevent telecommunication financial fraud.

Topic 3 (*regulatory innovation for economic growth*) explores regulatory innovation to support fintech industry and its economic implications. The discussion include terms such as "innovation", "regulation", "economy", "industry", and "investment", suggesting a comprehensive exploration of regulatory innovation for the growth of fintech sector and subsequent overall economic development. Key documents contributing to this theme include Ministerial Meeting for Fintech System Improvement (January 16, 2019.1.16), Strategy to Spread and Accelerate Innovative Growth (August, 21, 2019) which introduced the regulatory sandbox, Field-based Regulatory Innovation (February 20, 2019), Revision of the 3 Data Laws (December 4, 2019), and Introduction of Small Licenses (December 29, 2018).

The analysis of dominant themes in South Korean fintech policy documents provides valuable insights into the priorities of regulators. These three identified themes underscore the multifaceted nature of policy considerations. Leveraging these insights, policymakers can navigate the dynamic fintech landscape, fostering innovation, ensuring security, and driving sustainable economic growth. For institutional support (Topic 1), policymakers may consider continued backing through targeted initiatives, including financial incentives, incubation programs, and streamlined regulatory processes. Topic 2 (strengthening regulation) emphasizes the importance of security measures in establishing a trustworthy fintech landscape. Policymakers can implement robust cybersecurity frameworks and promote industry-wide best practices, and foster collaborations between regulatory bodies and industry stakeholders to ensure the reliability of digital financial services. Topic 3 (regulatory innovation) explores strategies for inducing fintech industrial development, pointing towards agile regulatory frameworks such as sandboxes. Continuous dialogue with industry ex-

perts to adapt regulations to technological advancements and strategic investments would propel the fintech industry's contribution to economic development.

Evolution of Fintech Policies in South Korea during the period between 2000-2023

Figure 9 reveals that fintech policy trends are shaped by the periodical context faced by the administration, rather than the political inclination of the ruling party. The initial focus on strengthening regulatory measures for secure fintech services (Topic 2) gradually shifted to the focus on institutional support for fintech businesses (Topic 1) during Park Geun-hye's administration (4th period). The Moon Jae-in administration (5th period) marked a significant turning point in prioritizing regulatory reliefs for economic growth (Topic 3) while concurrently supporting fintech businesses (Topic 1). The topic modeling results lead us to identify two major shifts in policy focus in 2015 and 2019.

The sudden shift in policy focus in 2015 can be attributed to an unexpected incident. As Baumgartner and Jones (1991) suggests, policies, usually characterized by long-term stability, experience "short bursts of rapid change" in response to external shocks or critical events. The incident stems from the popularity of the drama series 'My Love from the Star' among Chinese audiences. The attempts of overseas purchase of the main character's coat encountered difficulties in international payment processes due to the mandatory installation of security software, ActiveX. The government-led development of the digital certificate for financial transactions were less responsive to the innovations. The issue gained prominence at the Regulatory Reform Conference in March 2014 and, eventually, the Electronic Transaction Supervision Regulations were revised in order to do away with the requirement of the usage of ActiveX for financial institutions.

This shift played a pivotal role in reshaping fintech policy as it created a 'window of opportunity' for significant governmental policy implementation. In response, the Financial Services Commission (FSC) unveiled the 'Fintech Industry Revitalization Plan' in 2015. This marked a regulatory transformation, adopting a pre-approval-post-regulation approach. Positive regulation specifies what is allowed, while negative regulation allows everything unless expressly prohibited. This shift facilitated the entry of new products or services, contributing to the success of major

fintech services like Bank Salad, Toss, and Kakao pay.

The revitalization plan outlined three main goals: building fintech infrastructure, launching user-friendly fintech services, and establishing a foundation for early fintech enterprises. Key initiatives for first include easing entry regulations, encouraging investment in fintech, providing financial support, and addressing tech limitations. To enhance public access, the plan advocates for real name verification online, crowdfunding, internet-only banks, and online insurance sales channels. The third goal emphasizes vital fintech ecosystem, an autonomous security system, and IT-finance convergence using big data. Overall, the plan aimed to boost citizen access to diverse financial services, promote the growth of the fintech industry, and fortify the financial sector.

As a result of these policy efforts, Korea's Fintech adoption index, published by Ernst & Young⁶, doubled from 32% in 2017 to 67% in 2019. This placed Korea on par with Singapore (67%), HongKong (67%), and the financially advanced country, UK (71%). Notably, considering fintech adoption index is low in countries with traditionally mature financial markets, like the United States(46%) and Japan(34%), Korea's high fintech adoption index reflects strong policy will of the government.

During the policy shift in 2019, the government focused on supporting fintech businesses (Topic 1) and implementing regulatory innovations for economic growth (Topic 3). This shift was evident in policy documents, with a proportional rise in the prominence of these two key topics.

Global events significantly influenced South Korean governmental policies for institutional support in fintech. The introduction of GDPR (General Data Protection Regulation) in 2018 prompted the government to introduce support measures such as consulting for fintech companies to meet heightened data protection standards. The COVID-19 pandemic further accelerated digital financial services, resulting in a directive for the expansion of accessibility and openness in the financial payment system (Lyons et al., 2022). The policy document released in 2019 emphasizes fintech scale-up promotion, investment in major platform economy sectors including data and AI, and the development of the fintech ecosystem.

⁶https://www.ey.com/en_es/ey-global-fintech-adoption-index

Noteworthy initiatives include the introduction of a small license (temporary permission system) for fintech companies and the creation of a fintech-specific innovation investment fund worth KRW 300 billion over four years. The government aimed to promote fintech scale-up, targeting the emergence of global-level fintech unicorns in Korea. Support was extended for the development of fintech entrepreneurial services, particularly in robo-advisory and foreign exchange-related fintech fields.

The surge in regulatory innovation in South Korean fintech policies in 2019 can be attributed to the introduction of the financial regulatory sandbox by the Moon administration. Enacted in December 2018 and implemented in April 2019, the Special Act on Financial Innovation Support played a pivotal role in approving numerous innovative financial services. This regulatory sandbox provided a controlled environment for testing and launching new ideas, encouraging fintech entrepreneurs to push the boundaries of traditional financial services.

The legislative framework established by the financial regulatory sandbox not only facilitated the rapid introduction of new fintech services but also laid the groundwork for sustainable growth. Notably, 77 innovative new financial services were approved in 2019, followed by 58 in 2020, 50 in 2021, and 52 in 2022. This success reflects the effectiveness of the regulatory sandbox in fostering a dynamic and innovative fintech landscape, aligning with the government's strategic goal for economic development and technological advancement. The policy document issued in February 2019 emphasizes a preemptive response to the 4th Industrial Revolution, focusing on the innovative growth of new industries.

Intertopic Distance Map

The Intertopic Distance Map visually represents the distribution of topics, positioning them in a multidimensional space based on their similarity or distance from each other. Each bubble on the map corresponds to a distinct topic, with the size indicating its prevalence. Closer proximity between bubbles signifies greater similarity in the words they contain.

Topic 1, covering almost 60% of the dataset, focuses on promoting and innovating fintech businesses in South Korea (Figure 10). The overarching theme centers on leveraging technology to

transform the financial landscape. Government strategies aim to strengthen the fintech ecosystem, foster globally competitive services, and enhance financial infrastructure. This includes initiatives like financial-non-financial convergence, involving partnerships between banks and technology companies to provide innovative solutions for financial transactions or collaborations between financial institutions and companies from various industries.

Topic 2 represents one-third of the tokenized words and outlines plans to enhance regulation for reliable electronic transactions, a foundational aspect of the fintech sector (Figure 11). The main objectives include customer protection and ensuring the security of electronic systems. Documents in this topic highlight concerns about the rapid expansion of electronic financial transactions, emphasizing the need to strengthen monitoring on companies' electronic financial security measures. The Financial Supervisory Service plays a crucial role in implementing this initiative.

Topic 3, accounting for 5%, appears peripheral and diverges by encompassing human resource-related policies like job creation and expert nurturing (Figure 12). This deviation from the basic NMF topic result of Topic 3 (regulatory innovation) is attributed to the technical aspects of the NMF and Intertopic Distance Map processes. NMF is a mathematical technique for extracting topics from a document-term matrix, while the Intertopic Distance Map is a visualization tool that helps to interpret the relationships and separations between these topics. The unexpected inclusion of human resource policies highlights the nuanced outcomes resulting from these technical intricacies.

4.2 The LDA topic modeling results

While NMF has its merits, it has limitations in capturing the intricate inter-document dynamics (Blei et al., 2003). Recognizing the need for a more comprehensive view, we turn to Latent Dirichlet Allocation (LDA) model due to its probabilistic nature and adept handling of complexity across the entire dataset. The integration of LDA with TF-IDF provides top keywords which also accounts for the overall topic structure and intertopic distances.

The LDA analysis yields a refined categorization of fintech policies into three distinct areas, providing valuable insights for deeper research planning. The visualization of intertopic distance map unveils three main categories: *Regulation for Customer Safety*, emphasizing digital security and customer protection; *Support for Industrial Vitalization*, aligning with NMF result's main topic; and an intermediary cluster symbolizing themes that bridge these primary categories.

Regulation for Customer Safety (Topic 1, 48% of tokens) encompasses policies focused on ensuring digital security and safeguarding customers in the fintech sector (Figure 13). Noteworthy policies include educating fintech companies, presenting guidelines for service provision clauses, and the enactment of the Electronic Financial Transaction Act (EFTA).

Support for Industrial Vitalization (Topic 2, 34% of tokens) includes policies supporting the growth and development of the fintech industry (Figure 14). This category overlaps with the most prominent topic shown in NMF results, featuring terms like "Financial", "Technology", "Innovation", "Regulation", "Industry", and "Vitalize". Policies within this category encompass regulatory sandbox initiatives, support for fintech infrastructure such as financial big data and risk management system, and boosting investment for product development, overseas expansion, and market revitalization.

Between the two distinct big circles each representing opposite initiative directions, there are eight clustered small circles in the middle (Figure 15). These intermediate themes include economic compliance, collaboration between companies and human resource management. Specific initiatives within this realm involve nurturing fintech professionals, operating online education platforms like FinEDU, and supporting fintech entrepreneurs cloud system use costs and security consulting services. Additionally, events Fintech Demo Day and networking events between fintech companies and collaborative institutes contribute to the interconnectedness of these themes.

In conjunction with our exploration using LDA, it is crucial to acknowledge the robust foundation laid by NMF in uncovering key recurring themes within South Korean fintech policy documents. NMF's topic modeling results has elucidated three central themes: Institutional Support for Fintech Business, Strengthening Regulation for Safe & Secure Fintech Service, and Regulatory In-

novation for Economic Growth. In parallel, LDA has provided additional layers of insight, categorizing policies into three similar yet distinct areas — Regulation for Customer Safety, Support for Industrial Vitalization, and intermediary themes—enriching our understanding of the multifaceted inter-document dynamics within the regulatory landscape.

5 Conclusions

The integration of insights derived from Non-Negative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA) analyses provides a holistic understanding of the South Korean fintech policy landscape. By examining the evolution of fintech policy documents, we find that the correlation between regulatory adjustments, political events, and external influences underscores the complex interplay shaping regulatory dynamics in South Korea. Unraveling these dynamics not only provides valuable insights into the historical development of policies but also serves as a compass for future policy-making and strategic planning in the ever-evolving fintech landscape.

While NMF identifies three core themes, LDA contributes nuanced categorizations and interrelations between documents. The intertwining of these analyses weaves a comprehensive narrative of foundational pillars of fintech policies—institutional support, secure Fintech environment, and regulatory innovation—guiding policymakers through the multifaceted nature of policy considerations. We demonstrated that topic modeling proves to be a powerful tool for unveiling hidden structures and themes within extensive textual data, enabling the application of policy theories and concepts to larger datasets (Isoaho et al., 2021).

Despite its strengths, topic modeling has inherent limitations, primarily stemming from the nature of policy documents. Policy documents alone do not provide a complete coverage of policy challenges; instead, they offer a focused account of descriptive phases in the policy cycle (Knill and Tosun, 2008). The study does not explicitly capture stakeholder engagements, the real-world impact of implemented policies, and potential conflicts arising from regulatory changes. Future research endeavors could address these limitations by employing alternative methodologies. For example, network analysis could directly capture stakeholder perspectives, including those from

government agencies identified through this research, financial institutions, technology companies and customers. This presents an opportunity to enhance the holistic understanding of the policy landscape by incorporating the viewpoints of key actors involved in the fintech sector.

It is worth noticing that our analyses pertain to the published policy documents. Certain periods, despite being active in policy discussions, may be underrepresented due to limited publication of fintech policies. For instance, during the Lee Myung Bak administration, discussions on policies such as the relaxation of the separation of finance and industry, the establishment of megabanks, and privatization of the Korean Development Bank were prominent under the slogan of ‘advancement of the financial industry and globalization of the financial market’. However, external events such as the U.S subprime mortgage crisis and the bankruptcy of Lehman Brothers in 2008, which happened during the first year of the administration, limited the administration’s capacity to actualize discussed policies. This highlights the dynamic nature of policy formulation that may not be fully represented in the research.

Future research can explore the impact of fintech policies by analyzing economic indicators, industry growth, and innovation metrics over time through regression analysis. Additionally, incorporating sentiment analysis on media reactions from social media, news articles, or public forums could offer insights into public perceptions and sentiments toward fintech policies. Including sentiment analysis on media reactions could provide a more nuanced understanding of whether external pressure affected fintech policy or the trends of public perceptions to fintech policy and the industry.

Our results from LDA, narrowing topics to three distinct categories, could pave the way for further exploration, including the alignment of key fintech policy terms with media and academic terminology. A comparative analysis with global fintech policy trends that identifies similarities, differences, and areas where South Korea stands out, could also offer insights into its position in the global fintech ecosystem.

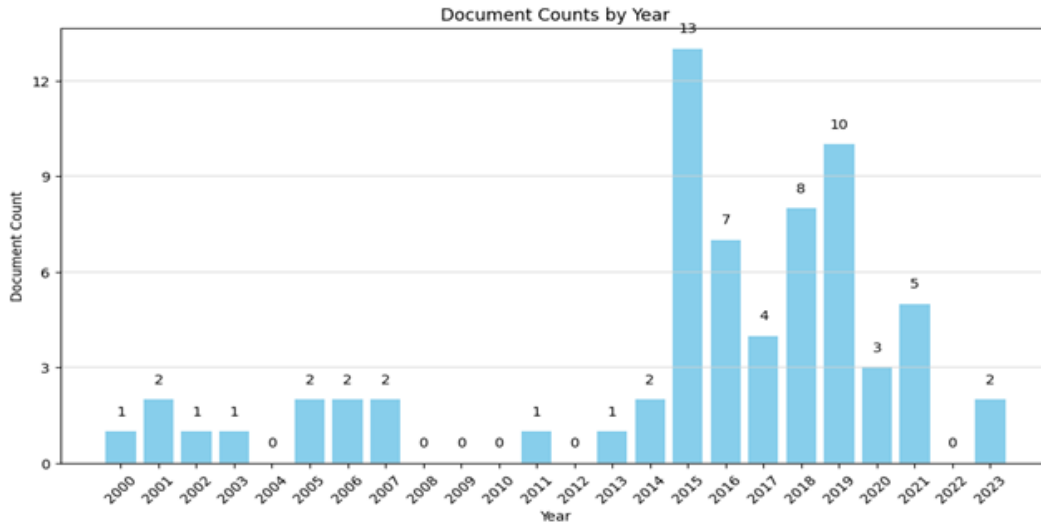
References

- [1] Batunanggar, S. (2019). Fintech development and regulatory frameworks in indonesia. Technical report, ADBI Working Paper Series.
- [2] Baumgartner, F. R. and Jones, B. D. (1991). Agenda dynamics and policy subsystems. *The journal of Politics*, 53(4):1044–1074.
- [3] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- [4] Cha, Y. and Cho, J. (2012). Social-network analysis using topic models. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 565–574.
- [5] Ehrentraud, J., Ocampo, D. G., Garzoni, L., and Piccolo, M. (2020). Policy responses to fintech: a cross-country overview. *BIS: Bank for International Settlements*.
- [6] Haddad, C. and Hornuf, L. (2019). The emergence of the global fintech market: Economic and technological determinants. *Small business economics*, 53(1):81–105.
- [7] Holtfort, T., Horsch, A., and Schwarz, J. (2021). Global fintech entrepreneurship and its influencing factors: An evolutionary economic analysis. *Risk Governance and Control: Financial Markets and Institutions*, 11(1):61–79.
- [8] Im, H. and Park, H. J. (2022). The use of policy narratives in policy changes on protection and utilization of personal information. *Journal of Regulation Studies*, 31(2):3–37.
- [9] Isoaho, K., Gritsenko, D., and Mäkelä, E. (2021). Topic modeling and text analysis for qualitative policy research. *Policy Studies Journal*, 49(1):300–324.
- [10] Kim, S.-W. and Gil, J.-M. (2019). Research paper classification systems based on tf-idf and lda schemes. *Human-centric Computing and Information Sciences*, 9:1–21.
- [11] Knill, C. and Tosun, J. (2008). Policy making.
- [12] Lee, D. and Seung, H. S. (2000). Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13.
- [13] Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *nature*, 401(6755):788–791.

- [14] Li, G., Dai, J. S., Park, E.-M., and Park, S.-T. (2017). A study on the service and trend of fintech security based on text-mining: Focused on the data of korean online news. *Journal of Computer Virology and Hacking Techniques*, 13:249–255.
- [15] Liu, J., Li, X., and Wang, S. (2020). What have we learnt from 10 years of fintech research? a scientometric analysis. *Technological Forecasting and Social Change*, 155:120022.
- [16] Lyons, A. C., Kass-Hanna, J., and Fava, A. (2022). Fintech development and savings, borrowing, and remittances: A comparative study of emerging economies. *Emerging Markets Review*, 51:100842.
- [17] Muganyi, T., Yan, L., Yin, Y., Sun, H., Gong, X., and Taghizadeh-Hesary, F. (2022). Fintech, regtech, and financial development: evidence from china. *Financial innovation*, 8(1):1–20.
- [18] Müller-Hansen, F., Callaghan, M. W., Lee, Y. T., Leipprand, A., Flachsland, C., and Minx, J. C. (2021). Who cares about coal? analyzing 70 years of german parliamentary debates on coal with dynamic topic modeling. *Energy Research & Social Science*, 72:101869.
- [19] Paatero, P. and Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126.
- [20] Pierson, P. (1993). When effect becomes cause: Policy feedback and political change. *World politics*, 45(4):595–628.
- [21] Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.
- [22] Saheb, T. and Saheb, T. (2023). Topical review of artificial intelligence national policies: a mixed method analysis. *Technology in Society*, page 102316.
- [23] Siek, M. and Sutanto, A. (2019). Impact analysis of fintech on banking industry. In *2019 international conference on information management and technology (ICIMTech)*, volume 1, pages 356–361. IEEE.
- [24] Whitford, A. B. and Anderson, D. (2021). Governance landscapes for emerging technologies: The case of cryptocurrencies. *Regulation & Governance*, 15(4):1053–1070.

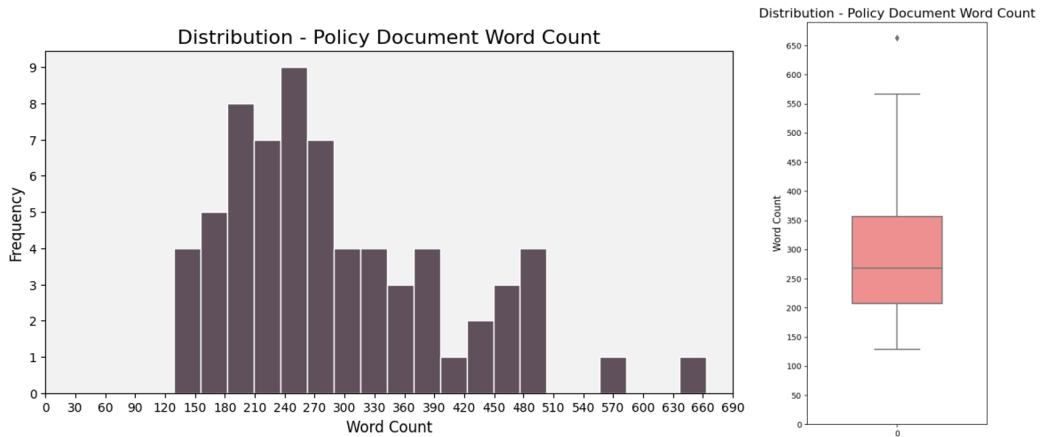
Figures and Tables

FIGURE 1 Frequency of Policy Documents by Year



Sources: The KDI Economic Policy Time Series.

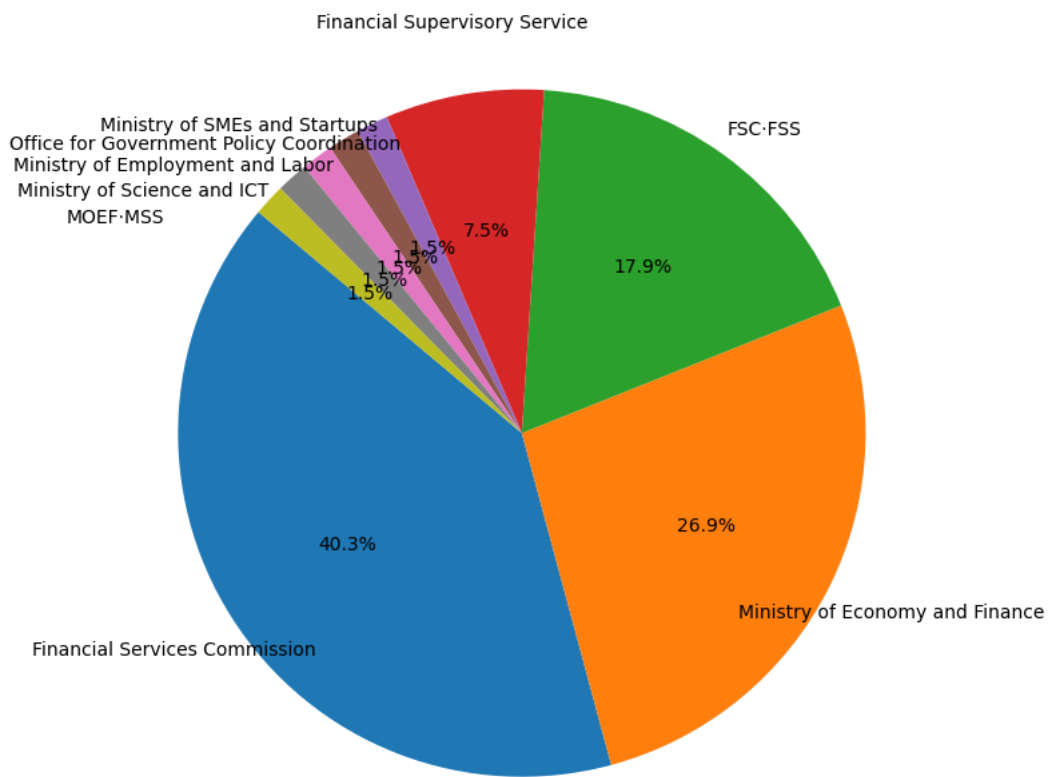
FIGURE 2 Word Count Distribution of Policy Documents



Sources: The KDI Economic Policy Time Series.

FIGURE 3 Share of Policy Document Producing Government Agencies

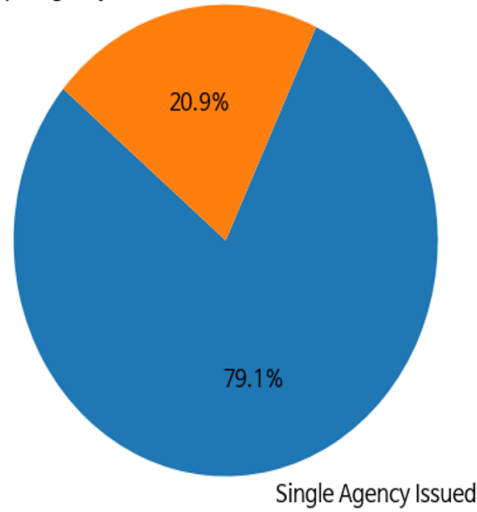
Policy Share by Government Agencies



Sources: The KDI Economic Policy Time Series.

FIGURE 4 Share of Single and Multiple Agency Issued Policy Documents

Proportion of Single and Multiple Agency Issued Policies
Multiple Agency Issued



Sources: The KDI Economic Policy Time Series.

FIGURE 5 Topic Modeling Process

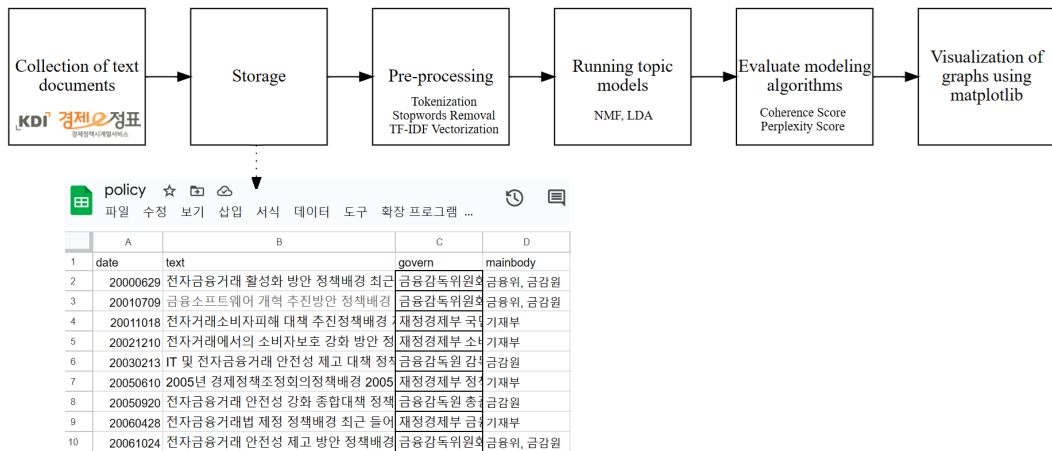
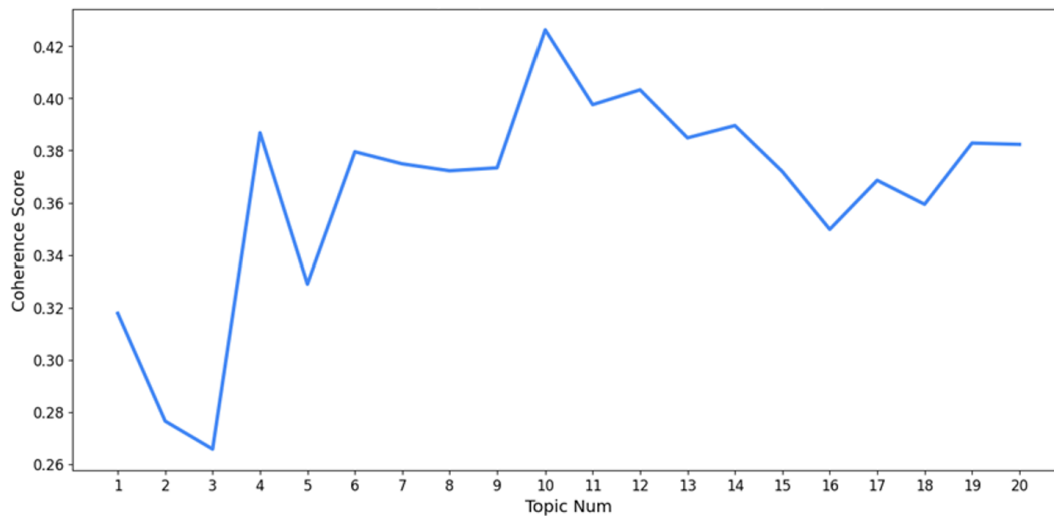


FIGURE 6 Word Cloud

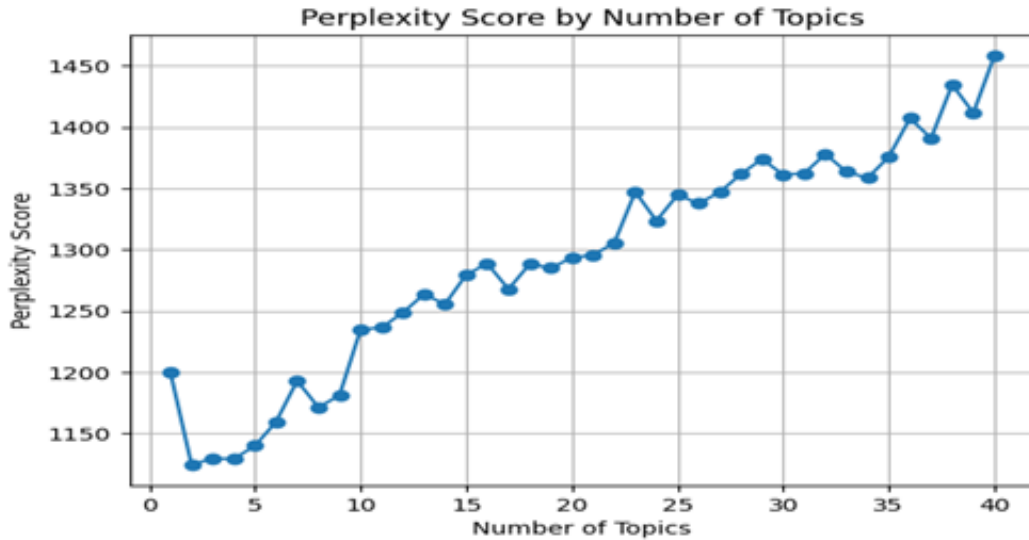


FIGURE 7 Coherence Score by Number of Topics



Sources: C_V score is computed using the Gensim library.

FIGURE 8 Perplexity Score by Number of Topics



Sources: Perplexity score is computed using Scikit-Learn (sklearn), a Python machine learning library.

TABLE 1 Fintech Policy Theme and Keywords derived by NMF Topic Modeling

Topic/Theme	Keywords (weights)	Core Documents (dates)
Topic 1 Institutional Support	Finance, Technology, Support , Business , Regulation, Innovation, Centers, Services, Industry, Overseas Expansion	핀테크 산업 활성화 방안('15.5.4) 금융중심지 조성·발전 기본계획 ('23.6.21, '17.10.12) 핀테크 스케일업 추진전략('19.12.5) 핀테크 혁신 활성화 방안('18.3.21)
Topic 2 Strengthen Regulation	Finance, Electronics , Transactions , Security , Strengthening , IT, Safety , Internet, Consumers , Measures	IT 및 전자금융거래 안전성 제고 대책('03.2.13), 전자금융거래 안전성 제고 방안('06.10.24), 전자금융거래 안전성 강화 종합대책('05.9.20) 전자금융거래 활성화 방안('00.6.29) 전자금융거래법 제정('06.4.28)
Topic 3 Relieving Regulation	Innovation, Industry, Regulation, Meetings, Economy , Promotion, Growth , Support, Investment, Strategy	핀테크 제도개선 장관회의('19.1.16) 혁신성장 확산·가속화 전략 ('19.8.21) - 규제샌드박스 도입 현장밀착형 규제혁신('19.2.20), 데이터 3법 개정 ('19.12.4), 스펙트럼 라이선스 도입('18.12.29)

FIGURE 9 Topic Proportions over Time

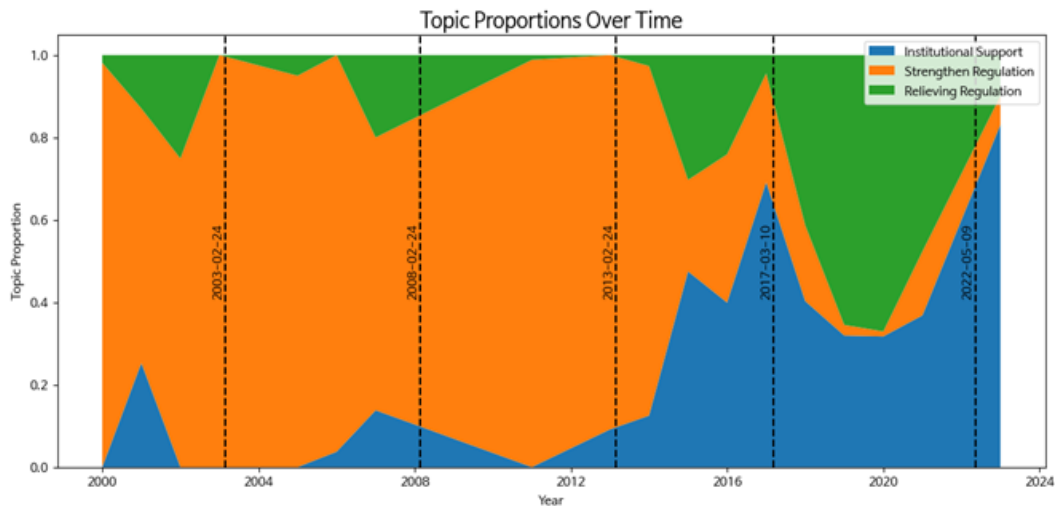


FIGURE 10 NMF Intertopic Distance Map for Topic 1

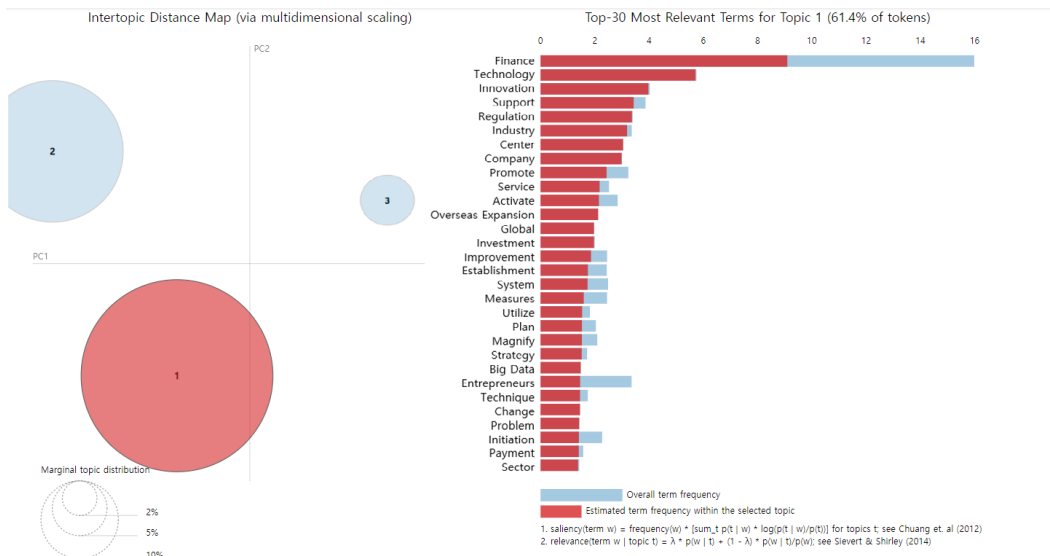


FIGURE 11 NMF Intertopic Distance Map for Topic 2

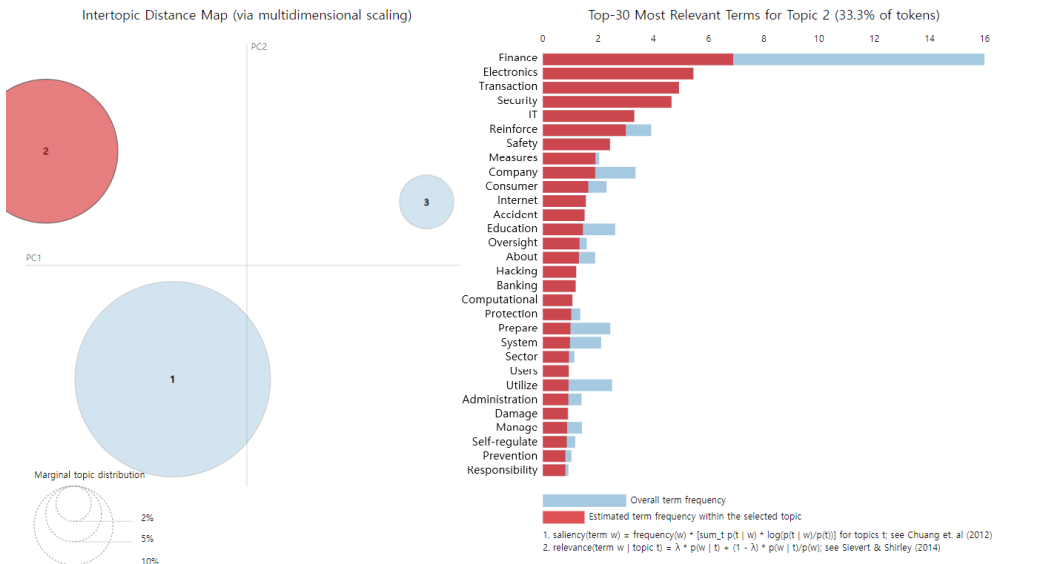


FIGURE 12 NMF Intertopic Distance Map for Topic 3

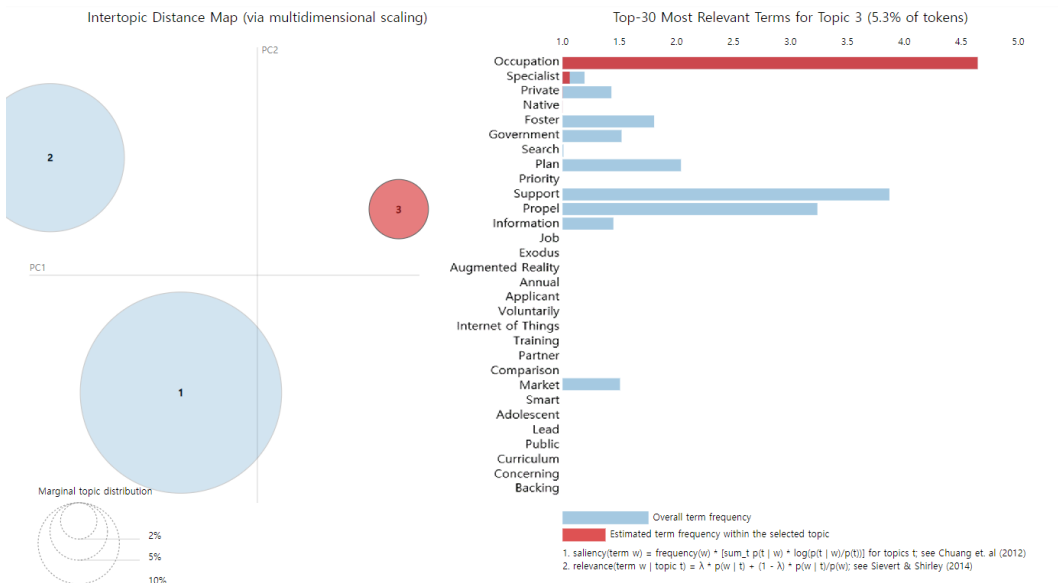


FIGURE 13 LDA Intertopic Distance Map for Topic 1

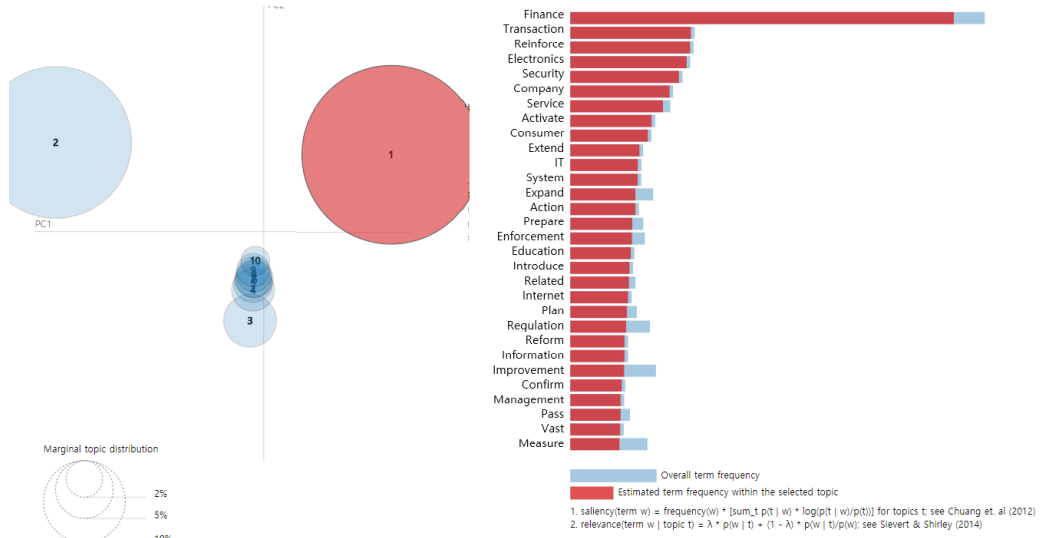


FIGURE 14 LDA Intertopic Distance Map for Topic 2

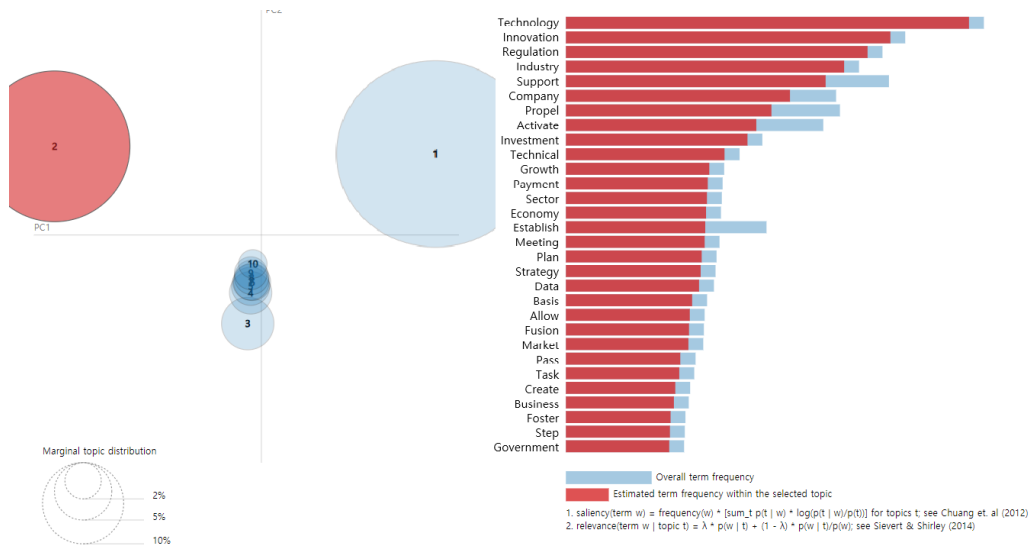


FIGURE 15 LDA Intertopic Distance Map for Topic 3

