Maoist Return or Continuation of Reform? A Structural Topic Modelling Analysis of Chinese Government Reports and Party Communiques.*

Dianyi Yang ID[1]

02 August 2024

Abstract

Motivated by the wide-spread speculation of a Maoist return under Xi Jinping, this project applies Structural Topic Modelling to the reports on the work of the government and communiques of the plenary sessions of the central committee of the CPC to identify the underlying themes and compare them vertically across different eras of Chinese leadership, and horizontally between the party and government. This project finds no evidence of a realignment with the Mao era in the party communiques or government reports under Xi Jinping. Instead, the reports under Xi Jinping are found to be more similar to those under his immediate predecessors. However, this project documents a divergence in the style of the party communiques and government reports under Xi Jinping, with the former becoming more abstract and ideological, and the latter becoming more concrete and policy-oriented. One source of this divergence is the downplaying of the five-year plans in the party communiques and government reports, which used to be the one of the overlapping themes of both documents.

Keywords — China, Quantitative Text Analysis, Political Ideology

[1] London School of Economics and Political Science

---

1. Introduction

As the largest existing Marxist-Leninist state, China's politics is often described as a "black box" for its opaqueness (Chen, Lu, and Wu 2023). As a result, researchers rely on official documents to understand the Chinese government's policy priorities and ideological shifts. Among them, the reports on the work of the government and communiques of the plenary sessions of the central committee of the Communist Party of China (CPC) are of great importance. The former are delivered by the Premier of the State Council at the annual meeting of Chinas legislature - the National People's Congress (NPC) and outline the government's achievements and goals for the coming year. The latter are issued after the plenary sessions of the central committee of the Party and summarise the decisions made during the meeting, which often set the tone for not only the government's policy agenda, but also every other aspect of Chinese society.

Nevertheless, Jiang (2021) points out that much of the existing literature merely translates theses government reports and rarely applies quantitative text analysis techniques to them. This project attempts to fill this gap by applying structural topic modelling to these reports to identify the underlying themes and compare them vertically across different eras of Chinese leadership, and horizontally between the party and government. Specifically, this project compares the reports under Xi Jinping's leadership with those under his predecessors to identify any significant differences in policy priorities and ideological emphasis on lexical and semantic grounds. Such differences between different stages of Xi's leadership are also documented.

This project finds no evidence of a realignment with the Mao era in the party communiques or government reports under Xi Jinping. Instead, the reports under Xi Jinping are more similar to those under his immediate predecessors. However, this project documents a divergence in the style of the party communiques and government reports under Xi Jinping, with the former becoming more abstract and ideological, and the latter becoming more concrete and policy-focused. One important factor in this divergence is the downplaying of the five-year plans in party communiques and government reports under Xi Jinping, which had been an important overlap between previous party communiques and government reports.

2. Motivation

The current Chinese leader, Xi Jinping, is often compared to Chairman Mao Zedong for his power centralisation, ambitions, and authority (The Economist 2023). Despite scholarly opposition to this claim (Matson 2022; Milanovic 2023; The Economist 2023), there is yet empirical evidence

for or against it. Moreover, Xi Jinping is said to have put an end to the export-driven and market-led "Reform and Opening-up" era in favour of a "New Era" of a more internal and state-focused model of economic development (Hsu 2023). Nevertheless, Xi himself stressed the importance of deepening reform and expanding *high-standard* opening-up recently (CPPCC 2023). Gaining empirical and quantitative evidence comparing Xi and his predecessors is crucial for understanding the current Chinese political landscape and forecasting its future.

This project focuses on two aspects in which the attitudes of and toward leaders have profound implications - the State Council (central people's government) and plenary sessions of the central committee of the Chinese Communist Party. This combination of the two aspects is intriguing as the State Council historically enjoyed a significant degree of autonomy from the party, albeit much less recently (Horsley 2023). This could be reflected in the government reports, which may not have necessarily echoed the prevailing ideological trends in the party.

This project compares government and party reports under Xi Jinping's leadership with those under his predecessors to identify any significant differences in policy priorities and ideological emphasis on lexical and semantic grounds. Specifically, such differences between different stages of Xi's leadership are also documented.

Unlike previous attempts to quantitatively analyse these government reports (Jiang 2021), this project goes beyond term frequency analysis and applies topic modelling to identify the underlying *themes* in these reports. Moreover, Jiang (2021) only analyses government reports from 2001 to 2020, this project expands the time span of the corpus to include those from 1954 to 2024. This allows for comparison across the Maoist (1954-1978), Reform (1979-2012) and the Xi (2013-2024) eras. Furthermore, this project also includes party reports to compare the government's policy priorities with those of the party, which could potentially be more insightful.

3. Corpus

3.1. Reports on the Work of the Government

The first part of the corpus consists of the reports on the work of the government delivered by the Premier of the State Council at the annual meeting of the National People's Congress (NPC) from 1954 to 2024, with discontinuities between 1961-1963, 1965-1974 and 1976-1977 due to political turmoil ($N = 56$).

The reports from 1954 to 2017 are sourced from a dropbox link in a Chinese text analysis example by Haiyan Wang (n.d.) for Quanteda. The reports from 2018 to 2024 are manually

collected from the Chinese government's official website. Here is an example link to the 2024 report.

3.2. Communiques of the Plenary Sessions of the Central Committee of the Communist Party of China

The second part of the corpus consists of the communiques of the plenary sessions of the central committee of the Communist Party of China from 1958 to 2023, with discontinuities in/between 1960, 1963-1965, 1967, 1969, 1961-1976, 1981-1987, 1999 due to no communiques produced, no plenary sessions held or communiques only containing personnel changes and no political information ($N = 46$).

All the communiques are sourced from the *Database of all national congresses of the Communist Party of China*, except for the 2023 communique. Here is an example link to the 2022 communique. The 2023 communique is sourced from the government website.

3.3. Summary Statistics

Table 1: Summary of the corpus

Year	Leader/Premier	Report by	Characters	Sentences	Tokens	Types
1954	周恩来 (Zhou Enlai)	gov	23609	453	14089	2231
1955	李富春 (Li Fuchun)	gov	57608	981	35120	3054
1956	李先念 (Li Xiannian)	gov	18966	347	10783	1867
1957	周恩来 (Zhou Enlai)	gov	35291	704	21391	2597
1958	薄一波 (Bo Yibo)	gov	24646	412	15160	2185
1959	周恩来 (Zhou Enlai)	gov	31626	577	19210	2513
1960	谭震林 (Tan Zhenlin)	gov	10039	164	6266	1302
1964	周恩来 (Zhou Enlai)	gov	19705	387	11668	1890
1975	周恩来 (Zhou Enlai)	gov	5417	125	3182	967
1978	华国锋 (Hua Guofeng)	gov	31672	659	19117	2965
1979	华国锋 (Hua Guofeng)	gov	28774	505	17330	2657
1980	姚依林 (Yao Yilin)	gov	13185	300	7789	1585
1981-1987	赵紫阳 (Zhao Ziyang)	gov	169989	3287	102854	5704
1988-1998	李鹏 (Li Peng)	gov	231534	5306	137708	6010
1999-2003	朱镕基 (Zhu Rongji)	gov	88210	2473	52350	3972
2004-2013	温家宝 (Wen Jiabao)	gov	202007	5299	118184	5359
2014-2023	李克强 (Li Keqiang)	gov	165580	4758	101347	5445
2024-2024	李强 (Li Qiang)	gov	17365	497	10694	2167
1958-1970	毛泽东 (Mao Zedong)	party	25779	463	15158	1945
1977-1977	华国锋 (Hua Guofeng)	party	3618	61	2166	584
1978-1989	邓小平 (Deng Xiaoping)	party	18191	293	10885	1882
1990-2002	江泽民 (Jiang Zemin)	party	30279	534	18016	1957
2003-2012	胡锦涛 (Hu Jintao)	party	38692	494	22720	2091
2013-2023	习近平 (Xi Jinping)	party	53695	678	32499	2735
1954-2024	Total	Total	1345477	29757	805686	13959

Table 1 provides summary statistics of the corpus. The corpus contains 102 reports from 1954 to 2024, with a total of 1,345,477 characters, 29,757 sentences, 805,686 tokens, and 13,959 types. The reports have been delivered by 18 leaders or premiers in total.

In terms of preprocessing, I removed punctuations, numbers (in Arabic numerals), symbols,

Table 2: Summary of the selected collocations

collocation	English	count	count_nested	length	lambda	z
社会主义	Socialism	1855	0	2	6.103365	142.02007
现代化	Modernisation	688	0	2	7.161395	92.07305
五年计划	Five-year plan	340	0	2	5.839201	78.61871
调控	Regulation and control	313	0	2	9.231175	71.20011
对外开放	Opening-up	206	0	2	6.797478	67.60806
基础设施	Infrastructure	257	0	2	7.103202	66.74228
改革开放	Reform and opening-up	306	0	2	5.044230	66.57277
精神文明	Spiritual civilisation	180	0	2	7.673868	63.59433
资产阶级	Bourgeoisie	158	0	2	7.678020	62.41967
党中央	Party Central Committee	181	0	2	5.825022	62.18833
资本主义	Capitalism	217	0	2	5.638804	56.51094
合作社	Cooperative	142	0	2	6.889551	54.09960
共产党	Communist Party	229	0	2	8.678416	52.27812
工人阶级	Working Class	73	0	2	7.106307	45.85869

urls, and stopwords in Chinese. I also padded the tokens with spaces to avoid the concatenation of words. I used the `textstat_collocations()` function to find collocations in the corpus. I selected some political collocations from the list and manually added some collocations as compound tokens for better analysis. Table 2 shows the selected collocations, their t-statistic results and their English translations. Table 3 shows the manually added compound tokens, their English translations and the reasons for inclusion.

Table 3: Manually added collocations for compound tokens

Collocation	English	Reason
习近平	Xi Jinping	Name of leader
江泽民	Jiang Zemin	Name of leader
中国特色社会主义	Socialism with Chinese Characteristics	Political concept
伟大复兴	Great Rejuvenation	Political slogan under Xi
深化改革	Deepen reform	Differs from reform and opening-up
共同富裕	Common prosperity	Political slogan under Xi

Collocation	English	Reason
中国式现代化	Chinese style modernization	Political slogan under Xi
中国梦	Chinese dream	Political slogan under Xi
新质生产力	New productivity	Political slogan under Xi
科学发展观	Scientific outlook on development	Political slogan under Hu
三个代表	Three representatives	Political slogan under Jiang
百家争鸣	A hundred schools of thought contend	Political slogan under Mao

4. Methods

4.1. Structural Topic Model

The structural topic model was introduced by Roberts et al. (2014) to include prevalance and content covariates into the traditional correlated topic model by Blei and Lafferty (2007). Unlike the earlier Latent Dirichlet Allocation (LDA) by Blei, Ng, and Jordan (2003), structural and correlated topic models allow for correlations between topics, which is more realistic and leads to better fit. Nevertheless, they retain the ability to allow multiple topics in a document and same words in multiple topics, similar to the LDA.

I first searched for the optimal number of topics (K) using a quantitative-metric-assisted approach combined with human judgement. Figure 1 presents the quantitative metrics of the K-search by the `stm` package. It can be seen that $K = 25$ has relatively high held-out likelihood and close-to-zero residuals without much compromise on semantic coherence. $K = 25$ also produces the most sensible and exclusive topics by human judgement.

4.2. Correspondence Analysis

As a robustness check, I use the Correspondence Analysis (CA) by Nenadic and Greenacre (2007) as a non-parametric alternative to compare the trends in Chinese government reports and party communiques over the past decades. The Correspondence Analysis performs singular value decomposition on the normalised document-feature matrix to project the matrix into lower dimensions. In this project, I reduce the matrix to 2 dimensions for the best visualisation.
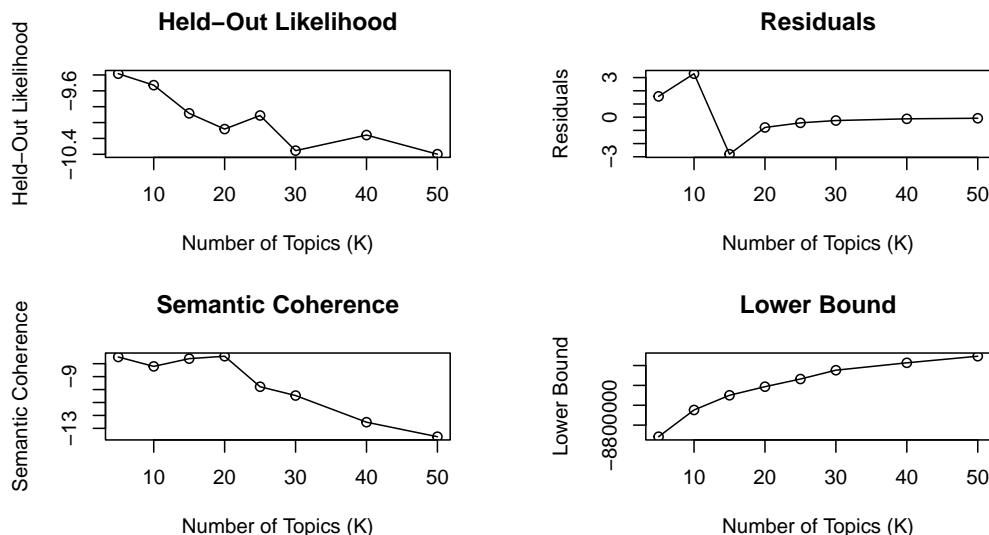
Figure 1: Result of the quantitative metric K-search

There are three reasons for using the Correspondence Analysis rather than the wordfish by Slapin and Proksch (2008). Firstly, the Correspondence Analysis allows us to keep more than one dimension, unlike the wordfish. Secondly, the Correspondence Analysis is much faster to run and can cope with larger matrices. Thirdly, the first component of Correspondence Analysis is usually similar to the wordfish results.

5. Results

5.1. Structural Topic Model - Topic Prevalence Effects

I ran a structural topic model with $K = 25$ topics on the Chinese government reports and party communiques with Leader/Premier names and Government/Party report as prevalence covariates. The reason for not including any content covariates is that this project focuses on what topics are mentioned more by each leader/premier rather than which words they use for each topic. Moreover, adding content covariates dramatically increases the time of running the model. The topic names and their Chinese keywords by FREX are presented in Figure 2 in the order of their prevalence in all documents. The FREX score is higher when the words are both frequent and exclusive to the topic. It is the harmonic mean of the rank by probability within topic (frequency)

and rank by distribution of topic given the word (exclusivity). An English version of the keywords with topic numbers is presented in Figure 3.
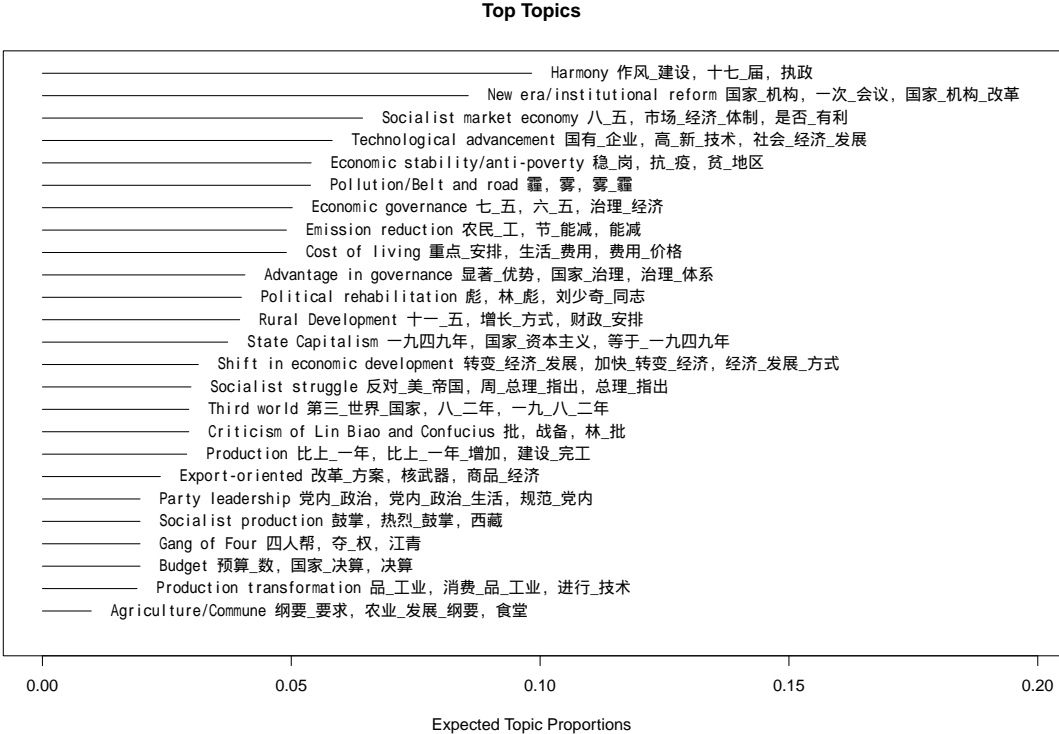
Top Topics

Figure 2: Topics in Chinese government reports and party communiques and related words by FREX

The results show that almost all topics correspond to a specific time period or even a specific leader/premier. However, there tended to be overlaps between party communiques and government reports when five-year plans or broad economic planning were of concern. For example, Topic 1, "Economic governance," is closely related to the 6th and 7th Five-Year Plans, as revealed by the FREX words in Figure 3. The leader associated with these plans were Deng Xiaoping and the corresponding premiers are Zhao Ziyang and Lipeng. This is reflected in Figure 4 which shows higher prevalence effects of them.

Similarly, Topic 4, "Rural Development," is associated with the 11th Five-Year Plan. The corresponding leader and premier were Hu Jintao and Wen Jiabao. This is reflected in Figure 5 which shows higher prevalence effects of them.

However, the interaction between the party and government on economic planning decreased under Xi Jinping's leadership. Neither the party communiques nor the government reports em-
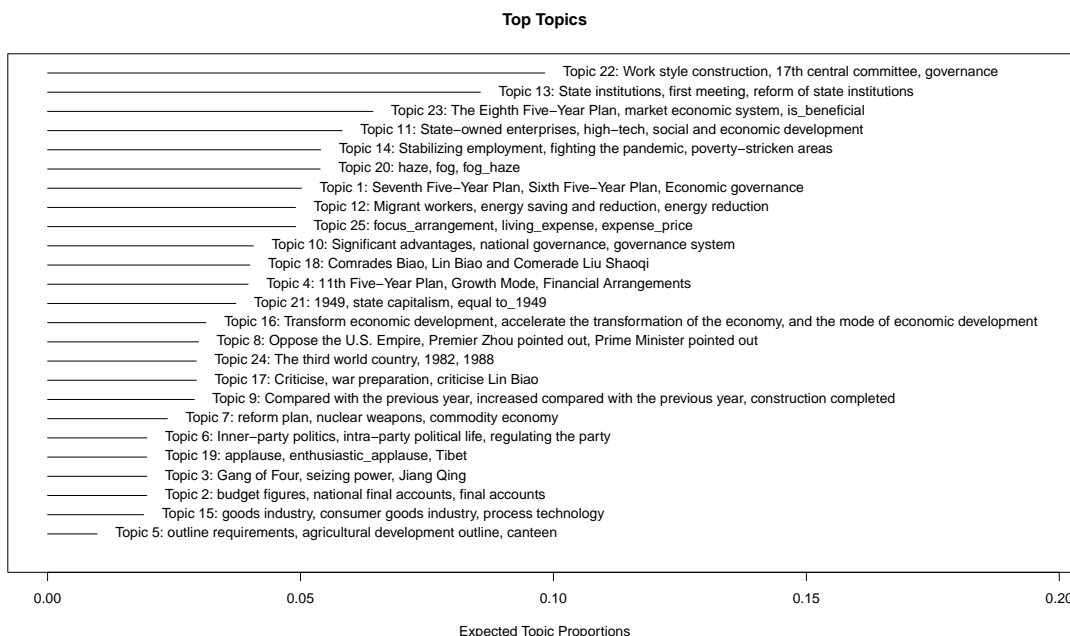
Top Topics

Figure 3: Topics in Chinese government reports and party communiques and related words by FREX (in English)

phasised the corresponding five-year plans. As a result, five-year plans under Xi failed to be recognised as a distinct topic by the model. The party communiques under Xi switched to party leadership and intra-party disciplines (Topic 6) and the government reports focused more on ex tempore economic problems such as the COVID-19 pandemic and pollution (e.g. Topic 14 and 20).

As shown in Figure 6, the leadership of the party and intra-party discipline has only been emphasized by Xi Jinping. This is not echoed by his prime ministers, Li Keqiang or Li Qiang.

On the other hand, Topic 14 (Economic stability/anti-poverty/pandemic) and Topic 20 (Pollution) are more related to the ex tempore economic problems that were only mentioned by Premier Li Qiang (Figure 7) and Li Keqiang (Figure 8), respectively. They are not echoed by party communiques under Xi Jinping, either.

In the topic related to the New Era and institutional reform (Topic 13), the prevalence has incrementally increased under the past party leaders since the Reform and Opening-Up, with the prevalence effect of Xi Jinping being the highest. This is shown in Figure 9. This implies that the recent political trends in the party were not necessarily intiated by Xi Jinping but were rather a continuation of the previous party leaders.
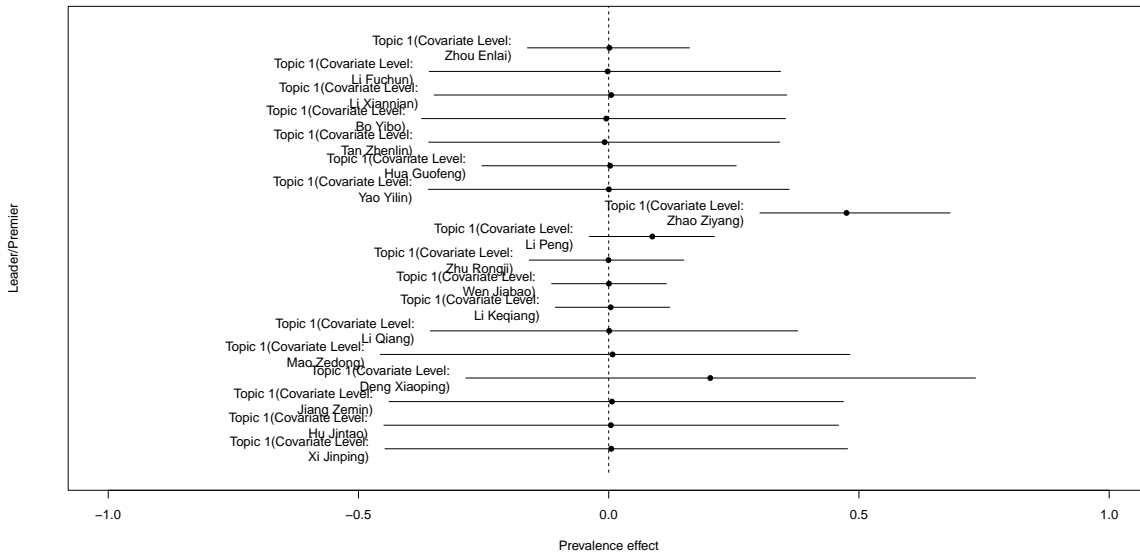
10

Figure 4: Topic 1, "Economic governance," in Chinese government reports and party communiques
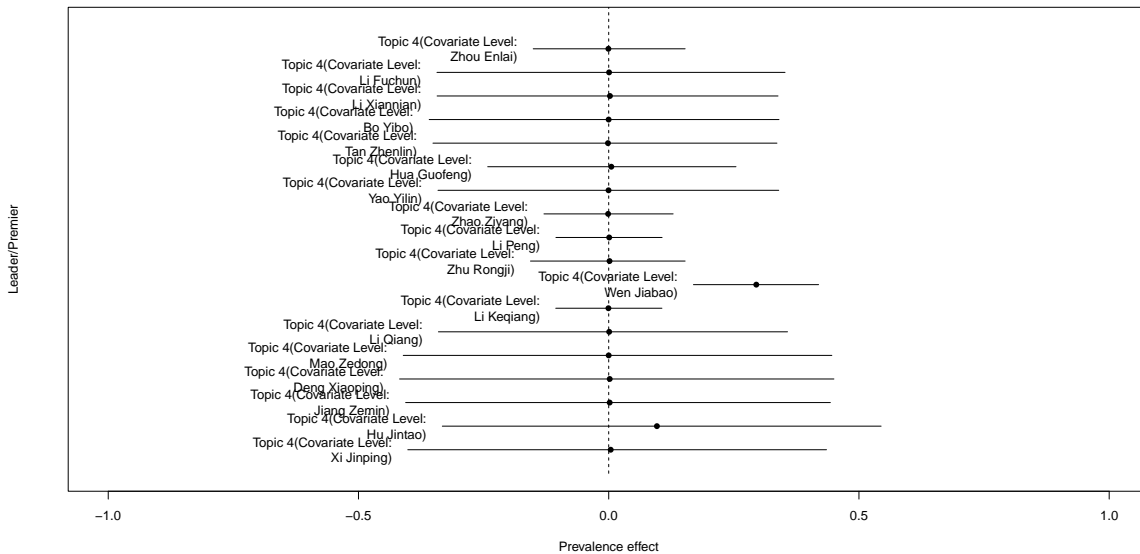

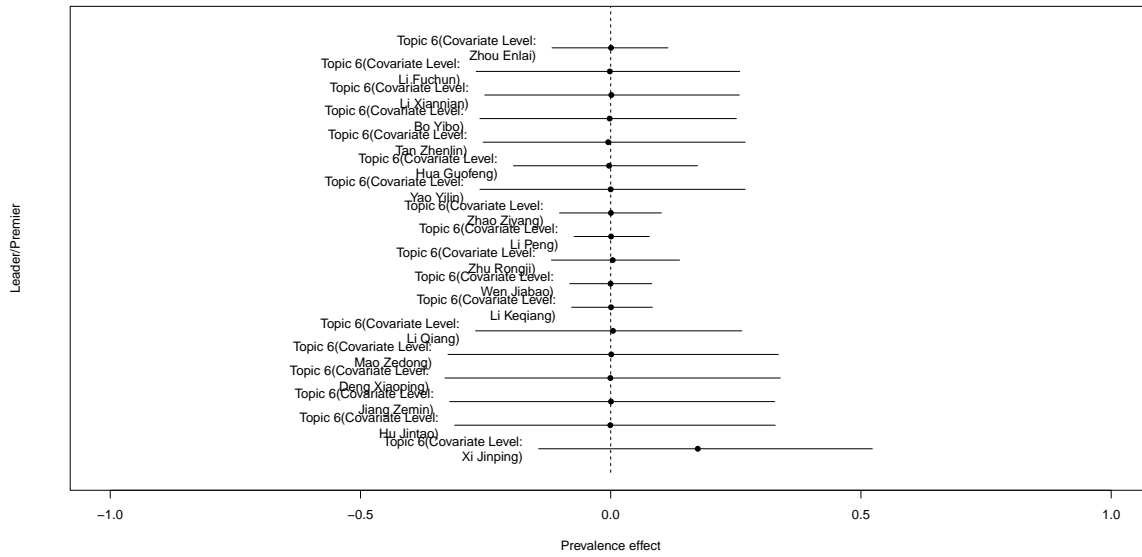Figure 5: Topic 4, "Rural Development," in Chinese government reports and party communiques

11

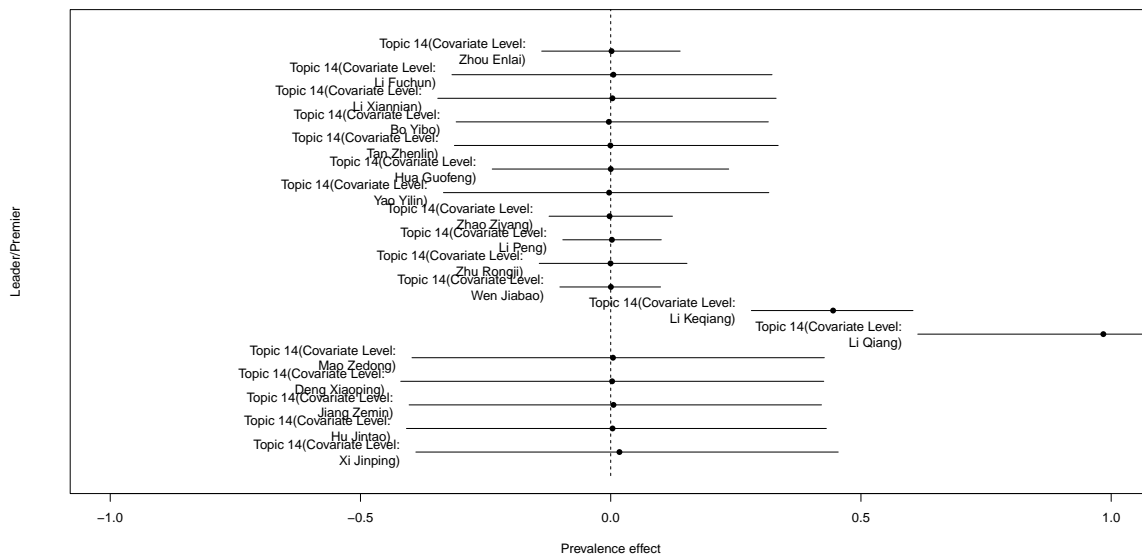Figure 6: Topic 6, "Party leadership," in Chinese government reports and party communiques

Figure 7: Topic 4, "Economic stability/anti-poverty/pandemic," in Chinese government reports and party communiques
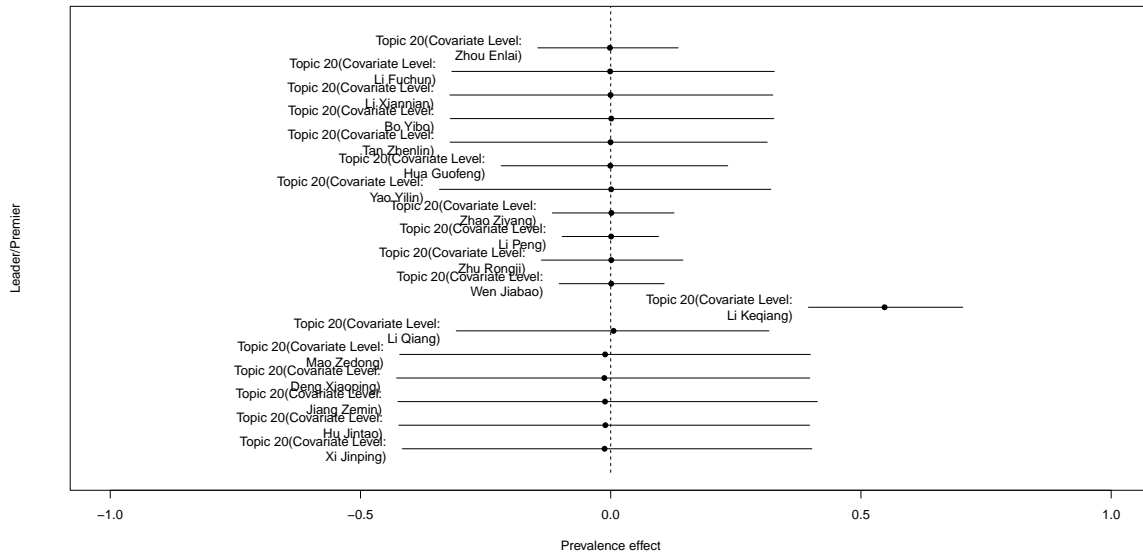
Figure 8: Topic 20, "Pollution," in Chinese government reports and party communiques
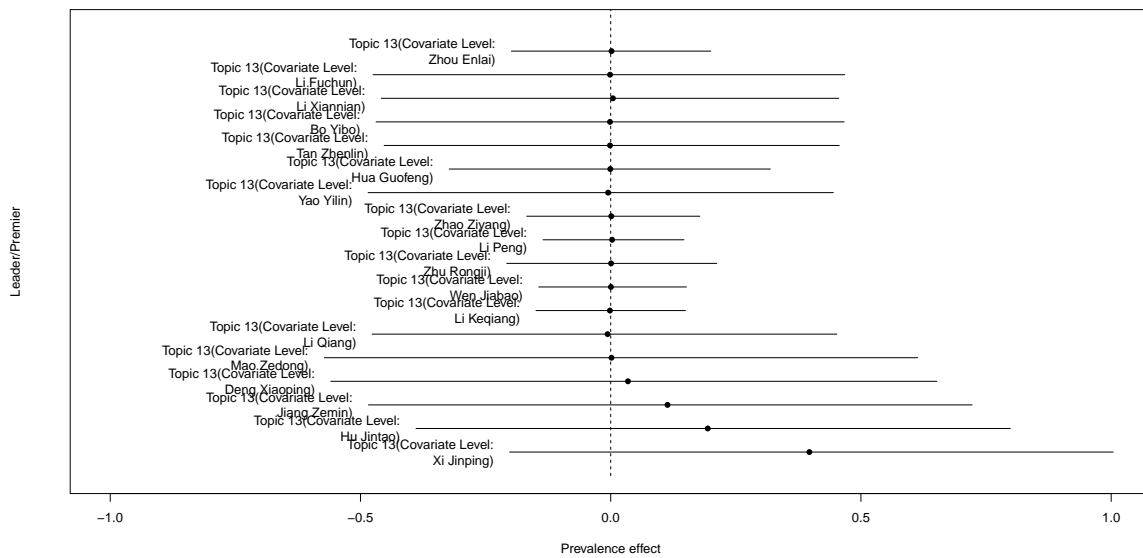

Figure 9: Topic 13, "New Era/institutional reform," in Chinese government reports and party communiques

In summary, the results of the topic prevalence effects show that the overlap between party communiques and government reports has decreased under Xi Jinping's leadership. The party communiques have shifted to party leadership and intra-party discipline, while the government reports have focused more on ex tempore economic problems. The recent political trends in the party were not necessarily initiated by Xi Jinping but were rather a continuation of the previous party leaders. These results will be further examined by other methods in the following sections.

Moreover, there is no evidence that the party communiques or the government reports under Xi have realigned with the Mao era. There is almost no overlap between the topics of the Mao era and the Xi era.

5.2. Structural Topic Model - Topic distribution agreement

One drawback of the prevalence effect analysis is that it only shows whether or not a topic is more prevalent under a specific leader/premier. It does not show the extent of similarity of topic distribution among different leaders/premiers.

The Structural Topic Model conveniently provides the distribution of topics in each document as a parameter (θ). We can use this information to calculate the similarity of topic distribution among leaders and their corresponding premiers. The similarity of topic distribution can be calculated using the Euclidean distance between the mean topic distribution of each leader and that of their corresponding premiers. The smaller the distance, the more similar the topic distribution.

Figure 10 shows the similarity of topic distribution among leaders and their corresponding premiers. The results show that the topic distribution of Xi Jinping and his premiers is the most dissimilar, followed by Hu Jintao and Wen Jiabao. The topic distribution of Mao Zedong and his premiers is the most similar. This confirms our previous results that the overlap between party communiques and government reports has decreased under Xi Jinping's leadership.

5.3. Correspondence analysis

The first dimension of the Correspondence Analysis scores by year and party/government is plotted in Figure 11. The Correspondence Analysis results confirm that the party communiques and government reports were largely overlapped under Mao, Deng and Jiang, but diverged under Xi.
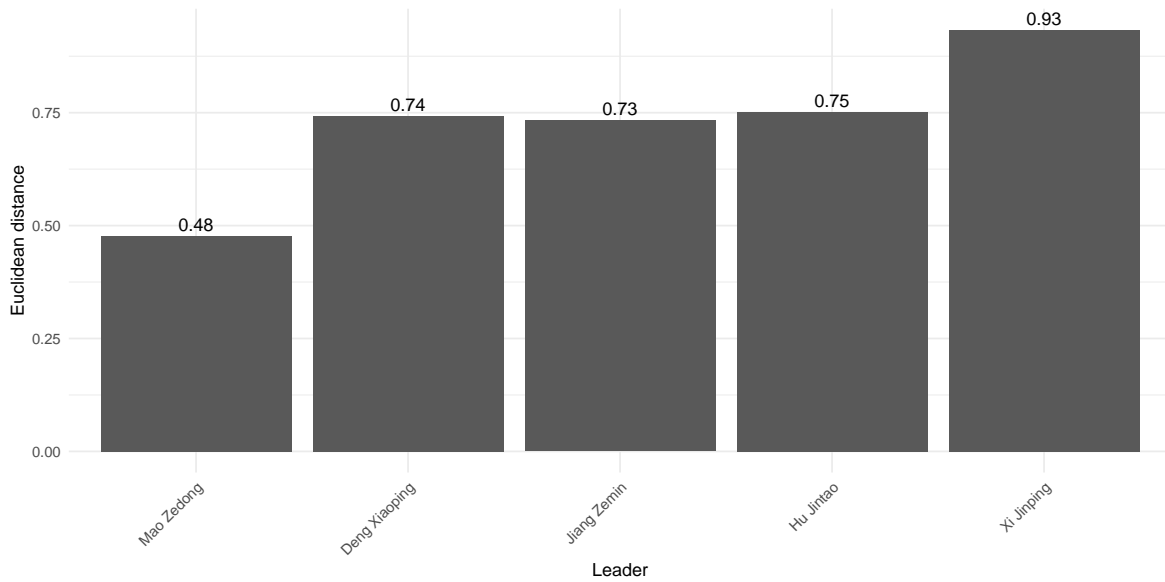
Figure 10: Euclidean distance of mean topic distribution among leaders and their corresponding premiers

However, the Correspondence Analysis results disagree with the STM results in that the former imply that the divergence started in the late Jiang or Hu era, whereas the latter suggest that the divergence has only been visible under Xi.

More interestingly, the first dimension of the Correspondence Analysis seems to capture the shift in the left-right political spectrum. A lower score on the first dimension indicates a more left-leaning political stance, while a higher score indicates a more right-leaning stance. Specifically, Reform and Opening-Up since 1978 marked a shift to the right in both party communiques and government reports. While the government communiques continued to shift to the right by emphasising economic development and entrepreneurship, the party communiques *ceased to* shift to the right by focusing on party leadership and intra-party discipline under Xi. This divergence in trends is unprecedented in the history of the PRC and is a promising area for further research.

However, the second component is not shown here as I have not found a meaningful interpretation for it. Nevertheless, the results for the second component are provided in Appendix A.

6. Conclusions

In conclusion, there is no evidence of a realignment with the Mao era in the party communiques or government reports under Xi Jinping. On the contrary, the style of the party communiques
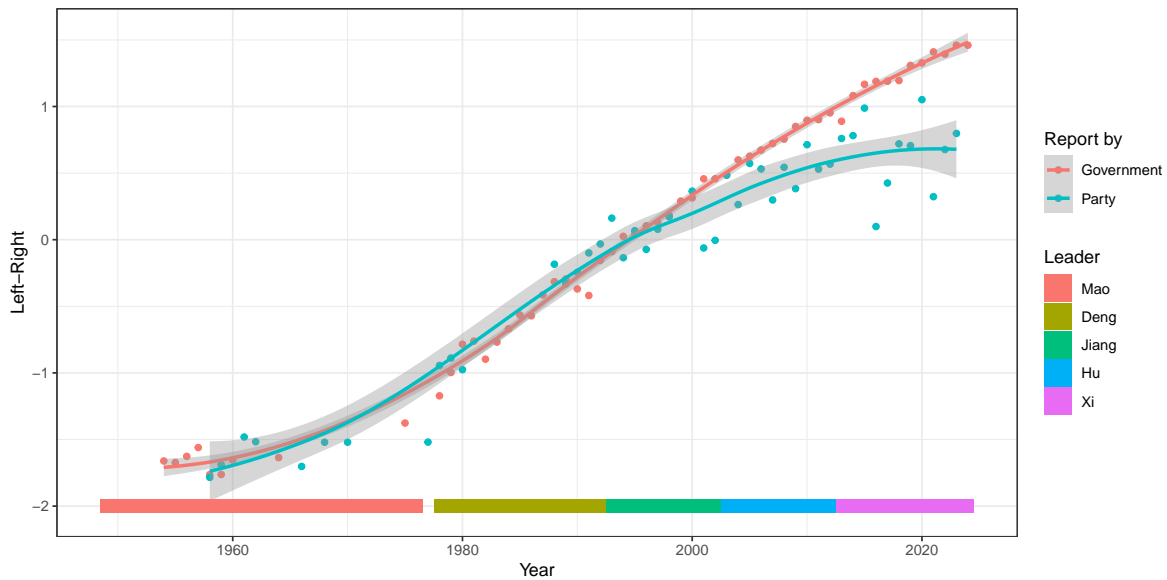
Figure 11: Correspondence analysis of Chinese government reports and party communiques (first dimension)

and government reports is more of a continuation of the Hu era. Therefore, the hypothesis that Xi Jinping is moving toward a more Maoist style and putting an end to the Reform and Opening-Up era is not supported by our data of party communiques and government reports. It is true that the topics in party communiques and government reports under Xi differ from those under his predecessors, but the topics in these reports differed among his predecessors after all. In terms of the left-right political spectrum, the party communiques merely *ceased* to shift to the right, rather than shifted to the left, under Xi Jinping. More interestingly, the government reports continued to shift to the right.

The divergence in the topic distribution of party communiques and government reports under Xi Jinping is unprecedented in the history of the PRC, and very counter-intuitive. This is a promising area for further research. The party communiques ceased to shift to the right by focusing on party leadership and intra-party discipline under Xi, while the government reports continued to shift to the right by emphasising economic development and entrepreneurship. This is counter-intuitive to the wide-spread belief that the party is increasingly intervening in the economy while the State Council is losing its autonomy. Nevertheless, the divergence should not be over-interpreted, as the institutional and personnel changes are much more substantial than public reports when assessing the party-government power dynamics in China.

One source of divergence is the downplaying of the five-year plans, which had been a key area of overlap between party communiques and government reports under Mao, Deng and Jiang.

16

This fall in economic planning is echoed by the fact that the government reports under Xi have been increasingly focused on ex tempore economic problems such as environmental pollution and COVID-19, which are more practical rather than strategic issues. This is in line with the speculation that the State Council is subdued and marginalised by the party under Xi.

References

Blei, David M., and John D. Lafferty. 2007. "A Correlated Topic Model of Science." *The Annals of Applied Statistics* 1 (1). https://doi.org/10.1214/07-aoas114.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Chen, Yuchen, Alex Jiahong Lu, and Angela Xiao Wu. 2023. "'China' as a 'Black Box?' Rethinking Methods Through a Sociotechnical Perspective." *Information, Communication & Society* 26 (2): 253–69. https://doi.org/10.1080/1369118x.2022.2159488.

CPPCC. 2023. "习近平在广东考察时强调：坚定不移全面深化改革扩大高水平对外开放在推进中国式现代化建设中走在前列 [During His Inspection in Guangdong, Xi Jinping Emphasized: Unswervingly Comprehensively Deepen Reform and Expand High-Level Opening up to the Outside World, and Be at the Forefront of Promoting Chinese-Style Modernization.]." https://www.gov.cn/yaowen/2023-04/13/content_5751308.htm.

Horsley, Jamie. 2023. "What Is the State of the Chinese State?" https://thediplomat.com/2023/09/what-is-the-state-of-the-chinese-state/.

Hsu, Sara. 2023. "Is China's Reform and Opening Era Over?" https://thediplomat.com/2023/02/is-chinas-reform-and-opening-era-over/.

Jiang, Yan. 2021. "Diachronic Lexical Features of Chinese Government Work Reports: A Text Mining Approach." *Advances in Social Science, Education and Humanities Research.* https://doi.org/10.2991/assehr.k.210407.140.

Matson, Emily. 2022. "Why It's Misleading to Call Xi Jinping the 'New Mao'." https://nuvoices.com/2022/10/03/why-its-misleading-to-call-xi-jinping-the-new-mao/.

Milanovic, Branko. 2023. "Xi Jinping Is Not Mao Reborn." *UnHerd.* https://unherd.com/2023/12/xi-jinping-is-not-mao-reborn/.

Nenadic, Oleg, and Michael Greenacre. 2007. "Correspondence Analysis in *R*, with Two- and Three-Dimensional Graphics: The **ca** Package." *Journal of Statistical Software* 20 (3). https://doi.org/10.18637/jss.v020.i03.

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): 1064–82. https://doi.org/10.1111/ajps.12103.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–22. https://doi.org/10.1111/j.1540-5907.2008.00338.x.

The Economist. 2023. "Why Xi Jinping Is Not Another Chairman Mao." https://www.economist.com/china/2023/04/05/why-xi-jinping-is-not-another-chairman-mao.

Wang, Haiyan. n.d. "Example: Chinese Text Analysis." Accessed April 9, 2024. https://quanteda.io/articles/pkgdown/examples/chinese.html.

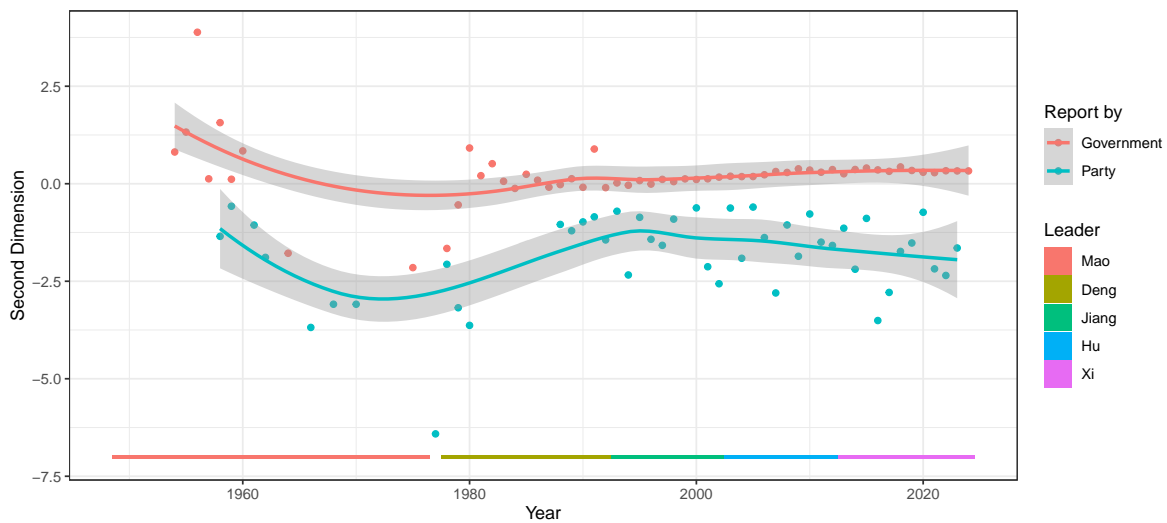# A. Results of the Second Component of the Correspondence Analysis

Figure 12: Correspondence analysis of Chinese government reports and party communiques (second dimension)