



RIETI Policy Discussion Paper Series 23-P-008

# **Policy Agenda and Trajectory of the Xi Jinping Administration Textual Evidence from 2012 to 2022**

**LIM, Jaehwan**

Aoyama Gakuin University

**ITO, Asei**

RIETI

**ZHANG, Hongyong**

RIETI



Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry

<https://www.rieti.go.jp/en/>

Policy Agenda and Trajectory of the Xi Jinping Administration  
Textual Evidence from 2012 to 2022<sup>1</sup>

Jaehwan LIM (Aoyama Gakuin University)

Asei ITO (The University of Tokyo/RIETI)

Hongyong ZHANG (RIETI)

Abstract

How many agendas has Xi Jinping put forth and promoted since taking office in 2012, and what are the types of agendas? What is the relationship between the agendas? How much of political attention has each agenda received, and how has the allocation of attention changed over time? Moreover, how do we know this? Despite the scholarly interest in policy development during the Xi era, few studies have systematically mapped the overall structure – the number, substance, and underlying relationship – of policy agendas pursued by the Xi administration. To address this research gap, we utilize a dataset of presidential statements, speeches, and reports from 2012 to 2022 and employed automated text analysis to identify major topics and terms associated with each topic. Our analysis reveals the identification of about 25 distinct policy agendas across diverse policy domains, with remarkable temporal variations between agendas in terms of the amount of leadership attention. We find a significant shift in both the substance and relative weight of policy agendas between the first and second term of Xi’s tenure, indicating his adaptation and responses to changing domestic and foreign policy environments.

Keywords: policy agenda, political attention, structural topic model, Xi Jinping, China  
Abbreviations— CCP: the Chinese Communist Party; LDA: Latent Dirichlet Allocation;  
STM: Structural Topic Model  
JEL classification: P20, O53

The RIETI Policy Discussion Paper Series was created as part of RIETI research and aims to contribute to policy discussions in a timely fashion. The views expressed in these papers are solely those of the author(s), and neither represent those of the organization(s) to which the author(s) belong(s) nor the Research Institute of Economy, Trade and Industry.

---

<sup>1</sup> This study is conducted as a part of a Chinese economy project, “Empirical Research on the Changing Chinese Economy: Upgrading, Expansion, Structural Reform,” in “Studies on Transformations of International Systems and their impact on Japan’s mid & long-term competitiveness” project undertaken at the Research Institute of Economy, Trade, and Industry (RIETI). We thank Shujiro Urata, Yasuhiko Yoshida, Haruhiko Ando, and participants in The Japan Association for Asian Studies (JAAS), Society for Computational Social Science of Japan, and RIETI DP research seminars for their valuable comments and suggestions. LIM: [t13025@aoyamagakuin.jp](mailto:t13025@aoyamagakuin.jp) (corresponding author); ITO: [asei@iss.u-tokyo.ac.jp](mailto:asei@iss.u-tokyo.ac.jp); ZHANG: [zhang-hongyong@rieti.go.jp](mailto:zhang-hongyong@rieti.go.jp).

## 1. Introduction

There is a general agreement that Xi Jinping's rise to power as the supreme leader of China has led to significant changes in the policy-making process of the country. Of particular note is Xi's deep involvement in the policy process in which the party leader now engages in nearly every stage of policy development, from agenda setting to implementation and evaluation (Heilmann and Stefan 2016; Lampton 2015). Certain scholars argue that Xi has expanded his influence beyond the traditional areas of focus of the General Secretary, such as economic, and social policies, which were previously the purview of the premier and the State Council. This has earned him the nickname "The Chairman of Everything" (Shirk 2018; Economy 2018).

Despite the widely held perception of the transformative leadership of the new leader, observers increasingly disagree over whether or not Xi has made significant progress in transforming the policy-making process and policy outcomes. For instance, scholars have expressed skepticism about the state of policy implementation and suggested that the reform efforts by Xi may have exerted a limited impact on pre-established norms and practices on the ground (Pei 2022; Jaros and Tan 2019; Ang 2022). Additionally, in light of the lack of tangible accomplishments, doubts are being raised regarding Xi's ability and willingness to remain committed even to his *signature* policy initiatives (Zhu 2022; McGregor 2019). In summary, as signs of policy deadlock and agenda shifts emerge under Xi's continued rule, reassessing the relationship between Xi's characteristic leadership and policy outcomes is important.

This study aims to contribute to the discussion by analyzing policy outcomes during the Xi era, with a special attention to the identification of the evolving policy agendas and priorities of Xi. The underlying assumption is that Xi has adapted to changes in domestic and external conditions, which leads to shifts in policy substance, priority, and enforcement strategy. To achieve this goal, the study uses probabilistic topic modeling on a vast collection of presidential statements, speeches, and reports from 2012 to 2022. The ultimate objective is to construct a comprehensive and reliable set of policy agenda and to analyze the structural and temporal dimensions of agenda dynamics throughout Xi's entire reign.

Our analysis on the dynamics of agenda in the Xi era in particular and authoritarian regimes in general presents two contributions. Despite the increased interest in policy development during Xi's

reign, only a few studies have systematically analyzed the overall structure of the policy agenda of the Xi administration. To address this research gap, we provide a comprehensive overview of the policy agenda of Xi, including both revealed and latent policies, by identifying their number, substance, and underlying relationships. Our findings not only improve the understanding of Xi's evolving policy priorities, which is particularly important given his extended tenure, but also shed light on the policy outcomes of the resurgence of strongman politics in China.

After applying the topic modeling method, we have identified approximately twenty-five unique policy agendas that span various policy areas. Among the most frequently occurring topics are China's relations with developing countries and global economic governance. Structurally, our analysis finds that each policy agenda is nested within one of three broad clusters that are only loosely connected, namely, foreign policy, domestic policy, and crisis response. We also demonstrate that China's relations with developing countries, which is the most dominant agenda at the corpus level, is less significant in relational terms, as programmatic agendas, such as Socialism, and Civilization, become salient due to their strong connections with a wide range of topics. Lastly, the findings indicate a significant shift in the relative amount of attention given to each agenda between Xi's first and second terms, which is particularly pronounced in agendas related to implementation strategy.

The remainder of this paper is structured as follows. Section 2 introduces the relevant literature and how this study enhances the existing research. Section 3 outlines the data sources and methodology used in this study. Section 4 displays the model output and presents the main findings supported by evidence from text analysis. Section 5 discusses various approaches to validating the findings. Finally, Section 6 summarizes the findings and draws conclusions.

## **2. Relevant Literature: Authoritarian Agenda Dynamics and China Under Xi**

Theories of policy agenda have traditionally emphasized democratic systems and focused on processes that ensure that policy priorities remain responsive to stakeholder demands. However, applying these theories to authoritarian regimes poses two challenges. First, in contrast to democracies, authoritarian policy processes are insulated from stakeholder activities external to the establishment (Han 2022). Second, under authoritarian regimes, representation *within bounds* is a more common form of agenda entrepreneurship instead of direct competition embedded in a broad network of policy venue (Truex 2016). Despite the growing interest in authoritarian responses to

societal demands, the majority of scholars agree that autocrats possess significant freedom in deciding on the extent and the form of external preferences to incorporate into government policy (Chen et al. 2016; Lorentzen 2014). Furthermore, autocrats are generally skeptical of greater public involvement in the policy process, as asserted by Wintrobe (2000).

Regarding internal processes, a significant discourse exists regarding the interplay between policy priorities, specifically the choice between the provision of public or private goods and regime survival (Mesquita, Smith, & Siverson, 2003). However, the intense interest of autocrats in mitigating threats to the survival of their regime does not necessarily indicate that such threats alone dictate all government priorities determined by officials dispersed throughout the broader political system. Despite the strong emphasis on top-down control, bureaucratic deviation from central priorities can be relatively prevalent in authoritarian regimes. As a response, autocrats may occasionally employ drastic measures to suppress non-compliant activities, which may also result in changes in policy priorities.

How then does China under the leadership of Xi fit into the existing literature? The prevalent view is that the Chinese Communist Party (CCP) under Xi's rule has shifted toward more repression and less responsiveness in managing external demands, and has been more skeptical of incorporating societal preferences into the policy-making process (Dickson, 2021). Many observers agree that Xi has significantly strengthened his personal control over the entire party-state bureaucracy using various methods, including anti-corruption campaigns. As a result, a popular characterization of the past decade is the emergence of the *strongman rule* in which a single leader, often aided by a small group of loyal supporters, wields disproportionate power over the party-state system (Shirk, 2018).

The concentration of power in Xi's hands has significant implications for the policy process and agenda outcomes. Notably, it has led to remarkable erosion of collective decision-making norms, which enables Xi to make decisions more quickly and with less input from other leaders or interest groups (Shirk, 2023). Policymakers in the broader political system, who face intense demands for loyalty, are reluctant to pursue independent agendas and instead focus on a smaller set of priorities within the boundaries of political acceptability (Stern and Hassid, 2012). Consequently, Chan et al. (2021) argue that Xi's tenure is associated with more frequent changes in policy agendas, greater unevenness in the distribution of attention across priorities, and more alignment in agenda focus across policy venues than did Hu Jintao's era (Chan et al. 2021).

In addition, Xi's consolidation of power also has implications for the substance of policy agendas. Scholars have pointed out that Xi has pursued an *elitist* or particularistic type of policy that satisfies the relatively narrow set of interests of his allies and trusted institutions (Li, 2016; Lam, 2015). If the case is true, it suggests a significant departure from the Hu era, during which significant attention was given to redistributive agendas such as fiscal rebalancing between urban and rural regions and the establishment of a basic social safety net (Kroeber, 2016). Moreover, the perceived need to centralize political authority in response to economic slowdown has led to a renewed focus on the role of ideology as a unifying mechanism (Cheek, 2021). Additionally, previous studies suggest that national policy priorities have reflected more of the personal political agenda of the leader with further consolidation of his personal power, particularly after the 19th Party Congress (Schatz, 2009; Fewsmith, 2021).

Despite the broad and potentially significant theoretical implications of Xi's concentration of power and authority on policy outcomes, only a few systematic, evidence-driven approaches are used to identify and explain the entire spectrum of constituent elements and their relationships in the policy agenda in the Xi era. Instead, many studies focus on specific issue areas (e.g., economics, security, or social issues) or individual policy initiatives (e.g., the Belt and Road Initiative [BRI], anti-corruption campaign, or common prosperity). Although a few analyses employ a comprehensive focus, the majority rely on a deep yet selective reading of Xi's speeches and writings, which results in an impressionistic portrayal of policy priorities. For example, Pei (2022) argues that Xi's agenda consists of three main components, namely, establishing personal political dominance, revitalizing the Leninist party-state, and expanding China's global power and influence. However, this type of account lacks falsifiability and does not enable empirical testing through alternative observations.

Furthermore, the existing research fails to address the level of leadership attention given to each policy agenda and its evolution over time<sup>2</sup>. A grasp of these factors is critical for understanding variations in policy outcomes, particularly in the Chinese context (Lieberthal and Oksenberg 1988;

---

<sup>2</sup> Notable exceptions include Chan et al. (2021) and Qin and Huang (2023). In addition to using different collections of policy documents, their research differs from ours in two crucial aspects. Firstly, those studies primarily aim to examine whether China's policy process is characterized by punctuations and, if so, to what extent. Secondly, while they focus on the dynamics of elite bargains, their main object of observation is the policy process within the central bureaucracy, rather than the policy preferences and priorities of central leaders.

Wang 2008). Therefore, a more comprehensive and evidence-based approach is required to identify and explain the entire spectrum of constituent elements and their relationships in the policy agenda of the Xi era.

In summary, this brief literature review highlights three main points. First, the authoritarian agenda setting frequently reflects the process and outcome of elite bargains but does not suggest that non-elitist societal preferences are irrelevant to the agenda outcome. Second, reports emerge that the concentration of power by Xi has introduced new dynamics to Chinese policy-making processes. Third, scholars have not reached a consensus on the specific agenda outcomes of the Xi era.

This study aims to fill the gap in the literature by answering three questions: (1) How many policy agendas has Xi Jinping proposed and promoted since taking office in 2012, and what are these agendas? (2) How much attention has each agenda received, and how has the allocation of attention evolved over time? (3) What is the relationship among policy agendas, and how has it changed? Notably, the objective of the study is to discover a comprehensive set of policy agendas proposed and pursued by Xi using political texts to explore their temporal evolution and structure. Although theoretical implications are inevitable and will be discussed, the immediate goal is not to test a particular hypothesis derived from the theories of agenda setting and authoritarian policy-making.

### **3. Data and Method**

#### **3.1. Data**

We collected textual data from the *Xi Jinping Xilie Zhongyao Jianghua Shujuku* (习近平系列重要讲话数据库; hereafter, XIS), which is a database of important speeches made by Xi Jinping<sup>3</sup>. XIS is a web-based source that mainly collects speeches and reports related to Xi Jinping from Chinese state media, including the People's Daily (domestic and international editions), Xinhua News Agency, local newspapers, and several state, or CCP publications, such as Qiushi. The database contains speeches, activity reports, field visits, press conferences, meetings, telegrams (e.g., ceremonial), and other types of articles. The articles date back to November 15, 2012, which is the day Xi Jinping took office as the General Secretary of the CCP.

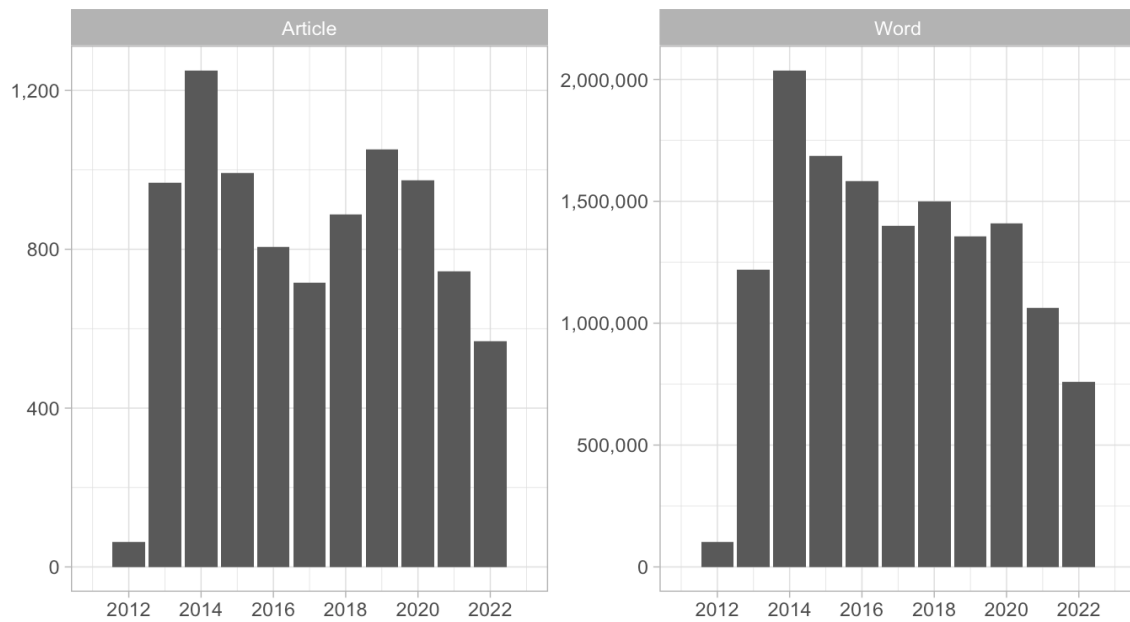
The pre-cleaning dataset included a total of 12,532 articles spanning from November 15, 2012, to

---

<sup>3</sup> The URL of the database: <http://jhsjk.people.cn/>

October 31, 2022, which comprises a total of 18,456,212 Chinese characters, with an average of 1,478 characters per article. To ensure the quality of the dataset, we eliminated duplicate articles, particularly those published by People’s Daily and Xinhua News Agency, and articles that were either very long (>30,000 characters) or too short (<200 characters). After cleaning, the final dataset for analysis consisted of 9,016 articles, which contain a total of 14,117,728 characters, with an average of 1,565 characters per article. Figure 0 provides a detailed summary of the dataset.

**Figure 0. Number of articles and words by year (2012~2022)**



Using these data poses advantages and limitations at the same time. One significant advantage is that the dataset contains a high concentration of policy statements made by Xi, allowing us to concentrate on his policy agenda and priorities, in contrast to collections of newspaper articles. However, the dataset's limitations arise from its direct compilation from articles published by state-owned media, particularly the People’s Daily and the Xinhua News Agency<sup>4</sup>. As a result, a possibility exists that the articles in the dataset were selected for propagandistic purposes, such as

---

<sup>4</sup> Previous research has demonstrated that media bias is more evident in news sources that are subject to strong state or party control. For example, using Chinese domestic newspaper data from 2000 to 2010, Piotroski et al. (2017) illustrates that news sources under the control of the central government or party organizations are more likely to display a favorable sentiment. Similarly, Qin et al. (2018) examines 117 newspapers across China to evaluate the degree of bias in each media outlet. The results indicate that the media at the central level tend to report more on highest political leaders, while topics emphasized by the anti-CCP media tended to be less frequent.



disseminating preferential messages for relevant organizations, instead of conveying true information about the policy priorities of the leader. Thus, Grimmer et al. (2021) stress that when using documents as a reflection of government inner workings, researchers should seek to understand the incentives that underlie document collection and the process through which the data are made available.

Nevertheless, given Xi's firm grip on the party-state system, it is unlikely that any organizational biases systematically influence document selection in ways that could conceal or distort the policy preferences of the leader. Notably, the policy priorities of a politician can frequently be inferred from their statements or speeches in meetings and congresses, which is particularly true in the highly controlled media environment of China. Furthermore, the current study focuses on policy priorities in the public domain, which must be effectively communicated to approximately 90 million CCP members and other stakeholders. Thus, assuming that media types, such as the propaganda apparatus of the Party, would not bias political priorities is reasonable, because it could be extremely misleading and costly given the magnitude of the potential audience.

### **3.2. Method**

Text has long been a crucial resource for social scientists, but advances in technology in recent years have rendered possible the analysis of vast amounts of text using quantitative methods. With the proliferation of the Internet, social scientists now have access to more data than ever before, and the research community has responded by creating user-friendly software packages. As a result, scholars are now able to draw on a wider range of data sources and use more diverse research methods (Wilkerson and Casas 2017).

To analyze our corpus of the XIS, this study utilizes topic modeling methods. This approach was selected, because topic models are well-suited to our analytic goals: identifying policy agendas and exploring their structural and temporal properties. Originally developed as an unsupervised machine learning algorithm for document classification and information retrieval (Blei et al. 2003), topic modeling not only saves significant labeling efforts but also enables researchers to summarize the underlying structure of large collections of documents. Notably, although topic modeling was initially intended for document classification, it has become a widely used tool in the social sciences for measuring latent concepts and generating new conceptualizations (Ying et al. 2021). Moreover, practitioners suggest that topic models are particularly effective for analyzing newspaper articles or

well-crafted statements from political or social elites (Grimmer et al. 2021), which makes them suitable for our objective.

In our topic modeling applications, we leverage a recently introduced framework, the structural topic model (STM). The STM is a mixed-membership model built on the popular latent Dirichlet allocation (LDA) model, which is particularly useful when documents likely contain multiple complex and overlapping topics.<sup>5</sup> In contrast to LDA, however, the STM provides a flexible approach for incorporating document metadata, such as author characteristics, writing location, and time of writing, into the analysis using document-level covariates (Lucas et al. 2015, Roberts et al. 2019). The inclusion of this information within the model can improve the quality of learned topics and facilitate hypothesis testing. Additionally, as Roberts et al. (2014) demonstrate in their application to open-ended survey responses, the STM is sensitive to the time period in which the text was created, which renders it advantageous for tracking changes in topic over time (Roberts et al. 2014, 17).

Prior to examining the applications of STM, an essential aspect to note is that unsupervised methods, such as topic modeling, require significant input from the researcher despite their name. The researcher must specify, among other things, the number of topics in advance and interpret their meanings. Furthermore, the preprocessing decisions made can exert a profound impact on the outputs of the topic model (Denny and Spirling 2018, Wilkerson and Casas 2017, Eshima et al. 2020). For instance, manipulating only one feature of a topic model (e.g., removing or retaining stop words) can produce vastly different results. Fundamentally, the agnostic perspective that no true, correct model exists for language analysis highlights the need for significant effort for devotion toward the validation of the model and outputs (Lucas et al. 2015, Quinn et al. 2010). This issue will be revisited in Section 5.

#### **4. Main Findings**

In this section, we conduct a detailed analysis of the dataset to present three sets of findings on the substance, structure, and temporal evolution of the policy agenda during the last decade under the Xi era.

---

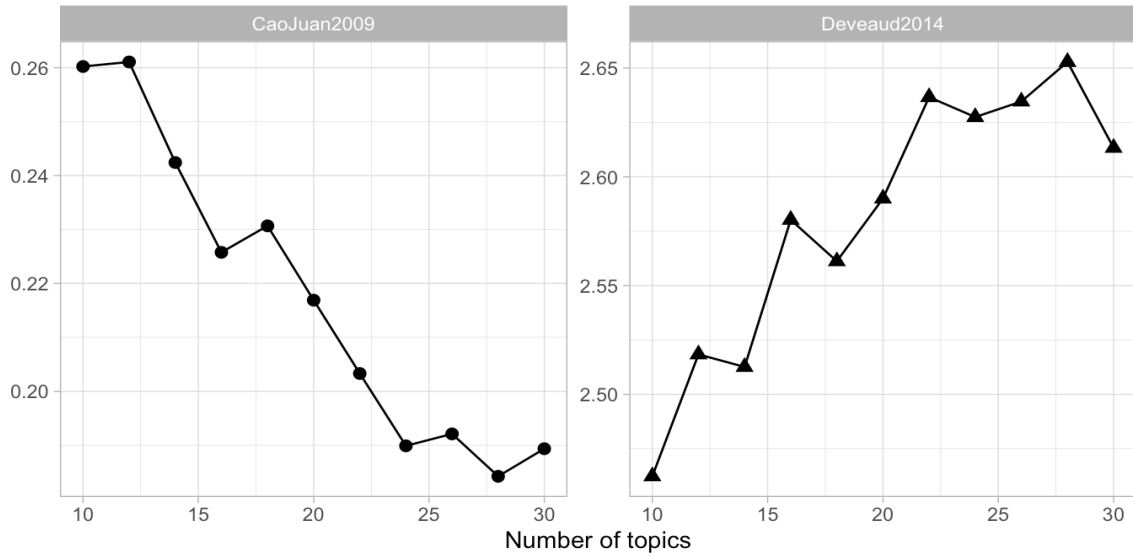
<sup>5</sup> Mixed membership models can provide more accurate and nuanced analysis of complex topics because they allow for a document to be associated with multiple topics, which facilitates generating the document-topic matrix and topic-word matrix simultaneously.

#### **4.1. Number and Substance of Policy Agendas**

As previously noted, topic modeling requires input from the researcher, including the specification of the number of topics, which is one of the most important and challenging tasks from the methodological standpoint. The optimal number of topics (parameter  $K$ ) is dependent on various factors. If  $K$  is very small, then the corpus will be divided into a few general semantic contexts, whereas if  $K$  is very large, then the collection will be divided into numerous topics, a few of which may overlap, such that topics may be difficult to interpret. No definitive answer exists to the appropriate number of topics for a given corpus, and human judgment is used to select the final model (Grimmer and Stewart 2013). Political scientists who have used topic models typically select the number of topics on the basis of whether or not the discovered topics are *substantively* meaningful (Grimmer 2010; Moser and Reeves 2015; Catalinac 2016). However, the existing software packages for topic modeling offer multiple options for measuring topic quality and assisting with model selection. Given that the quantities of interest to our study include the number of topics, a brief discussion of model selection is necessary.

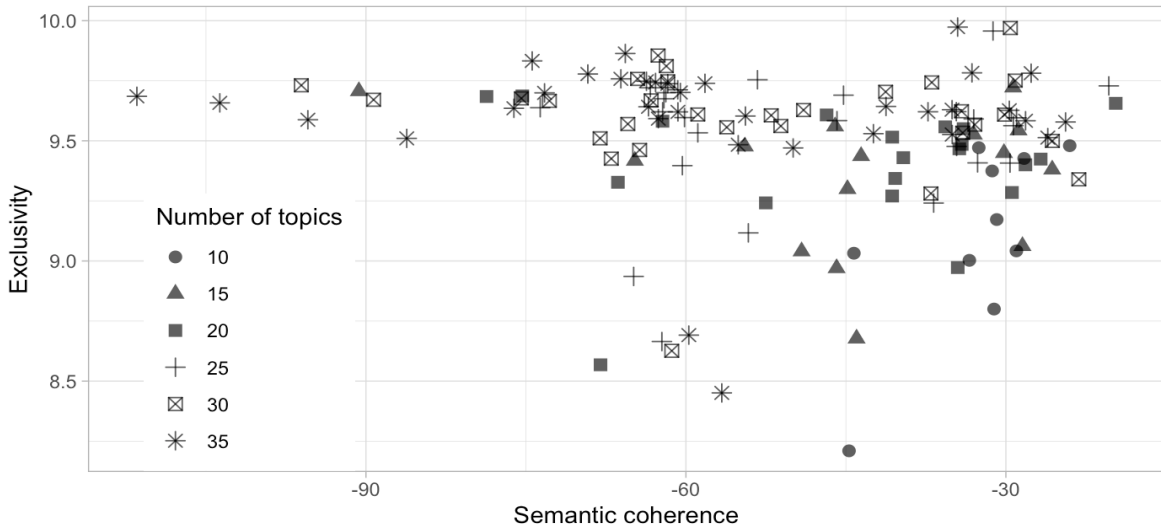
**Figure 1. Finding the optimal number of topics**

Notes: CaoJuan2009 computes similarities between pairs of topics in which the optimal number of topics is reached when the overall similarity (or distance) between topics achieves its minimum value, while Deveaud2014 calculates the information divergence between pairs of topics.



**Figure 2. Measuring topic quality**

Notes: Semantic coherence is maximized when the most probable words in a given topic co-occur together and is known to correlate well with human judgement of topic quality. Often times users will select a model with desirable properties in both dimensions (i.e., models with average scores towards the upper right side of the plot).



Figures 1 and 2 depict the results of the application of the existing built-in metrics to our cleaned text corpus and multiple model outputs, respectively, to assess the quality of topic structures. A good topic structure should consist of understandable, meaningful, and distinct semantic clusters, with each topic being mutually exclusive. Figure 1 illustrates that two metrics were used to calculate intra-topic density and inter-topic distance, which, when taken together, indicate that the

optimal number of topics falls in the range of 20 to 30. Another common strategy for evaluating topic quality is by combining semantic coherence and the exclusivity of words to topics (Roberts et al. 2019), as shown in Figure 2. The results demonstrate that models tend to perform well with 20 to 25 topics.

After applying various assessment metrics, we selected 25 topics that produce an acceptable tradeoff across the metrics and a relatively interpretable result. Table 1 presents the STM output in which each topic is labeled on the basis of the highest probability and the frequent and exclusive (FREX) words in each topic.<sup>6</sup> In a number of instances, we further examined exemplar documents to better understand a topic, that is, the documents that contain the highest proportion of words pertaining to the topic.

Our model generated a broad array of topics across diverse policy domains, except for a limited number of topics related to Xi's official activities as the head of the party and state, such as Foreign and Local Visits (topics 5 and 16), Press Conference (topic 13), and Congresses and Meetings (topic 17). This finding is notable, because it exceeds the original classification of policy areas of the CCP (Shao 2019).<sup>7</sup> To provide an overview of topical prevalence in the entire corpus, Figure 3 displays the expected proportion of each topic in the corpus. The figure shows that China and Developing Countries (topic 9) is the most prevalent theme in the corpus, which indicates that foreign economic policy issues, along with those pertaining to Global Economy and Governance (topic 3), were the major aspects of Xi's policy agenda.

Table 1 highlights an agenda that merits attention: Poverty Alleviation (topic 12). It is the only agenda in the model output that has a redistributive appeal. A close examination of the most frequently occurring words and FREX words suggests that the inclusion of this agenda may reflect the traditional emphasis of the CCP on the so-called agricultural issues instead of indicating a preference for Xi's personal policy priorities. However, this agenda received significantly more attention

---

<sup>6</sup> Frequent and exclusive (FREX) words combine the frequency of a term in the corpus with its exclusivity to the topic, as measured by the ratio of the frequency of the term within a topic to its frequency in the corpus as a whole (Roberts et al. 2019).

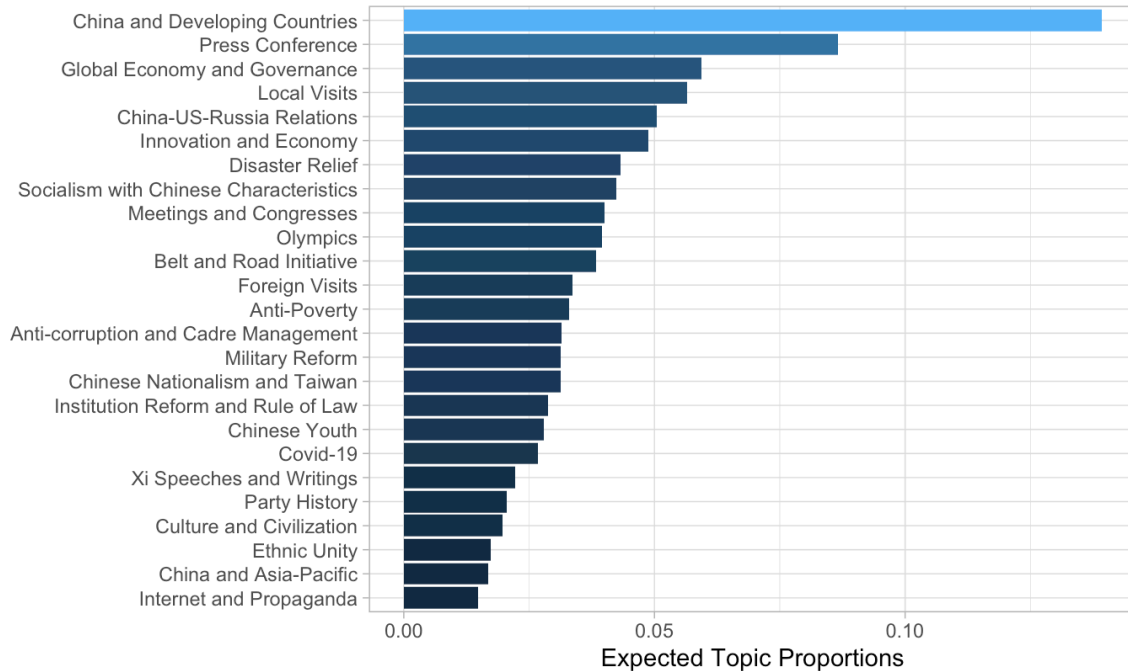
<sup>7</sup> The official categories are economics, politics, culture, society, ecology, party-building, military, and diplomacy.

**Table 1. Topics and Top Words in the Structural Topic Model**

#	Topic Labels	Top words
1	<b>Culture and Civilization</b>	Highest Prob: 文化, 中国, 文明, 历史, 教育, 发展, 世界 FREQ: 社会科学, 哲学, 中华文明, 文物, 文化遗产, 思政课, 学术
2	<b>Ethnic Unity</b>	Highest Prob: 工作, 发展, 民族, 群众, 坚持, 社会, 人民 FREQ: 西藏, 新疆, 民族团结, 妇女, 宗教, 各族群众, 工人阶级
3	<b>China and Asia-Pacific</b>	Highest Prob: 中国, 经济, 领导人, 亚太, 增长, 会议, 亚太经合组织 FREQ: 澳大利亚, 亚太, 亚太经合组织, 中澳, 马来西亚, 亚太地区, 新西兰
4	<b>China and Youth</b>	Highest Prob: 青年, 习近平, 中国, 广大, 时代, 希望, 实现 FREQ: 青春, 广大青年, 青年, 回信, 共青团, 老师, 人生
5	<b>Foreign Visits</b>	Highest Prob: 中国, 习近平, 主席, 访问, 友谊, 两国, 人民 FREQ: 比利时, 荷兰, 中法, 墨西哥, 驻华大使, 法国, 到访
6	<b>Rule of Law</b>	Highest Prob: 改革, 制度, 法治, 国家, 全面, 推进, 工作 FREQ: 法治, 司法, 立法, 宪法, 法律, 依法治国, 政法
7	<b>Covid-19</b>	Highest Prob: 疫情, 防控, 工作, 安全, 肺炎, 新冠, 加强 FREQ: 防控, 疫情, 肺炎, 新冠, 公共卫生, 抗疫, 物资
8	<b>Speech and Writing</b>	Highest Prob: 讲话, 习近平, 十八, 问题, 学习, 不能, 没有 FREQ: 论述, 中央文献出版社, 年版, 文献选编, 摘编, 讲话
9	<b>China and Developing Countries</b>	Highest Prob: 合作, 发展, 关系, 两国, 习近平, 中方, 中国 FREQ: 中非, 建交, 非洲, 传统友谊, 中方, 战略伙伴, 两国关系
10	<b>Socialism with Chinese Characteristics</b>	Highest Prob: 社会主义, 发展, 全面, 中国, 党, 坚持, 特色 FREQ: 特色, 社会主义, 现代化, 十九, 改革开放, 四个, 小康社会
11	<b>Innovation and Economy</b>	Highest Prob: 发展, 创新, 经济, 科技, 改革, 我国, 推动 FREQ: 供给, 科技, 驱动, 结构性, 侧, 结构, 实体
12	<b>Poverty Alleviation</b>	Highest Prob: 生态, 发展, 脱贫, 建设, 农村, 工作, 保护 FREQ: 脱贫, 扶贫, 生态环境, 乡村, 贫困人口, 攻坚, 生态
13	<b>Press Conference</b>	Highest Prob: 习近平, 主席, 国家, 记者, 摄, 出席, 举行 FREQ: 抵达, 摄, 记者, 机场, 兰红光, 古巴, 杜尚泽
14	<b>Anti-Corruption and Cadre management</b>	Highest Prob: 干部, 政治, 党, 问题, 工作, 领导, 建设 FREQ: 党内, 党风廉政, 纪律, 八项, 党章, 规矩, 从严治党
15	<b>China and Major Powers</b>	Highest Prob: 合作, 发展, 组织, 上海, 金砖, 共同, 双方 FREQ: 中美, 中拉, 成员国, 中美关系, 金砖, 上海, 俄
16	<b>Local Visits</b>	Highest Prob: 习近平, 总书记, 考察, 来到, 情况, 群众, 了解 FREQ: 正定, 乡亲们, 村民, 浙江, 福州, 察看, 村里
17	<b>Congresses and Meetings</b>	Highest Prob: 代表, 习近平, 会议, 委员, 工作, 大会, 李克强 FREQ: 全国人大常委会, 表决, 主席团, 赵乐际, 汪洋, 全国工商联, 最高人民检察院
18	<b>Party History</b>	Highest Prob: 党, 同志, 人民, 革命, 学习, 历史, 精神 FREQ: 初心, 长征, 共产党人, 党史, 红军, 革命, 毛泽东
19	<b>Military Modernization</b>	Highest Prob: 军队, 建设, 习近平, 部队, 强军, 主席, 中央军委 FREQ: 全军, 部队, 强军, 我军, 打仗, 军队, 官兵
20	<b>Olympics</b>	Highest Prob: 习近平, 北京, 中国, 香港, 澳门, 国家, 主席 FREQ: 冬奥会, 澳门, 冬, 香港, 残奥会, 筹办, 冰雪
21	<b>Internet and Propaganda</b>	Highest Prob: 工作, 文化, 人民, 互联网, 中国, 文艺, 宣传 FREQ: 文艺, 互联网, 舆论, 新闻, 网络空间, 文艺工作者, 网络安全
22	<b>Chinese Nationalism and Taiwan</b>	Highest Prob: 中国, 人民, 伟大, 中华民族, 发展, 实现, 历史 FREQ: 两岸, 两岸关系, 抗日战争, 台湾同胞, 同胞, 抗战, 近代

23	<b>Global Economy and Governance</b>	Highest Prob: 发展, 世界, 中国, 国际, 全球, 经济, 合作 FREX: 全球, 发展中国家, 联合国, 全球化, 气候变化, 复苏, 主义
24	<b>Disaster Relief</b>	Highest Prob: 习近平, 同志, 表示, 人民, 主席, 国家, 仪式 FREX: 遇难者, 哀悼, 防汛, 人员伤亡, 逝世, 受灾, 颁授
25	<b>Belt and Road Initiative</b>	Highest Prob: 合作, 中国, 发展, 一路, 一带, 亚洲, 建设 FREX: 亚洲, 东盟, 一带, 一路, 丝绸之路, 互联互通, 博览会

**Figure 3. Estimated Topic Proportions in the XIS**



during the second term of Xi’s tenure, as evidenced below (Figure 6), which suggests the presence of a different motivation behind its presentation.

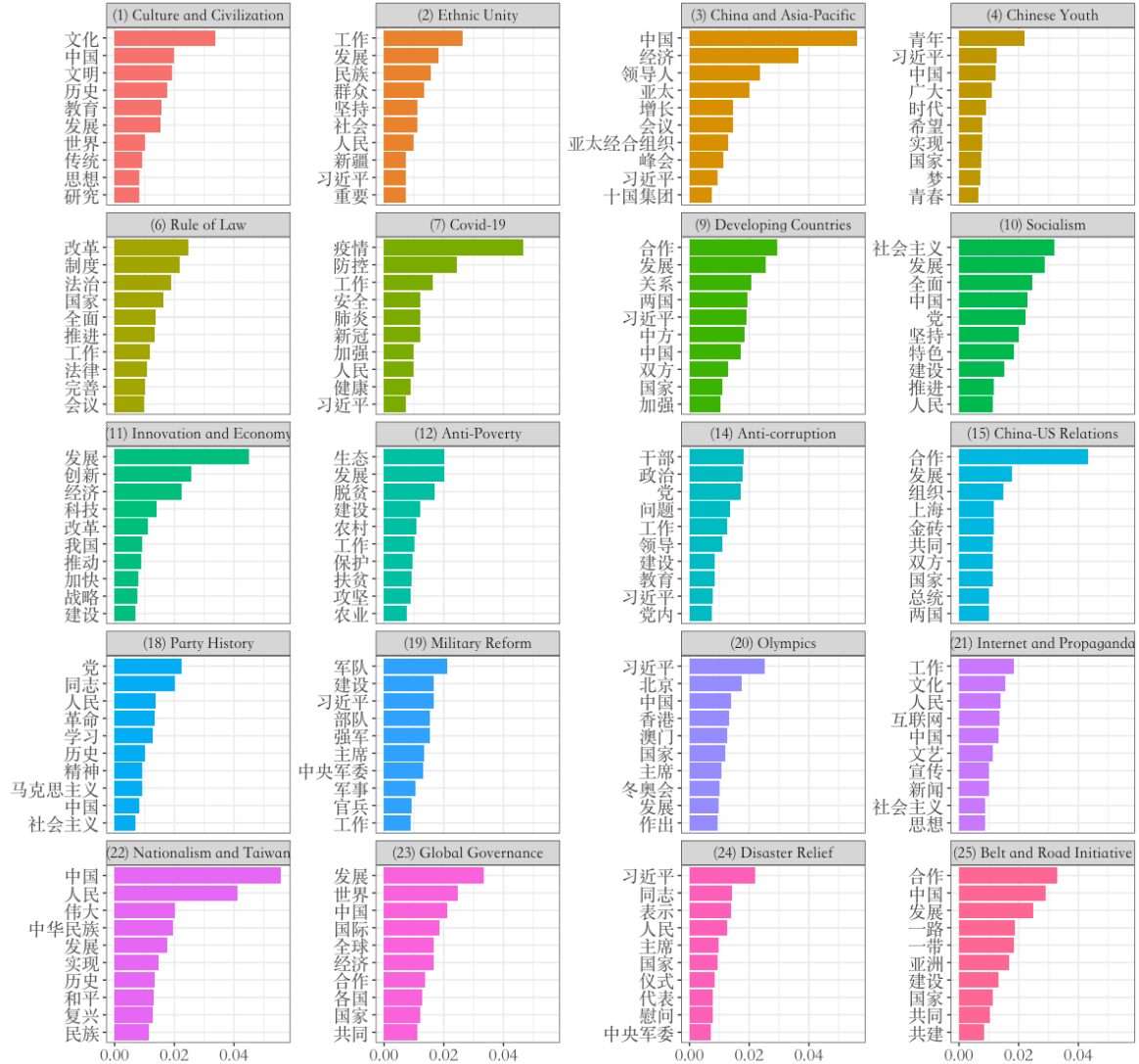
To gain a better understanding of the content of each policy agenda, Figure 4 displays the results of the calculation of the probability of each term being generated from a particular topic. For instance, Military Modernization (topic 19) generates not only general terms, such as “军队 military” or “军事 military affairs,” but also specific terms, such as “中央军委 Central Military Commission” or “强军 building a strong and capable military,” which indicates the central agenda-setter and major policy objectives. Notably, we observe a significant variation in the probability of the term “习近平 Xi Jinping” appearing as the most common word in each topic, which is of particular interest to our research. In general, foreign policy agendas are more likely to include this term than domestic policy agendas do, although variations can also be observed among foreign policy topics, such as those between China and Developing Countries (topic 9) and China and Major Powers (topic 15).

## 4.2. Correlations and Structures of Policy Agendas

In addition to the inclusion of covariates, another distinctive feature of the STM is the explicit

**Figure 4. Ten most probable words for each topic and expected proportions**

Notes: Topics 5, 8, 13, 16, and 17 are excluded



estimation of correlation between topics.<sup>8</sup> The model output allows the analyst to produce a network graph of topics in which each topic is a node (point), and two nodes are connected when they are highly likely to co-occur in a document, which offers insight into the organizational structure at the corpus level and enables researchers to identify large themes that transcend topics.

<sup>8</sup> Correlations are estimated by replacing the Dirichlet distribution in the standard LDA framework with a logistic normal distribution as in the correlated topic model (Blei and Lafferty 2007). When no covariates are specified, the STM reduces to an instance of the correlated topic model.



**Figure 5. The network of correlated topics for a 25-topic STM output.**

Notes: Node size is proportional to centrality and edge width is to the strength of the correlation between topics.

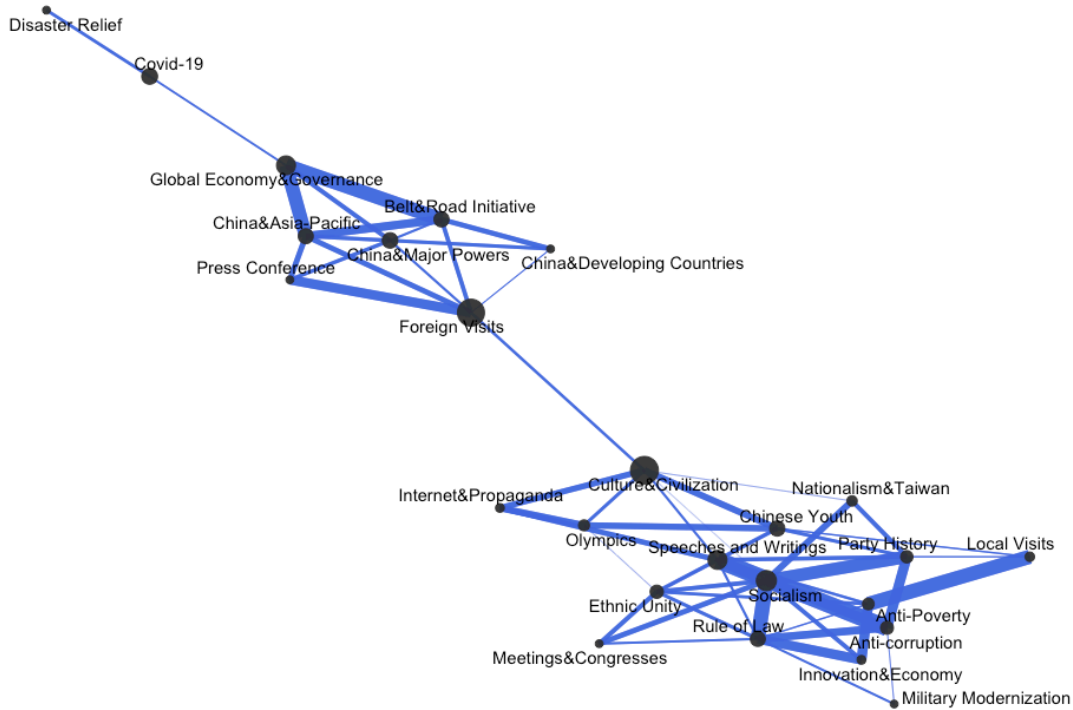
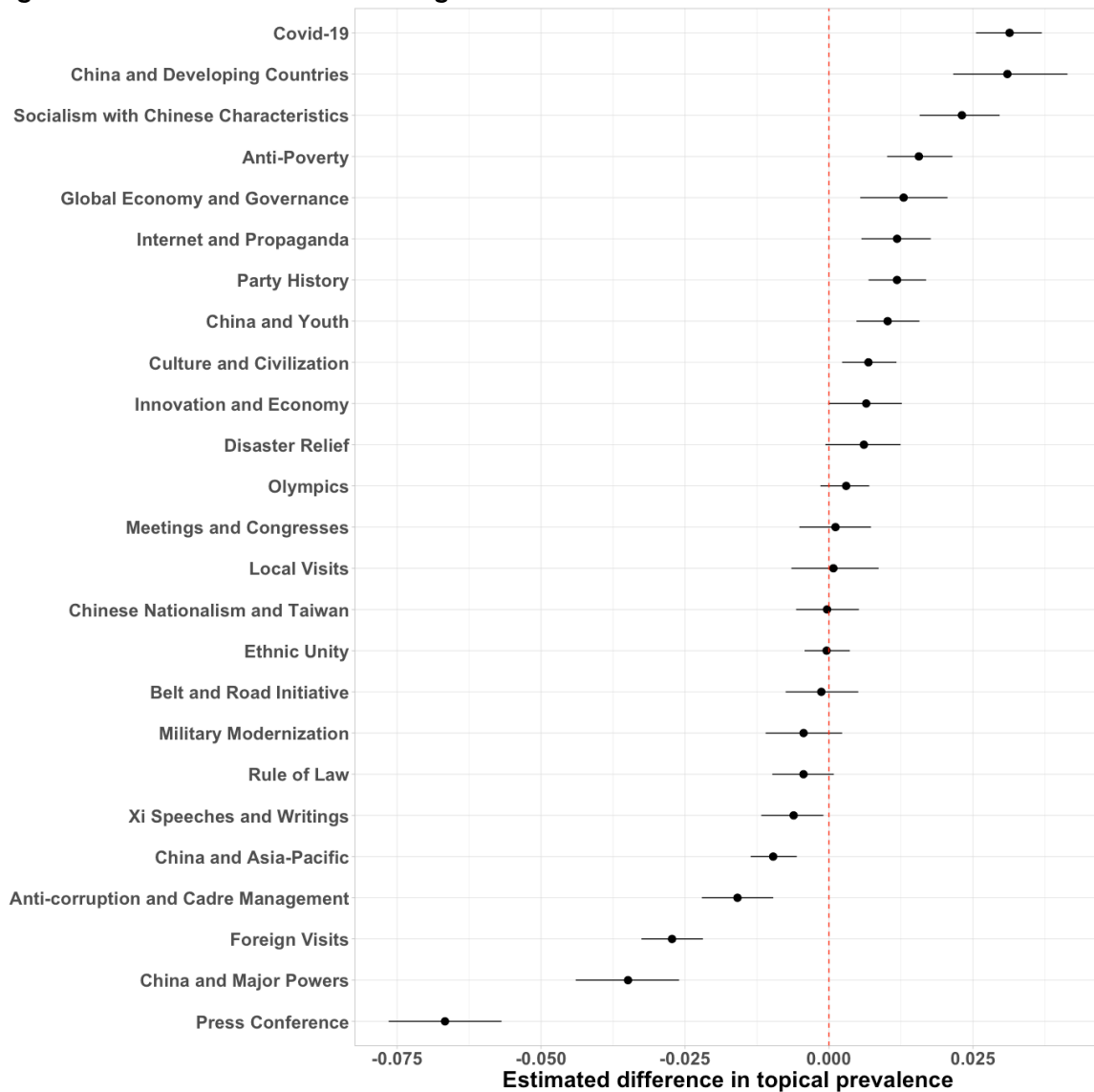


Figure 5 illustrates the structure of correlations among the discovered agendas. The size of each node represents betweenness centrality, a metric frequently used to capture the importance of nodes that act as intermediaries or bridges in a network, while the edge (line) width is proportional to the strength of correlation between the two connected topics. In general, the more strongly connected any two topics are (as measured by the thickness of the connected line), the more similar are their word use patterns are in Xi’s political text.

Our model has identified three broad policy clusters that are loosely connected to one another, namely, foreign policy, domestic policy, and crisis response. The foreign policy and domestic policy clusters are linked by connecting topics, such as Foreign Visits (topic 5) in the former and Culture and Civilization (topic 1) in the latter. A few topics in each cluster present obvious connections, such as the relationship between global economic governance topics (e.g., Global Economy and Governance and BRI) or between topics related to Xi’s ideological contribution (e.g.,

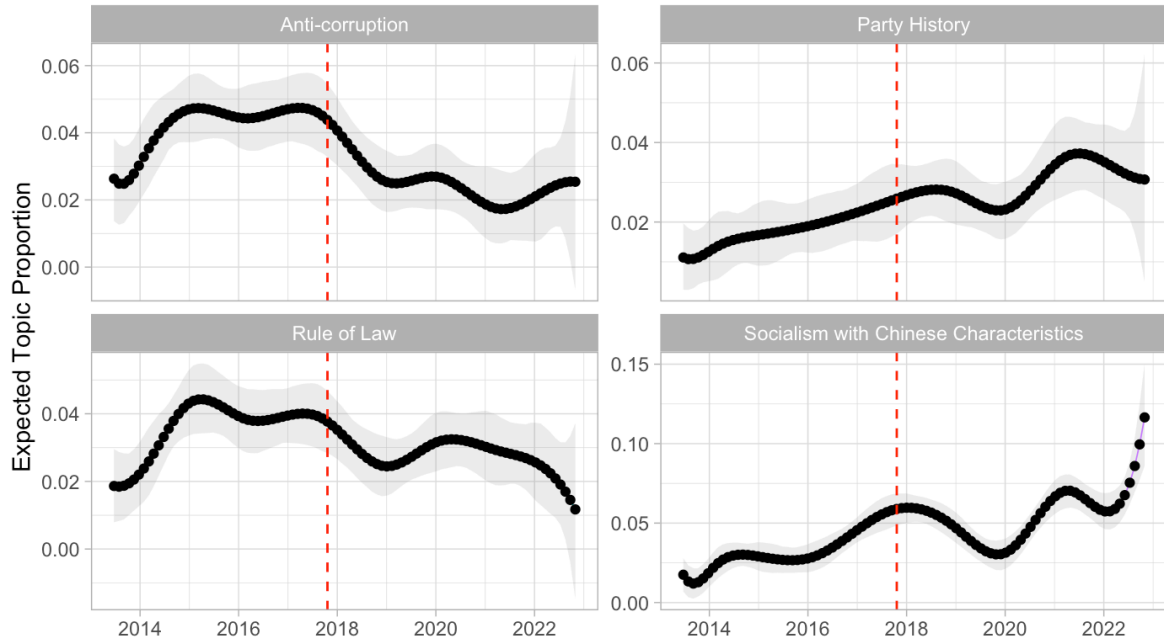
**Figure 6. Estimated difference in agenda outcomes between Xi's first and second term**



Socialism with Chinese Characteristics, and Speeches and Writings). However, certain connections (or the lack thereof) are subtle and require further investigation. For instance, Military Modernization is firmly situated in the domestic policy cluster and only linked to the topics of Anti-Corruption and Rule of Law, while Chinese Nationalism and Taiwan lack a connection to any foreign policy agendas.

Furthermore, it's worth noting that China and Developing Countries, which is the most prevalent topic at the corpus level, loses much of its significance in relational terms, which suggests that the

**Figure 7. Topical prevalence over time for a set of topics**



agenda might be located at a lower level in the latent agenda hierarchy. In contrast, programmatic agendas, such as Socialism, Culture and Civilization, and Global Governance, become structurally salient due to their strong connection with a wide range of topics. This finding not only aligns with our previous discussion on authoritarian agenda politics, in which the intense need of an autocrat for political survival does not solely dictate all policy priorities but also prompts further inquiry into the conditions under which this particular type of policy agendas emerges and its influence on the overall distribution of agendas.

#### **4.3. Temporal Evolution of Policy Agenda in the Xi Era**

As previously noted, STM distinguishes itself from other topic modeling methods, such as LDA, by allowing document-level covariates to be incorporated into the model as a means of pooling information. A covariate may influence either the prevalence or the content of a topic. Building on these advances, we introduced a time covariate into the model to explore the temporal dimension of the agenda outcomes during the Xi era and examined whether and how the weight of policy agendas in the XIS has shifted.

Figure 6 displays the estimated difference in topic prevalence for each topic between Xi's first (2012-2017) and second (2017-2022) terms. Among the top agendas that demonstrated a significant

increase in prevalence during Xi's second term are COVID-19 (topic 7), China and Developing Countries (topic 9), Socialism with Chinese Characteristics (topic 10), Poverty Alleviation (topic 12), and Global Economy and Governance (topic 23). In contrast, topics, such as Press Conference (topic 13), China and Major Powers (topic 15), Foreign Visits (topic 5), and Anti-Corruption (topic 14) have witnessed a significant decrease in estimated prevalence. Although these policy agendas suggest a possible discontinuity or agenda shift after Xi articulated and confirmed his intention to break the pre-existing party norm regarding leadership retirement, other issues also demonstrate significant continuity in policy development throughout Xi's tenure, including Military Modernization (topic 19), BRI (topic 25), Ethnic Unity (topic 2), Chinese Nationalism and Taiwan (topic 22), and Local Visits (topic 16), which have consistently received attention during this period.

One intriguing discovery is the change in the agenda within a group of categories that may fall under the rubric of enforcement strategy. In contrast to another group of agendas that focus on policy content, these agendas outline how the leader seeks to unify and cohere the bureaucracy to ensure effective policy implementation, which is a significant concern for the General Secretary. Figure 7 illustrates the trajectories of the four relevant agendas. It clearly demonstrates that although Xi emphasized an institutional approach to policy implementation during his first term, his second term was marked by a more ideological approach that seemingly placed greater emphasis on the sense of loyalty of individuals.

Additionally, the analysis provides insight into the effect of the COVID-19 outbreak on the distribution of Xi's policy attention. It reveals that, in general, the arrival of COVID-19 has diverted significant attention away from many but not all of the existing agendas. Specifically, we observe that the policy agendas that have experienced a similar surge in attention to COVID-19 include Global Economy and Governance and Innovation and Economy. This finding is largely consistent with the broad observation that the earlier response of the CCP to the pandemic emphasized the superiority of the Chinese system and the need to reform global institutions to align with the policy priorities of China (Huang 2021; Yang and Chen 2021).

However, the finding suggests that the adoption and strict adherence of the CCP to the Zero Covid-19 policy, which resulted in increased discontent among the population, has led to the increased reliance on upholding Xi's authority and symbolically justifying it through appeals to Chinese Culture and Civilization in which the youth is the main target of this effort (Wing-Chung 2022). This stance

is evident in the recent increase of attention to agendas such as Socialism with Chinese Characteristics, Culture and Civilization, and China and Youth. In summary, although COVID-19 has diverted attention from many policy agendas, it has also reinforced the emphasis of the CCP on the unique system and culture in China, which presents implications for the future of policy priorities.

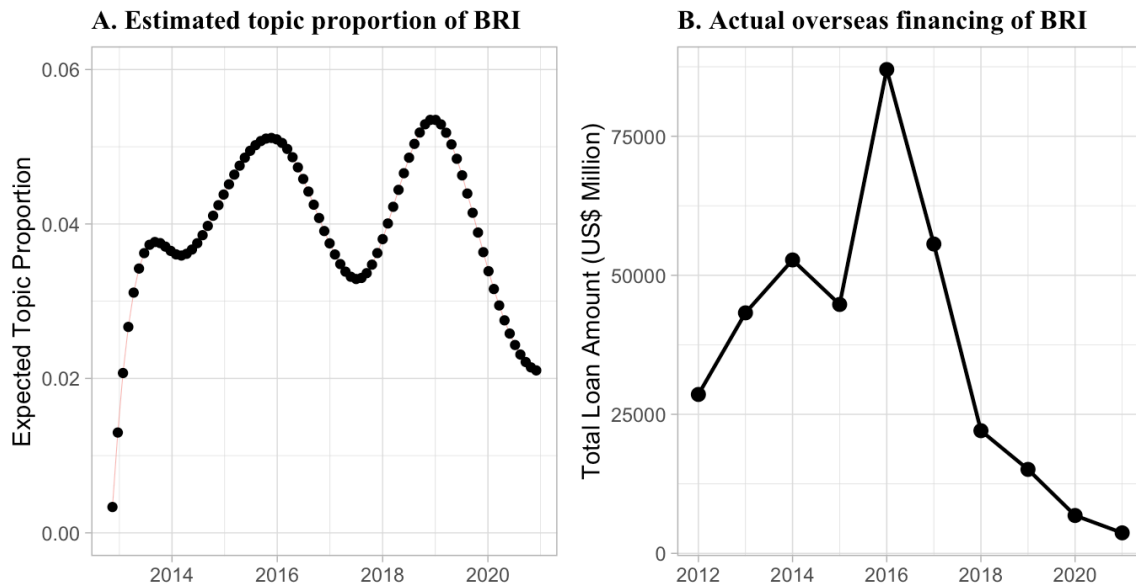
## **5. Validation: Tracking the Belt and Road Initiative and Anti-Corruption Campaign**

We have used the model fit from STM to measure the prevalence of various categories of policy agenda and sought to understand the relationship with one another and their evolution over time. To establish how many and which topics capture these categories, we *already* engaged in extensive validation. For example, in Section 4.1, we pursued several possible approaches for identifying the optimal number of topic categories by utilizing a set of evaluation metrics for topic quality. Another avenue for exploring the quality of topics involved the examination of the highest probability and exclusivity words assigned to the topic along with the documents that the model indicates are most closely associated with each topic.

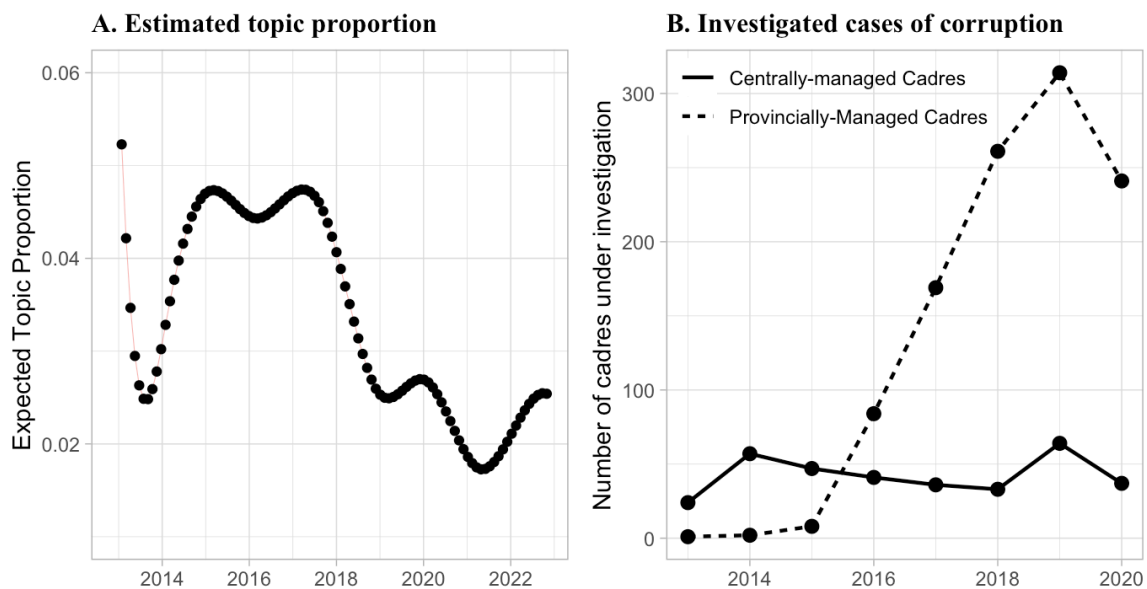
Notably, the validation conducted by the study differs from the commonly employed validation techniques in the machine learning literature, which tends to focus on the degree to which a model can predict unseen words in a document. For the objective of the study, it is not important that the model can predict unseen words – we have all the words! Instead, our validation is designed to ensure that the agenda categories learned from the XIS are not idiosyncratic or do not reflect a wildly different process than that examined by other relevant work.

With these caveats in mind, this section presents another approach for validating the reliability and usefulness of the discovered topics. Following the practices of Grimmer et al. (2013) and Catalinac (2016), we demonstrate that our measures align with well-known facts about policy development during the Xi’s era. This validation approach is designed to assess the notion that Quinn et al. (2010) refer to as “predictive validity,” which represents an expected correspondence between a

**Figure 8. Agenda setting and policy implementation for Belt and Road Initiative**  
 Plot B is based on China's Overseas Development Finance Database (2023).



**Figure 9. Agenda setting and policy implementation for Anti-corruption Campaign**  
 Plot B is based on Tiger's Corruption Dataset (Yang, Milanovic, and Lin 2023).



measure and exogenous events uninvolved in the measurement process.<sup>9</sup>

<sup>9</sup> However, as the authors caution, the correspondence need not be a pure forecast of events from measures, but can be concurrent or postdictive, and causality can run from events to measure (Quinn et al. 2010, 27).

Figure 8 illustrates the relationship between the topical prevalence of the BRI (topic 25) and China's distribution of overseas development finance, which we sourced from the Chinese Overseas Development Finance Database of Boston University.<sup>1 0</sup> The graph demonstrates a close correlation between the topical coverage of BRI and China's investment in BRI-related projects, particularly during Xi's first term. However, since 2019, the actual implementation of BRI has not aligned as closely with Xi's discursive patterns, which could be due to the outbreak of COVID-19. Nevertheless, the study observed a significant increase in the proportion of the Global Economy and Governance agenda since the pandemic (see the Appendix), which indicates that resources may have been redirected to other areas or that BRI has been integrated into a broader agenda (Ray 2023). Consequently, further investigation is required to determine whether or not changes can be observed in the language used to discuss BRI by examining the effect of the time covariate on topical content.

Additionally, Figure 9 illustrates the correlation between the topical prevalence of Anti-corruption (topic 14) in the XIS and the actual implementation of a nation-wide anti-corruption campaign.<sup>1 1</sup> The figure reveals that Xi's focus on anti-corruption aligns well with the progress of the anti-corruption campaign targeted at centrally-managed cadres. The number of central leaders under investigation peaked during the early years of Xi's leadership and has since steadily decreased. However, there exists a noticeable time delay between the leader's agenda setting and policy implementation at the provincial level, as evidenced by the number of provincially-managed cadres accused of corruption. Nevertheless, this time inconsistency does not indicate that the leader's agenda failed to attract the attention of provincial leaders; rather, they presumably required more time to assess the genuineness and sustainability of Xi's commitment to combatting corruption (Ding et al. 2020). Notably, despite the observed decrease in the topical proportion of Anti-corruption during the second term, Xi continued to exhibit rhetorical commitment to the agenda, which in turn explains the ongoing implementation of the campaign both at the central and provincial levels.

---

<sup>1 0</sup> <https://www.bu.edu/gdp/chinas-overseas-development-finance/>

<sup>1 1</sup> The data is sourced from Yang, Milanovic, and Lin 2023, which is derived from official information published by the Central Commission for Discipline Inspection (ccdi.gov.cn). Several studies have explored the anti-corruption campaign using datasets compiled from various sources, such as Gang and Yano (2017), Lorentzen and Lu (2018), and Aidt et al. (2020), and Wang and Dickson (2020). However, the data utilized in this study represents the most recent, and the only variant that encompasses the period of Xi's second term.

## 6. Conclusion

This study leveraged probabilistic topic modeling on a vast collection of presidential speeches and conducted qualitative and quantitative interpretations of the discovered topics. This process enabled the presentation of a comprehensive, structural, and dynamic overview of agenda outcomes during the entire period of Xi's leadership.

The analysis revealed approximately 25 distinct policy agendas, which exhibit significant temporal variations in terms of leadership attention. We estimated the difference in topical prevalence between the first and second terms of Xi's tenure, which indicates the shifting policy priorities and different enforcement strategies of the leader. Structurally, our model output implies that foreign and domestic agendas separately cluster. Agendas that pose implications for foreign affairs, such as the military and Taiwan, are discussed in the domestic context, which suggests that Xi appeals to international and domestic audiences with different themes and rhetoric. Additionally, we validated the reliability and usefulness of the model outputs and topic labeling by examining the relationship between the discursive patterns and the actual policy execution of the BRI and anti-corruption campaign, which is widely known as Xi's flagship policy initiatives.

The empirical findings of this study shed new light on previously overlooked aspects of the policy agenda of the Xi Jinping administration. While prior studies have produced valuable insights into Xi's key policy agendas, most of them have focused on the initial years of his tenure or specific policy domains, offering only a partial and fragmented account of agenda dynamics. Instead, this study aims to present a comprehensive and reliable list of policy agendas that covers the entire duration of Xi's leadership, while also seeking to investigate the evolving priorities associated with each agenda. For instance, the estimation results reveal that in addition to the three major agendas identified by Pei (2020) – establishing personal political dominance, revitalizing the party-state system, and expanding China's global power and influence – poverty alleviation and domestic innovation have also gained significant rhetorical attention, a tendency that has further intensified during Xi's second term.

From a theoretical perspective, the findings highlight the importance of examining the dynamics of authoritarian agendas. The analysis suggests that a dictator's personal accumulation of power and authority can potentially influence policy outcomes by shaping their interactions with regime elites.



This dynamic may lead to a shift from agenda items driven by particularistic goals to those characterized by programmatic or ideological connotations, particularly when the dictator further centralizes power. It is worth noting that assuming a fixed preference for a dictator is both challenging and undesirable since dictatorial preferences are prone to change, thereby impacting agenda outcomes.

## Bibliography

- Aidt, Toke S., Arye L. Hillman, and Qijun Liu. 2020. "Who Takes Bribes and How Much? Evidence from the China Corruption Conviction Databank." *World Development* Vol. 133: 1-12.
- Ang, Yuen Yuen. 2022. "How Resilient the CCP?" *Journal of Democracy* 33 (3): 77-91.
- Blei, David M. and John D. Lafferty. 2006. "Correlated Topic Models." Proceedings of Advances in Neural Information Processing Systems.
- Blei, David M., Andrew Y. Ng, and Micheal J. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993-1022.
- Cao, Juan, Tian Xia, Jintao Li, Yongdong Zhang, and Sheng Tang. 2009. "A Density-based Method for Adaptive LDA Model Selection." *Neurocomputing* 72: 1775-1781.
- Catalinac, Amy. 2016. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." *Journal of Politics* 78 (1): 1-18.
- Chan, Kwan Nok, Wai Feng Lam, and Shao Wei Chen. 2021. "Elite Bargains and Policy Priorities in Authoritarian Regimes: Agenda Setting in China under Xi Jinping and Hu Jintao." *Governance* 34: 837-854.
- Cheek, Timothy. 2021. "Xi Jinping's Counter Reformation: The Reassertion of Ideological Governance in Historical Perspective." *Journal of Contemporary China* 30 (132): 875-887.
- Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2016. "Sources of Authoritarian Responsiveness: A Field Experiment in China." *American Journal of Political Science* 60 (2): 383-400.
- Denny, Matthew J., and Arthur Spirling. 2018. "Text Processing for Unsupervised Learning: Why It Matters, When It Misleads, and What To Do About It." *Political Analysis* 26: 168-189.
- Deveaud, Romain, Eric Sanjuan, and Patrice Bellet. 2014. "Accurate and Effective Latent Concept Modeling for Ad Hoc Information Retrieval." *Document Numerique* 17: 61-84.
- Dickson, Bruce. 2021. *The Party and the People: Chinese Politics in the 21st Century*. Princeton: Princeton University Press.
- Ding, Haoyuan, Hanming Fang, Shu Lin, and Kang Shi. 2020. "Equilibrium Consequences of Corruption on Firms: Evidence from China's Anti-corruption Campaign." Working Paper 26656, National Bureau of Economic Research.
- Economy, Elizabeth. 2018. *The Third Revolution: Xi Jinping and the New Chinese State*. New York: Oxford University Press.
- Eshima, Shusei, Kousuke Imai, and Tomoya Sasaki. 2020. "Keywords Assisted Topic Models ." 1-92.
- Fewsmith, Joseph. 2021. *Rethinking Chinese Politics*. Cambridge: Cambridge University Press.
- Gang, Xu and Go Yano. 2017. "How Does Anti-Corruption Affect Corporate Innovation? Evidence from Recent Anti-corruption Efforts in China." *Journal of Comparative Economics* 45: 498-519.
- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24: 395-419.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21: 267-297.
- Grimmer, Justin. 2010. "A Bayesian Hierarchical Topic Model for Political Texts." *Political Analysis* 18 (1): 1-35.
- Han, Kangwook. 2022. "Political Budgetary Cycles in Autocratic Redistribution." *Comparative Political Studies* 55 (5): 727-756.
- Haug, Yanzhong. 2021. "The Myth of Authoritarian Superiority: China's Response to Covid-19 Revisited." *China Leadership Monitor* 68: 1-15.
- Heilmann, Sebastian, and Metthias Stepan, eds. 2016. *China's Core Executive: Leadership Styles, Structures, Processes under Xi Jinping*. Mercator Institute for China Studies.
- Jones, Bryan D., and Frank R. Baumgartner. 2005. *The Politics of Attention: How Government Prioritizes Problems*. Chicago: The University of Chicago Press.
- Kroeber, Arthur R. 2016. *China's Economy: What Everyone Needs to Know*. Oxford: Oxford University Press.
- Lam, Willy Wo-Lap. 2015. *Chinese Politics in the Era of Xi Jinping: Renaissance, Reform, or*

- Retrogression*. New York and London: Routledge.
- Lampton, David M. 2015. "Xi Jinping and the National Security Commission: Policy Coordination and Political Power." *Journal of Contemporary China* 24(95): 759-777.
- Li, Cheng. 2016. *Chinese Politics in the Xi Jinping Era: Reassessing Collective Leadership*. Washington D.C.: Brookings Institution Press.
- Lieberthal, Kenneth, and Michel Oksenberg. 1988. *Policy Making in China: Leaders, Structures, and Processes*. Princeton: Princeton University Press.
- Lorentzen, Peter and Xi Lu. 2018. "Personal Ties, Meritocracy, and China's Anti-Corruption Campaign." SSRN (Available at <http://dx.doi.org/10.2139/ssrn.2835841>)
- Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58(2): 402-414.
- Lucas, Christopher, Richard A. Nielsen, Margaret E. Roberts, Brandon M. Stewart, Alex Storer, and Dustin Tingley. 2015. "Computer-Assisted Text Analysis for Comparative Politics." *Political Analysis* 23: 254-277.
- McGregor, Richard. 2019. *Xi Jinping: Backlash*. Penguin.
- Mesquita, Bruce Bueno De, Alastair Smith, and Randolph M. Siverson. 2003. *The Logic of Political Survival*. Cambridge: MIT Press.
- Moser, Scott, and Andrew Reeves. 2015. "Taking the Leap: Voting, Rhetoric, and the Determinants of Electoral Reforms." *Legislative Studies Quarterly* 39 (4): 467-502.
- Pei, Minxin. 2022. "Xi Jinping's Political Agenda and Leadership: What Do We Know from His Decade in Power." *China Leadership Monitor* 73: 1-12.
- Piotroski, Joseph D., T. J. Wong, and Tianyu Zhang. 2017. "Political bias in corporate news: the role of conglomeration reform in China." *The Journal of Law and Economics*, 60(1), 173–207.
- Qin, Bei, David Strömberg, and Yanhui Wu. 2018. "Media bias in China." *American Economic Review*, 108(9), 2442–2476.
- Qin, Xiaolei, and Jing Huang. 2023. "Policy Punctuations and Agenda Diversity in China: A National Level Analysis from 1980 to 2019." *Policy Studies*.
- Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael S. Crespin, and Dragomir, R. Radev. 2010. "How to Analyze Political Attention with Minimal Assumptions and Costs." *American Journal of Political Science* 54 (1): 209-228.
- Ray, Rebecca. 2023. "'Small is Beautiful' A New Era in China's Overseas Finance?" Global Policy Brief 17, Boston University Global Development Policy Center.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "STM: An R Package for Structural Topic Models." *Journal of Statistical Software* 91 (2): 1-40.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. "stm: R Package for Structural Topic Models." *Journal of Statistical Software* 91 (2): 1-40.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 773-1091.
- Schatz, Edward. 2009. "The Soft Authoritarian Tool Kit: Agenda-Setting Power in Kazakhstan and Kyrgyzstan." *Comparative Politics* 41 (2): 203-222.
- Shao, Hsuan-Lei. 2019. "Machine Learning: An Application of Text Mining to Xi's Grand External Propaganda Strategy." *Zhongguo Dalu Yanjiu* 62 (4): 133-157.
- Shirk, Susan L. 2018. "The Return to Personalistic Rule." *Journal of Democracy* 29 (2): 22-36.
- Shirk, Susan L. 2023. *Overreach: How China Derailed Its Peaceful Rise*. New York: Oxford University Press.
- Stern, Rachel E., and Jonathan Hassid . 2012. "Amplifying Silence: Uncertainty and Control Parables in China." *Comparative Political Studies* 45 (10): 1230-1254.
- Truex, Rory. 2016. *Making Autocracy Work: Representation and Responsiveness in Modern China*. Cambridge : Cambridge University Press.
- Wang, Shaoguang. 2008. "Changing Models of China's Policy Agenda Setting." *Modern China* 34 (1): 56-87.

- Wang, Yuhua and Bruce J. Dickson. 2022. "How Corruption Investigation Undermine Regime Support: Evidence from China." *Political Science Research and Method* 10: 33-48.
- Wilkerson, John, and Andreu Casas. 2017. "Large-scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20: 529-544.
- Wing-Chung, Ho. 2022. "The Surge of Nationalist Sentiment among Chinese Youth during the COVID-19 Pandemic." *China: An International Journal*, 20(4): 1-22.
- Wintrobe, Ronald. 2000. *The Political Economy of Dictatorship*. Cambridge: Cambridge University Press.
- Yang, Yifan and Xuechen Chen. 2021. "Globalism or Nationalism? The Paradox of Chinese Official Discourse in the Context of the Covid-19 Outbreak." *Chinese Journal of Political Science* 26: 89-123.
- Yang, Li, Branco Milanovic, and Yaoqi Lin. 2023. "Anti-corruption Campaign in China: An Empirical Investigation." World Inequality Lab, Working Paper.
- Ying, Luwei, Jacob M. Montgomery, and Brandon M. Stewart. 2021. "Topics, Concepts, and Measurement: A Crowdsourced Procedure for Validating Topics as Measures." *Political Analysis* 30 (4).
- Zhu, Jiangnan. 2022. "The Rise and Fall of Ruling Oligarchs: Fighting Political Corruption in China." *The China Review* 22 (2): 49-79.
- Zhu, Tingxuan. 2023. "The Not So Humble Brag: Foreign Aid Propaganda and Donor Image." (Working Paper).

## Appendix: Topical prevalence over time (for all topics)

