Party Supervision, Localized Features, and the Report of Inequality among Chinese Newspaper, 2003 - 2019*

YE, Yizhou ZHANG, Yi

December 24, 2020

Abstract

In this paper, we combine more than 170,000 articles from Wisenews during 2003 to 2019 and official statistics to investigate how localized features affect the report of inequality, and the role of party supervision in this process, for which we adopt sentiment analysis as well as topic model in this research. Our findings show that party supervision would directly bring more positive tone and higher concern of macro topics, while the pattern of localized features' effects and party moderation effects are significant but complicated. This study contributes to both the understanding of perception of inequality from a media perspective, and the exploration of connecting large scale text data with official statistics.

Keywords: Inequality, Media, Sentiment Analysis, Topic Model, China Communist Party

*In this project, Yizhou Ye and Yi Zhang scrape and clean the data together. Zhang mainly works on sentiment score analysis, and Ye mainly works on topic model analysis. In the writing part, Ye is responsible for introduction, literature review and topic model relevant parts, and Zhang works on general data and method part as well as sentiment score parts. Two authors think they equally contribute to this project, and the authorship is alphabetical.

1 Introduction

After 1978's reform, China experiences rapid economic growth alongside with the fast increase of multidimensional social inequality, including wealth, education, gender, rural-urban etc. (Xie & Zhou, 2014; Zhou & Xie, 2019; Piketty et al., 2019; Wu, 2019). However, it is a puzzle that why China keeps social stability successfully under the severe and increasing inequality. Whyte (2010) argued that inequality perception is one of the key explanations, saying the level of inequality felt by people is lower than the factual inequality because they mainly assess the inequality based on localized information and a great part of China's inequality could be decomposed into regional and rural-urban inequality which are hard to be detected through life experiences (Xie & Zhou, 2014). However, except from the ordinary life, people could also receive information about inequality through media. There should be two steps in the flow model (Katz & Lazarsfeld, 1966), the first of which is from the so-called real world to media and the second one is from media to people. We could break out the perception of inequality in China following this perspective, based on which the puzzle of Chinese inequality and stability could be decomposed to two parts: whether people really have idea about the inequality and whether they care about the level of inequality.

In this paper, we mainly want to answer the first half of this perception chain: how Chinese media report its inequality. Specifically, we would investigate how localized inequality in different aspects affect the local newspaper's report of inequality, in the meaning of sentiment score and topic preference. Given Chinese context, we would pay special attention to the role of party supervision. The remaining part of this paper contains literature review, data and method, results, and conclusion.

2 Literature Review

In this part, we would review the literature relevant to inequality perception, and the role of China Communist Party (CCP) as well as the marketization of Chinese media.

Literatures discussing perception of inequality could be divided into three strands: social position theory, mobility expectation theory and reference group theory (Kluegel & Smith, 1981; Kelley & Evans, 1993; Wegener, 1991; Guillaud, 2013). The first one claims that people's perception of inequality is determined by their position, which means invested interest tend to believe society more equal and vice versa (Kluegel & Smith, 1981; Robinson & Bell, 1978; Erikson & Goldthorpe, 1987). The second one holds the view that people would believe society more equal when they have higher mobility expectation and vice versa (Guillaud, 2013). The third one explains people's perception through micro and local environment rather than macro situation (Wegener, 1991), which is also the most effective explanation under Chinese context (Whyte, 2010; Zhang et al., 2018). However, three mainstream explanations of inequality perception have not taken the role of media report into consideration. In addition to direct life experience, media is also an essential channel for people to sense the society and it provides the possibility to break out the limitation of everyday life, which might reshape all three mechanisms listed above. In this research, we tend to contribute to the understanding about the report of inequality.

Chinese media market is significantly different from western in many aspects, one of which is the supervision from CCP. Given the censorship institution of publishment, Chinese media are supervised by CCP less or more, especially for those formal one, including newspaper which this research focus on (King et al., 2013; Qin et al., 2018). We could summary the main power in Chinese media market into two kinds: market preference and political censorship.

Because many newspapers in China are required to be responsible for their own finance, media must consider readers' preference like ordinary media in democracy market (Mullainathan & Shleifer, 2005; Gentzkow et al., 2014; Cagé et al., 2014). In Chinese context, the readers of newspaper are mainly urban resident, especially those who have the reading habit, which leads the reports of inequality would more likely discuss topics these groups interested in. Another essential stakeholder of Chinese media market is CCP. CCP supervision would affects both the topic media talking about and the tone used by them (Zhao, 2008; Qin et al., 2018). However, because of the financial support from local government or local party branch, party newspaper would rely on readers less, resulting to the possibility that they would report more various inequality compared with marketized media. We would examine the influence of party supervision in this research in both sentiment and topic aspects. Moreover, we would also investigate whether party supervision moderate the effect of localized features of inequality.

To sum up, past studies relevant to the perception of inequality mainly focus on the explanation based on people's own experiences, and existing media analysis has not talked about inequality and mainly focused on political issues. Our research could benefit people's understanding of how Chinese newspaper report inequality, unpacking the role of CCP supervision and market preferences. This study could also contribute to the explanation why China could remain stable under a high level of social inequality.

3 Data and methods

3.1 Data

3.1.1 Main Dataset

Our newspapers data comes from Wisenews. Wisenews is a real-time updated Chinese media information database provided by Wisers Information Limited. Wisenews contains content from large newspapers, magazines, news agencies and websites from the Greater China region (Mainland China, Hong Kong, Macao and Taiwan) and the United States. Moreover, the database is well applied in existing research (Qin et al., 2018).

We build up our dataset following several steps:

1. Search keywords *inequality* (不平等) and *unfairness* (不公平) from 2003 to 2019 in Wisenews. We add the keyword *unfairness* since it has highly similar meaning as *inequality*[1]. We choose 2003 as the start year since the data in Wisenews since President Hu Jintao came to power, and the result is not be confounded by the government turnover (The year of news in Wisenews during era of President Jiang Zemin is not complete). The searching system return us a list of links, and news are stored in remote server.

2. Request the links and scrape the website.

3. Extract title, content, date and media source of the news from the website.

We obtain a dataset with 198200 articles. In the process of data cleaning, we remove not completed, too short, and duplicated articles. We also remove articles from media with too few

[1]According to Chinese Word Vectors, the largest pre-trained Chinese word vectors, the word with the highest similarity of *equality* (平等) is *fairness* (公平).See details of Chinese Word Vectors in https://github.com/Embedding/Chinese-Word-Vectors. To increase available observations, we think add unfairness as another keyword is reasonable.

articles and too professional content (see the complete process of data cleaning in Appendix A.1). After cleaning, there are 172474 articles in our dataset.

3.1.2 Placebo Dataset

For the reliability of the result, we also build up a placebo dataset. We choose *summer* (夏季) as the placebo keyword for several reasons: first, summer is a natural event and will not be affected by human activities, so we can regard the variation in reports about summer are unlikely to be influenced by media's preference or other socio-economic factors. Second, there is no direct relationship between summer and inequality, so reports with keyword *summer* are independent with reports with keyword *inequality*. Third, media's preference for reporting varies in different areas and times and reports about *summer* are unlikely to be regulated. Consequently, placebo dataset can help us capture the change. Placebo dataset consists of 305581 articles.

3.1.3 Data of Outer Source

Our data of outer source comes from two sources. The media information comes from Qin et al. (2018), consisting the province, city, supervisor and other information of media. We define party newspaper as the media is supervised directly by national/local party committee. Data of local socio-economic features mainly comes from Chinese Statistic Yearbook. Provincial Statistic Yearbooks are used as supplement. The details of variables in outer source are in Appendix A.2.

3.2 Methods

3.2.1 Sentiment Analysis

We apply on existing well-developed algorithm in computing sentiment score of each article. The toolkit we utilize is PaddleHub developed by Baidu, a leading company in deep learning. PaddleHub is a toolkit with abundant pre-trained models built on PaddlePaddle, a deep learning platform.

In order to choose the best built-in algorithm for our dataset, we build up a validation set by random sampling the main dataset, consisting of 500 articles. The sentiment label (positive or negative) of each article is coded manually. We follow the same rule while coding, and the sentiment of articles is judged by the entire article. To make sure the accuracy of labeling, we also do cross-validation while labeling articles with indistinct sentiment.

Next, we utilize several built-in algorithms (BOW, CNN, GRU, LSTM, and Bi-LSTM) to label our validation set. After running the algorithms, we evaluate the performance of these five algorithms by comparing their ROC curve (see ROC curve of each algorithm in Appendix B.1). CNN and GRU perform best in the evaluation. We also compare the confusion matrix and several indicators of two the best algorithms in initial setting ($\tau = 0.5$) and best threshold according to ROC curve (see the result in Appendix B.2). Since our validation set is an imbalanced dataset (this feature is like other usual news dataset) with few positive news, the sensitivity of algorithms in initial setting is relatively low. To balance TPR and TNR, we choose CNN with best threshold ($\tau = 0.204$) as the algorithm in labeling sentiment score.

The application of CNN (convolutional neural networks) in sentiment classification is proposed by Kim (2014). He applied CNN in pre-trained word vector for the task and find that the application outperformed than other existing algorithms. The pre-trained CNN in

PaddleHub is trained by Baidu's own dataset.

Each article is classified as positive or negative, and 45.44 percent articles are classified as positive. We use logistic regression to investigate the factors affecting sentiment score.

3.2.2 Topic Model

To analyze the topics discussed by Chinese newspaper, we adopt LDA approach to serve this goal, based on which we could construct topics with specific words and compose articles by several topics (Blei et al., 2003). Besides clustering topics, we also aim to figure out how would localized features and party supervision affect topics' prevalence, for which R package Structal topic model (STM) is finally used for analysis (Roberts et al., 2019). Since what we analyze are articles published by media, it is naturally that media characteristics and spatiotemporally localized features generated based on the media location and published date could be treated as covariates.

One of the essential settings of topic model is number of K. To search the most appropriate K, we firstly adopt the data driven approach, from which we get 95 topics. Nevertheless, many of them could be classified into same category and we finally get 20 topics. To obtain the locally optimized K, we run several models with K around 20. After the comparison of exclusivity and sematic coherence as well as well discussion about topic's differences, we believe K equal to 22 is best choice (Mimno et al., 2011; Airoldi & Bischof, 2012). It should be noted that STM package only consider the top 10,000 frequent words to overcome sparse problem.

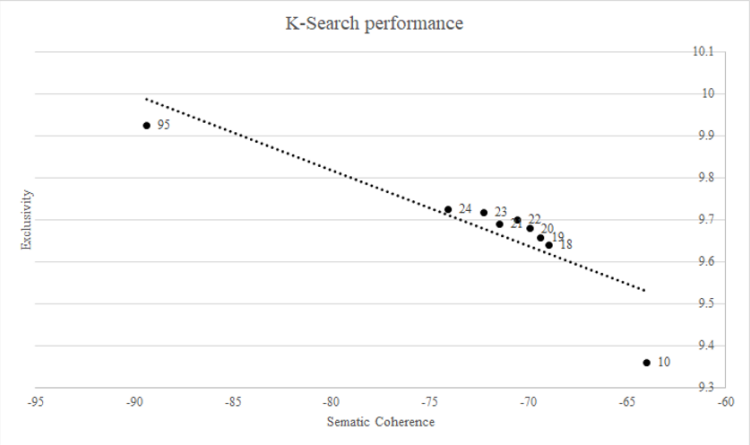


Figure 1: Sematic Coherence and Exclusivity Performance of Different Ks

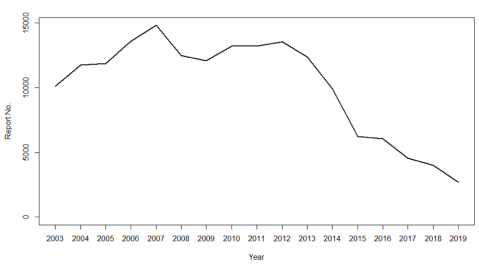
In our main analysis, we follow these steps: clustering the topic, analyzing topic correlation, investigating the direct effect of party supervision, and exploring the moderation effect of party supervision.

4 Results

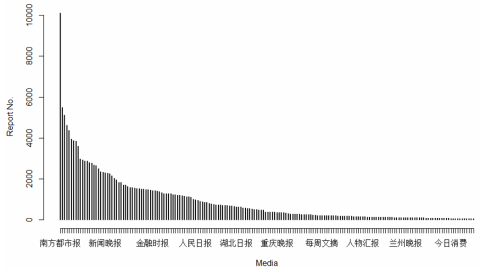
4.1 Data Description

In the beginning of results, we would briefly introduce the descriptive characteristics. The article number per year ranges from 3,000 to 15,000. Before 2013, the article number keeps stable, while after 2013, the number suffers dramatically flop. To examine whether it is resulted from censorship, we draw the curve of placebo dataset and find the pattern is same (see Appendix C.1), for which we believe it is caused from other factors. The distribution of newspaper is long tail, among which Southern Metropolis Daily occupies the first place with 10,091 articles. Articles published by top ten medias occupy over 20 percent among all articles.

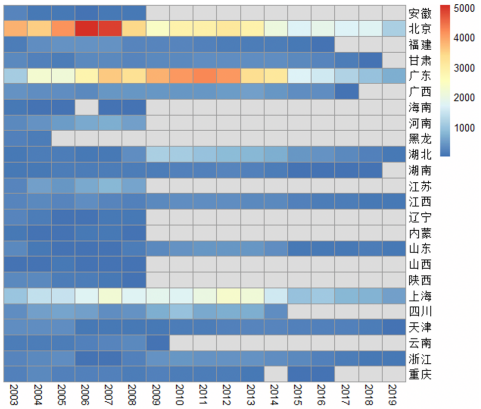
We provide the heat map to show the spatiotemporal distribution, for which we could observe the attrition (*grey* for attrition) of many provinces indeed begin from 2008 but not 2013. Beijing, Guangdong and Shanghai hold the top three places with most report of inequality every year.



(a) Articles Number, 2003-2019



(b) Distribution of Articles among Media



(c) Spatiotemporal Distrib. of Articles No.

Ten Media with Most Articles	
Southern Metropolis Daily	10091
Yangcheng Evening Post	5488
The Beijing News	5109
New Express Daily	4618
Guangzhou Daily	4380
Oriental Morning Post	3952
21st Century Business Herald	3884
Nanfang Daily	3840
Xinmin Evening News	3597
Beijing Youth Post	2884

(d) Ten Medias with Most Articles

Figure 2: Descriptive Results of Articles about Inequality

4.2 Sentiment Analysis

4.2.1 Pattern of Time and Party Newspapers

First, our research investigates the change in sentiment score among different years. The result is in Figure 3:

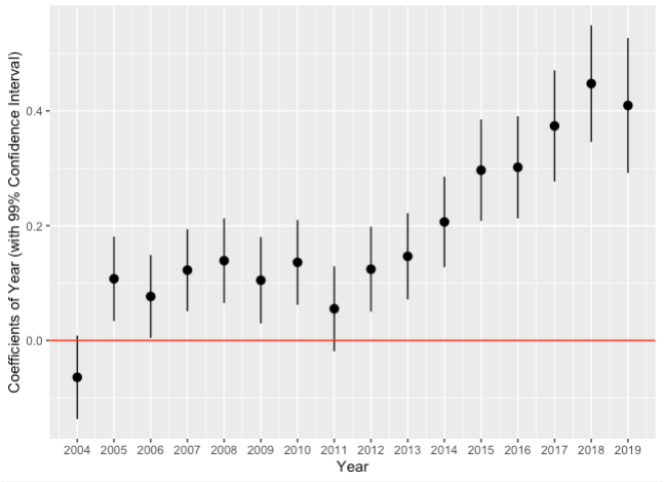
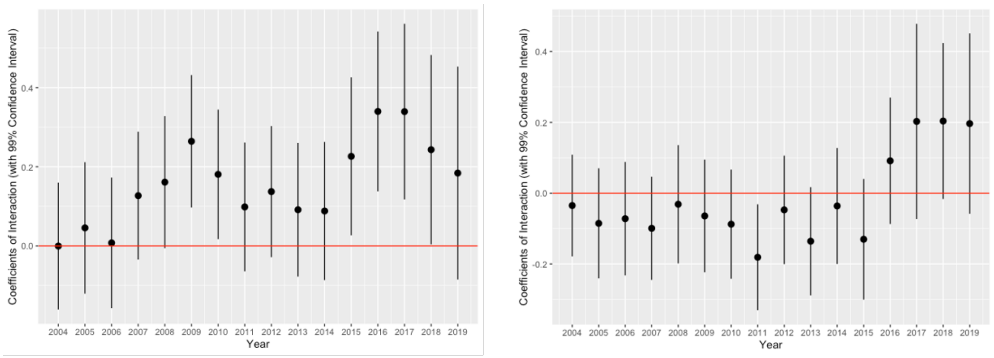


Figure 3: Coefficients of Year with 99% CI

Note: Reference group is year 2003. Media are fixed. Standard errors are clustered at the media level.

The result shows that the pattern is similar among 2005-2013. Articles at that time are slightly more positive compared to 2003. After 2014, the coefficients become much larger, which means articles related to inequality are more positive after 2014.

Since party newspapers are more strictly controlled by CCP. Consequently, the preference for reporting and commenting is more likely to be influence by government turnover and preference. Talking about inequality reflects socio-economic status and government efficiency, so articles about inequality will be affected by wills of states. Here, we show the coefficients of interaction term of party newspapers and year in Figure 4.



(a) Articles about *Inequality* (b) Articles about *Summer*

Figure 4: Coefficients of Interaction of Year and Party Newspapers with 99% CI

Note: Reference group is the interaction of party newspapers and 2003. The coef. of party supervision and year effects are in models but not reported. Media are fixed. Standard errors are clustered at the media level.

The pattern between 2004 and 2019 is bimodal. Between 2009 and 2010, articles related to inequality in party newspaper are more likely to be positive compared to their non-party newspaper counterpart. After 2014, the coefficient begins to increase and come to the peak between 2016 and 2017. Combined with the pattern in Figure 4, we can find that party newspapers are more likely to report positively among 2009 and 2010. The pattern later is different. Articles related to inequality are overall more likely to be positive after 2014, and party newspaper reports more positively between 2015 to 2017. The probability to report positively continue to increase after 2017, while the effect of interaction terms declines. We regard that party supervision plays a role in the process. There may exist the probability that non-party newspapers catch up with party newspaper. It may come from that party supervision disseminate. Overall, the finding is preliminary, more explanations are under exploration.

As we mention in the section of data, our placebo dataset with keyword *summer* is unlikely

to be censored or supervised intentionally, so the pattern of the interaction of year and party newspapers should not be significant. We use the same model setting in placebo dataset.

We can see that the coefficients are not significant overall (except one of 17 years). Though the coefficients are different, the standard errors are relatively large.

4.2.2 Variation in Local Features

We calculate the mean of sentiment score of each province in different years, and we show it in Figure 5. Overall, the variation in different years of each province is similar to the pattern overall. The heatmap also cluster the sentiment score of each province by K-Means. The pattern of cluster is related to local development status. Well-developed provinces[2] (Beijing, Shanghai, Tianjing, Guangdong) are clustered together, and they tend to report inequality more positively. We can also observe that less-developed provinces are more likely to report inequality in a contradictory way.

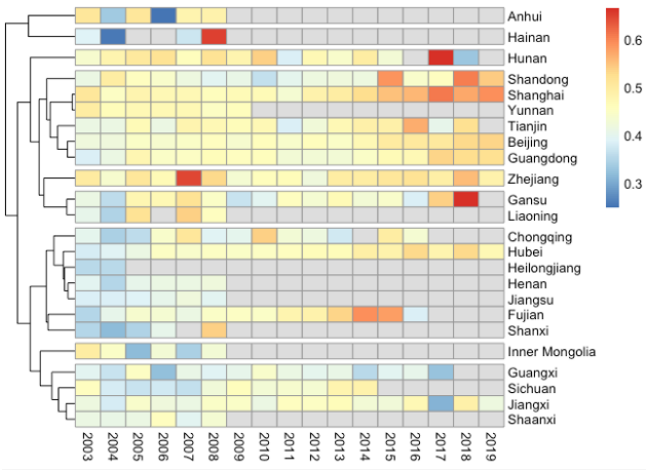


Figure 5: Mean of Sentiment Score of Provinces, 2003-2019

Note: Missing values are set grey.

[2]Yunnan is also in this cluster, but it is missing in later year, which may confound the mean of the whole period.

In this part, we investigate the association of sentiment of articles and local features. We select some socio-economic factors as independent variables of each provinces in each year and use positive or not of the article as dependent variables (see the result in Appendix C.1). We can observe that with the growth of population and urbanization, local media are more likely to report inequality negatively. The enlarged family size and more medical provision will also add to the negative report. However, it exists the concern that local features will affect the general preference of local media, but not only inequality. To overcome this problem, our research utilizes analytical strategy making use of placebo dataset.

As we mention above, our placebo dataset is a reliable source for reflecting the general pattern of media articles, since it is unlikely to be influenced intentionally and its topics are not affected by socio-economic status. We can combine our main dataset and placebo dataset and find out what local features affecting the preference for report. The result is in Table 1.

We can observe that social-economic factors, public service provision and consumption contribute to the sentiment of articles while reporting inequality uniquely. From the result, media in provinces with high GDP per capita is likely to report negatively, and urbanization has contradictory effects. The effect of socio-economic factors may come from that the development of economic make inequality become a more critical concern. Moreover, urbanization fill the gap between urban and rural areas and reduce the inequality among two types of district, which reflects on the articles in media. Two public service factors have effects from different directions. It shows that provision in medical service still cannot solve the problems of inequality in distribution and become more severe, but the education is different. Moreover, the increase in goods price also induce media' negative perceptions of inequality.

Table 1: Local Features and Sentiment of the Articles

Variables Type	Independent Variables	Dependent Variables: Sentiment of the Articles
Socio-economic Development	GDP per capita × Inequality	-0.126**
		(0.047)
	Population × Inequality	0.005
		(0.025)
	Urbanization × Inequality	0.568***
		(0.128)
Public Service Provision	Student-Teacher Ratio of Senior High School × Inequality	-0.019**
		(0.006)
	Number of Medical Bed per capita × Inequality	-0.050***
		(0.008)
Consumption	CPI × Inequality	-0.026***
		(0.005)
Family	Cases of Divorce per Million People × Inequality	-0.144
		(0.117)
	Family Size × Inequality	-0.102*
		(0.042)
	Local Features	Controlled
	Articles from Inequality Dataset	Controlled
	Party Newspaper	Controlled
	Year FE	Controlled
	Media FE	Controlled
	Observations	492749
	AIC	431502

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the media level. GDP per capita and Population are logarithm. Local Features include variables in the interaction terms.

4.3 Topic Model

In this part, we mainly discuss the prevalence of topics, the direct effect of party supervision and the moderation effect of party supervision.

4.3.1 The prevalence of topics

As mentioned above, we use LDA approach to figure out what kind of topics are discussed among Chinese newspapers. Given the selected K equal to 22, we find that the topics of

inequality could be clustered into four categories: everyday life, business, entertainment, and international issues. The expected proportion that a certain topic would be covered among all articles discussing inequality varies from 2 percent to 8 percent. We provide Figure 6 showing ten most used words in each topic below and name these topics based on the most frequent words. We also compute the covariant correlation between each two topics, which is also shown below, to make the clusters among topics clear. Briefly, topics relevant to inequality among media reports are various, while they could be easily clustered into some categories.

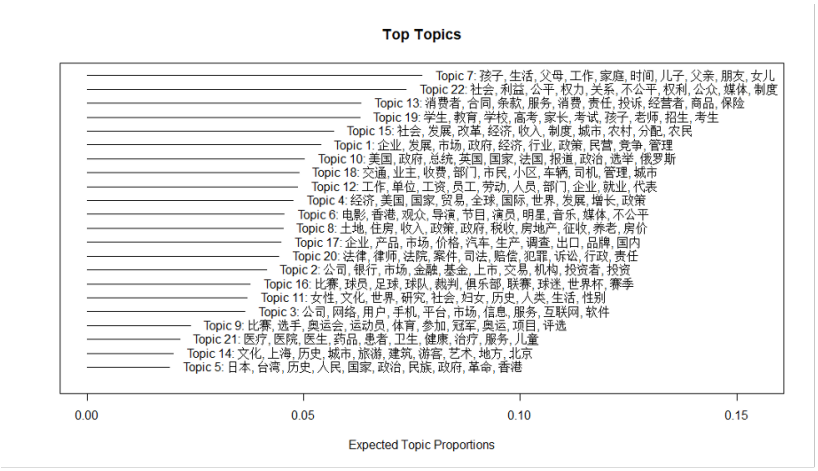


Figure 6: Ten Most Frequent Words and Expected Proportion of 22 Topics

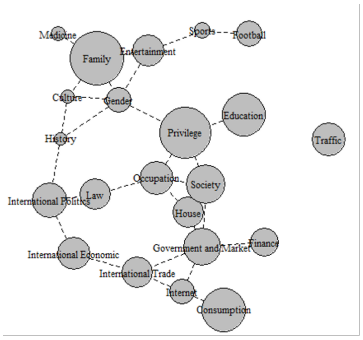


Figure 7: The Correlative Pattern among 22 Topics (Larger the size, higher the expected proportion)

4.3.2 The direct effect of party supervision

Compared with simply describing the prevalence, we are more interested in the consequences of party supervision. Given the 22 topics classified above, party newspaper and non-party newspaper show significant difference in the preference. As what we could learned from Figure 8, the general pattern is that party newspapers show higher preference on the macro topics and what non-party newspapers care are more micro. Interestingly, society, which is original topic 15 and discussed most by party newspapers, is contrast with the topic 7, family, discussed by

non-party newspapers most. To provide more straightforward comparison between these two topics, Figure 9 shows 25 most frequent words used by topic society and family, respectively. Even within a same topic, the words adopted by party and non-party newspapers are quite different, which could be observed from Figure 10. Corresponding with the hypothesis we raised before, when discussing about house and land, non-party newspapers mainly reflect what urban residents, their target readers, might be interested in, but party newspapers would spend more effort on the report relevant to rural issues. We think this might be a side of party supervision effect, with which they could less or more defend themselves from market preferences and speak for the voiceless.

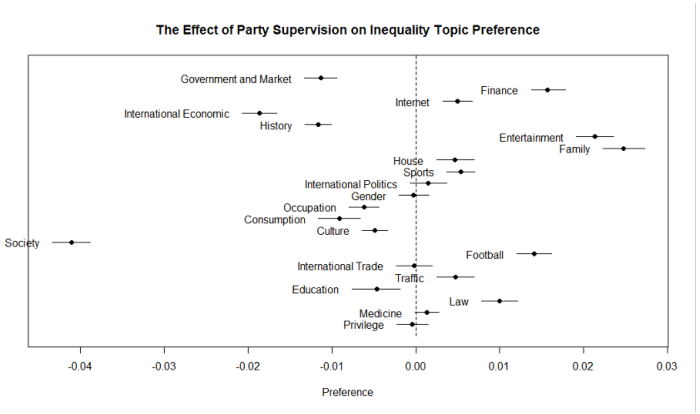


Figure 8: The Difference of Topic Preference Between Party and non-Party Newspaper

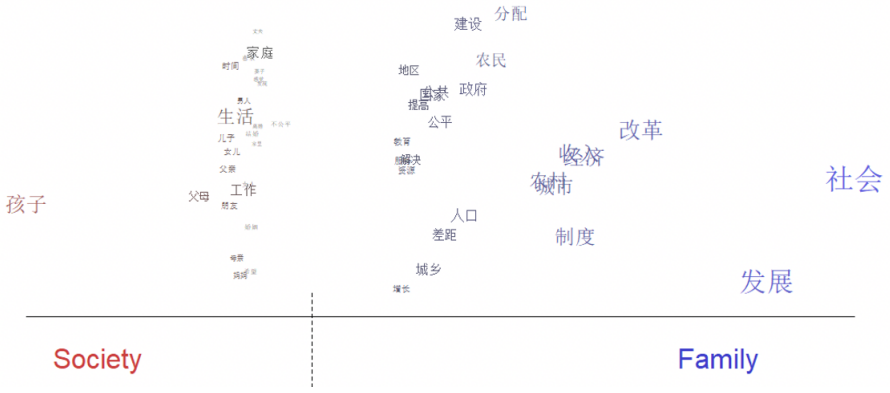


Figure 9: Difference of Words between Topics Preferred by Party and non-Party Newspapers

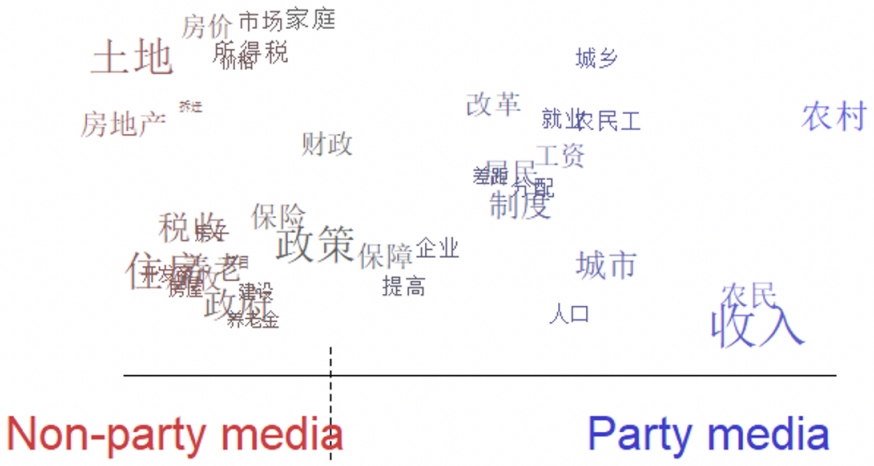


Figure 10: Difference of Words Preferred by Party and non-Party Newspapers in Same Topic

4.3.3 The moderation effect of party supervision

We also investigate the indirect effect, moderation, of party supervision. As the database used by us ranges from 2003 to 2019, topics prevalence might also change during this period. Based on the regression results, we draw estimated time effect of topics and report those experience substantial change (see Figure 11). However, we also observe significant moderation effect of party supervision. Take topic *society* as the example (see Figure 12), the prevalence of non-party media remained, but the party media's prevalence of this topic went down.

To systematically consider the correlation between localized features and reports of inequality, we use STM to estimate the effect or covariates as well as the moderation effect of party supervision. The simplest model only control province effect, and investigate year effect, party effect and moderation effect for different topics. The results are heterogeneous and when interaction term added, the pattern of party effect changes, implying that some observed effects of party indeed come from its moderated role.

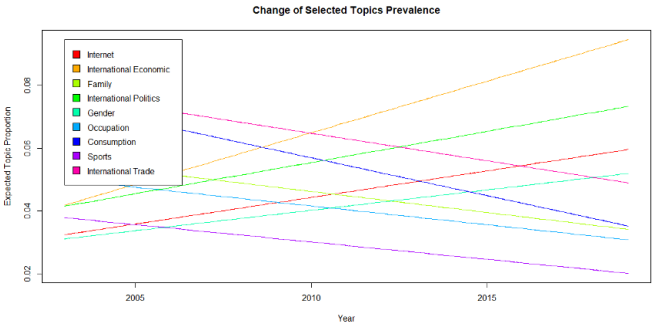


Figure 11: Change of Selected Topics' Prevalence

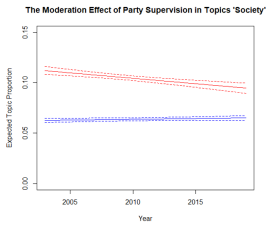


Figure 12: The Moderation Effect of Party Supervision on Topic *Society*

The larger regression results with covariates of local economic and local family situation shows the moderation effect of party supervision is complicated. On the one hand, party newspapers are censored more strongly, and they must work for propaganda; on the other hand, without or with less market pressure, the influence of readers' preference on party newspapers' report is smaller, for which their reports contain inequality related to those with less power to speak. Hence, the moderation effect of party supervision would be mixed and complex. We plan to decompose them in future studies.

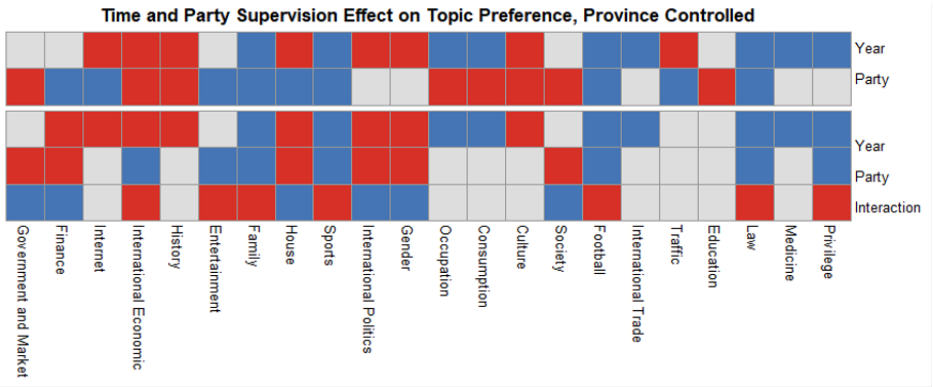


Figure 13: Time and Party Supervision Effect on Topic Prevalence, Province Controlled

Note: Cell is set to *grey* if the P-value is larger than 0.001; cell is set to *red* if the coefficient is significantly positive; cell is set to *blue* if the coefficient is significantly negative

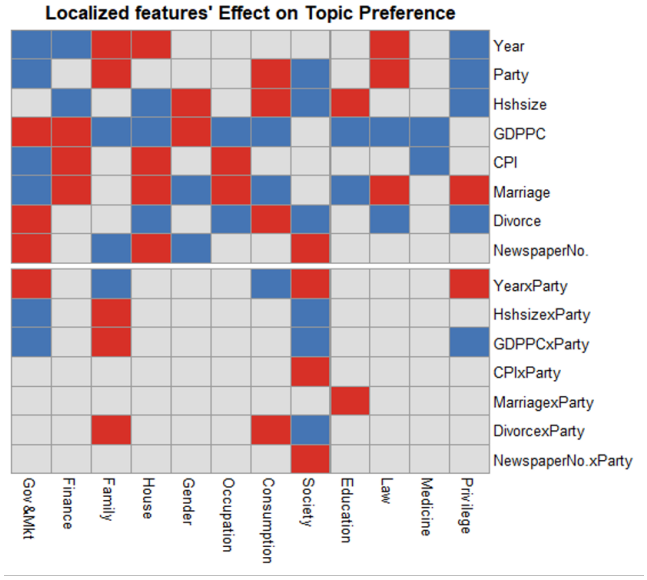


Figure 14: Localized Features and Party Moderation Effect on Topic Prevalence

Notes: This figure adopts same standard with Figure 13. These models are designed to contain as more observations as possible, and for models sacrifice observations for more covariates, please see Appendix C.3. The pattern between these two sets of models is different.

5 Conclusions

In this paper, we analyze report of inequality among Chinese newspapers, and find localized features related to inequality would significantly affect the report of inequality, both the topics chosen, and the tone used. Besides localized features, the general pattern of tone goes more positively after 2013 and the topic interests keep changing. In the analysis of party supervision, we observe both direct and moderation effect in two aspects. For the direct effect, Party supervision would raise the possibility that an article adopts positive tone to discuss inequality, and party supervision would also lead to preference of macro topics. The moder-

ation patterns are complicated, but obviously party supervision would adjust the way media reports inequality even under equivalent local conditions. The findings of this paper show Chinese media do reflect localized inequality, and party plays the vital role in this process.

This study mainly contributes to the understanding of perception of inequality from a media perspective. It also shows a possibility how researchers could connect large scale text data with official statistics. The limitations of this paper are causality and pattern recognition. We would try to decompose party supervision's suppression and amplification effect, and cluster effects of localized features in future studies.

References

Airoldi, E. M., & Bischof, J. M. (2012). A poisson convolution model for characterizing topical content with word frequency and exclusivity. *arXiv preprint arXiv:1206.4631*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Cagé, J., et al. (2014). Media competition, information provision and political participation. *Unpublished manuscript, Harvard University*.

Erikson, R., & Goldthorpe, J. H. (1987). Commonality and variation in social fluidity in industrial nations. part i: A model for evaluating the 'fjh hypothesis'. *European sociological review*, *3*(1), 54–77.

Gentzkow, M., Shapiro, J. M., & Sinkinson, M. (2014). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review*, *104*(10), 3073–3114.

Guillaud, E. (2013). Preferences for redistribution: an empirical analysis over 33 countries. *The Journal of Economic Inequality*, *11*(1), 57–78.

Katz, E., & Lazarsfeld, P. F. (1966). *Personal influence, the part played by people in the flow of mass communications*. Transaction Publishers.

Kelley, J., & Evans, M. D. (1993). The legitimation of inequality: Occupational earnings in nine nations. *American journal of sociology*, *99*(1), 75–125.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, 326–343.

Kluegel, J. R., & Smith, E. R. (1981). Beliefs about stratification. *Annual review of Sociology*, *7*(1), 29–56.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing* (pp. 262–272).

Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review*, *95*(4), 1031–1053.

Piketty, T., Yang, L., & Zucman, G. (2019). Capital accumulation, private property, and rising inequality in china, 1978–2015. *American Economic Review*, *109*(7), 2469–96.

Qin, B., Strömberg, D., & Wu, Y. (2018). Media bias in china. *American Economic Review*, *108*(9), 2442–76.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An r package for structural topic models. *Journal of Statistical Software*, *91*(1), 1–40.

Robinson, R. V., & Bell, W. (1978). Equality, success, and social justice in england and the united states. *American Sociological Review*, 125–143.

Wegener, B. (1991). Job mobility and social ties: Social resources, prior job, and status attainment. *American Sociological Review*, 60–71.

Whyte, M. (2010). *Myth of the social volcano: Perceptions of inequality and distributive injustice in contemporary china*. Stanford University Press.

Wu, X. (2019). Inequality and social stratification in postsocialist china. *Annual Review of Sociology*, *45*, 363–382.

Xie, Y., & Zhou, X. (2014). Income inequality in today' s china. *Proceedings of the National Academy of Sciences*, *111*(19), 6928–6933.

Zhang, Z., Yeung, J. W., & Yeun Kim, T. (2018). Rural to urban migration and distributive justice in contemporary china. *Asian and Pacific Migration Journal*, *27*(1), 80–100.

Zhao, Y. (2008). *Communication in china: Political economy, power, and conflict*. Rowman & Littlefield Publishers.

Zhou, X., & Xie, Y. (2019). Market transition, industrialization, and social mobility trends in postrevolution china. *American Journal of Sociology*, *124*(6), 1810–1847.

Appendix A Data

A.1 Data Cleaning

We clean our main dataset following several steps, and we list the number of articles removed.

1. Remove articles with length less than 10 characters (-1033)
2. Remove articles with not complete content (-4868)
3. Remove articles with duplicated content or duplicated date and title (-727)
4. Remove media containing less than 51 articles during 17 years (-1633)
5. Remove too professional or sport media (-17455)

We clean the placebo dataset following the previous 3 steps. After that, we only keep media are also in the main dataset. The sample size of raw placebo dataset is 383956, and the final sample size is 305581.

A.2 Variables

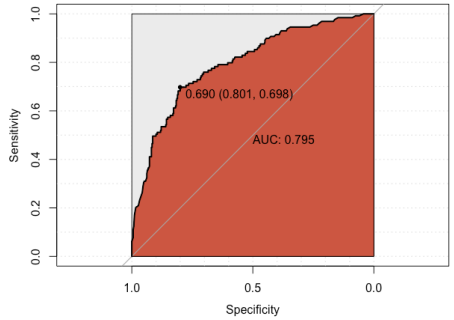
We list the variables from outer source as below:

Table 2: Variables List

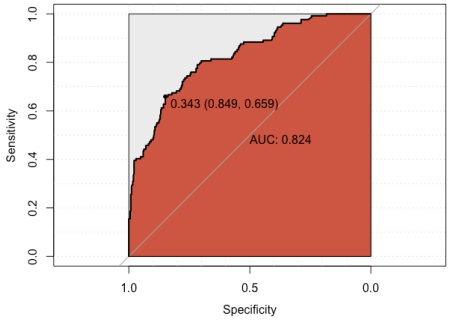
Variable	Description	Main Datasource
Province	Province of Media Location	Qin et al. (2018)
Party newspapers	Media is Supervised by Party Committee or not	Qin et al. (2018)
GDP per capita	GDP per capita (Yuan) of each Province in each Year (below is the same)	*Chinese Statistical Yearbook*
Population	Population size (ten thousand)	*Chinese Statistical Yearbook*
Urbanization	Proportion of Urban Residents	*Chinese Statistical Yearbook*
Student-Teacher Ratio of Senior High School	Number of Students/Number of Teachers of Senior High School	*Chinese Statistical Yearbook*
Number of Medical Bed per thousand people	Number of Medical Bed per thousand people	*Chinese Statistical Yearbook*
CPI	Consumer Price Index	*Chinese Statistical Yearbook*
Cases of Divorce per Million People	Cases of Divorce/ (Population Size $\times$ 100)	*Chinese Statistical Yearbook*
Family Size	Average of Household Size	*Chinese Statistical Yearbook*
Cases of Marriage	Cases of Marriage	*Chinese Statistical Yearbook*
Number of Newspaper	Number of Local Newspaper	*Chinese Statistical Yearbook*

Appendix B Method

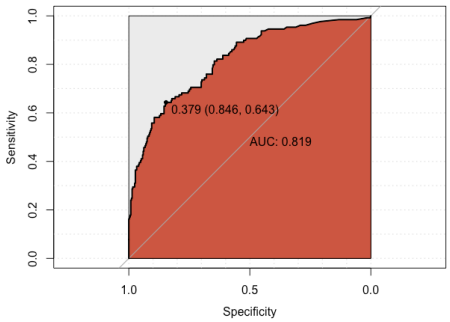
B.1 ROC Curve of five algorithms



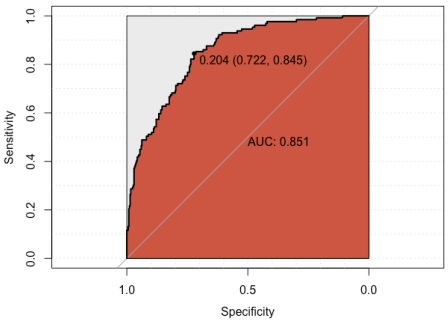
(a) Bi-LSTM



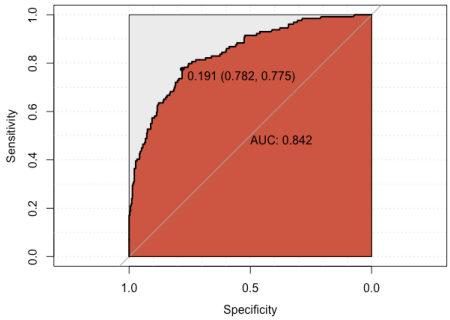
(b) LSTM



(c) BOW



(d) CNN



(e) GRU

Figure 15: ROC Curve

B.2 Performance of CNN and GRU

Table 3: Performance of CNN and GRU

Algorithm	CNN ($\tau = 0.5$)	CNN ($\tau = 0.204$)	GRU ($\tau = 0.5$)	GRU ($\tau = 0.191$)
Sensitivity	0.488	0.837	0.543	0.775
Specificity	0.935	0.722	0.911	0.782
Accuracy	0.820	0.752	0.816	0.780

Appendix C Data Analysis

C.1 Comparison of the Articles Number

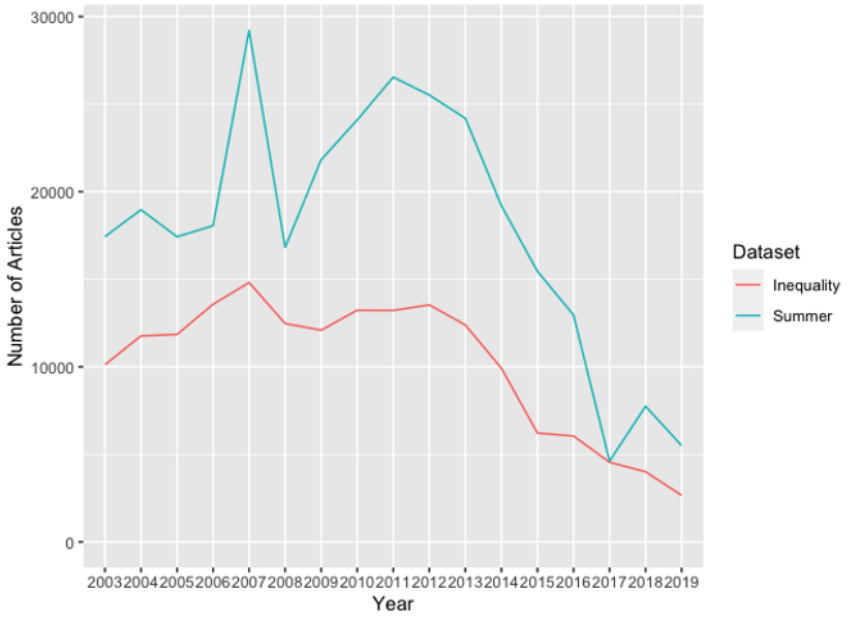


Figure 16: Number of Articles of Main Dataset and Placebo

C.2 Table1's Result in Main Dataset

Table 4: Local Features and Sentiment of the Articles

Variables Type	Independent Variables	Dependent Variables: Sentiment of the Articles
Socio-economic Development	GDP per capita	0.258
		(0.134)
	Population	-0.904**
		(0.267)
	Urbanization	-2.351***
		(0.581)
Public Service Provision	Student-Teacher Ratio of Senior High School	-0.010
		(0.009)
	Number of Medical Bed per capita	-0.034**
		(0.011)
Consumption	CPI	0.021
		(0.013)
Family	Cases of Divorce per Million People	0.064
		(0.247)
	Family Size	-0.321***
		(0.087)
	Party Newspaper	Controlled
	Year FE	Controlled
	Media FE	Controlled
	Observations	197806
	AIC	195670

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors are clustered at the media level. GDP per capita and Population are logarithm.

C.3 Localized Features and Party Moderation Effect on Topic Prevalence, with More Covariates and Less Observations

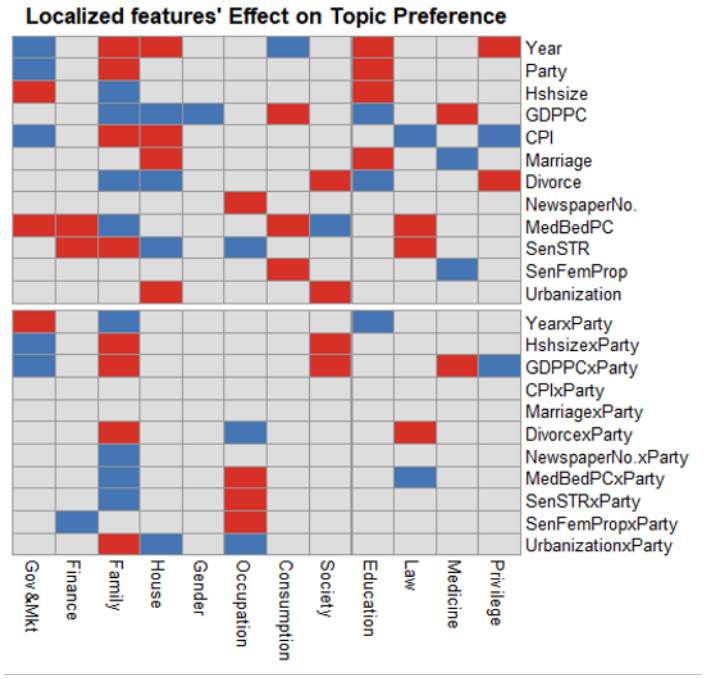


Figure 17: Localized Features and Party Moderation Effect on Topic Prevalence