Is Propaganda Front-Page News?*

Philine Widmer[1]

[1]*ETH Zürich, University of St.Gallen*

Abstract

In the online age, autocracies' censors must manage abundant information. At the same time, they face consumers with varying valuations for investigative content. This paper reveals a strategic pattern in the placement of news within Chinese online newspapers (2020-22). Studying over a million articles from 53 news outlets, I show that front-page articles are more likely to feature content favored by the government relative to articles published in other locations of news websites. Different text-based measures are employed to determine each article's alignment with the government's preferred content: Citing the government's press agency, Xinhua, makes an article approximately eight percentage points more likely to feature on the front-page. Similarly, a one-standard-deviation *decrease* in the resemblance with foreign content on China increases the front-page placement probability by 1.1 percentage points. Both theoretical and empirical evidence suggest that foreign information sources – costly but not impossible to access for an investigative minority of readers – influence domestic censorship strategies.

Draft as of 15 November 2023.

*I thank the Basic Research Fund (GFF) of the University St.Gallen for a research grant, as well as the Paris School of Economics for hosting me while working on this project. In particular, I thank for helpful comments by Elliott Ash, Roland Hodler, Molly Roberts, Arturas Rozenas, Dmitriy Vorobyev, Maiting Zhuang, and Ekaterina Zhuravskaya, as well as participants of the Monash-Warwick-Zurich Text-As-Data Workshop, the Democratic Backsliding and Electoral Autocracies SITE Conference at SSE, seminar participants at ETH Zurich, the Ludwig Maximilian University of Munich, the University of St.Gallen, the Paris School of Economics, and Sciences Po.

Email address: widmerph@ethz.ch.

1. Introduction

Throughout the twentieth century, many autocrats engaged in overt censorship.[1] Recently, however, subtler forms of censorship and repression, in general, have become more widespread (Guriev and Treisman, 2019). Broad access to information technologies has challenged traditional approaches to censorship (Guriev et al., 2021): Not only can users generate and disseminate content themselves, but also access foreign information, even in countries with rather sophisticated censorship infrastructure (NPR, 2017).

In such a context, keeping censorship as invisible as possible can be a viable strategy from an autocrat's perspective. News consumers who are not (too) aware of censorship or not particularly interested in investigative content may search for information beyond what's readily available.[2] As information gets less expensive and more abundant, individual attention becomes the binding constraint (Gleick, 2011). Moreover, if readers consider domestic news fully uninformative, they might consume less.[3]

Also, censorship is costly. It may lead to productivity-enhancing resources being suppressed (Egorov and Sonin, 2020). Hence, while achieving complete control may be difficult, the internet age offers plenty of opportunity for systematic and potentially covert manipulation (see Edmond, 2013). Accordingly, many current censorship methods resemble a "tax" on information, forcing news consumers to spend more time (or money) to access critical information (Roberts, 2018). Such notions of subtle censorship are not new but arguably more salient in the internet age.[4]

These considerations raise the question of what strategies regimes can employ to engage in subtle censorship. This paper sheds light on such strategies in *online newspapers* in China. Online newspapers are a relevant source through which people obtain political information in China.[5] The focus on China is warranted by its standing as the largest autocracy in the world and one of the countries with the highest censorship levels. In

[1] Examples are publicized killings of reporters or demonstrative burnings and bans of books.

[2] See, e.g., Chen and Yang (2019).

[3] See Kamenica and Gentzkow (2011), Gehlbach and Sonin (2014), and Qin et al. (2018). Also, if no bad news is published at all, citizens could lose trust in the domestic media (see, e.g., Shadmehr and Bernhardt, 2015) – lowering the reach of the propaganda.

[4] Roberts (2018) uses the term "porous censorship".

[5] Adults in China spend more minutes on the consumption of digital media than on traditional media (Statista, 2020). Hoelig et al. (2021) asked a representative sample of 1,617 internet users which type(s) of source(s) they would most likely consult for fact-checking if they came across controversial information about a leading politician in their country. Respondents named news websites the second-most consulted source (31%), after TV (58%).

the World Press Freedom Index by the Reporters Without Borders, China ranked fourth to last in 2020. Specifically, I show that front-page news stories are systematically less likely to feature content that is potentially politically sensitive relative to stories on the back-pages. I define front-pages as the landing pages and back-pages as other locations of the news websites.[6]

Comparing front-page to back-page news offers a suitable setting to study subtle censorship. Front-page placement is an intuitive proxy for an article's accessibility. If attention is limited relative to the available news, there is little incentive to search for slightly less accessible news items actively. In line with this intuition, Fedyk (2018) finds that equivalent financial news stories on Bloomberg have more impact on financial markets when they appear on the front-page. The eminence of front-pages also generalizes beyond newspapers. 90% of Google searchers do not go past page one of the search results (Buddenbrock, 2016). In the context of censorship, a consumer unaware of subtle censorship possibly occurring might even expect front-page news to be more informative as (s)he trusts the newspaper to place information correctly: Beyond its literal meaning, the term "front-page news" also stands for particularly salient news. Such a consumer likely reads more front-page rather than back-page news. The consumer could also be aware of potential censorship (on some level of consciousness) but not value more critical information enough to incur the cost of searching beyond the most accessible news items.

This intuition is reflected in the theoretical model (Section A). It describes a country with low media freedom. Readers choose between domestic front-pages, domestic back-pages, or foreign news. They incur the lowest cost by consulting the most readily available source: domestic front-pages. Reading domestic back-pages comes with slightly higher costs – due to higher attention costs, for instance. The highest costs are those for foreign news (for instance, because they must download a VPN). Importantly, not all readers value investigative (and thus potentially sensitive) news equally. Investigative content is understood as news that could be politically sensitive – that is, news that deviates from pure propaganda.[7]. The readers' different valuations are empirically plausible: Chinese users willing to make an effort to jump the Great Firewall tend to be a well-educated minority more likely to be interested in politics than the general popula-

[6]See Section B.2 for a screenshot of a front-page.

[7]Mere assertions of the leader's competence are an example of pure propaganda (see, e.g., Gehlbach et al., 2022).

tion (Roberts, 2018).[8] Specifically, two reader types are assumed: those with low and those with high valuations. Readers maximize their utility, taking into account their valuation of investigative news and the cost of accessing it. The propagandist controls the domestic media market and chooses the level of potentially sensitive news on domestic front-or back-pages. Its goal is to minimize the overall consumption of investigative news because it could have destabilizing effects (e.g., by challenging the worldview curated by the government). The model suggests that content differences on front- and back-pages occur because of two main forces acting together. First, with porous censorship, high-type readers have the outside option of consulting critical sources, such as foreign news. This pressures the propagandist to provide content on issues it otherwise would prefer not to. Second, consumers are heterogeneous, and many less-investigative readers only consume readily available news. This gives rise to an equilibrium where the propagandist provides the less attentive majority with highly government-aligned news on the front-pages. At the same time, some potentially more sensitive news is provided on back-pages, to cater to high-valuation readers.

My empirical analysis is based on one million articles from 53 Chinese online news outlets.[9] For every outlet, I compile all article links from 2020 through 2022. For each article, I collect information on the mentioned individuals, organizations, geographic locations, or events. That is, I know which "named entities" are mentioned. These annotations come from state-of-the-art neural network algorithms.[10] Moreover, I query whether an article was featured on the front-page or in a different location of the news website. The front-page is defined as a given outlet's landing page. Then, I measure how much the article content is aligned with the regime's preferred content in different ways.

The first and simplest approach is an indicator of whether an article cites Xinhua. Xinhua is the official press agency and a key propaganda instrument (see Qin et al., 2018). The second approach is a generalization of the Xinhua indicator – it builds on

[8] Along these lines, Guriev and Treisman (2018) argue that an autocrat's repression strategy depends on the size of the informed elite (and how easily the regime can control the media). With a large informed elite, democracy may become the only cost-effective solution, while outright repression may be cost-effective in undeveloped contexts. At intermediate elite size levels, the autocrat's strategy will depend on how effectively they can control what information ordinary citizens obtain.

[9] The list of news outlets is compiled by combining information from different sources (notably Qin et al., 2018; ABYZ News Links, 2021, and own web research; see Section B.1) for details.

[10] In Natural Language Processing (NLP), Named Entity Recognition (NER) identifies, in a given text, entities that belong to named categories.

the intuition that some entities' mentions are more predictive of news the government favors. For every article, I predict whether an article resembles the domestic or the foreign perspective (i.e., how the New York Times or the BBC report on China). This prediction comes from a supervised machine learning model, where the input is a matrix of the aforementioned named entities. One crucial advantage of using the named entities as tokens for prediction is that they can very tractably be linked across languages: One can easily compare the domestic news in Mandarin to the foreign news in English. That is, whether an entity like Xi Jinping is mentioned in an article is relatively easy to assess across language borders. With other techniques, like N-grams or embeddings, it is challenging to establish whether "the same" content was referred to since translations can be more ambiguous. The machine learning model confirms that mentioning Xinhua is highly predictive of the domestic perspective. Meanwhile, frequent mentions of "Hong Kong" or "Taiwan" are indicative of foreign news. For China, one can interpret content strongly resembling foreign news as content that the regime generally does not want citizens to read – it takes costly measures to block access to most foreign news sources (Roberts, 2018). The New York Times and the BBC were entirely inaccessible from Mainland China without using foreign tools to circumvent the Great Firewall during the study period (GreatFire.org, 2022). At the same time, Chinese governments, throughout all levels (from national to county), directly own all general-interest newspapers – rendering the domestic perspective relatively aligned with the government. A potential concern is that *whether* an entity is mentioned only captures part of the story: two articles could discuss "Hong Kong" or "Taiwan" very differently. It is reassuring that the mentions of the entities alone (abstracting from framing) already predict whether an article is from a domestic or foreign outlet with 91% accuracy. This figure suggests that my prediction-based measure captures relevant information about an article's content, as plausibility and robustness checks further confirm. One can interpret this such that mentioning certain entities is, per se, sensitive. That is, this paper does not claim that an article on an entity frequently mentioned in foreign news would *necessarily* challenge the government's preferred view on the related policy issue. For instance, back-page news on "Taiwan" could easily fully embrace the One China principle.[11] However, the government may prefer to keep the issue's salience low, not pushing information to readers unaware

[11]This principle emphasizes that Taiwan is an inalienable part of China.

of or not interested in it.[12] The third approach to measure an article's alignment with the government's preferred content investigates whether articles mentioning the leader ("Xi Jinping") are more likely to appear on the front-page if their overall sentiment score is more favorable.[13]

My results reveal that articles featured on front-pages have higher values of these alignment measures. This association is already visible in the raw data for Xinhua mentions and the machine learning-based measure. For all three measures, my main results – where I exploit variation within days, topics, and outlets – confirm it. That is, for a given newspaper and day, I compare articles that cover similar topics (e.g., "international business") but differ in how strongly they align with the regime's perspective. This approach is meant to render articles otherwise comparable. One remaining concern is that some unobserved dimension makes news items "breaking", such that they appear in international outlets and on the front-pages of domestic outlets. However, this would impose a downward bias; my estimates would represent a lower bound.

In my main specification, citing Xinhua makes an article approximately eight percentage points more likely to feature on the front-page. Similarly, when turning to the machine learning-based measure, moving from a complete predicted resemblance of the foreign perspective to a full resemblance of the domestic perspective increases an article's front-page probability by seven percentage points. The extreme counterfactual of complete foreign vs. domestic alignment is rare. The median of the alignment measure in domestic news lies at 92.6%. So, put differently, a one-standard-deviation increase in resemblance with the domestic perspective increases the front-page placement probability by 1.1 percentage points. Finally, for the sentiment-based measure, an article featuring the leader with the maximum sentiment score of one is 16.3 percentage points more likely to appear on the front-page, relative to an article with a minimum sentiment score of zero. Empirically, articles about the leader never come with a sentiment below 0.5. Thus, note that a one-standard-deviation increase in sentiment increases the front-page probability by 1.4 percentage points. A heterogeneity analysis investigates regional differences in entities' political sensitivity. It replicates the sentiment-based analysis from the main results but focuses on provincial governors and Communist Party Secretaries. It shows that articles covering a local leader from the province where the newspaper is

[12] This question is further discussed in Section 3.2.

[13] The sentiment measure also comes from a state-of-the-art neural network algorithm.

based are more likely to feature on the front-page if the sentiment is higher. There is no similar front-page premium for positively toned articles when they cover a local leader from another province.

While the main results establish that Chinese online outlets provide the least politically sensitive news items on their front-pages, they also raise the question of why state-controlled media outlets do not simply provide equally sensitive content on the front- and the back-pages. Turning to mechanisms, I further investigate the hypothesis from the theoretical model: that the differential is influenced by the availability of outside options, such as foreign news, especially for investigative readers. In the first mechanism analysis, I build a dictionary of entities that Chinese consumers read about on foreign platforms but that have explicitly been subject to censorship endeavors by the Chinese government. As a proxy for the (investigative) consumers' interest, I use the most-visited Chinese-language Wikipedia entities. During the observation period, Wikipedia was only accessible from Mainland China via a VPN. To identify possibly sensitive entities, I combine keyword searches on the entities' Wikipedia pages and manual annotations. These searched-for but sensitive entities are more likely to feature on domestic back-pages if they feature in the domestic news at all.

In a second analysis, I investigate a shock to the political sensitivity of certain entities. Specifically, I present an event study around a major geopolitical crisis falling into my observation period: the crisis in Ukraine that led to the invasion by Russia on 24 February 2022. The inception of hostilities increased the sensitivity of actors such as "Russia" or "Vladimir Putin" from the Chinese government's perspective: Commentators and scholars posit that the conflict introduces considerable political and economic complications for China, thus making information management challenging (e.g., deLisle, 2022). At the same time, news pressure on such entities has increased considerably in international media. Thus, the crisis outbreak is a suitable setting to study because more investigative individuals could easily find plenty of information pieces, including some on China's role, on foreign websites (e.g., criticism of China's support for Russia or allegations that China was informed about the invasion in advance). I show that content covering "Russia" "Ukraine," "Vladimir Putin," and "Volodymyr Zelenskyy" becomes more likely to be back-page news after the deployment of Russian troops to the Ukrainian border in 2021. This pattern emerges when only looking at the Chinese news – when comparing the front-page probability of articles mentioning a related entity before and after the escalation. It is also confirmed in an event study where articles in the New

York Times and the BBC serve as controls.

This paper makes several contributions. First, I add to a broad body of work on censorship. There is a broad consensus that media freedom matters for democratic decision-making in the spirit of a "fourth estate". Consistent with this view, numerous contributions suggest that censorship is associated with worse outcomes for citizens.[14]

Recent work on censorship has addressed information-rich settings.[15] Specifically, I complement this work by providing evidence of subtle censorship and its "mechanics" in Chinese online newspapers. So far, systematic evidence of specific strategies for online news outlets has, to my knowledge, been absent. There are several pieces of empirical evidence for such content manipulation by political elites on social media. Popular techniques include drowning out critical content by distracting posts (King et al., 2017), creating fake accounts (e.g., to promote pro-regime content while pretending to be a regular citizen), or amassing fake followers (Bradshaw and Howard, 2019). For newspapers, anecdotes of more regime-critical articles placed on the last pages of newspapers have been reported (Roberts, 2018). I also contribute to broader debates on autocratic regimes shifting away from or complementing overt censorship and repression strategies. Recent work argues that autocrats are increasingly subverting political institutions and managing information flows in more covert ways (instead of publicizing brutality to deter opponents). Besides information technologies, increasing education levels and economic development likely contribute to this shift (Guriev and Treisman, 2019; Egorov and Sonin, 2020). My paper highlights how (subtle) censorship can be shaped by access to foreign information sources.

Second, my work documents product differentiation and placement *within* newspapers in a censored context. The literature on censorship has been focused on cross-country differences, regional differences, or market-driven differentiation between newspapers. For example, prominent measures of media freedom rank countries (see the Freedom of the Press Index by Freedom House and the Press Freedom Index by Reporters Without Borders).[16] Qin et al. (2018) show that Chinese newspapers diversify into more

[14] See, e.g., Brunetti and Weder, 2003; Djankov et al., 2003; McMillan and Zoido; 2004; Choi and James, 2007; Leeson, 2008; Bhattacharyya and Hodler, 2015.

[15] For example, see the discussion of Gleick (2011), Edmond (2013), King et al. (2017), Roberts (2018), Guriev and Treisman (2019), Bradshaw and Howard (2019), or Guriev et al. (2021) earlier in this Section.

[16] Estimates suggest that more than 100 (200) organizations worldwide engage in some form of media freedom assessment, evaluation, or promotion (Becker and Vlad, 2009 and Schneider, 2014, respectively).

propaganda-oriented (mouthpiece) outlets and outlets reporting more freely due to economic competition around advertising revenues. Similarly, Zhuang (2022) finds that local newspapers under-report corruption scandals concerning politicians from their own province, especially when they do not rely on advertising revenue. Some contributions analyze which contents are subject to censorship. For example, apart from preventing criticism, elites may be concerned about other content, e.g., related to social mobilization (see King et al., 2013; Qin et al., 2017; Wu et al., 2021).[17] Similarly, regimes may allow critical reporting on lower-level officials to keep them in check (Egorov et al., 2009; Lorentzen, 2014; Qin et al., 2017; Wu et al., 2021). My work provides some specifics on censored content (i.e., on entities considered sensitive in China in 2020 through 2022).

Lastly, this paper connects to limited attention – a concept from cognition suggesting that individuals do not adequately process all pieces of information once exposure to information surpasses a certain threshold. Psychologists and neuroscientists have long studied it – starting with Miller (1956). In economics, there is both theoretical and empirical work showing that (limited) attention can drive various outcomes related to competition (e.g., Falkinger, 2008), inequality (Banerjee and Mullainathan, 2008), the financial markets (beginning with Hirshleifer and Teoh, 2003), or foreign aid (Eisensee and Strömberg, 2007). Also, several contributions highlight how political (or military) action is timed to be taken while the groups with whom these actions are not popular are distracted (Balles et al., 2018; Durante and Zhuravskaya, 2018; Djourelova and Durante, 2021). Specific to censorship, King et al. (2017) show that the Chinese government fabricates news stories on social media to distract from an engaged debate.

The remainder of the paper is structured as follows. Section 2 describes my data. Then, Section 3 provides more details on measuring the alignment of an article's content with the regime's perspective. Section 4 explains the empirical framework. Next, Section 5 presents the main results, while Section 6 illuminates the mechanisms. Finally, Section 7 concludes.

[17] To a certain degree, content on mobilization may also be allowed. Such content can inform the authorities on upcoming protests (and, thus, help them to understand citizens' dissatisfaction or to police such events more effectively).

2. Data

2.1. Observation period and units of observation

The observation period covers 17 January 2020 through 31 December 2022. The beginning of the observation period is given by the availability of large-scale Chinese online news article annotations (see Section 2.3 below). The end of the observation period marks the beginning of the data collection for this working paper version.

In the main specification, the observation unit is article i published on date d, on topic t, and in news outlet j. Observations can, thus, be thought of as $idtj$.

2.2. Seed list of Chinese online news outlets

I combine two sources to compile a list of relevant Chinese news outlets. Given that I look at online news, I define an outlet as a domain, such as people.com.cn. First, I collect all news links assigned to the Chinese market by ABYZ News Links (2021). ABYZ News Links is a collection of links to websites with "news" content; they define "news" broadly (some of their listed websites cover, for example, primarily sports). I cross-check this initial list with the Chinese outlets considered by Qin et al. (2018).[18] All in all, I proceed with a list of 53 outlets. Appendix B.1 details the sample construction.

2.3. Article links and their content

Then, I query the article links published by any of the outlets in my seed sample from the Global Database on Events, Language, and Tone (GDELT).[19]

Named Entity Annotations Specifically, I first query GDELT's Global Entity Graph (GEG).[20] Through the GEG, I obtain, for every article published by one of the Chinese outlets, the *named entities* mentioned in this article. Named Entity Recognition (NER) is an approach from Natural Language Processing (NLP). A text piece is parsed automatically to find entities that belong to named categories. These can be (prominent)

[18] As they do not specifically focus on online news, Qin et al.'s replication folder does not include the domains of the outlets. Accordingly, for every outlet from their replication folder, I confirm whether it has an online presence and identify the corresponding domain (by manual web search).

[19] The Global Database on Events, Language, and Tone (GDELT) monitors news from around the world by scraping content from online newspapers, TV station websites, and other online news sources.

[20] For all the outlets that GDELT tracks, they scrape all news articles. The GEG takes a large random sample of all articles.

individuals, organizations, geographic locations, or events. Consider "Many countries have seen virus outbreaks, and so has China." Here, the named entity is "China". Of course, named entities can also be more localized (such as the "Huoshenshan Hospital" in Wuhan). The entity annotations queried through the GEG are based on state-of-the-art neural network algorithms, as implemented by Google's Cloud Natural Language API. A convenient feature of these annotations is that, for frequent entities, ambiguities are resolved such that alternative names ("President Xi" and "Xi Jingping") are assigned to one reference and connected to the corresponding Wikipedia entry. By exploiting this feature, I focus on entities with a Wikipedia entry (in any language).[21] The annotations for Chinese have been available since 17 January 2020. I collect information from this date onward, but data is available rather sporadically before 1 February 2020.[22] In sum, for each news article, I know the named entities mentioned therein.

Front Page Information Next, I query GDELT's Global Front-Page Graph (GFG). Like other GDELT functionalities, it indexes news links. However, the GFG exclusively scrapes front-pages. Therefore, I match the links collected from the GEG to the GFG index. Like this, for every article, I know if it ever appeared on the front-page of the corresponding outlet (or only in another website location). The GFG comes with a time resolution of one hour. Hence, I will surely capture an article's front-page placement if it remains there for at least one hour. The outlets covered by GDELT change over time. Thus, I only keep outlet-dates for which at least one front-page article is available. Thereby, I rule out that outlets not (yet) covered by GDELT on a given day would falsely have all their articles labeled as back-page articles.[23] Figure A1 shows a screenshot of a front-page.

Topics For each news article, I wish to infer its topic. To this end, I build a topic model using the named entities (see paragraph on named entity annotations above) as input tokens. Most articles only cover a handful of entities. The average number of

[21] Likely, with this filter, I miss some highly localized entities (i.e., those without any Wikipedia entry). As emphasized in the Results Section, my results are robust to excluding local news.

[22] Thus, I cannot conduct event studies around the much-debated Wuhan lockdown in January 2020 because of data unavailability.

[23] For instance, the outlet "yangtse.com" is only covered from 26 October 2020 onward. Hence, if I did not drop outlet-days with no front-page articles, I would falsely classify all articles by this outlet from before 26 October 2020 as back-page articles.

entities is 8.1. Standard topic models (such as the popular latent Dirichlet allocation algorithm, Blei et al., 2003) come with limitations when the number of tokens per textual observation (here, per article) is low. Therefore, I use a Dirichlet multinomial mixture model optimized for sparse (yet high-dimensional) use cases (Yin and Wang, 2014). I assign every article to one (main) topic. The optimal topic number is 95.[24] As the model clusters the articles into topics but does not label the topics, I manually add labels. Almost all the topics can be labeled intuitively (see Appendix B.3 for details on the topic labels). The topic model allows a given named entity to be part of different topics. For instance, the "United States" is a frequent token in both the topic on "International Relations" and the (postponed) "Olympics 2020". Taken together, for every article, I predict its main topic. It will be used as a control in the main analyses.

Also, for robustness and heterogeneity analyses, I distinguish between international news and other news. I separate these news types via the topic model annotations. Specifically, I annotate all topics as international news prominently featuring foreign entities. All other topics are labeled as (primarily) relevant at the national or local level. Appendix B.3 shows the topic annotations and provides plausibility checks regarding the news levels derived from the topics.

Regional Leaders To analyze entities whose political sensitivity varies regionally, I create indicators for articles mentioning the governor or the secretary of the first subnational administrative unit (typically a province) where the newspaper is based: $OwnGovernor_{idtj}$ and $OwnSec_{idtj}$. Moreover, I build an indicator for mentions of the governors or secretaries of other provinces: $OtherGovernor_{idtj}$ and $OtherSec_{idtj}$. Table A2 shows the names of the relevant personalities for each province in my data.

2.4. Outlet metadata

Traffic Data I obtain traffic information (specifically, estimates of website reach) for each outlet from Amazon's Alexa Web Information Services (AWIS). Specifically, I have access to the average reach for 2018 through 2020 for each domain. AWIS provides the web traffic data normalized by one million active website visitors (on all domains tracked by AWIS) during the measurement time. Consider people.com.cn and its average

[24] The topic number is passed as a hyperparameter in this approach. I started with 120 topics. However, the number of clusters that actually contain documents can vary. Here, 95 topics are populated.

reach figure of 1,911. For every million unique visitors present on any website worldwide between 2018 and 2020, an average of 1,911 went to people.com.cn. For comparison, the page reach figure for google.com is 574,893. I use the traffic data in robustness checks – to weight each article by the circulation of its outlet.[25]

3. Measuring Alignment of Content with Regime Narratives

To understand whether political motives drive within-outlet product placement (front-page vs. back-page news), I need to measure how well an article aligns with the regime's perspective. First, I use a simple proxy for regime-friendly reporting used in the literature: mentions of the official state press agency Xinhua. Second, I devise a machine learning-based measure that captures how likely an article is to come from a Chinese outlet instead of a foreign outlet (specifically, the New York Times or the BBC). I also present a third measure, namely the sentiment of an article that mentions the leader (Xi Jinping). I provide more details on these measures for article alignment in the following.

3.1. Mentions of Xinhua

Xinhua is the official (i.e., state-run) press agency in China. It is a key instrument in enforcing the communist party's propaganda (Qin et al., 2018). Building on the named entity annotations, I construct an indicator for whether an article mentions the "Xinhua News Agency," $Xinhua_{idtj}$. I assume that mentions of Xinhua proxy for articles more aligned with regime narratives (following Qin et al., 2018).

3.2. Machine learning: similarity with Chinese vs. foreign news

The machine learning-based model generalizes the intuition of the Xinhua-based indicator. Mentions of some entities might be more indicative of a foreign perspective (e.g., a cite of Reuters). In contrast, references to entities like the aforementioned Xinhua News Agency (or the second largest state-owned news agency, China News Service) might indicate a view more aligned with the Chinese regime. For every article, the model

[25] In using web traffic figures, I follow Matter and Widmer (2021). Using traffic figures (as opposed to revenues, for example) is advantageous because online news is often freely accessible. Also, inferring reach from advertisement spending would not be straightforward. The data availability (the aforementioned average for 2018-2020) is also given by Matter and Widmer.

provides a probability that the news is Chinese (as opposed to foreign). I assume that articles with a higher predicted probability of being Chinese news are (in expectation) more regime-friendly. For China, distinguishing between the domestic and the foreign perspective can be seen as an expression of what the regime does (not) want people to read since the government blocks access to most foreign news sources (through the "Great Firewall"; Roberts, 2018). The predicted probabilities are built as described in the following.

3.2.1. Balanced dataset of Chinese and foreign articles

Similar to the Chinese article links and their content, I also collect all article links published from 2020 to 2022 by the New York Times and the BBC. As of the writing of this article, both sources were entirely inaccessible from Mainland China without using foreign tools aimed at circumventing the Great Firewall (such as virtual private networks, VPN).[26] As for the Chinese articles, I retrieve the links and their named entity annotations from GDELT (see Section 2.3).

In total, there are 46,946 articles on China. Articles on China are considered as such if they mention either "China" or "Chinese". For a balanced sample to train the predictor, I randomly draw the same number of articles from the Chinese news articles. The dataset to build the classifier contains $M = 93{,}892$ articles. Every article comes with a true label, D_{idtj}, indicating whether it was featured by a domestic ($D_{idtj} = 1$) or foreign ($D_{idtj} = 0$) source.

Let E_{idtj} give the entities mentioned in article $idtj$. In total, 110,300 entities are mentioned at least in one of the 93,892 articles used to build the classifier. The nth element in vector E_{idtj} is 1 if the nth entity is mentioned in the article once or more and 0 otherwise.[27] I filter for entities mentioned in at least 50 articles (3,018 unique entities). I also impose that an entity must feature *both* in Chinese and foreign news (984 entities): Entities that never appear in the Chinese news are not useful to establish, across Chinese articles, the relative similarity with foreign perspectives. I drop mentions of the names of the newspapers in my sample to avoid mechanical predictions.[28] Finally, the input for

[26] Regular checks were made between early 2021 and mid-2022 (GreatFire.org, 2022). For more on VPNs in China, see, e.g., NPR (2017).

[27] E is similar to a document-term-matrix but captures *whether* an article mentions an entity (rather than how many times).

[28] "The New York Times" is necessarily mentioned frequently by the Times itself.

the classifier is E, a $M \times 968$ matrix, and the target is D, a vector of length M. I split the data into 75% training data and 25% test data to train the classifier.

3.2.2. Classification model

My classification method is a penalized logistic regression (Hastie et al., 2009). I parametrize the probability that an article i is from a domestic outlet (as opposed to a foreign one) as:

$$\widehat{D}_i = \Pr[Di = 1|E_i] = \frac{1}{1 + \exp(-\psi' E_i)}$$

where ψ is a vector of coefficients for each entity. The penalized logistic regression model picks ψ to minimize the cost objective

$$J(\psi) = -\frac{1}{M^*} \sum_{m=1}^{M^*} \Big(D_i \log(\widehat{D}_i) + (1 - D_i) \log(1 - \widehat{D}_i) \Big) + \lambda |\psi|_2 \tag{1}$$

where M^* gives the number of documents in the training sample.

The rightmost term in Equation (1) is the regularization penalty. Specifically, it is the L2 (Ridge) regularization, as indicated by the L2 norm $|\cdot|_2$. The penalty addresses over-fitting of the training set by shrinking less predictive coefficients towards zero. Regularization strength is calibrated by the hyperparameter $\lambda \geq 0$. λ is selected using a five-fold cross-validated grid search in the training set. Next, I calculate the accuracy of the classifier in the test data. It amounts to 91% (+/- 0.01).[29]

Figure 1 plots all entity coefficients with an absolute value larger than one. Entities shown in red are predictive of being from a Chinese outlet (positive coefficient), and those shown in blue resemble foreign news. Each entity's font size is proportional to its frequency in the Chinese news – to illustrate how often an entity is actually used to predict an article's alignment. Among the entities most predictive of Chinese news are the "Xinhua News Agency" and the "China News Service". The data, thus, validates the Xinhua measure discussed in Section 3.1. Conversely, mentions of "Reuters" or "Agence France-Presse" are indicative of foreign news. Among the rather frequent entities predictive of the Chinese perspective is also the "Huoshenshan Hospital". Anecdotal evidence suggests that the fast construction of this hospital (and the "Leishenshan Hospital") was presented as a miracle in Chinese Covid-19 reporting – to underscore the authorities'

[29] I calculated the error through five-fold cross-validation in the test set.

Figure 1: Entities Predictive of Domestic vs. Foreign News and their Occurrence in Domestic News

Notes: The entities with the largest absolute coefficient to predict similarity with the domestic vs. the foreign perspective (see Section 3.2). Specifically, I filtered for entities with an absolute coefficient size > 1 and that appear at least 100 times in the Chinese corpus. Red entities are most predictive of the domestic perspective (positive coefficient); blue entities of the foreign perspective (negative coefficients). The entity's font size correlates with its frequency in domestic news – that is, how often the entity is used to predict an article's similarity $\widehat{D}_{idtj}$.

swift and competent reaction (Molter and DiResta, 2020). Along these lines, mentions of "Zhong Nanshan" also predict the Chinese perspective. Nanshan, a medical professional, was celebrated as "the People's Hero" in the fight against the pandemic (Xinhuanet, 2020). Moreover, mentions of the "Great Hall of China" (a government building of high symbolic value) or the "Economy of China" are emphasized in domestic news. Both "Hong Kong" and "Taiwan" mentions indicate a resemblance to foreign news.

The high prediction accuracy suggests that the entities discussed in the news carry information. That is, even abstracting from framing, some entities are inherently sensitive. Take the example of "Taiwan". Writing a convincing propaganda piece about the One China principle is more involved than a laudation of the leader (e.g., potential questions about the government's narrative have to be anticipated). Also, more generally, the government might not want to draw attention to certain issues, especially if people would not think about them unless actively informed about them. Of course, whether an entity is considered sensitive in a given context is an empirical question, and the machine learning model provides a data-driven approach to infer such entities.

Lastly, I use the classifier to predict $\widehat{D}_{idtj}$ for the entire sample of Chinese articles.

Figure B1 shows the distribution of these predictions. The distribution shows more mass around higher probabilities in line with the high prediction accuracy. In sum, for every article, I now have a measure of how much it resembles foreign as opposed to domestic perspectives on China.

3.3. Sentiment of articles mentioning the leader

Building on the named entity annotations, I construct an indicator for whether an article mentions the country leader Xi Jinping: $Leader_{idtj}$. I will interact this indicator with the article sentiment (also available from the Global Entity Graph). It is a state-of-the-art neural network-based sentiment measure lying between 0 and 1: $Sentiment_{idtj} \in [0, 1]$.[30] Higher values indicate more positive sentiment. I assume that more positive articles on the leader reflect higher alignment with regime narratives (all else equal; see the outlet, day, and topic controls in the main analysis).

3.4. Summary statistics

Table 1 shows the summary statistics. The sample contains 1,090,601 articles. Nearly half, specifically 49.2%, of articles appear on the front page at some point. 9.5% of articles cite Xinhua. The predicted similarity with domestic news ($\widehat{D}_{idtj}$) is 89.2% on average. The leader features in almost 11% of articles. Across the corpus, sentiment averages at 59.2%.

The number of distinct entities discussed in each article is, on average, 8.06; it is used as a control. An article typically comes from an outlet with a reach of approximately 1580 visitors (per one million active website visitors).

Local news constitutes a significant portion: 39.3% of articles belong to this category. Similarly, international news is represented in 27.7% of articles. Accordingly, the shares of the three news levels (local, national, and international) are relatively balanced. Around 2.8% of articles appear during high-level political meetings. Of all articles, 0.2% cover the governor of the province where the respective newspaper is based. The same figure also emerges for the local Communist Party secretary. In comparison, 0.5% and 0.8% of articles cover governors or CCP secretaries of other regions, respectively.

Sensitive or partly censored (as defined by Wikipedia) appear in 1.6% of articles, as $Sensitive\_Entity_{idtj}$ conveys. Mentions of Russia and Ukraine are relatively low, with

[30] As the other information available through the GEG, the sentiment measure is based on Google's Cloud Natural Language API.

articles featuring Russia in 3.1% of cases and Ukraine in 0.7%. Putin appears in 0.5% of stories, Zelenskyy in 0.03% (rounded to zero in the Table). More details on the variables in this paragraph are provided when discussing the mechanisms (Section 6).

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
$Frontpage_{idtj}$	0.492	0.5	0	1	1090601
$Xinhua_{idtj}$	0.095	0.293	0	1	1090601
$\widehat{D}_{idtj}$	0.892	0.155	0	1	1090601
$Leader_{idtj}$	0.107	0.309	0	1	1090601
$Sentiment_{idtj}$	0.592	0.089	0.05	0.950	1090601
$Entity\_Count_{idtj}$	8.061	6.55	1	30	1090601
$Reach_j$	1580.335	3400.375	0.489	11557.433	1090601
$Local_{idtj}$	0.393	0.488	0	1	1090601
$International_{idtj}$	0.277	0.447	0	1	1090601
$Political_{idtj}$	0.028	0.165	0	1	1090601
$OwnGovernor_{idtj}$	0.002	0.043	0	1	1090601
$OtherGovernor_{idtj}$	0.005	0.07	0	1	1090601
$OwnSec_{idtj}$	0.002	0.041	0	1	1090601
$OtherSec_{idtj}$	0.008	0.09	0	1	1090601
$Sensitive\_Entity_{idtj}$	0.016	0.124	0	1	1090601
$Russia_{idtj}$	0.031	0.173	0	1	1090601
$Putin_{idtj}$	0.005	0.07	0	1	1090601
$Ukraine_{idtj}$	0.007	0.084	0	1	1090601
$Zelenskyy_{idtj}$	0	0.016	0	1	1090601

Notes: See Section 3.4 for details.

4. Empirical Framework

4.1. Specification

I use a linear probability model to estimate whether an article that aligns (more) with the domestic perspective is more likely to be put on the front-page.

$$Frontpage_{idtj} = \alpha_j + \beta_d + \gamma_t + \phi\ Aligned_{idtj} + \Omega Controls_{idtj} + \vartheta_{idtj} \tag{2}$$

$Frontpage_{idtj}$ is an indicator for whether article i from outlet j published on day d covering topic t is ever on the front-page. α_j are outlet fixed effects (FE), β_d day-year FE, and γ topic FE. In robustness checks, I replace the day-year FE by week-year FE.

$Aligned_{idtj}$ captures whether an article is relatively more aligned with the government view/the domestic perspective. In particular, $Aligned_{idtj}$ will be $Xinhua_{idtj}$ (see Section 3.1) or D_{idtj} (see Section 3.2). Ω is a coefficient vector for control variables (mostly the number of entities mentioned in the article, as a proxy for whether it is a more complex, possibly longer article). ϑ_{idj} is the error term. I multi-way cluster errors for dates, topics, and outlets.

Along the same logic, the equation for the sentiment-based alignment measure is:

$$\begin{aligned} Frontpage_{idtj} = \alpha_j + \beta_d + \gamma_t + \delta\ Leader_{idtj} + \zeta\ Sentiment_{idtj} + \\ \phi\ Leader_{idtj} \times Sentiment_{idtj} + \Omega Controls_{idtj} + \vartheta_{idtj} \end{aligned} \tag{3}$$

Likely, numerous factors determine whether an article ends up on the front-page, most notably its "newsworthiness". In expectation, the newsworthiness of different entities changes over time (for example, the Olympic Games are arguably more salient right before or while they take place). By zooming in on time windows as narrow as a day (via the time FE), I try to keep the pool of potentially newsworthy entities constant. Moreover, whether articles align with foreign news might vary across topics (e.g., domestic news covering international politics might be more aligned with foreign news on China and global politics by referring to entities such as the United Nations). Also, different outlets might vary in what they generally deem newsworthy (e.g., specialization in more localized or international news). Therefore, I also include topic FE and outlet FE.

4.2. Interpretation

We would like to interpret ϕ as the marginal effect of the consistency with the regime's perspective on the front-page placing of articles – excluding other characteristics potentially correlated with resembling foreign news. That is, we want to compare otherwise similar articles, only differing in how well they align with the regime's preferred content. The three fixed effects render articles comparable (on the same day, on the same topic, and from the same outlet). Yet, a remaining concern is that, among similar articles, other characteristics drive both their foreign perspective alignment and their front- or back-page placing. There are two scenarios.

One confounder is the degree to which a story is considered "breaking" or newsworthy – it could influence both front-page placements and foreign interest. If domestic and

foreign outlets have the same understanding of newsworthiness, this would lead to a downward bias on my ϕ estimate: more breaking stories could be on the front-page and also appear in foreign news (i.e., only major stories make it beyond the country border). Then, any evidence of foreign-resembling news stories on the back-page (this project's main result) would represent a lower bound.

One news category where domestic and foreign newsworthiness perceptions could differ is local news, which might be picked up by domestic outlets prominently but not by foreign news. To address this, robustness checks show that excluding local news does not change the results. Also, a simple Chi2 analysis on which entities are most associated with back-page appearance provides suggestive evidence that local stories tend to be back-page news even domestically: As Figure C.2 shows, mentions of entities containing the phrase "district" are particularly frequent on back-pages (in China, districts are administrative units below the provincial level).

Section 6 on Mechanisms provides evidence that differences in the news content measures are consistent with differences in political sensitivity and that the front- vs. back-page placements reflect, indeed, censorship (see Section 6).

5. Main Results

Table 2 presents estimates of ϕ in Equation 2 (columns 1 and 2) and Equation 3 (column 3). In all columns, the dependent variable captures whether an article ever appeared on the front-page: $Frontpage_{idtj}$. The right-hand side variable of interest indicates whether article $idtj$ cites Xinhua. In the second column, it is the predicted probability of article content being domestic (as opposed to foreign): $\widehat{D}_{idtj}$. The third column focuses on the sentiment of articles mentioning the leader, $Leader_{idtj} \times Sentiment_{idtj}$ (relative to the sentiment of those not mentioning the leader). All columns include day, topic, and outlet fixed effects and control for the number of entities mentioned in the article. Standard errors are multiway-clustered for dates, topics, and outlets.[31]

The coefficient signs of both $Xinhua_{idtj}$ and $\widehat{D}_{idtj}$ suggest that citing the state-run press agency or being more resembling of the domestic perspective increases an article's probability to feature on the front-page. In terms of coefficient size, a Xinhua mention

[31] When using the $Xinhua_{idtj}$ indicator in column 1, the Xinhua website itself is excluded (Xinhua citing itself does not have the same interpretation as other outlets citing Xinhua).

Table 2: Alignment with the Government Perspective and Front-Page Placement

	Dep. variable: $frontpage_{idtj}$		
	(1)	(2)	(3)
$Xinhua_{idtj}$	0.075***		
	(0.024)		
$\widehat{D}_{idtj}$		0.073**	
		(0.031)	
$Leader_{idtj}$			-0.007
			(0.055)
$Sentiment_{idtj}$			-0.123
			(0.075)
$Leader_{idtj} \times Sentiment_{idtj}$			0.163**
			(0.067)
N observations	984621	1090596	1090596
Day-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest captures whether article *idtj* cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

increases the front-page placement probability by 7.5 percentage points. Moving from perfectly resembling foreign news to perfectly resembling domestic news increases the front-page probability by 7.3 percentage points. Domestic articles rarely ever fully resemble the foreign perspective (see Figure B1). Thus, a more intuitive interpretation is the following: a one-standard deviation increase in alignment with the domestic perspective increases the probability of an article being on the front-page by by 1.1 percentage points. Both sources used to represent the foreign perspective are entirely censored in China (see Section 3.2). So, resembling foreign news implies resembling censored content.

Moving to the third column, the results show that articles with a more negative tone are more likely to feature on the front-page (the estimate is insignificant, though) – unless they mention the leader.[32] Specifically, when interacting the leader indicator with the sentiment, the coefficient sign implies that for articles mentioning Xi Jinping, *more positive* ones tend to be on the front-page. Concretely, if the sentiment changes from completely negative to completely positive (i.e., from 0 to 1), an article featuring him is 16.3 percentage points more likely to appear on the front-page. Empirically, articles talking about the leader never come with a sentiment below 0.5. Thus, note that a one-standard-deviation increase in sentiment comes with a 1.4 percentage points higher front-page probability.

Robustness checks Tables C1 to C3 (for the Xinhua indicator, the predicted probability of being from a domestic outlet and the sentiment of leader articles, respectively) add the fixed effects sequentially.[33] The coefficient signs and their significance are fully robust for the Xinhua indicator and the machine learning-based measure. For the sentiment-based measure, adding the fixed effects for topics seems to matter (with no or only date-year fixed effects, there is a null result; see columns 1 and 2). This is plausible since sentiment is a strongly context-dependent measure that might vary greatly from topic to topic, even for heavily aligned news. Along the same lines, for Xinhua mentions and the machine learning-based measure, the front-page differential is already visible in the raw data (see Figure C1).

[32] The finding that negativity can drive news consumption aligns with previous research (e.g., Robertson et al., 2023).

[33] Column 1 is without any FE, column 2 with date-year FE, and column 3 with date-year and topic FE. Column 4 also includes outlet FE, thus replicating the main specification from Table 2 (it is included again for an easy within-table comparison).

When week-year instead of day-year fixed effects are included in Table C4, the results largely reflect Table 2. Next, Table C5 weights every article by the circulation of its outlet (to account for the differences in size among the outlets). A similar qualitative pattern arises for the Xinhua indicator and the similarity with the domestic perspective (highly significant positive coefficients). For the coefficient size, the weights diminish the Xinhua indicator's effects but reinforce the prediction-based measure's effect. For the interaction of sentiment and the leader indicator, a null effect is measured, suggesting less robustness of this result. Table C6, which excludes articles covering local news, again confirms the main results.

Heterogeneity analyses Table C7 interacts the three alignment measures with an indicator for international news (as annotated by the topic model, see Appendix B.3) and shows that more alignment leads to higher front-page placement probabilities for both international and national/subnational news. For the effect sizes, no clear pattern emerges. Xinhua mentions appear to matter slightly less for the front-page placement of international news. For the machine learning-based measure, there is no difference. For articles on the leader, a positive tone matters more for international news. Table C8 interacts the main explanatory variables with an indicator for the days the high-level political meetings ("Two Sessions") occur. The main effects remain unchanged, and all interaction terms with the political meeting indicator are insignificant, implying no detectable changes in the front- vs. back-page differential during the Two Sessions.

Finally, Table C9 looks at regional heterogeneity. The three measures used for the main results (Table 2) are not well-suited to capture differences in entities' political sensitivity across provinces (Xinhua is a national agency, the machine learning-based measure is geared towards domestic vs. foreign differences, and the leader is a crucial figure across the country). Therefore, Table C9 focuses on entities with varying political sensitivity across regions: provincial governors and provincial Communist Party Secretaries. The assumption is that news items mentioning a *local* governor or secretary are more politically sensitive than those mentioning other governors or provincial secretaries. For the analysis, I interact the provincial leader indicator with the article sentiment (along the lines of the main results for $Leader_{idtj} \times Sentiment_{idtj}$). The front-page premium of positive sentiment should be larger for articles on the province's own leader relative to another province's leader. Indeed, this hypothesis is confirmed in Table C9, both when looking at provincial secretaries (column 1) and provincial governors (column 2):

While there is an association between more positive sentiment and front-page placement for articles on a region's own leadership, no similar association is found for mentions of other provinces' leadership.

Bottom line These results show that outlets prioritize Xinhua content on their front-pages and content favored by the domestic (government-owned) outlets. I also find some (slightly less robust) evidence that, among articles that mention the leader, more positively toned ones tend to be put on the front-page.

6. Mechanisms

So far, I have established that Chinese online outlets provide more Xinhua content and content closely aligning with news preferred by the government on front-pages, relative to back-pages. Similarly, for a given topic, articles mentioning the leader tend to feature on the front-page if they are more positively toned. These findings raise the question of why state-controlled media outlets engage in such strategies – instead of providing equally sensitive content on the front- and the back-pages. The conceptual framework in Section A suggests that outside options, in particular foreign news outlets, can influence domestic reporting when censorship is porous and not all consumers value uncensored content equally. This is because some news consumers are willing to incur the cost of accessing more investigative, less propaganda-oriented news. In this Section, I further investigate the hypothesis that the front vs. back-page differential can, at least partly, be explained by censorship motives interacting with the availability of foreign news.

6.1. Searched-for but explicitly sensitive entities

Censorship prevents people from consuming information they would be interested in if available (contrasting with purely market-driven mechanisms where certain contents are not shown or shown less prominently because interest in them is lower). Thus, I seek to identify entities that Chinese consumers can and do read about in foreign sources but that are considered sensitive by the Chinese government. To measure Chinese users' interests, I need data on country-specific page views for news covering specific entities. To my knowledge, country- and article-specific page views are unavailable for any news outlet. Fortunately, however, one information source provides fine-grained, entity-specific page views: Wikipedia. Thus, as a proxy for entities Chinese users are

potentially interested in, I consider all entities among the top 1,000 visited Chinese-language (Mandarin) Wikipedia pages on at least one day in 2020.[34]

Among these entities of interest, I seek to label those considered sensitive by the Chinese government. To this end, I scrape each entity's Wikipedia page. Since the entities annotated in the articles by the GEG (see above) are most often resolved to English Wikipedia pages, I scrape, for every Chinese Wikipedia page, its English counterpart (exists for 84% of Chinese pages). This leaves me with 17,199 Wikipedia pages with available scrapes. Based on a keyword search, I seek to establish whether the entity discussed on the Wikipedia page was (partly) censored in China: I check whether the keywords "censor" or "propaganda" appear along with "China" or "Chinese" *in the same paragraph*. The within-paragraph search aims at reducing false positives.[35] I tag all pages as potentially censored where at least one paragraph contains a relevant keyword combination (applies to 233 pages). Then, I manually inspect each of these pages to establish whether they describe, indeed, censored entities: 106 (45%) of them do. These pages include mentions of cultural productions that were (partly) censored in Mainland China, such as the page on "Nomadland (film)". While some foreign cultural productions are highly demanded in China, the Chinese authorities have been restrictive regarding the public's exposure to non-domestic culture, or only allow censored versions of some productions (e.g., Screen Daily, 2021; Park, 2020; The Initium, 2020). Another censored entity, "Someday or One Day", refers to a Taiwanese production that was adapted for the Mainland due to symbols of Taiwanese independence and depictions of homosexuality (The Initium, 2020). There are also examples of censored domestic productions (e.g., "A Touch of Sin"). "Animal Crossing", an online game, was banned in Mainland China because Hong Kong democracy activists used it in 2020 to spread politically sensitive

[34]Ideally, I would get page view statistics based on *users from China*. However, only aggregate country-specific views are available. Having page-specific views is essential for this exercise – to link the Wikipedia views to the entities mentioned in the domestic news. Despite the Great Firewall, one can assume that most Mandarin speakers on the free internet are from China: According to survey evidence, there are more than 200 million Chinese internet users with VPNs (Statista, 2017b; Statista, 2017a). VPN users can access Wikipedia. This figure is large relative to Mandarin speakers from other countries (e.g., the entire population of Taiwan is around 23.5 million; National Statistics – Republic of China, 2017). Hence, in all likelihood, Chinese Wikipedia page views correlate with Chinese users' interests.

[35]For instance, the Wikipedia page on "Vietnam" contains references to "censorship" and "propaganda" since Vietnam itself engages in censorship. At the same time, the article discusses Vietnam's relations with "China" (or the "Chinese"). However, the two need not be linked such that China censors mentions of "Vietnam".

images and slogans (The Guardian, 2020). Other sensitive entities refer to controversial individuals (e.g., "Anson Chan", a pro-democracy Hong Kong politician, "Chai Jing", an environmental activist, or "Chen Qiushi", a citizen journalist documenting the Covid-19 outbreak in China who then disappeared for 600 days, see The Wall Street Journal, 2021). Among the 55% of pages manually annotated as not covering censored entities, many are about entities *engaging in censorship*, for example, "Tuo Zhen" (the chief editor and president of the CCP newspaper People's Daily) or the "Ministry of Public Security (China)". Others are not explicitly about China.[36] Table D1 shows all 233 entities whose pages were flagged through the keyword search (along with their manual annotation).

Next, for every newspaper article, I create an indicator of it mentioning at least one of these searched-for but censored entities. As a plausibility check, it is reassuring that all three previously used alignment measures are significantly lower in articles with a sensitive entity.[37]

Unsurprisingly, given their (partly) censored nature, these entities are featured relatively infrequently – 160 times per entity on average (between 2020 and 2022). Overall, 17,020 articles contain at least one sensitive entity. Table D1 also lists the number of articles on a given entity: Entities that never feature in the domestic news are not helpful here because one cannot calculate a front-page vs. back-page difference for those. Interestingly, most of the 106 sensitive entities appear at least once in the domestic news.

However, their appearance is more likely on back-pages than front-pages, as Figure 2 shows. The front-page placement probability is 49% for articles without a sensitive entity mention. In contrast, it is 41% for those with such a mention. The difference is statistically significant (see the 95% confidence intervals). Table D2 documents the same result in a regression format: It replicates Tables C1 to C3 and shows that sensitive entities are less likely to feature on the front-page, irrespective of the specific fixed effects that are included. Precisely, in the preferred specification with all the fixed effects, shown in the last column of Table D2 (this column reflects the main specification as used in Table 2), a censored entity mention makes an article 2.1 percentage points less likely to

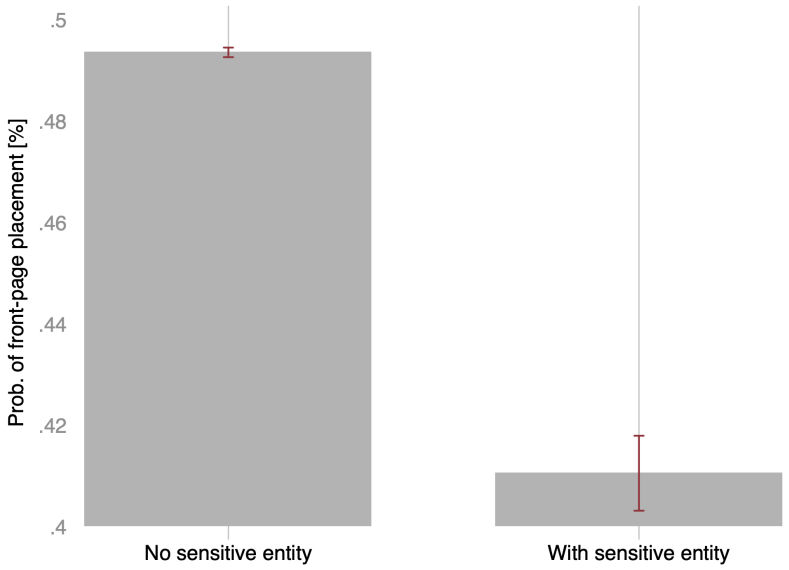
[36] As a case in point, the Wikipedia page on the "Tunisian Revolution" cites critiques that the Western media provided less sympathetic coverage relative to other protests or relative to "censorship in China".

[37] While the probability of a Xinhua mention for articles not mentioning such an entity is 9.5%, it is 7.9% in those mentioning one. Similarly, the prediction-based measure is 89.5% vs. 67.3%. The average score of articles mentioning the leader but no sensitive entity is 0.28, while articles on the leader and a sensitive entity are only at 0.24. All differences are significant with $p<0.01$.

feature on the front-page.

This Section's evidence corroborates the hypothesis that the back-page placing of certain news stories reflects subtle censorship in a context where a subset of the population consumes foreign information.

Figure 2: Front-Page Placements of Articles With or Without a Censored Entity Mention

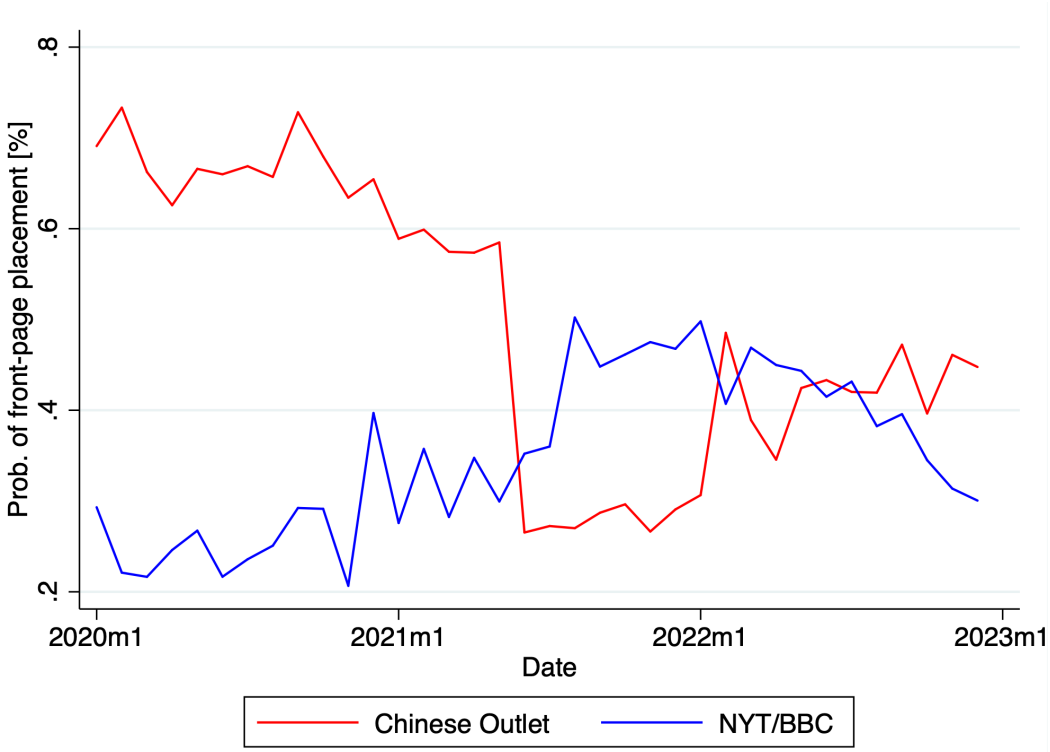


Notes: The articles' raw front-page placement probability (in percent, on the vertical axis), with a 95% confidence interval. This probability is plotted depending on whether they mention a censored entity (see articles with or without such an entity on the horizontal axis).

6.2. A shock to an entity's political sensitivity

My conceptual framework does not only predict that outside options for readers who value investigative news matter. It also predicts that the front vs. back-page differential should widen when the potentially unaligned content on foreign platforms increases. In this Section, I investigate this hypothesis by looking at the outbreak of the crisis in Ukraine that led to Russia invading Ukraine on 24 February 2022. In February 2021, Zelenskyy's government froze the assets of opposition leader Viktor Medvedchuk, the Kremlin's most prominent ally in Ukraine. Subsequently, Russia started amassing troops at the Ukrainian border in March and April 2021 in what Russia claimed to be training exercises. These developments were the most decisive escalation since the Crimean War in 2014 (Reuters, 2022).

Figure 3: The Front-Page Probability of Ukraine war-related Entities in Chinese Outlets vs. the NYT &BBC



Notes: The Figure shows the probability of a Ukraine war-related entity (Putin, Zelenskyy, Russia, Ukraine) to feature on Chinese front-pages (red line) or the NYT and the BBC front-pages (blue line) by month. These are raw probabilities.

With the crisis outbreak, entities like "Russia", "Ukraine", "Vladimir Putin", and "Volodymyr Zelenskyy" have become (more) sensitive from the Chinese regime's viewpoint almost literally overnight. Many commentators argue that the war in Ukraine poses significant challenges to China, both from a political and economic view. Western media has heavily criticized China's support for Russia or drawn parallels to the situation around Taiwan (e.g., deLisle, 2022). Thus, the crisis outbreak is a suitable event study candidate: The government might not want to push related information to those who would not search for it. However, a more interested minority might be aware of the conflict, thus forcing the government to comment on the situation. Figure 3 plots the raw front-page probability of articles mentioning Putin, Zelenskyy, Russia, or Ukraine for both Chinese outlets (red) and the NYT and the BBC (blue). It does not only show an abrupt drop in Chinese outlets after Russia deploys troops but also an increase in the NYT and the BBC.

Before turning to the event study, I look at the development of these sensitive entities' front-page probabilities in domestic news. That is, I compare the front-page probabilities of articles covering Russia, Putin, Zelenskyy, and Ukraine before and after the escalation. Table 3 presents the results. The first column shows that, initially, articles on Putin were 17.4 percentage points more likely to appear on the front-page relative to articles not mentioning him. However, after the escalation, such articles' front-page placement probability significantly decreases by 30.2 percentage points. For an effect size comparison with the main results, recall that mentioning Xinhua comes with an increase of 7.5 percentage points. Similarly, column 2 shows that mentioning Zelenskyy after the war lowers the front-page probability by 9.3 percentage points: As the precise null of the interaction with $After\_War_{idtj}$ hints at, Zelenskyy was not mentioned at all, or at least not detectably so, before the war.[38] Accordingly, the non-interacted $Zelenskyy_{idtj}$ indicator reflects the post-escalation effect. In column 3, references to Russia show the same pattern as those for Putin (-26.4 percentage points). Column 4 shows that Ukraine also becomes back-page news after the escalation (-19.3 percentage points).

Next, since the Ukraine war-related outbreak is an international event for both the United States/Europe and China, I can use the New York Times and BBC articles as controls in my event study:[39] I can compare how Ukraine war-related entities' front-page probability evolves in Chinese outlets relative to the NYT and BBC. While these entities have different front-page probability levels in the NYT/BBC and Chinese outlets, they follow statistically indistinguishable trends before the conflict outbreak (this can be seen in the event study graph, as discussed shortly). I estimate the following equation:

[38] Even after February 2021, he is only mentioned 276 times (relative to 6,361 mentions of Ukraine).

[39] I use the same articles as when building the machine learning-based model; see Section 3.2.1. When building the machine learning-based model, I filtered for articles on China. Here, I use all 318,017 NYT/BBC articles from 2020-22 as controls, not just the 46,946 ones covering China.

Table 3: Ukraine War-Related Entities and Front-Page Placement

	Dep. variable: $frontpage_{idtj}$			
	(1)	(2)	(3)	(4)
$Putin_{idtj}$	0.174***			
	(0.035)			
$Putin_{idtj} \times After\_War_{idtj}$	-0.302***			
	(0.080)			
$Zelenskyy_{idtj}$		-0.093***		
		(0.020)		
$Zelenskyy_{idtj} \times After\_War_{idtj}$		0.000		
		(0.000)		
$Russia_{idtj}$			0.132**	
			(0.053)	
$Russia_{idtj} \times After\_War_{idtj}$			-0.264***	
			(0.091)	
$Ukraine_{idtj}$				0.085*
				(0.043)
$Ukraine_{idtj} \times After\_War_{idtj}$				-0.193***
				(0.061)
N observations	1090596	1090596	1090596	1090596
Same FE as Main Results	X	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest is an interaction between (i) an indicator of whether an entity salient in the war in Ukraine is mentioned and (ii) an indicator for dates after February 2021 (i.e., after Russia's deployment of troops close to the Ukrainian border). In column 1, the salient entity is Vladimir Putin. In column 2, it is Volodymyr Zelenskyy. Columns 3 and 4 focus on Russia and Ukraine as salient entities, respectively. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

$$
\begin{aligned}
Frontpage_{idtj} = \alpha_j + \beta_d + \kappa\ & Ukr\_Entity_{idtj} \\
& + \sum_{q=2020-02}^{2022-12} \mu_q\ Ukr\_Entity_{idtj} \times I[Quarter = q]_{idtj} \\
& + \sum_{q=2020-02}^{2022-12} \tau_q\ Chin\_Outlet_{idtj} \times I[Quarter = q]_{idtj} \times Ukr\_Entity_{idtj} \\
& + \Omega Controls_{idtj} + \vartheta_{idtj} \qquad (4)
\end{aligned}
$$

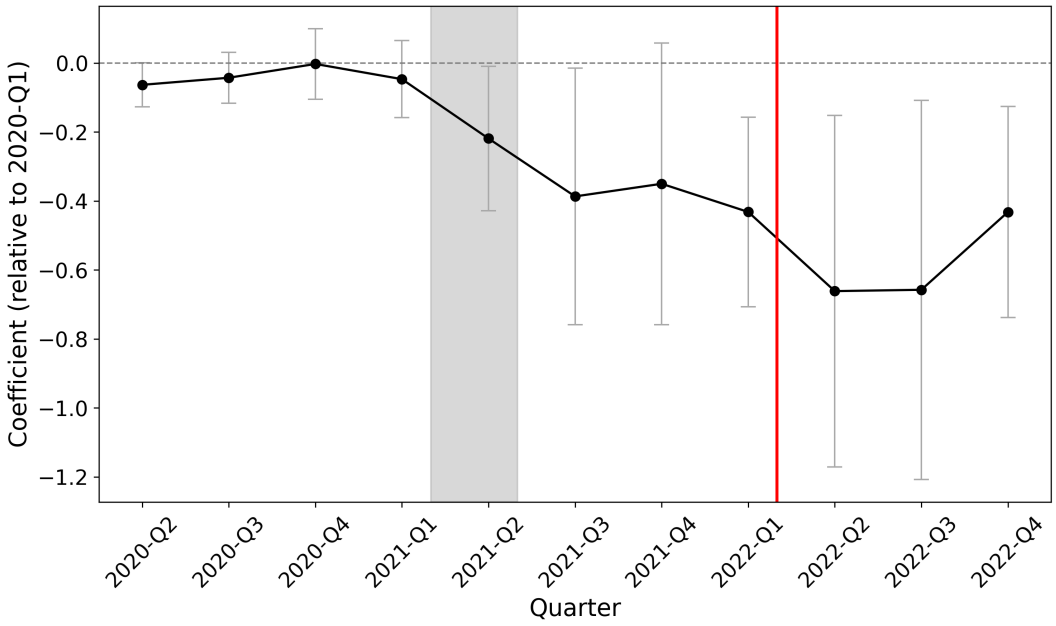
$Frontpage_{idtj}$ and the two fixed effects α_j and β_d are defined as above. Ukr_Entity_{idtj} indicates whether Putin, Zelenskyy, Russia, or Ukraine are mentioned in an article. $Chin\_Outlet_{idtj}$ equals one for Chinese outlets (and zero for the NYT/BBC). $I[Quarter = q]$ captures in which quarter from 2020-Q2 to 2022-Q4 the article appeared. Articles appearing in 2020-Q1 are the reference group. The coefficients of interest are τ_q, where the expectation is that quarters before the beginning of the crisis are statistically indistinguishable from zero, and those after are negative.

Figure 4 shows that the trends between war-related articles from the NYT/BBC and those from Chinese outlets start to diverge early in 2021, consistent with the escalation around Viktor Medvedchuk and the deployment of Russian troops (see above). Over the subsequent months, the front-page probability of the articles keeps dropping relative to the New York Times and BBC (as well as in absolute terms; recall the raw data in Figure 3).[40]

In sum, Table 3 and Figure 4 document how a shock to an entity's political sensitivity due to foreign forces can widen the front- vs. back-page differential. In an additional case study (see Section E.3), I look at the differential during the Beijing Winter Olympics in 2022. It is assumed that during the Olympics, China was in the spotlight, thus making potentially sensitive information on China more salient. The results suggest that, relative to other times, article alignment is more strongly associated with front-page placement.

[40] Section E.2 documents that my results are unlikely driven by problems that can arise in difference-in-differences settings going beyond the classical 2 × 2 design (as described, for example, by De Chaisemartin and D'Haultfoeuille, 2020).

Figure 4: The Marginal Front-Page Probability of Ukraine War-related Entities
(with NYT &BBC articles as controls)



Notes: The Figure shows the marginal probability of a Ukraine war-related entity (Putin, Zelenskyy, Russia, Ukraine) to feature on Chinese front-pages, relative to the NYT and the BBC front-pages, over time. Specifically, the vertical axis captures the τ_q coefficient (see event study specification in Equation 4) for each quarter from 2020-Q2 to 2022-Q4 (2020-Q1 is the reference). The first and second quarter of 2021 mark the beginning of the escalation, especially due to the deployment of Russian troops close to the Ukrainian border (area shaded in grey). The red line is the date of the Russian invasion (24 February 2022).

7. Conclusion

This paper reveals a strategic pattern in the placement of news within Chinese online newspapers (2020-22). Studying over a million articles from 53 news outlets, it shows that front-page articles are more likely to feature government-endorsed content relative to articles published in other locations of news websites. Specifically, front-page articles are more likely to cite the official press agency, Xinhua. Similarly, they resemble more heavily the perspective favored by the government-owned domestic news market relative to major foreign news sources that are blocked in Mainland China. This measure of resemblance is built using text-as-data methods and relies on named entities. Named entities' mentions are relatively easy to track across languages – to compare domestic news in Mandarin vs. foreign news in English. Along the same lines, front-page articles on the leader, Xi Jinping, tend to convey a more favorable overall sentiment than back-page articles. These differences in article placement can be observed even when holding the date, the news outlet, and the broader topic constant. Theoretical evidence suggests that these patterns reflect subtle censorship where attention is scarce and information abundant. In particular, only a minority of readers actively search for more investigative content. In the internet age, the government cannot entirely prevent this investigative minority from accessing foreign news. By featuring slightly more politically sensitive news on back-pages, the government caters to this minority. At the same time, the less attentive majority is provided with the government's most favored content on front-pages. Empirical analyses corroborate this theory, showing that content that Chinese users consume on Wikipedia (which is only accessible to a VPN-utilizing minority) tends to appear, if at all, on domestic back-pages. Also, if an entity's political sensitivity changes due to foreign forces, the front- vs. back-page differential widens. These findings contribute to broader debates on autocracies' information management, especially in the online age.

References

ABYZ News Links (2021). Newspaper and News Media Guide. abyznewslinks.com. Accessed: 2021-10-06.

Balles, P., Matter, U., and Stutzer, A. (2018). Special Interest Groups Versus Voters and the Political Economics of Attention.

Banerjee, A. V. and Mullainathan, S. (2008). Limited Attention and Income Distribution. *American Economic Review*, 98(2):489–93.

Becker, L. B. and Vlad, T. (2009). Freedom of the Press around the World. In *Global Journalism. Topical Issues and Media Systems.*, pages 65–85.

Bhattacharyya, S. and Hodler, R. (2015). Media Freedom and Democracy in the Fight Against Corruption. *European Journal of Political Economy*, 39:13–24.

Blei, D., Ng, A., and Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bradshaw, S. and Howard, P. N. (2019). *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation.* Oxford Internet Institute.

Brunetti, A. and Weder, B. (2003). A Free Press is Bad News for Corruption. *Journal of Public Economics*, 87(7-8):1801–1824.

Buddenbrock, F. (2016). *Search Engine Optimization: Getting to Google's First Page*, pages 195–204. Springer New York, New York, NY.

Chen, Y. and Yang, D. Y. (2019). The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review*, 109(6):2294–2332.

Choi, S.-W. and James, P. (2007). Media Openness, Democracy and Militarized Interstate Disputes. *British Journal of Political Science*, pages 23–46.

De Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

deLisle, J. (2022). China's Russia/Ukraine Problem, and Why It's Bad for Almost Everyone Else Too. *Orbis*, 66(3):402–423.

Djankov, S., McLiesh, C., Nenova, T., and Shleifer, A. (2003). Who Owns the Media? *Journal of Law and Economics*, 46(2):341–382.

Djourelova, M. and Durante, R. (2021). Media Attention and Strategic Timing in Politics: Evidence from US Presidential Executive Orders. *American Journal of Political Science*.

Durante, R. and Zhuravskaya, E. (2018). Attack when the World is not Watching? US News and the Israeli-Palestinian Conflict. *Journal of Political Economy*, 126(3):1085–1133.

Edmond, C. (2013). Information Manipulation, Coordination, and Regime Change. *Review of Economic Studies*, 80(4):1422–1458.

Egorov, G., Guriev, S., and Sonin, K. (2009). Why Resource-Poor Dictators allow Freer Media: A Theory and Evidence from Panel Data. *American Political Science Review*, pages 645–668.

Egorov, G. and Sonin, K. (2020). The Political Economics of Non-Democracy. Technical report, National Bureau of Economic Research.

Eisensee, T. and Strömberg, D. (2007). News Droughts, News Floods, and U.S. Disaster Relief. *Quarterly Journal of Economics*, 122(2):693–728.

Falkinger, J. (2008). Limited Attention as a Scarce Resource in Information-Rich Economies. *The Economic Journal*, 118(532):1596–1620.

Fedyk, A. (2018). Front Page News: The Effect of News Positioning on Financial Markets. Technical report, Working Paper.

Freedom House (2019). Press Freedom. https://freedomhouse.org/report-types/freedom-press. Accessed: 2019-07-03.

GDELT (2019). The GDELT Project. gdeltproject.org. Accessed: 2019-05-23.

Gehlbach, S., Luo, Z., Shirikov, A., and Vorobyev, D. (2022). A Model of Censorship, Propaganda, and Repression. Technical report.

Gehlbach, S. and Sonin, K. (2014). Government Control of the Media. *Journal of Public Economics*, 118:163–171.

Gleick, J. (2011). *The information: A History, a Theory, a Flood*. Vintage.

Graham-Harrison, E. and Ni, V. (2022). Sport, Politics, and Covid Collide at the Beijing Winter Olympics. https://www.theguardian.com/world/2022/jan/30/sport-politics-and-covid-collide-at-the-beijing-winter-olympics. Accessed: 2023-08-17.

GreatFire.org (2022). Online Censorship In China. https://en.greatfire.org/https/www.nytimes.com. Accessed: 2022-08-06.

Guriev, S., Melnikov, N., and Zhuravskaya, E. (2021). 3G Internet and Confidence in Government. *Quarterly Journal of Economics*.

Guriev, S. and Treisman, D. (2018). Informational Autocracy: Theory and Empirics of Modern Authoritarianism. *Available at SSRN*.

Guriev, S. and Treisman, D. (2019). Informational Autocrats. *Journal of Economic Perspectives*, 33(4):100–127.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

Hirshleifer, D. and Teoh, S. H. (2003). Limited Attention, Information Disclosure, and Financial Reporting. *Journal of Accounting and Economics*, 36(1-3):337–386.

Hoelig, S., Hasebrink, U., and Behre, J. (2021). Keeping on Top of the World: Online News Usage in China, the United States and Five European Countries. *New Media & Society*, 23(7):1798–1823.

Kamenica, E. and Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, 101(6):2590–2615.

King, G., Pan, J., and Roberts, M. E. (2013). How Censorship in China allows Government Criticism but silences Collective Expression. *American Political Science Review*, pages 326–343.

King, G., Pan, J., and Roberts, M. E. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument. *American Political Science Review*, 111(3):484–501.

Leeson, P. T. (2008). Media Freedom, Political Knowledge, and Participation. *Journal of Economic Perspectives*, 22(2):155–169.

Lorentzen, P. (2014). China's Strategic Censorship. *American Journal of Political Science*, 58(2):402–414.

Matter, U. and Widmer, P. (2021). Who Owns the Online Media? *Available at SSRN 3969253*.

McMillan, J. and Zoido, P. (2004). How to Subvert Democracy: Montesinos in Peru. *Journal of Economic Perspectives*, 18(4):69–92.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, 63(2):81–97.

Molter, V. and DiResta, R. (2020). Pandemics & Propaganda: How Chinese State Media Creates and Propagates CCP Coronavirus Narratives. *Harvard Kennedy School Misinformation Review*, 1(3).

National Statistics – Republic of China (2017). Latest Indicators: Total Population. https://eng.stat.gov.tw/point.asp?index=9. Accessed: 22-06-20.

NPR (2017). Behind China's VPN Crackdown, A Game Of Cat And Mouse Continues. https://text.npr.org/541554438. Accessed: 2021-01-31.

Park, N. S. (2020). China Backs Off From Fight With K-Pop Fans. https://foreignpolicy.com/2020/10/20/china-south-korea-bts-kpop-nationalism-soft-power/. Accessed: 2022-06-10.

Qin, B., Strömberg, D., and Wu, Y. (2017). Why does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda. *Journal of Economic Perspectives*, 31(1):117–40.

Qin, B., Strömberg, D., and Wu, Y. (2018). Media Bias in China. *American Economic Review*, 108(9):2442–76.

Reporters Without Borders (2021). Reporters Without Borders. https://rsf.org/en. Accessed: 2021-02-15.

Reuters (2022). Timeline: The Events Leading up to Russia's Invasion of Ukraine. https://www.reuters.com/world/europe/events-leading-up-russias-invasion-ukraine-2022-02-28/. Accessed: 2023-10-15.

Roberts, M. E. (2018). *Censored: Distraction and Diversion inside China's Great Firewall*. Princeton University Press.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., and Feuerriegel, S. (2023). Negativity Drives Online News Consumption. *Nature Human Behaviour*, 7(5):812–822.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*.

Schneider, L. (2014). Media Freedom Indices. www.dw.com/downloads/28985486/mediafreedomindices.pdf. Accessed: 2020-08-03.

Screen Daily (2021). Is China Finally Opening to Korean Content as Political Relations Improve? www.screendaily.com/features/is-china-finally-opening-to-korean-content-as-political-relations-improve/5157997.article. Accessed: 2022-03-14.

Shadmehr, M. and Bernhardt, D. (2015). State Censorship. *American Economic Journal: Microeconomics*, 7(2):280–307.

Statista (2017a). Leading Markets for VPN Usage among Internet Users Worldwide as of 2nd Quarter 2017. www.statista.com/statistics/301204/top-markets-vpn-proxy-usage/#:~:text=This%20statistic%20presents%20the%20online,network%20in%20the%20past%20month. Accessed: 2022-06-20.

Statista (2017b). Number of Internet Users in China from 2008 to 2021. www.statista.com/statistics/265140/number-of-internet-users-in-china/. Accessed: 22-06-20.

Statista (2020). Average Daily Time Spent on Traditional and Digital Media by Adults in China from 2016 to 2019. www.statista.com/statistics/1061951/china-traditional-digital-media-usage-time/. Accessed: 2022-08-04.

Statista (2023). Penetration Rate of Internet Users in China from 2012 to H1 2023. www.statista.com/statistics/236963/penetration-rate-of-internet-users-in-china/. Accessed: 23-08-21.

The Guardian (2020). Animal Crossing Game Removed From Sale in China over Hong Kong Democracy Messages. www.theguardian.com/world/2020/apr/14/animal-crossing-game-removed-from-sale-in-china-over-hong-kong-democracy-messages. Accessed: 2022-06-10.

The Initium (2020). Roundtable Cinematheque Someday or One Day. https://theinitium.com/roundtable/20200218-roundtable-cinematheque-someday-or-one-day/. Accessed: 2022-06-03.

The Wall Street Journal (2021). Chinese Citizen Journalist Who Documented Covid-19 in Wuhan Resurfaces After 600 Days . www.wsj.com/articles/chinese-citizen-journalist-who-documented-covid-19-in-wuhan-resurfaces-after-600-days-11633077956. Accessed: 2022-06-13.

Wu, Y., Strömberg, D., and Qin, B. (2021). Social Media and Collective Action in China.

Xinhuanet (2020). Zhong Nanshan: Outspoken Doctor Awarded China's Top Honor. www.xinhuanet.com/english/2020-09/08/c_139352929.htm. Accessed: 22-06-16.

Yin, J. and Wang, J. (2014). A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 233–242.

Zhuang, M. (2022). Intergovernmental Conflict and Censorship: Evidence from China's Anti-Corruption Campaign. *Journal of the European Economic Association*, 20(6):2540–2585.

A. Theoretical Framework

Consider a country with low media freedom. On the one hand, there is a propagandist who supports the government. The propagandist controls the domestic media – which comprises front- and back-pages – and decides what information (articles) to feature in each location. Information is characterized by its level of freedom $f \in [f_{min}, f_{max}]$. For example, freer content could cover more politically sensitive news items. Assuming that the propagandist cannot offer negative levels of free information, let $f_{min} = 0$. Empirically, we can think of information at f_{min} as pure propaganda, which is assumed to be uninformative, such as mere displays of the leader's competence (Gehlbach et al., 2022). Taken together, the propagandist's choice variables are f_{front} and f_{back}.

On the other hand, there are readers who decide where to read their news. They can visit domestic front-pages, domestic back-pages, or foreign sites.[41] Foreign news consumption in a censored context is plausible where the government cannot fully shield all consumer types from the outside world (e.g., due to the internet).[42] For readers, consuming news comes at a cost $C(\ell)$ that depends on the location $\ell = front, back, foreign$. So, for example, reading on front-pages comes with $C(front) = c_{front}$. It is assumed that $c_{front} < c_{back} < c_{foreign}$. To recapitulate, the readers' choice variable is the location ℓ where they consume their news. The model assumes that the propagandist cannot influence the readers' cost of reading front-pages, back-page news, or foreign news – at least not in the short run, when it is decided if a story goes to the front- or the back-page.

The cost-related assumptions are empirically plausible: Regarding $c_{front} < c_{back}$, reading back-pages is more costly because the reader pays a cognitive cost for navigating to the back-page.[43] Similarly, the relationship $c_{back} < c_{foreign}$ is also influenced by the reader's human nature: They pay cognitive costs to search and access suitable foreign sources, such as by downloading a VPN. Foreign actors and technology can also impact it. For example, how much potentially sensitive information is available in foreign outlets

[41] I assume that readers from all geographical locations within the country can freely access domestic front- and back-pages (i.e., we are looking at one market). Thereby, I abstract from local news.

[42] The option of reading foreign news – that is, news that the government may not approve of – could be replaced by any other type of "revolting behavior" that the regime may deem destabilizing.

[43] Fedyk (2018) finds that financial news on Bloomberg's front-page impacts markets more than equivalent stories elsewhere on the website. Similarly, 90% of Google users don't go beyond the first page of search results Buddenbrock (2016).

on news items relevant for citizens of the censored country?[44] Lastly, $c_{foreign}$ contains a share determined by the degree to which the government invests in repression (e.g., by setting up firewall infrastructure or cracking down on VPNs). Arguably, the propagandist cannot easily influence this last component in the short term. It likely results from a mid- or long-term strategy. Therefore, for a short period (such as a given day), the readers' cost of accessing foreign news can be seen as a given.[45]

Next, the model posits that readers value the freedom of the information they consume but with varying intensity. When choosing source $\ell = front, back, foreign$, a reader r derives utility

$$U(\alpha_r, \ell_r) = V(\alpha_r, F(\ell_r)) - C(\ell_r) \tag{5}$$

α captures the intensity of preferences over information freedom – that is, the consumption of content F that is sensitive from the government's perspective. Readers obtain an amount of F by choosing ℓ. For instance, $F(front) = f_{front}$.[46] Going forward, let there be two types α_L and α_H, with $\alpha_L < \alpha_H$. I assume readers will always read some news: they do not have an outside option and will necessarily read on one of the three locations $\ell = front, back, foreign$.

It is assumed that $V()$ is differentiable with respect to all its arguments. Higher types value non-propaganda news strictly more than lower types, so $\frac{dV}{d\alpha_r} > 0$. $V()$ is also non-decreasing in f_r, such that $\frac{dV}{df_r} \geq 0$. That is, nobody can be made worse off by having access to non-propaganda content. At first, this assumption may appear slightly restrictive from an empirical viewpoint. While extreme supporters may, in reality, dislike content unaligned with the propagandist, empirical evidence suggests that the majority of media consumers positively value information freedom (Qin et al., 2018).[47] Therefore, I assume that people prefer non-propaganda news to propaganda or, at worst, are indifferent between the two.

As argued in the literature and this paper's empirical part, many countries with limited media freedom are porous systems. That is, with no free information in their

[44] Another example is whether companies offer VPN services.

[45] Note that the paper's empirical part looks at within-day (as well as outlet and topic) variation.

[46] Similarly, $F(back) = f_{back}$, and $F(foreign) = f_{foreign}$.

[47] Specifically, propagandist-imposed bias shows a strong negative correlation with advertising revenue, which is, arguably, a valid proxy for reader preferences.

country, high types will read foreign news.[48] Formally, it is characterized as

$$V(\alpha_H, f_{foreign}) - c_{foreign} = V(\alpha_H, 0) - c_{front} \tag{6}$$

for any value of $f_{foreign}$. That is, if the propagandist does not provide some level of freedom anywhere (neither on the front- nor back-pages), the high types will necessarily read foreign news, even if the free foreign news relevant to the high types is at its lowest possible value.[49]

Next, another characteristic of a porously censored system is that not everyone will consume foreign news as soon as it comes at a cost higher than for any of the domestic locations. Put differently, if the barriers were so low that even the lowest types would consume foreign news, it would – de facto – be an open system. In this model, it implies that for any value of $F(foreign)$, the low types will choose the domestic outlets

$$V(\alpha_L, f_{max}) - c_{foreign} < V(\alpha_L, f_{min}) - c_{front} \tag{7}$$

I will operationalize this by setting a_L to 0. This might appear restrictive. However, it can be plausible if we think of low types being busy to make ends meet, so no time remains to consider whether their government provides them with true information. They could also simply be interested in entertainment, such that the degree of media freedom in the content does not matter to them.[50] It could also reflect a lack of information, where individuals with limited resources (financial, cognitive) who were educated in a censored system since birth are not even aware of different qualities of information, such that they would simply go for the cheapest option. As an analogy, think of someone in a supermarket who is unaware or does not believe that the actual quality of different types of flour vary (organic, full grain, local brands, etc.) and would, thus, simply buy the cheapest option.

The other player, the propagandist, minimizes S, the total amount of potentially unaligned content conveyed, since it could cause instability. This amount is influenced by the (potentially) sensitive information made available in the bundle chosen by the

[48] For example, in the context of the internet, it can be prohibitively costly to prevent everyone from consuming any foreign news.

[49] Formally, it holds that $f_{\text{foreign}} \in [f_{foreign}^{lower}, f_{max}]$. As a side note, in Equation 6, $V(\alpha_H, 0) - c_{front} \geq V(\alpha_H, 0) - c_{back}$ is trivially satisfied since $c_{front} < c_{back}$.

[50] In this simple framework, I do not model the entertaining value of news.

high and the low type, respectively, and the share of the high types s_H (with $s_L = 1 - s_H$):

$$S(F(\ell_L), F(\ell_H), s_H) = (1 - s_H) \times F(\ell_L) + s_H \times F(\ell_H) \tag{8}$$

Let us now characterize the equilibrium. There could be a pooling equilibrium in which all readers choose the same source of information or a separating equilibrium in which low and high types choose differently.[51] First, consider a potential separating equilibrium. The propagandist sets $F(front)$ to its minimum 0 since the low types have no better option than choosing $\ell = front$, anyways, as implied by 7 and $c_{foreign} > c_{back} > c_{front}$.

Then, the propagandist would set $F(back)$ such that the high types will choose *back* over $foreign$:

$$V(\alpha_H, f^{sep}_{back}) - c_{back} = V(\alpha_H, f_{foreign}) - c_{foreign} \tag{9}$$

Note that the high types have no incentive to switch to the front-page, as implied by equation 6. This would lead to a separating equilibrium where

$$S(F(\ell_L), F(\ell_H), s_H) = (1 - s_H) \times 0 + s_H \times f^{sep}_{back} \tag{10}$$

The only possible pooling equlibrium is the one in which both the low and high types read the front-pages (since we have established that low types always read front-pages). The propagandist would would set f_{front} such that

$$V(\alpha_H, f^{pool}_{front}) - c_{front} = V(\alpha_H, f_{foreign}) - c_{foreign} \tag{11}$$

That is, the propagandist chooses $f_{front} > 0$ to satisfy the high types' constraints by offering them some unaligned content on the front-page to discourage them from the

[51] The model employs concepts from classical screening models, often used to set prices for product bundles, to understand information dissemination strategies. Unlike Perfect Bayesian Equilibria, my approach involves a propagandist who creates content that encourages readers to self-select based on their preferences, revealing their information type. This self-selection process is the basis of my equilibrium concept. It is a strategy of information control by segmentation rather than belief updating.

$foreign$ choice.[52] The total consumption of potentially destabilizing content is

$$S(F(\ell_L), F(\ell_H), s_H) = (1 - s_H) \times f^{pool}_{front} + s_H \times f^{pool}_{front} = f^{pool}_{front} \tag{12}$$

Intuitively, whether a separating or a pooling equilibrium is more attractive to the propagandist depends on s_H. If s_H is low, there is no need to over-inform the large masses of low types,[53] and a separating equilibrium is more attractive for the propagandist. Conversely, if s_H is high, the over-information of the few low types does not raise $S()$ much, and the propagandist can lower $S()$ by incentivizing the large masses of high types to consume slightly less free content on the front-pages: $f^{pool}_{front} < f^{sep}_{back}$.[54] Accordingly, a separating equilibrium arises if

$$s_H \times f^{sep}_{back} < f^{pool}_{front} \tag{13}$$

In the corner case where $s_H = 0$, f is 0 on all domestic platforms. Conversely, with $s_H = 1$, there is no more product differentiation and everyone reads the front-page which provides some degree of free content.

Let us illustrate some comparative statics now. Considering the potential equilibria conditions, the content level in a separating equilibrium is defined by:

$$f^{sep}_{back} \geq f_{foreign} + \phi(c_{foreign}, c_{back}, \alpha_H) \tag{14}$$

The function ϕ captures how changes in the costs ($c_{foreign}$, c_{back}) and type-dependent valuation α_H influence the equilibrium value of f^{sep}_{back}: It increases with $f_{foreign}$ and c_{back} but decreases with $c_{foreign}$. For the pooling equilibrium, the content level f^{pool}_{front} is characterized by:

$$f^{pool}_{front} \geq f_{foreign} + \phi(c_{foreign}, c_{front}, \alpha_H) \tag{15}$$

From the perspective of the propagandist, the amount of potentially destabilizing

[52] f^{pool}_{back} would satisfy $V(a_H, f^{pool}_{front}) - c_{front} = V(a_H, f^{pool}_{back}) - c_{back}$.

[53] They are over-informed in a pooling equilibrium since $f^{pool}_{front} > f^{sep}_{front} = 0$.

[54] These arguments assume that the value function $V()$ is well-behaved, such that an equilibrium actually exists.

content in the separating and pooling equilibria is, respectively:

$$S^{sep} = s_H[f_{foreign} + \phi(c_{foreign}, c_{back}, \alpha_H)] \tag{16}$$

$$S^{pool} = f_{foreign} + \phi(c_{foreign}, c_{front}, \alpha_H) \tag{17}$$

These equations emphasize the influence of $f_{foreign}$ on the equilibrium content level. In particular, all else equal, a positive shock to $f_{foreign}$ can increase the differential between front- and back-pages in a separating equilibrium. For a pooling equilibrium, such a shock has a one-to-one impact on $S()$. In a separating equilibrium, the effect is scaled by s_H. This distinction implies that, under certain conditions, a surge in $f_{foreign}$ could tip a pooling equilibrium toward a separating one.

For an example with a specific function form, assume that

$$U(\alpha_r, \ell_r) = \alpha_r \times F(\ell_r) - C(\ell_r) \tag{18}$$

Then, f^{sep}_{back} is defined as follows (again highlighting it increases in $f_{foreign}$ and c_{back} but decreases in $c_{foreign}$).

$$f^{sep}_{back} \geq f_{foreign} - \frac{1}{\alpha_H} c_{foreign} + \frac{1}{\alpha_H} c_{back} \tag{19}$$

Similarly, f^{pool}_{front} is

$$f^{pool}_{front} \geq f_{foreign} - \frac{1}{\alpha_H} c_{foreign} + \frac{1}{\alpha_H} c_{front} \tag{20}$$

For the propagandist, it means that

$$S^{sep} = s_H[f_{foreign} - \frac{1}{\alpha_H} c_{foreign} + \frac{1}{\alpha_H} c_{back}] \tag{21}$$

and

$$S^{pool} = f_{foreign} - \frac{1}{\alpha_H} c_{foreign} + \frac{1}{\alpha_H} c_{front} \tag{22}$$

This conceptual framework provides some intuition for several stylized facts in the empirical part. It suggests that content differences on front- and back-pages occur with porous censorship because of two main forces that act together. The first force is porosity: In the internet age (and, to some extent, likely even before that), no country is an informational island. The second force is that there is some share of the population

with a relatively high valuation of free content. Hence, for this type, obtaining foreign perspectives is worth the additional cost – relative to not obtaining free news in their own country. These forces put pressure on the government to cater to this group. In light of this framework, one can interpret my empirical findings on a differential between front- and back-page news as evidence of a separating equilibrium. The model predicts such an equilibrium to arise when the share of high types is not too high, which is, again, empirically plausible.[55] Furthermore, the model also provides intuition on why an (exogenous) increase in potentially sensitive foreign news can lead to more such content being available on back-pages, relative to front-pages.

[55] As an example proxy for the high shares, around one-third of Chinese internet users have a VPN (Statista, 2017a). At the same time, internet users amount to 60 to 70% of the population (Statista, 2023), suggesting that around one-fifth of the population uses a VPN.

B. Data Appendix

B.1. Outlets included in the analysis

I identify 157 outlets that are either (i) in ABYZ News Links (2021) or (ii) are in Qin et al.'s (2018) sample and have an online presence. Out of the 39 outlets that are only covered by ABYZ, 12 links are not reachable anymore (i.e., the outlets do not exist anymore). The 118 outlets that are (also) in Qin et al.'s sample are necessarily reachable – since I only included outlets from Qin et al. that I could find online. When searching for an online presence of Qin et al.'s outlets, I consulted with a China-based Chinese native speaker to ensure proper matches. Out of the 145 outlets that are reachable, 53 are covered by GDELT (see Section 2.3 for details on this data source) such that I can include them in my sample. The traffic of the outlets not covered by GDELT is, relative to the traffic of those that are covered, significantly smaller on average – by a factor of 37 ($p<0.01$). Similarly, the median is smaller by a factor of 23 (see Section 2.4 for details on the traffic data).

B.2. Screenshot of a front-page

Front-pages are defined as the landing page of a website. Hence, front-page articles are visible to a website visitor without further clicks. Figure A1 shows a screenshot of the People's Daily newspaper (as of 13 November 2023). People's Daily (Rénmín Rìbào) is an official newspaper of the Central Committee of the Chinese Communist Party (CCP). Back-page stories are defined as those that can only be accessed by navigating to another part of the website, typically listed under one of the subcategories (see tabs in the white area under the title section) but not linked on the landing page.

Figure A1: Screenshot of the People's Daily Front-page

Notes: The front-page (landing page) of the newspaper People's Daily, accessed as people.com.cn on 13 November 2023.

B.3. Details on the topic model

This Appendix shows the most frequent entities per topic (along with their frequency) and the manually chosen topic label – see the first and second column of Table A1, respectively. I use a Dirichlet multinomial mixture model optimized for sparse (yet high-dimensional) use cases (Yin and Wang, Yin and Wang). These are use cases with few tokens (in this case, entities) per observation (in this case, articles). Only three topics come with an inconclusive label, and they typically concern very few articles, as highlighted by the low entity frequencies in the first column.

In the third column, I also show the news level (local, national, or international). The news level is to analyze the results' robustness to excluding local news (Table C6) and heterogeneity regarding international vs. other news (Table C7). To confirm the plausibility of the news level annotations, consider, for example, that mentioning the United States is much more likely (27.0%) in articles labeled as "International" than those labeled as "Local" or "National" (only 1.7%). The same pattern arises for the United Nations: they appear in 6.1% of international articles but only in 0.3% of other articles. Prominent Western press agencies (Reuters, AP, AFP, DPA) appear in 4.1% of international articles but in a rounded 0.0% of other articles. Conversely, a mention of the sub-string "District" (districts are typical subnational units in China) is more likely in the "Local" class: 25.6% vs. 5.8%. These differences are significant with at least $p < 0.05$.

Table A1: Topic Model and Manual Topic Labels

Top entities	***Manual label***	***News level***
Tibet Autonomous Region 1416 Qinghai 843 China ...	Tibetan Autonomous Region	Local
National Basketball Association 53 Los Angeles ...	Sports	International
Xi Jinping 6 Chinese Communist Party 6 Yunnan 6...	Chinese Politics	National
China 9 Vladimir Putin 5 Russia 5 Xi Jinping 5 ...	International Relations	International
Taiwan 1549 China 1497 Democratic Progressive P...	Taiwan	International
Beijing 7 Yangzhou 6 Xinsheng Subdistrict, Yang...	Local News Yangzhou	Local

Table A1: Topic Model and Manual Topic Labels

Top entities	***Manual label***	***News level***
William Li 10 Pingquan 10 Truong Loi 10 Lingyua...	Undefined	NaN
Chinese Basketball Association 403 Liaoning Fly...	Sports	National
Yunnan 2472 Kunming 1646 China 898 Xi Jinping 4...	Yunnan Province	Local
Xinhua News Agency 11 Xi Jinping 8 Beijing 8 Av...	Science And Technology	National
Shanxi 1410 Taiyuan 1144 China 299 Xi Jinping 2...	Local News Shanxi	Local
China 2 Beijing 2 Xinhua News Agency 2 Chaoyang...	Beijing News	National
Guangzhou 7591 Guangdong 7506 Shenzhen 5000 Chi...	Local News Guangdong	Local
Inter Rhone 6 Syrah 6 Gigondas Aoc 6 Grenache 6...	Culture and Entertainment	International
Shenyang 3405 Liaoning 2507 Dalian 1284 China 9...	Local News Northeast	Local
Henan 2270 Zhengzhou 1449 China 522 Hebi 512 Xi...	Local News Henan	Local
Hebei 4374 Shijiazhuang 2221 Hebei Daily 1597 B...	Local News Hebei	Local
Xi Jinping 16842 China 14258 Beijing 10111 Chin...	Chinese Politics	National
Nan Jing Sen Lin Jing Cha Xue Yuan 3 Yun Nan Ji...	Police	National
Federal Government Of The United States 2 Alask...	United States	International
Wuhan 3 Beijing 3 Xiamen 3 Japan 3 Xinhua News ...	Covid	International
Guangzhou 2 Dongguan 2 Shenzhen 2 Guangdong-Hon...	Local News Guangdong-Hong Kong-Macau Greater Ba...	Local
Shanghai 4 Shenzhen 4 Hai Nan Zi You Mao Yi Gan...	Chinese Politics	National
Russian Railways 1 Hypertensive Crisis 1	Undefined	International
Xi'An 1979 Shaanxi 1935 Gansu 1172 Lanzhou 889 ...	Chinese Provinces And Cities	National

Table A1: Topic Model and Manual Topic Labels

Top entities	***Manual label***	***News level***
North Korea 483 People'S Volunteer Army 460 Chi...	North Korea	International
Beijing 8140 Chaoyang District, Beijing 1391 Ha...	Local News Beijing	Local
Guangxi 2997 Nanning 2306 China 994 Liuzhou 702...	Local News Guangxi	Local
Shanghai 2009 China 662 Pudong 546 Huangpu Dist...	Local News Shanghai	Local
Hong Kong 2535 China 1230 Xinhua News Agency 10...	Hong Kong	International
Anhui 1175 Jiangxi 1163 China 988 Yangtze 955 X...	Chinese Provinces And Cities	National
China 4407 Beijing 3504 People'S Bank Of China ...	Economics and Business	National
United States 6733 Russia 6342 United Kingdom 3...	International Relations	International
Anhui 4 People'S Daily 4 Haikou 4 Hainan 4 Heil...	Weather	Local
Xi Jinping 3 Hebei 3 Nanning 3 Shanxi 3 Shaanxi...	Chinese Provinces And Cities	National
Hainan 8899 Haikou 5234 Nanguo Metropolis Daily...	Local News Hainan	Local
Europe 13 Ludwig Van Beethoven 13 Central Conse...	Culture and Entertainment	International
China 1311 2020 Summer Olympics 1008 Tokyo 828 ...	Olympics 2020	International
China 4 Changchun 4 Jilin 4 Hunchun 4 East Asia...	Environment	NaN
Si Teng 2 Vin Zhang 2 Couple (Droit Et Sociolog...	Culture and Entertainment	National
China 1375 Beijing 589 Tang Dynasty 368 Song Dy...	History	National
Wuhan 6911 China 5266 Hubei 5138 Beijing 4481 X...	Covid	International
Delta Ursae Majoris 3 Beta Ursae Majoris 3 Gamm...	Space Exploration	National

Table A1: Topic Model and Manual Topic Labels

Top entities	***Manual label***	***News level***
High-Speed Rail 6 Kunming 6 Hangzhou 6 Chengdu ...	Transportation	National
United States 10663 China 3487 Joe Biden 2122 D...	United States	International
China 1111 Beijing 745 Shanghai 355 United Stat...	International Relations	International
Hohhot 1316 China 525 Hulunbuir 450 Ordos City ...	Inner Mongolia	Local
China 1361 Beijing 938 Yangtse Evening Post 755...	Culture and Entertainment	National
Japan 2441 South Korea 1006 United States 661 T...	Japan	International
Jilin 1605 Changchun 1074 China 454 Jilin City ...	Local News Jilin	Local
Italy 324 Spain 290 Premier League 280 Xinhua N...	Sports	International
Chinese Super League 544 Yangtse Evening Post 5...	Sports	National
Mao Zedong 2 Vereinigte Staaten 2 Vereinte Nati...	History	International
Xichou County 2 Fred Li 2 Xi Sa Zhen 2 Soil 2 D...	Food Safety	National
Beijing 2668 2022 Winter Olympics 2399 China 19...	Olympics 2022	International
India 2 China 2 Shanghai 2 Guangzhou 2 Peking O...	Culture and Entertainment	International
China 16408 Beijing 7230 Xi Jinping 7025 Xinhua...	International Relations	International
China 13 Xi Jinping 10 China Media Group 10 Gua...	Hong Kong	International
Beijing 12596 Shanghai 8462 China 6990 Zhejiang...	Chinese Provinces And Cities	National
Nanjing 2883 Jianye District 2095 Yangtse Eveni...	Culture and Entertainment	Local
Zhejiang 3381 Ningbo 2095 Hangzhou 1983 China 1...	Local News Zhejiang	Local

Table A1: Topic Model and Manual Topic Labels

Top entities	***Manual label***	***News level***
Xinjiang 1177 Urumqi 459 China 346 Xinhua News ...	Xinjiang and Uyghus	International
Ada Choi 3 Wang Lin (Yan Yuan) 3 Youku 3 He J...	Culture and Entertainment	Local
China 3 Guangzhou 3 Macau 3 Chinese Dream 3 Fay...	Culture and Entertainment	Local
Guizhou 1984 Guiyang 1274 Zunyi 449 China 427 X...	Chinese Provinces And Cities	Local
Chile 20 Pacific Ocean 19 Argentina 19 South Am...	South America	International
Guangdong 9 Guangzhou 6 Beijing 5 Tokyo 5 Leaf ...	Sports	Local
Heilongjiang 1136 Harbin 1031 China 404 Heihe 2...	Local News Heilongjiang	Local
China 847 Alternative Fuel Vehicle 546 Tesla, I...	Automotive Industry	International
China 3247 Beijing 3090 5G 1509 Internet Of Thi...	Economics and Business	National
China 13 Xi Jinping 11 Gansu 11 Sanjiangyuan 11...	Environment	National
Zhejiang 3 China 3 Beijing 3 Xinhua News Agency...	Chinese Politics	National
Fujian 6131 Fuzhou 3622 Quanzhou 3044 Xiamen 28...	Local News Fujian	Local
Chongqing 620 Logo 296 Chongqing Morning Post 1...	Local News Chongqing	Local
Losang Jamcan 3 Wang Wei Dong (1968Nian 5Yue)...	Chinese Politics	National
Xi Jinping 7 Beijing 7 Xinhua News Agency 7 Chi...	Chinese Politics	National
Jiangsu 6010 Nanjing 3513 Yangtse Evening Post ...	Local News Jiangsu	Local
North China 600 National Meteorological Center ...	Weather	National
Peking Opera 2 Qilu Evening News 2 Pear Garden ...	Culture and Entertainment	National

Table A1: Topic Model and Manual Topic Labels

Top entities	***Manual label***	***News level***
China 2711 United States 1939 Beijing 1276 Shan...	Science And Technology	International
China 8 Beijing 8 Xi Jinping 8 Xinhua News Agen...	Chinese Politics	National
Beijing 5 Shenzhen 5 China 5 Xi Jinping 5 Chine...	Covid	National
Volksrepublik China 860 Peking 801 Xinhua 510 X...	Undefined	International
Tianjin 564 Binhai 317 Nankai District 226 Wuqi...	Local News Tianjin	Local
China 3 Xi Jinping 3 Chongqing 3 Central Commit...	Chinese Politics	National
Mongolia 5 Sukhbaatar Province 5 Dornogovi Prov...	Mongolia	International
Sichuan 1407 Chengdu 1106 Liangshan Yi Autonomo...	Local News Sichuan	Local
China 2 China Network Television 2 Tang Guoqian...	Culture and Entertainment	Local
China 4 Xinhua News Agency 4 Beijing 4 Xi Jinpi...	Chinese Politics	International
China 5951 Chinese Communist Party 4787 Xi Jinp...	Chinese Communist Party	National
Beijing 3 China 3 Daxing District 3 Tang Dynast...	History	Local
Ningbo 2 Jiang Shan Zhan 2 Shi Shan Zhan (Zhu ...	Transportation	Local
Hunan 4263 Changsha 3536 Jpeg 2182 Wechat 1386 ...	Local News Hunan	Local
United States 130 Hollywood 99 China 82 United ...	Culture and Entertainment	International
Beijing 165 Jiangsu 113 Shanghai 112 Guangdong ...	Chinese Provinces And Cities	Local

B.4. Local Leaders

To analyze entities with varying political sensitivity across first-level subnational administrative units (typically provinces), I create indicators for articles mentioning the governor or the Communist Party secretary of the place where the newspaper is based: $OwnGovernor_{idtj}$ and $OwnSec_{idtj}$. I also build an indicator for mentions of the governors or secretaries of other provinces: $OtherGovernor_{idtj}$ and $OtherSec_{idtj}$. Table A2 shows the local leaders for each subnational unit represented in my data. These personalities were identified through manual web search. Some provinces are not represented because their (major) news outlets are unavailable in the relevant GDELT products: Chongqing (32M), Guizhou (39M), Hainan (10M), Henan (99M), Jilin (27M), Ningxia (7M), Qinghai (6M), Shanxi (35M), Tibet (4M), Xinjiang (26M). Cumulatively, the provinces covered in my data reflect 80% of the Chinese population. If the governor or the secretary changed between 2020 and 2022, I take the person in office during the larger part of that period.

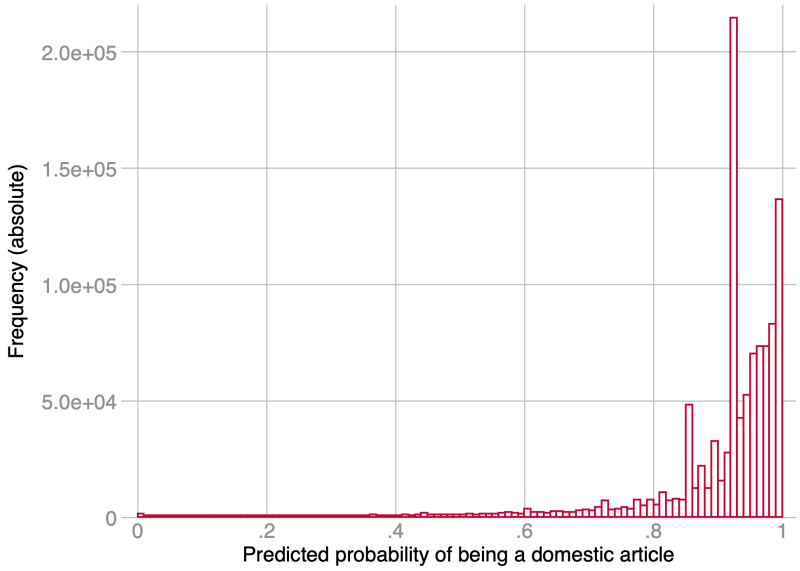
Table A2: Administrative Units and Local Leaders

Administrative unit	***Governor***	***CCP Secretary***
Beijing	Chen Jining	Cai Qi
Guangdong	Ma Xingrui	Li Xi
Fujian	Wang Ning	Yin Li
Guangxi	Chen Wu	Lu Xinshe
Jiangsu	Wu Zhenglong	Lou Qinjian
Inner Mongolia	Bu Xiaolin	Li Jiheng
Heilongjiang	Zhang Qingwei	Zhang Qingwei
Shaanxi	Zhao Yide	Liu Guozhong
Zhejiang	Yuan Jiajun	Che Jun
Shanghai	Li Qiang	Li Qiang
Hunan	Xu Dazhe	Du Jiahao
Liaoning	Chen Qiufa	Chen Qiufa
Shandong	Liu Jiayi	Li Ganjie
Hubei	Wang Xiaodong	Jiang Chaoliang
Tianjin	Zhang Guoqing	Li Hongzhong
Yunnan	Ruan Chengfa	Chen Hao
Hebei	Xu Qin	Wang Dongfeng
Jiangxi	Yi Lianhong	Liu Qi
Sichuan	Yin Li	Peng Qinghua
Gansu	Tang Renjian	Lin Duo
Anhui	Li Guoying	Li Jinbin

C. Measuring Alignment: Additional Material

C.1. Predicted similarity with domestic vs. foreign content

Figure B1: Distribution: Predicted Prob. of Being a Domestic (vs. Foreign) Article ($\widehat{D}_{idtj}$)



Notes: The distribution of the predicted probability of being domestic (as opposed to foreign) content, $\widehat{D}_{idtj}$, for the sample of Chinese news. The histogram shows absolute frequencies.

C.2. Chi2 association of entities and back-page placements

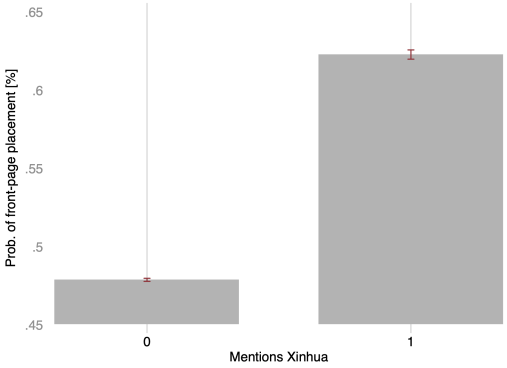
Figure B2: Entities Associated with Back-Page Placement in Domestic News (Chi2 Association)

Notes: The 1,000 entities with the highest frequency ratio for back-page placement. Only entities with a total frequency above 100 were considered, and those whose Chi2 association with back-page placement is significant with $p<0.01$. The entity's font size correlates with its frequency in domestic news.

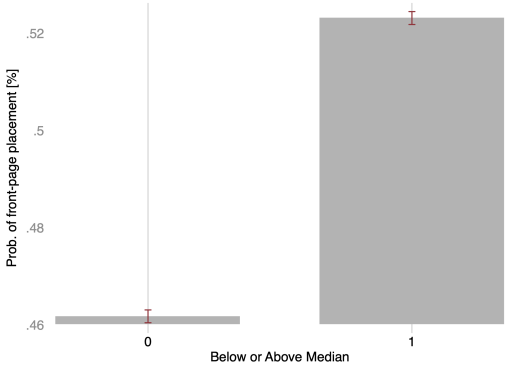
D. Main Results: Additional Material

Figure C1: Raw Front-Page Placement Probabilities by Content

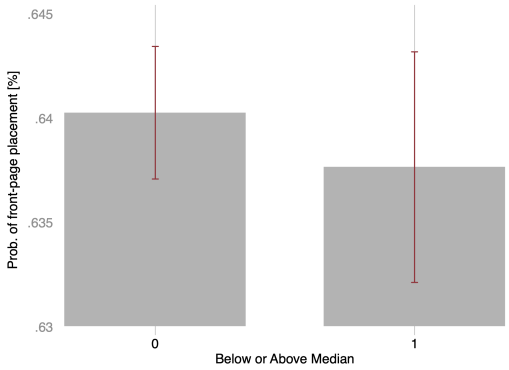
(a) Xinhua mention



(b) Predicted prob. of being domestic



(c) Leader × Sentiment



Notes: All Subfigures show the front-page probability on their vertical axis. Regarding the horizontal axis, Subfigure (a) distinguishes between articles that do or do not mention Xinhua, Subfigure (b) between articles with below or above median values of the machine learning-based measure, and Subfigure (c) between articles on the leader with sentiment scores below or or above the median value. The red bars show the respective 95% confidence interval.

Table C1: Xinhua Mentions and Front-Page Placement – Different FE

	Dep. variable: $frontpage_{idtj}$			
	(1)	(2)	(3)	(4)
$Xinhua_{idtj}$	0.126***	0.115***	0.135***	0.075***
	(0.040)	(0.039)	(0.033)	(0.024)
N observations	984626	984626	984622	984621
Day-Year FE	-	X	X	X
Topic FE	-	-	X	X
Outlet FE	-	-	-	X
Entity Count Control	X	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest captures whether article *idtj* cites Xinhua. All columns control for the number of entities mentioned in the article. The fixed effects (day-year, topic, and outlet) are added sequentially, as shown in the Table. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C2: Similarity with Domestic News and Front-Page Placement – Different FE

	Dep. variable: $frontpage_{idtj}$			
	(1)	(2)	(3)	(4)
$\widehat{D}_{idtj}$	0.143***	0.139***	0.079***	0.073**
	(0.044)	(0.044)	(0.029)	(0.031)
N observations	1090601	1090601	1090597	1090596
Day-Year FE	-	X	X	X
Topic FE	-	-	X	X
Outlet FE	-	-	-	X
Entity Count Control	X	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable captures the predicted probability of article content being domestic (as opposed to foreign). All columns control for the number of entities mentioned in the article. The fixed effects (day-year, topic, and outlet) are added sequentially, as shown in the Table. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C3: Leader Articles' Sentiment and Front-Page Placement – Different FE

	Dep. variable: $frontpage_{idtj}$			
	(1)	(2)	(3)	(4)
$Leader_{idtj}$	0.154	0.155	0.043	-0.007
	(0.126)	(0.125)	(0.079)	(0.055)
$Sentiment_{idtj}$	-0.016	-0.046	-0.116	-0.123
	(0.195)	(0.185)	(0.116)	(0.075)
$Leader_{idtj} \times Sentiment_{idtj}$	0.011	-0.000	0.149	0.163**
	(0.170)	(0.169)	(0.101)	(0.067)
N observations	1090601	1090601	1090597	1090596
Day-Year FE	-	X	X	X
Topic FE	-	-	X	X
Outlet FE	-	-	-	X
Entity Count Control	X	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest captures the sentiment of articles mentioning the leader (relative to those not mentioning the leader). All columns control for the number of entities mentioned in the article. The fixed effects (day-year, topic, and outlet) are added sequentially, as shown in the Table. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Alignment with the Government Perspective and Front-Page Placement – Week-Year FE

	Dep. variable: $frontpage_{idtj}$		
	(1)	(2)	(3)
$Xinhua_{idtj}$	0.076***		
	(0.024)		
$\widehat{D}_{idtj}$		0.011**	
		(0.005)	
$Leader_{idtj}$			-0.008
			(0.055)
$Sentiment_{idtj}$			-0.123
			(0.075)
$Leader_{idtj} \times Sentiment_{idtj}$			0.165**
			(0.068)
N observations	984621	1090596	1090596
Week-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest captures whether article *idtj* cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. All columns include week-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C5: Alignment with the Government Perspective and Front-Page Placement – Circulation Weights

	Dep. variable: $frontpage_{idtj}$		
	(1)	(2)	(3)
$Xinhua_{idtj}$	0.016***		
	(0.006)		
$\widehat{D}_{idtj}$		0.158***	
		(0.045)	
$Leader_{idtj}$			0.074
			(0.051)
$Sentiment_{idtj}$			-0.237***
			(0.080)
$Leader_{idtj} \times Sentiment_{idtj}$			-0.006
			(0.054)
N observations	984621	1090596	1090596
Day-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic, weighted by the publishing outlet's traffic (specifically, the reach). The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest captures whether article $idtj$ cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C6: Alignment with the Government Perspective and Front-Page Placement – Excluding Local News

	Dep. variable: $frontpage_{idtj}$		
	(1)	(2)	(3)
$Xinhua_{idtj}$	0.083***		
	(0.025)		
$\widehat{D}_{idtj}$		0.056**	
		(0.023)	
$Leader_{idtj}$			0.021
			(0.066)
$Sentiment_{idtj}$			-0.067
			(0.079)
$Leader_{idtj} \times Sentiment_{idtj}$			0.157*
			(0.088)
N observations	984621	1090596	1090596
Day-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. Articles covering local news (as annotated by the topic model, see Appendix B.3) are excluded. The dependent variable is an indicator of whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest captures whether article *idtj* cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C7: Alignment with the Government Perspective and Front-Page Placement – International vs. Other News

	Dep. variable: $frontpage_{idtj}$		
	(1)	(2)	(3)
$Xinhua_{idtj}$	0.086***		
	(0.022)		
$Xinhua_{idtj} \times International_{idtj}$	-0.024*		
	(0.014)		
$\widehat{D}_{idtj}$		0.090**	
		(0.034)	
$\widehat{D}_{idtj} \times International_{idtj}$		-0.023	
		(0.020)	
$Leader_{idtj}$			0.016
			(0.043)
$Sentiment_{idtj}$			-0.161*
			(0.086)
$Leader_{idtj} \times Sentiment_{idtj}$			0.117**
			(0.047)
$Sentiment_{idtj} \times International_{idtj}$			0.140*
			(0.080)
$Leader_{idtj} \times Sentiment_{idtj} \times International_{idtj}$			0.108***
			(0.038)
N observations	984621	1090596	1090596
Day-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The main right-hand side variable captures whether article *idtj* cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. This respective main right-hand side variable is interacted with an indicator for international news (as annotated by the topic model, see Appendix B.3) in all columns. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C8: Alignment with the Government Perspective and Front-Page Placement – During Political Meetings

	(1)	(2)	(3)
$Xinhua_{idtj}$	0.075***		
	(0.024)		
$Xinhua_{idtj} \times Political_{idtj}$	0.002		
	(0.026)		
$\widehat{D}_{idtj}$		0.074**	
		(0.031)	
$Political_{idtj} \times \widehat{D}_{idtj}$		-0.038	
		(0.045)	
$Leader_{idtj}$			-0.010
			(0.054)
$Sentiment_{idtj}$			-0.122
			(0.075)
$Leader_{idtj} \times Sentiment_{idtj}$			0.167**
			(0.067)
$Leader_{idtj} \times Political_{idtj}$			0.099
			(0.121)
$Political_{idtj} \times Sentiment_{idtj}$			-0.038
			(0.120)
$Leader_{idtj} \times Political_{idtj} \times Sentiment_{idtj}$			-0.143
			(0.175)
N observations	984621	1090596	1090596
Day-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The main right-hand side variable captures whether article $idtj$ cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. In all columns, this respective main right-hand side variable is interacted with an indicator for days during the Chinese high-level political meeting, the "Two Sessions". All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.

Table C9: Alignment with the *Local* Government Perspective and Front-Page Placement

	Dep. variable: $frontpage_{idtj}$	
	(1)	(2)
$Sentiment_{idtj}$	-0.064	-0.065
	(0.081)	(0.082)
$OwnSec_{idtj}$	-0.264	
	(0.207)	
$OwnSec_{idtj} \times Sentiment_{idtj}$	0.660*	
	(0.364)	
$OtherSec_{idtj}$	0.228	
	(0.149)	
$OtherSec_{idtj} \times Sentiment_{idtj}$	-0.329	
	(0.228)	
$OwnGovernor_{idtj}$		-0.304
		(0.183)
$OwnGovernor_{idtj} \times Sentiment_{idtj}$		0.649**
		(0.305)
$OtherGovernor_{idtj}$		0.250*
		(0.147)
$OtherGovernor_{idtj} \times Sentiment_{idtj}$		-0.344
		(0.246)
N observations	1090596	1090596
Day-Year FE	X	X
Topic FE	X	X
Outlet FE	X	X
Entity Count Control	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The first right-hand side variable of interest captures the sentiment of articles mentioning the province's own leader (relative to those not mentioning them). The second right-hand side variable of interest captures the sentiment of articles mentioning any provincial leader from another province (relative to those not mentioning them). In column 1, the provincial leader indicator focuses on provincial Communist Party Secretaries. In column 2, it captures the provincial governors. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E. Mechanisms: Additional Material

E.1. Annotations for sensitive Wikipedia entities

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
18th National Congress of the Chinese Communist Party	0	0
2008 Chinese Milk Scandal	1	2
2014 Hong Kong Protests	1	0
A City Of Sadness	1	3
A Touch Of Sin	1	0
Ai Fen	1	1483
Alexandria Ocasio-Cortez	1	11
Anarchism	0	1
Animal Farm	1	1
Anson Chan	1	22
Anthony Wong Yiu-Ming	1	3
Apple Daily	1	95
Azerbaijan	0	1175
Bad Genius	1	3
Bai Chongxi	0	5
Bai Yansong	1	283
Baidu	0	2887
Beauty And The Beast (2017 Film)	1	12
Bei Dao	1	34
Belt And Road Initiative	0	169
Bet365	1	1
Bilibili	0	568
Blind Mountain	1	2
Boxer Rebellion	0	113
Brokeback Mountain	1	3
Bytedance	0	69
Cai Xia	1	142
Caixin	1	46
Chai Jing	1	60
Charles Heung	0	5

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
Chen Qiushi	1	21
China	0	368061
China Central Television	0	3766
China Radio International	0	72
Chinese Civil War	0	13
Chinese Communist Party	0	53053
Chow Yun-Fat	1	20
Cold War	0	109
Confucius Institute	0	694
Cult	0	10420
Cultural Revolution	0	60
Cyberpunk 2077	1	9
Dai Xianglong	0	6
Daryl Morey	1	38
Death Of Lei Yang	1	9
Death Of Li Wangyang	1	1
Deng Xiaoping	0	3021
Deng Yujiao Incident	1	1
Deyunshe	0	68
Doctor Strange (2016 Film)	1	1
Edo Period	0	2
Education Bureau	0	86
Edward Leung	1	5
Eileen Chang	0	72
Eric Tsang	1	44
Eritrea	0	117
Facebook	1	2462
Falun Gong	1	13
Fang Fang	1	126
Farewell My Concubine (Film)	1	89
Fidel Castro	0	76
First Sino-Japanese War	0	18
Fort Detrick	0	1002
Franklin D. Roosevelt	0	84
Freedom Of Speech	1	3
Game Of Thrones	1	41

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
Gao Gang	0	21
Genshin Impact	0	54
Github	1	12
Global Times	0	9306
Go Princess Go	1	9
Gong Li	1	374
Google	1	1792
Government Of The Republic Of China	0	22
Grand Order	1	6
Great Chinese Famine	1	0
Greater East Asia Co-Prosperity Sphere	0	0
Hannibal (Tv Series)	1	2
Headliner (Tv Programme)	1	3
Ho Chi Minh	0	204
Hong Kong Alliance in Support of Patriotic Democratic Movements of China	1	58
Hong Kong Human Rights And Democracy Act	1	14
Hong Kong National Security Law	0	1680
Hu Lancheng	0	8
Hu Yaobang	1	24
Hua Guofeng	0	2
Huang Ju	0	230
Hundred Regiments Offensive	0	49
Hungarian Revolution Of 1956	0	1
Iwane Matsui	0	10
Ji Bingxuan	0	47
Jiang Qing	0	18
Jiang Zemin	0	473
Jin Xing	1	85
Jing Junhai	0	246
Johor	0	55
Julian Assange	0	116
Jurchen People	0	8
Kai-Fu Lee	1	23
Kangxi Emperor	0	114
Keith J. Krach	1	20

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
Kim Jong-Un	0	507
Korean War	0	377
Koxinga	0	97
Kublai Khan	0	4
Lee Teng-Hui	1	91
Lei Feng	0	327
Li Hongzhi	1	42
Li Rui (Politician)	1	112
Li Wenliang	1	150
Li Xi (Politician, Born 1956)	0	308
Li Zongren	0	14
Lianhe Zaobao	1	760
Liberate Hong Kong, Revolution Of Our Times	1	9
Lin Biao	0	30
Line (Software)	1	41
Liu Xiaobo	1	187
Logan (Film)	1	2
Long March	0	1537
Lu Xun	0	1088
Luo Huining	0	109
Luo Yonghao	1	315
Lust, Caution	1	6
Ma Sicong	1	7
Manchukuo	0	80
Mao Sui	0	13
Mao Zedong	0	8439
March Of The Volunteers	0	453
Memoirs Of A Geisha (Film)	1	1
Ming Dynasty	0	502
Ministry Of Public Security (China)	0	91
Ministry Of State Security (China)	0	16
Monster Hunter (Film)	1	0
Mulan (2020 Film)	1	145
Music	0	2537
My Love From The Star	1	7
Netease	1	327

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
New Tang Dynasty Television	1	4
Nomadland (Film)	1	27
Northern And Southern Dynasties	0	44
Occupy Central With Love And Peace	1	221
Ouyang Nana	0	87
Peng Dehuai	0	335
Pewdiepie	1	0
Pinterest	1	18
Prague Spring	0	0
Publicity Department of the Chinese Communist Party	0	170
Puyi	0	46
Qing Dynasty	0	389
Radio Free Asia	1	14
Red Guards	1	66
Regina Ip	0	57
Regional Comprehensive Economic Partnership	0	3702
Reporters Without Borders	1	2
Republic Of China Air Force	0	0
Russo-Japanese War	0	14
Sansha	0	156
Second Sino-Japanese War	0	383
Second Taiwan Strait Crisis	0	6
Sharon Cheung	1	6
Shinjuku Incident	1	0
Sina Weibo	0	2848
Sino-Soviet Conflict (1929)	0	1
Someday Or One Day	1	38
South China Morning Post	0	492
Story Of Yanxi Palace	1	87
Sun Zhengcai	1	29
Sundar Pichai	0	26
Taiwan Affairs Office	0	569
Taiwan Under Japanese Rule	0	3
Telegram (Software)	1	339
Tencent	0	2112

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
Tencent Video	0	169
Teresa Teng	0	1095
Tfboys	0	25
The Beijing News	0	1825
The Eight Hundred	1	11
The Empress Of China	1	10
The Epoch Times	1	16
The Flowers Of War	1	9
The Founding Of A Republic	0	72
The Karate Kid (2010 Film)	1	2
The New York Times	1	4019
The Three-Body Problem (Novel)	1	52
The Untamed (Tv Series)	1	67
Thucydides Trap	0	53
Tian Zhuangzhuang	1	29
Tianjin	0	17346
Tim Cook	0	125
To Live (1994 Film)	1	13
Traditional Chinese Medicine	0	1483
Tunisian Revolution	0	4
Tuo Zhen	0	44
Turkey	0	4000
Twitter	1	1823
Uyghurs	1	778
Vanguard (Film)	0	90
Vietnam	0	5050
Vivi Miao	0	185
Wang Chen (Politician)	0	299
Wang Dan (Dissident)	1	262
Wang Jingwei	0	27
Warriors Of The Rainbow: Seediq Bale	1	1
Weathering With You	0	2
Wechat	0	27429
Wen Jiabao	1	25
Wenzhou Train Collision	1	4
Western Xia	0	76

Table D1: Sensitive Entities and their Mentions in Domestic Outlets

Entity	**Censored**	**Mentions**
Identified as potentially censored through keyword search on its Wikipedia page	***Manual label***	***In domestic news***
Wikipedia	1	89
William Barr	1	67
Winnie-The-Pooh	1	7
World Journal	0	440
Wuhan Diary	1	3
Xi Jinping	0	116664
Xi Zhongxun	0	164
Xiao Zhan	1	96
Xinhua News Agency	0	99055
Xu Lin (Born 1963)	0	101
Yan Xishan	0	23
Yang Jia	1	718
Yellow Emperor	0	180
Yongle Emperor	0	58
Yoshiko Kawashima	0	2
Yoshiko Yamaguchi	1	2
Yuan Dynasty	0	125
Zhang Binglin	0	16
Zhang Xianzhong	0	18
Zhang Zhehan	1	67
Zhang Zhixin	0	16
Zhao Ziyang	1	3
Zhongnanhai	0	733
Zhou Enlai	0	1182
Zhu Yi (Figure Skater)	0	23

Table D2: Censored Entity Mentions and Front-Page Placement – Different FE

	Dep. variable: $frontpage_{idtj}$			
	(1)	(2)	(3)	(4)
$censored\_entity_{idtj}$	-0.099***	-0.093***	-0.060***	-0.021***
	(0.032)	(0.028)	(0.022)	(0.007)
N observations	1090601	1090601	1090597	1090596
Day-Year FE	-	X	X	X
Topic FE	-	-	X	X
Outlet FE	-	-	-	X
Entity Count Control	X	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The right-hand side variable of interest is an indicator of whether an explicitly censored entity is mentioned in the article: $censored\_entity_{idtj}$ (see Section 6.1 for details on the definition). All columns control for the number of entities mentioned in the article. The fixed effects (day-year, topic, and outlet) are added sequentially, as shown in the Table. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E.2. Checks around the event study research design

It is, to the best of my knowledge, unclear from the current literature on difference-in-differences (DID) how to ensure that my results based on Equation 4 do not suffer from the problems described by De Chaisemartin and D'Haultfoeuille (2020). De Chaisemartin and D'Haultfoeuille demonstrate that linear regressions with period and group fixed effects estimate weighted sums of the average treatment effects (ATE) in each group and period. Thereby, weights can be negative. Negative weights could reverse the coefficient sign (e.g., the linear regression coefficient could be negative while all the ATEs are positive).

My observations are at the article level, but I do not follow articles over time. Instead, I look at unique articles that appear in outlets that are followed over time. To the best of my knowledge, the DID literature does not address potential problems and, accordingly, solutions that may apply to this data structure (which is *not* a traditional panel).

I transform my data into a more traditional panel structure to ensure that my results are not driven by idiosyncracies in the data structure. Specifically, I switch from article-level observations to entity-level observations. I use the 984 entities according to the filtering for the machine learning-based model, see Section 3.2.1: They have to appear 50 times and in both foreign and Chinese news. Entities' presence in domestic and foreign news is also essential for the event study since the NYT/BBC articles are the control group.

So, I follow every entity over the entire observation period. For each day, I calculate the entity's front-page probability in domestic and foreign news, respectively. This gives rise to a fully balanced panel allowing this specification:

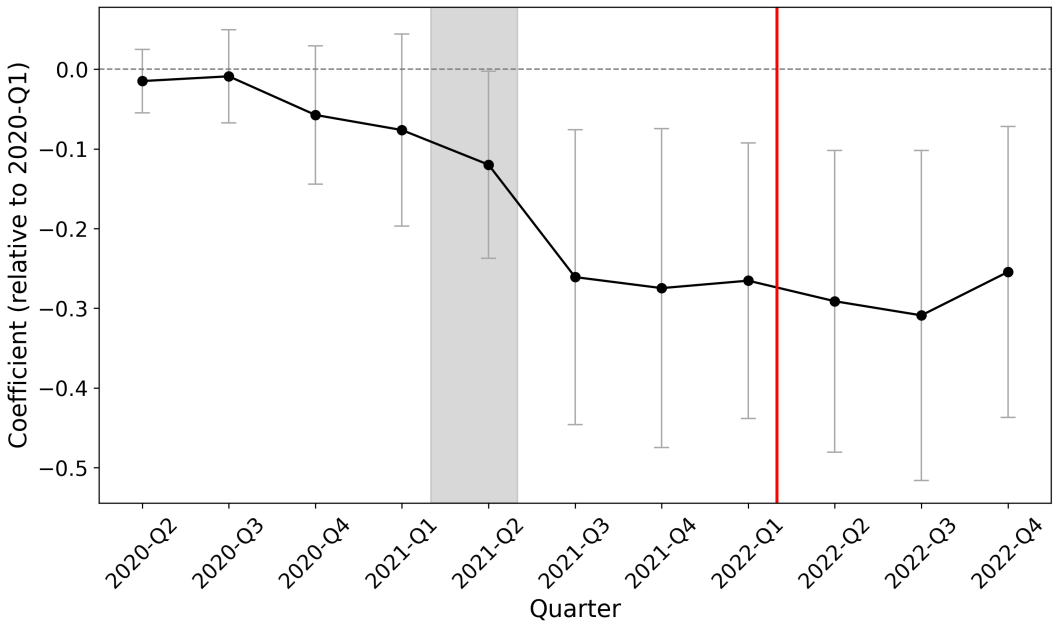
$$\begin{aligned} Frontpage_{ed} = \nu_e + \kappa\ Ukr\_Entity_e \\ + \sum_{q=2020-02}^{2022-12} \mu_q\ Ukr\_Entity_e \times I[Quarter = q]_{ed} \\ + \sum_{q=2020-02}^{2022-12} \tau_q\ Chin\_Outlet_{ed} \times I[Quarter = q]_{ed} \times Ukr\_Entity_e \\ + \vartheta_{ed} \end{aligned} \tag{23}$$

$Frontpage_{ed}$ now captures if entity e's front-page probability on date d (calculated as

its front-page appearances on that day divided by its total appearances). ν_e is an entity fixed effect. Ukr_Entity_e equals one if the entity is Putin, Zelenskyy, Russia, or Ukraine. $Chin\_Outlet_{ed}$ indicates whether the figures refer to the entity's Chinese or foreign appearances on that day. $I[Quarter = q]$ captures in which quarter from 2020-Q2 to 2022-Q4 the respective date lies. Again, Articles appearing in 2020-Q1 are the reference group. The coefficients of interest are τ_q, where the expectation is that quarters before the beginning of the crisis are statistically indistinguishable from zero, and those after are negative.

Since my entity-level panel is strongly balanced and all units are treated at the same time (with the escalation around 2021-Q1/Q2), a standard DID research design is valid to estimate τ_q (see Roth et al., 2023). Figure D1 presents the τ_q coefficients from Equation 23, confirming the findings from the Main Part's Figure 4.

Figure D1: The Marginal Front-Page Probability of Ukraine War-related Entities (with NYT &BBC articles as controls)



Notes: The Figure shows the marginal probability of a Ukraine war-related entity (Putin, Zelenskyy, Russia, Ukraine) to feature on Chinese front-pages, relative to the NYT and the BBC front-pages, over time. Specifically, the vertical axis captures the τ_q coefficient (see event study specification in Equation 23) for each quarter from 2020-Q2 to 2022-Q4 (2020-Q1 is the reference). The first and second quarter of 2021 mark the beginning of the escalation, especially due to the deployment of Russian troops close to the Ukrainian border (area shaded in grey). The red line is the date of the Russian invasion (24 February 2022).

E.3. Beijing Winter Olympics case study

Complementing the case study on the outbreak of the war in Ukraine (February 2022), the following case study investigates how the Beijing 2022 Olympic Games influence the front- vs. back-page differential.

The Olympic Games are a useful setting: Anecdotal evidence (e.g., Graham-Harrison and Ni, 2022) suggests that the reporting differed quite drastically between Chinese outlets and foreign (Western) outlets. While the domestic narrative emphasized national pride and Beijing's remarkable ability to rally against the Coronavirus, Western media highlighted concerns over human rights abuses. For example, the latter discussed the treatment of Uyghurs, athletes like tennis star Peng Shuai, and potential risks and challenges related to the pandemic.

Table D3: Entities Most Frequently Mentioned in Articles on the Winter Olympics 2022

Foreign News Sources		***Chinese Outlets***	
Entity	Count	Entity	Count
United States	245	Xi Jinping	5085
Russia	132	Xinhua News Agency	3666
Joe Biden	115	Zhangjiakou	3508
Xi Jinping	104	Hebei	2326
COVID-19	86	Winter	2040
United Kingdom	85	United States	1967
Vladimir Putin	83	Chinese Communist Party	1862
Ukraine	82	Yanqing District	1789
Uyghurs	79	Russia	1267
Hong Kong	74	2008 Summer Olympics	1070

Notes: The 10 most frequently mentioned entities in articles that also refer to the "2022 Winter Olympics". The left-hand side shows those entities for foreign news sources (The New York Times, BBC), and the right-hand side for Chinese outlets.

The anecdotal evidence is supported by Table D3, which shows the ten most frequent entities in Olympic Games articles in the foreign news (The New York Times, BBC) vs. those most frequent in Chinese outlets.[56] In foreign news, both mentions of the "Uyghurs" and "Hong Kong" feature prominently. Also, the mentions of Russia, Ukraine, and Putin suggest the ongoing crisis in Ukraine was also discussed prominently (recall that the

[56]Entities that are "mechanically" frequent in Olympic Games articles, such as "Winter Games" or "Beijing", are removed.

invasion of Ukraine by Russia took place immediately after the Olympics ended, on 24 February 2022).

This case study assumes that potentially sensitive content on China in foreign outlets becomes more salient during the Olympic Games. At the same time, as a sports mega-event, the Olympic Games likely attract a lot of entertainment-oriented interest, making it a plausible candidate for market differentiation between more and less investigative readers.

Table D4 compares how the alignment metrics relate to front-page placement during the Olympic Games – relative to other dates. $Olympics_{idtj}$ indicates whether an article was published during the Olympics (4 to 20 February 2022) and interacted with the alignment measures. In line with Table 2's main result, column 1 shows that articles citing Xinhua are 7.5 percentage points more likely to end up on the front page. The interaction term between $Xinhua_{idtj}$ and $Olympics_{idtj}$ has a positive coefficient. However, both the machine learning-based measure and the leader indicator interacted with the sentiment come with relatively large and significant positive effect sizes. During the Olympics, being more aligned in terms of $\widehat{D}_{idtj}$ or sentiment of articles mentioning the leader results in more than double the impact on front-page placement (with $p<0.05$ and $p<0.10$, respectively).

All in all, these results are consistent with more aligned articles having an enhanced likelihood of featuring prominently during the Olympics.

Table D4: Alignment with the Government Perspective and Front-Page Placement – Olympic Games

	Dep. variable: $frontpage_{idtj}$		
	(1)	(2)	(3)
$Xinhua_{idtj}$	0.075***		
	(0.024)		
$Xinhua_{idtj} \times Olympics_{idtj}$	0.022		
	(0.030)		
$\widehat{D}_{idtj}$		0.011**	
		(0.005)	
$Olympics_{idtj} \times \widehat{D}_{idtj}$		0.019**	
		(0.009)	
$Leader_{idtj}$			-0.005
			(0.055)
$Score_{idtj}$			-0.124
			(0.075)
$Leader_{idtj} \times Score_{idtj}$			0.160**
			(0.067)
$Leader_{idtj} \times Olympics_{idtj}$			-0.165
			(0.104)
$Olympics_{idtj} \times Score_{idtj}$			0.153
			(0.135)
$Leader_{idtj} \times Olympics_{idtj} \times Score_{idtj}$			0.267*
			(0.159)
N observations	984621	1090596	1090596
Day-Year FE	X	X	X
Topic FE	X	X	X
Outlet FE	X	X	X
Entity Count Control	X	X	X

Notes: OLS estimates. Cross-section with article-newspaper observations by day and topic. The dependent variable indicates whether an article is featured on the front-page: $frontpage_{idjt}$. The main right-hand side variable captures whether article *idtj* cites Xinhua in column 1, the predicted probability of article content being domestic (as opposed to foreign) in column 2, and the sentiment of articles mentioning the leader (relative to those not mentioning the leader) in column 3. This respective main right-hand side variable is interacted with an indicator for the Olympics (4 to 20 February 2022) in all columns. All columns include day-year, topic, and outlet fixed effects, as well as a control for the number of entities mentioned in the article. In column 1, the media outlet Xinhua is dropped. Standard errors are multiway-clustered at the outlet, day-year, and topic level (in parentheses): * p < 0.1, ** p < 0.05, *** p < 0.01.