Assessing Mass Media Coverage of China in Four OECD Countries: Using CATA to Compare Trends and Topics

Juraj Medzihorsky[1] and Erin K. Jenne[1]

[1]Central European University

29/08/2016

Abstract

This paper outlines the preliminary descriptive results and planned next steps for a Minerva Initiative-funded computer-aided text analysis (CATA) of journalistic texts related to China in four major English language media markets from the end of the Cold War to the present. To assess China's success in presenting itself to the world in a positive light (what we loosely call Chinese soft power), we have mapped journalistic treatments of China in those countries where Beijing has invested considerable resources attempting to curry public favor, indicating Beijing sees its perception in these countries as a foreign policy priority. To construct these datasets, we used a keyword search of the LexisNexis database to collect all articles in the four countries' newspapers that referenced China. A secondary aim of the paper is to outline how we will drill down in the data to arrive at a more fine-grained analysis of how China is portrayed. Our research design describes initial explorations of this large newspaper corpus to see if the overall news coverage of China in a given country (and thus public opinions on China) varies depends on the subject of the story and attendant resonance of the subject with the audience nation–in making this determination, we examine salient China-related stories ranging from human rights, Tibet, the South China Sea to the Beijing Consensus. In a later stage, we intend to parse the data using topic models and sentiment/tonality dictionaries to assess the overall positivity of media coverage of China over time and across markets. Ultimately, we aim to dissect the connection between media coverage and public opinion in respective societies to determine whether differential media coverage across countries and time corresponds with meaningful variation in public opinion.

Paper prepared for the 112th APSA Annual Meeting in Philadelphia, September 1-4, 2016

Word count: 5,556

Introduction

China has finally arrived as a contender for global primacy. Following two decades averaging ten percent GDP growth per year, China now boasts the second biggest economy in the world after that of the United States.[1] China is also the most important trading partner of the US, has the world's largest foreign currency reserves, and is the single largest holder of US treasuries. Although Chinese power is almost entirely based on its economic leverage, Beijing has also begun to build up its military power while tensions have grown across the Taiwan Strait and with Japan over the Senkaku/Diaoyudao islands in the South China Sea. Majorities of Western societies now believe that China will inevitably overtake the US as the global superpower,[2] and Western analysts have grown concerned with the "rise of China," particularly in light of the country's poor human rights record and totalitarian character as well as its increasingly bellicose rhetoric in its conflicts with neighbors. Scholars and pundits alike have sounded the alarm about China's growing capabilities and attendant threat to the West.

Eager to allay international fears of Chinese global dominance, Beijing has struck a largely accommodationist tone on the international stage.[3] At the same time, it has used economic instruments such as financial aid, trade and direct foreign investment to leverage the country geopolitically by building relations of economic dependence with other countries. At the same time, Beijing has rather self-consciously attempted to offer itself as an alternative model of global hegemony–one that respects the sovereignty and self-determination of countries around the world by refraining from interfering in other state's domestic affairs.

At least as important as Beijing's economic relations and international diplomacy has been the government's efforts to manage China's public image in rich democracies. From the early 2000s, then-Chinese President Hu Jintao announced a policy of soft power toward western countries, establishing the hundreds of so-called "Confucius Institutes" mostly affiliated with local universities;

---

[1] http://www.tradingeconomics.com/china/gdp-growth-annual.

[2] Pew Research Center. Global Publics Back U.S. on Fighting ISIS, but Are Critical of Post-9/11 Torture: Asian Nations Mostly Support TPP, Defense Pivot âĂŞ but Also Value Economic Ties with China. June 23, 2015. http://www.pewglobal.org/files/2015/06/Balance-of-Power-Report-FINAL-June-23-20151.pdf.

[3] Baohui Zhang. "Chinese foreign Policy in Transition: Trends and Implications," *Journal of Current Chinese Affairs* 2 (2010): 39-68.

these institutes promote Chinese values and culture and seek to cultivate positive lateral relations with foreign publics. These developments have led to dire warnings that China could soon overtake the US in propagating its values. Toshi and Holmes stated that "Chinese soft power is growing in all three of Nye's dimensions [culture, political values, diplomatic behavior], bolstering Beijing's diplomatic fortunes throughout Asia."[4] In Beijing's view, the utility of soft power "lies in its ability to foster an external environment conducive to China's rise as an economic and military power," and to mitigate the notion that China poses a threat to the world.[5]

Many scholars have greeted China's soft power policies with alarm, while others argue that they are unlikely to yield much in the way of influence due to the totalitarian government's inherently self-limited appeal. Many others note that the success of these strategies depend on the target country, its demographic characteristics, culture, ethnic traits and political beliefs. This paper describes a first-cut analysis of these data and explains how we intend to drill down further into media coverage of China in these four countries by exploring whether differential coverage of China across different foreign media markets helps to account for variable views of China in the different populations as revealed in survey research. The tentative question is twofold: whether there is a relationship between the type of coverage of China in a cross-section of newspapers in a given society, and if so, whether the type of media coverage maps onto popular "feelings" about China, as measured through survey research.

This paper presents an overview of a large text database containing news articles on China in four of the world's most important English language media markets: the United States, the United Kingdom, Canada and Australia. This paper mostly describes trends in coverage of themes related to China in the four countries over time. In the present exploratory analysis, we offer suggestive evidence as to the ways in which national media can limit the success of a country's soft power strategy. To assess this possibility, we conduct a preliminary analysis on trends in how China has been portrayed in four English-speaking countries that are targeted by Chinese diplomatic outreach. We gathered all relevant articles on China in each of these four countries from 1989 to 2015. Later, we perform a more fine-grained examination of whether and how coverage of China varies (in valent

---

[4]Yoshihara, Toshi and James R. Holmes, "China's energy-driven 'soft power'." *Orbis* 52(1) (2008), pp. 126.
[5]Wuthnow, Joel. "The Concept of Soft Power in China's Strategic Discourse." *Issues and Studies* 44(2) (2008), p. 7.

and/or content) across different media markets within the same country, and whether this variation maps onto differential perceptions of China within their respective readerships. This may shed light into the role of the media in facilitating the success of soft power strategies.

This paper is structured as follows. We first provide a brief overview of Joseph Nye's concept of "soft power." We then describe why it is of particular importance to rising hegemons such as China. Third, we present an overview of our datasets, including descriptive statistics on how much and how China is covered in the four media markets and discuss how these results can be interpreted. We close with a frank summary of the analytical hurdles we face in analyzing these massive databases and explain how we intend to address them.

Soft Power and Rising Hegemons

Joseph Nye first coined the term "soft power" to refer to the ways in which a state might influence other states or audiences beyond military means. Soft power, in Nye's formulation, lies beyond coercive tools such as economic compellence or military force, but instead refers to the nation's intangible abilities to get others "to want what you want." Nye later elaborated on the sources of a nation's soft power, which he claims is rooted in "its culture (in places where it is attractive to others), its political values (when it lives up to them at home and abroad), and its foreign policies (when others see them as legitimate and having moral authority)." Writing at the end of the Cold War in response to predictions of declining American material power, Nye argued that the United States can continue to lead the world despite its waning material advantages due to America's intangible power of attraction (or "soft" power) that leads other nations to follow due to the overwhelming appeal of its institutions and values. In his 1990 volume, Nye explained that the US would continue to exercise influence on the international stage even as its hard power waned because American soft power is still so vital, engendering many allies despite the US' decreasing ability to compel others to follow its lead.

Soft power is vital not just to waning global hegemons, but also to rising global powers. Beijing has sought to exercise greater international influence despite the fact that it does not yet have a preponderance of material power. No one recognizes this fact more than China, which has attempted

to use soft power to allay other countries' concerns about China's growing power. Since nations are continually seeking influence over others (the conventional definition of power according to international relations theory), achieving influence through persuasion rather than coercion is both cheaper and safer.

Much has been written upon China's attempts to cultivate soft power as a foreign policy resource.[6] Since the early 2000s, China has deliberately employed cultural diplomacy to curry positive relations within the populations of the prevailing global powers (the US and its allies) by attempting to appeal to the west and sell itself as a benevolent rising hegemon that poses no real risks to the democracies of the world. Over a decade ago, Chinese Premier Hu Jintao announced its intent to win over other nations through charm. This includes a program inaugurated in 2004 to establish hundreds of cultural outreach centers, Confucius Institutes, which partner with local universities and institutes around the world. In depicting itself as a benign peaceful power, China

> has taken the lead in promoting ancient cultural icons whom it thinks might have global appeal...At the forefront is Confucius. Few Westerners can quote a saying of Confucius. But most at least regard him as a bearded, wise dispenser of aphorisms, far more profound than America's superficial consumerism. The party is promoting him as a kind of Father Christmas without the undignified jolliness.[7]

Having settled on their soft power brand, Beijing has invested at least USD 500 million to found 530 CIs at universities and colleges and 631 CCs in primary and secondary schools in 127 countries.[8] Likewise, China has promoted what is called the "Beijing Consensus," "a combination of mixed ownership, basic property rights, and heavy government intervention" that stands as a deliberate contrast to the still dominant neo-liberal "Washington Consensus."[9] In the recent years, Xi Jinping has propagated the ideal of "Chinese Dream," which is taken to be part of a "national rejuvenation" involving the return of China's lost international status.[10] Their working hypothesis is that if western

---

[6] Joseph N. Nye, *Bound to Lead: The Changing Nature of American Power* (New York: Basic Books, 1990).

[7] http://www.economist.com/node/21541714.

[8] "China Says No Talking Tibet as Confucius Funds U.S. Universities", http://www.bloomberg.com/news/articles/2011-11-01/china-says-no-talking-tibet-as-confucius-funds-u-s-universities.

[9] Yang Yao. "The End of the Beijing Consensus: Can China's Model of Authoritarian Growth Survive?" *Foreign Affairs* (online, February 2, 2010).

[10] Zheng Wang. "The Chinese Dream: Concept and Context." *Journal of Chinese Political Science* 19 (2014), pp. 1-13.

populations are well-disposed to China, they will pressure their political leaders to maintain largely accommodationist positions with China.

The Debate over Chinese Soft Power

Since the end of the Cold War, China has begun to close the gap with the United States as a viable contender for superpower status. China's low labor costs and undervalued currency have allowed the government to build up the kind of industrial base and export-driven economy that can allows the emerging power to compete with the US and its allies for influence around the world. The ruling party has invested heavily in infrastructure, human capital, and other things aimed at making the country even more competitive internationally. Despite its considerable and growing economic clout, China's military remains small, which is partly the reason that China is not viewed with greater distrust. Despite this, the West is viewing China with increasing alarm as Beijing has provoked conflict with neighbor states in the South China Sea and begun to build up its military apparatus, while the country is again engaged in cross-border disputes with Taiwan over the latter's international status.

Seeking to ameliorate western concerns about China's intentions in the early 2000s, Beijing has launched a multi-decade information campaign aimed at selling the west on a benevolent Chinese power. These policies were at first met with alarm in the west. Ramo, for instance, argued that China's growth "contradicts every single principle formulated in the Washington Consensus," and that the key to China's success its "ideological and operating dexterity."[11]

Proponents of the "China as threat" thesis have grown less strident somewhat in recent years, as Beijing's soft power strategy has so far yielded disappointing returns. Popular perceptions of the country stagnated or even dipped in countries around the world. This has led some to argue that China's soft power is inherently self-limited, largely due to its totalitarian system of government and poor track record on human development as well as the fact that China's appeal is based on a fairly narrow criterion–high levels of economic growth.[12] China's global clout could easily go south

---

[11]Ramo, quoted in Huang, 2011, p. 3.

[12]Zhao, Suisheng. "The China Model: can it replace the Western model of modernization?" *Journal of Contemporary China* 19 (2010), 419-436.

if its economy begins to slow or crash, particularly since China is in many ways a "dysfunctional," "paradoxical" or "weak developmental state."[13] China is also said to suffer from a significant "cultural deficit," undermining its performance in the world cultural-industry market.[14] Further hampering its ability to exercise soft power is China's overly propagandistic approach, which fails to resonate with western publics.[15]

Contemporary analysts have taken a more nuanced view of Beijing's ability to influence foreign audiences. Scholars have written extensively on cross-national variation in outside perceptions of China. Their underlying premise is that the culture, values and ethnicity of the audience country serve as crucial intervening variables in the potency of Chinese soft power "resources." In short, Chinese soft power is more likely to resonate in countries with greater ethnic affiliation with China. Still another argument is that positive opinions of China are directly proportional to the level of economic and cultural ties between China and the country in question. Holyk uses survey data to argue that Chinese soft power has little purchase with democratic publics, such as the US and Japan, due to their vastly diverging values.[16] The conclusion is that China can best manage public perceptions in countries where "cultural and economic interfaces have tempered the "aggressive" flavor associated with the rise of China as a great power."[17]

We seek to contribute to an understanding of the conditions under which Chinese soft power is likely to find receptive audiences. To investigate these conditions, we draw attention to an understudied component of public receptivity to another nation's soft power diplomacy, which is the intervening impact of the audience country's national media on the perceptions of China within that public. Perhaps the closest to what we attempt here is the work of Ambrosio, who conducted content analysis on journalistic stories on China in newspapers both in and outside of China over time.[18] He

---

[13]Stephan Ortmann. "The 'Beijing Consensus' and the 'Singapore Model': Unmasking the Myth of an Alternative Authoritarian State-Capitalist Model." *Journal of Chinese Economic and Business Studies* 10 (4) (2012), p. 342. See also Yang 2010; and Ramo, Joshua Cooper. *The Beijing Consensus*. (London: The Foreign Policy Centre, 2004).

[14]Huang, Yanzhong and Sheng Ding. "Dragon's Underbelly: An Analysis of China's Soft Power." *East Asia* 23(4) (2006), p. 35.

[15]David Shambaugh. *China Goes Global: The Partial Power*. Oxford: Oxford University Press, 2013.

[16]Gregory G. Holyk. "Paper Tiger? Chinese Soft Power in East Asia." *Political Science Quarterly* 126(2) (2011), pp. 223-254.

[17]Parama Sinha Palit. "China's Soft Power in South Asia." *SIS Working Paper*, No. 200. Singapore: Nanyang Technical University, 21.

[18]Thomas Ambrosio conducts a similar exploratory analysis to what is offered here. He conducts content analysis on articles on China in newspapers outside of China. He finds that there was an increased interest in China around 2010, but

concludes that the publicized fears about the appeal of the "Beijing consensus" are not supported by textual evidence, and that while some see the Chinese model of authoritarian development positively, their numbers are few.

Compared to Ambrosio, what we aim to do is far more ambitious; we undertake computer-aided content analysis (CATA) on a complete archive of LexisNexis articles on China from 1990 to 2015 that were published in four English-speaking countries where China has focused much of its soft power efforts. Below are the summary results of how this coverage varied over time and across space, as well as suggested explanations for this variable coverage. Subsequently, we detail how we intend to use this data to drill down into why certain China stories received more sympathetic treatments in certain mass media markets than others. We hope to find that these treatments are associated with different perceptions of China in each of these countries, demonstrating that national media serve as a vital intervening variable in the effectiveness of China's soft power strategies.

The Corpus

Our research assistants have collected a vast corpus of textual data consisting of all articles published in four English-speaking OECD countries (Australia, Canada, UK, and USA) from the LexisNexis newspaper database containing the word "China." To ensure semantic homogeneity of our textual data, and because Chinese engaged in little soft power diplomacy before 1990, we collected only data from the post-Cold-War period, restricting the publication dates from 1 January 1990 to 31 December 2015. (However, for the countries of UK and USA we also included articles from all of 1989.) This procedure yielded over 4 million entries.

A non-negligible fraction of the collected documents were perfect or near-perfect duplicates. This is because frequently the same text was released in several slightly different versions by the same outlet or by multiple outlets. To handle this issue, we have used the Jaro-Winkler distance as implemented in R package Record Linkage[19] to deduplicate the data.[20] The measure ranges

---

that negative stories increasingly outnumbered positive stories.

[19]Borg, A., & Sariyar, M. (2016). RecordLinkage: Record Linkage in R. R package version 0.4-10, `http://CRAN.R-project.org/package=RecordLinkage`. Sariyar, M., and Borg, A. (2010). The RecordLinkage package: Detecting errors in data. *The R Journal*, 2(2), 61-67.

[20]M. A. Jaro. 1989. "Advances in record linkage methodology as applied to the 1985 census of Tampa Florida". *Journal*

from 0 (completely different documents) to 1 (identical documents). Specifically, we have used the following algorithm:

> 1) We arranged the documents from each of the four countries chronologically within each country.
>
> 2) For each document, except the last three, we computed the pairwise Jaro-Winkler distances to the three documents that follow it according to the ordering in (1). The span of three documents was selected after testing a span of 20 on a subset of the documents first, which indicated that the serious duplicates are nearly always within three places according to the order in (1).
>
> 3) We removed all documents with pairwise distance larger than 0.85 with at least one of the three documents that precede it.

The de-duplication procedure yielded a corpus of 3,548,727 documents. Of those, 266,196 were from Australia, 529,132 from Canada, 1,933,493 from the UK, and 819,196 from the US. The oldest document was originally published on 1 January 1989 and the newest on 31 December 2015. Figure 1 depicts the count of texts by country and year.[21]

What is immediately apparent in these graphs is that mass media coverage of China is greater in the UK than the other three countries, particularly from the early 2000s. Assuming that this is not an artefact of overrepresentation of British media in the LexisNexis database during these dates or simply that news articles in the UK are overall more numerous than articles in the US, Canada, or Australia (likely to be the case for Canada and Australia, but unlikely to be the case for the US), then we can tentatively conclude that a greater proportion of British media has been devoted to stories about China than in the other three mass media markets. If true, this is at least partly due to stronger historical ties between UK and China as well as a British sense of moral responsibility toward Hong Kong since its handover of sovereignty to China in 1997.[22] However, what we also

---

of the American Statistical Association 84 (406): 414-20; M. A. Jaro. 1995.. "Probabilistic linkage of large public health data file". *Statistics in Medicine* 14 (5âĂŞ7): 491-8; W. E. Winkler. 1990. "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage". Proceedings of the Section on Survey Research Methods (American Statistical Association): 354-359.

[21]The gap in China stories in the US in 1997 is puzzling, as are the two shorter gaps for the UK in the mid 2000s. Some articles miss dates, since they were not parsed correctly. We will correct these problems in later analyses.

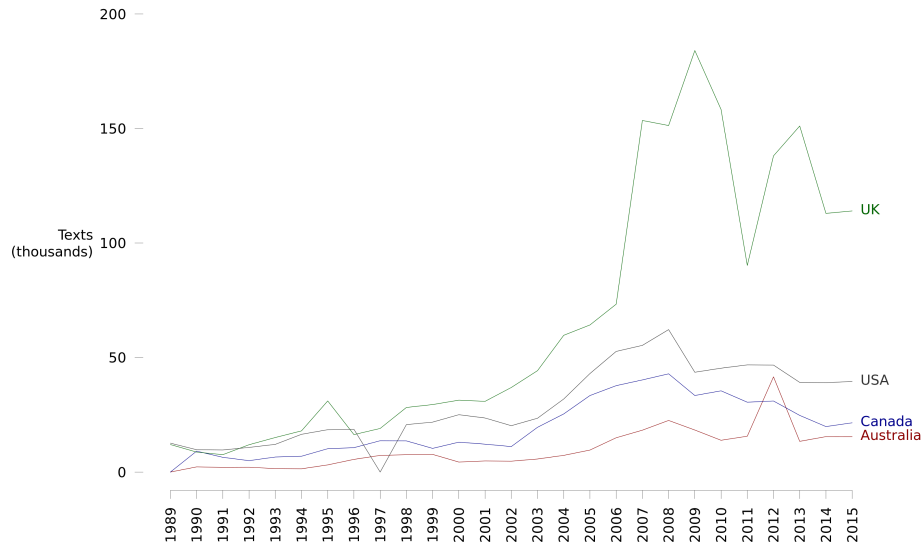[22]Correspondence with Victoria Hui, August 27, 2016.

Figure 1: Texts in the corpus by country and year of publication

see in this data is that the coverage of China in the British media has increased considerably from the mid-2000s. Given that much of this coverage is economic in nature, and given the increasing dependence of the conservative government on Chinese investment in recent years in particular, we believe that at least some of this rise in coverage is an outgrowth of growing dependence of the British economy on China.

These figures also indicate upticks in coverage of China (overall and within each country) in 2008 and 2012. For the most part, this was a response to events that were of considerable interest to western audiences. In 2008, China hosted the Olympics in Beijing, and many western journalists wrote stories about China's insufficient observance of human rights. Also in that year, there was unrest in Tibet on the annual commemoration of Tibet Uprising Day, with street protests led by monks escalating into violence by Tibetans against Hui and Han citizens. The Chinese government put down the unrest quickly, and foreign media have focused on human rights violations committed by Chinese authorities in the process.

The 2012 uptick in coverage was partly driven by the 2012 anti-Japanese protests over the disputed islands in the South China Sea. These protests, held in cities across China, escalated into violence and many protesters were detained and demonstrations banned. Also that year, there was
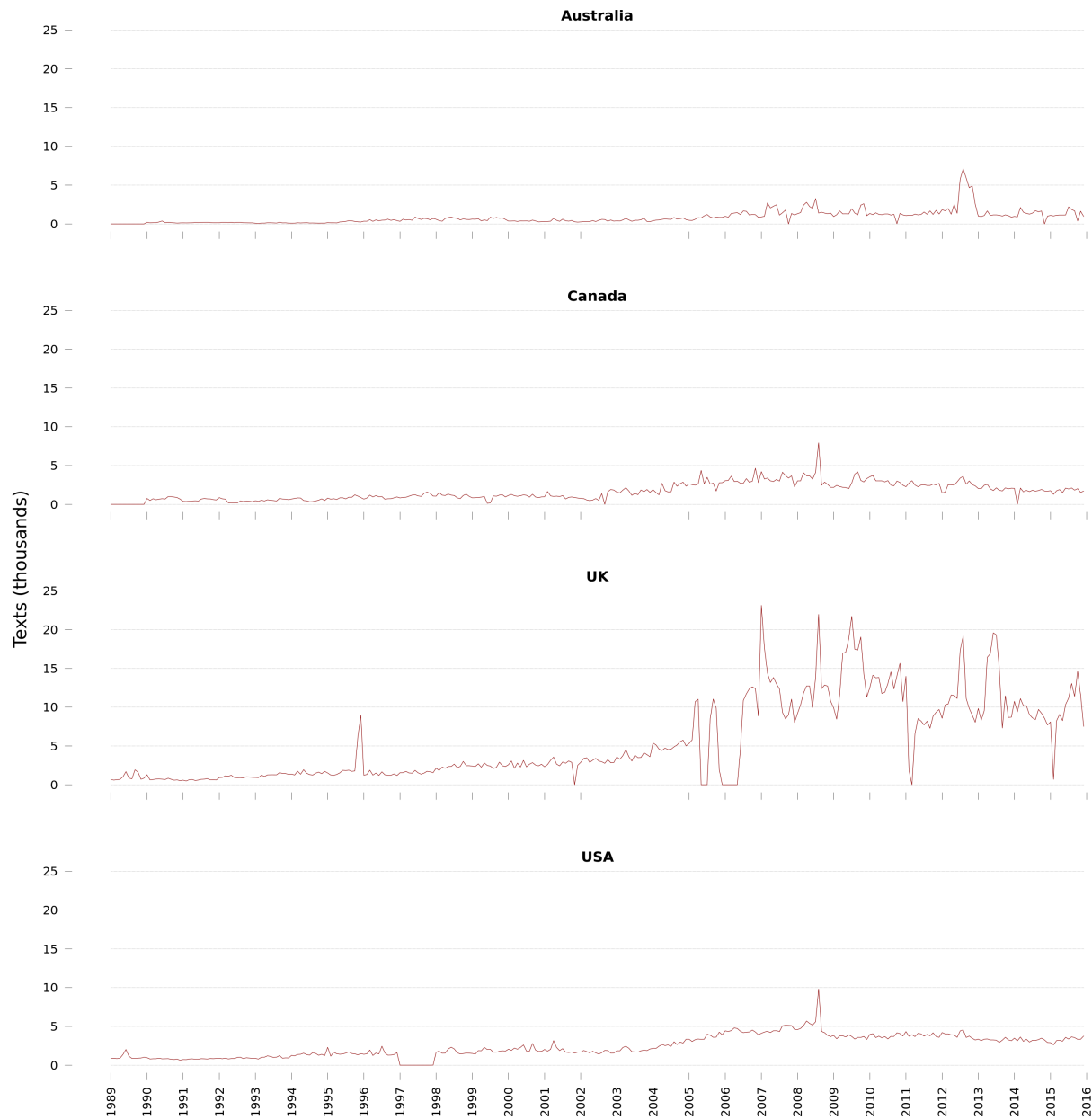
Figure 2: Texts in the corpus by country and month of publication

a change in leadership from Hu Jintao to current Chinese President Xi Jinping in the context of China's rising stature on the international stage. We acknowledge, however, that these are merely preliminary guesses based on macro-trends, which will be examined more closely on further parsing of the data.

Given the size of the dataset, we opted to explore it by inspecting subsets defined by keyword

11

lists. Each subset was produced by taking all documents that contained at least one word from a list of keywords attached to a theme from a set thereof. The themes are as follows:

1. Human Rights

2. Beijing consensus

3. Tibet

4. South China Sea

5. China-led efforts to establish new international banks, the Asian Infrastructure Investment Bank and the New Development Bank (formerly known as the "BRICS Bank") (AIIB and BB) that can serve as alternatives to western-dominated World Bank and International Monetary Fund (IMF).

Human Rights

The "Human Rights" sub-corpus was produced by selecting all the documents in the corpus that lowercased contain the expression "human rights" (hereafter, HR corpus). The oldest of the selected documents was published on 2 January 1989 and the newest on 31 December 2015. 10,088 of the documents are from Australia, 20,964 from Canada, 42,627 from the UK, and 28,294 from the US. Figure 3 shows the numbers of documents and Figure 4 the numbers of words by years. Figure 5 shows the counts of documents by country and month. Figure 6 shows the ten most common words in the sub-corpus by year and Figure 7 the ten most common words by country-year.

Based on the above graphs, we conclude that the the peaks in human rights coverage (in the overall corpus and in individual mass media markets) largely correspond to peaks in overall coverage of China, suggesting that stories about China in each of the four markets usually feature a human rights angle–when western journalists write about China, they often write about their observance of human rights. This is certainly true for the uptick in coverage associated with the Beijing Olympics, where stories often centered around China's failures to improve their substandard record on human rights, and on China's handling of the unrest in Tibet that same year. Indeed, this is the only major
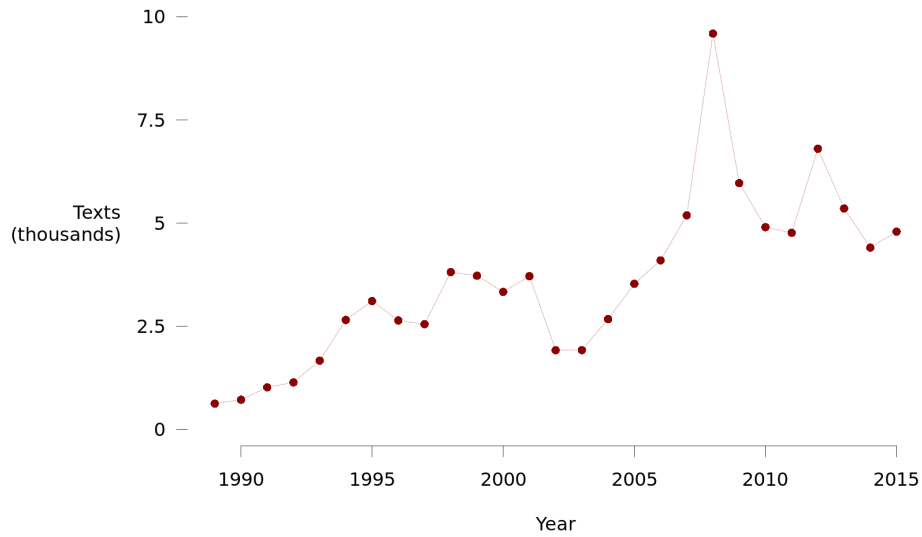
Figure 3: Texts in the HR corpus by year of publication

Figure 4: Words (stemmed) in the HR corpus by year of publication

peak of human rights coverage on China in Australia and, to a lesser extent, Canada. The 2008 peak also showed up in the UK and the US, but the latter two featured additional upticks in human rights coverage of China at other times (early 2000s for the US and post-2010s for the UK). Here again, the UK is the outlier–it is the only country whose media featured considerable human rights coverage on China after 2008. As noted above, the 2012 uptick was most probably due to western

Figure 5: Texts in the HR corpus by year of publication

journalists' critical coverage of China's repressive handling of the 2012 anti-Japanese protests.

Beijing Consensus

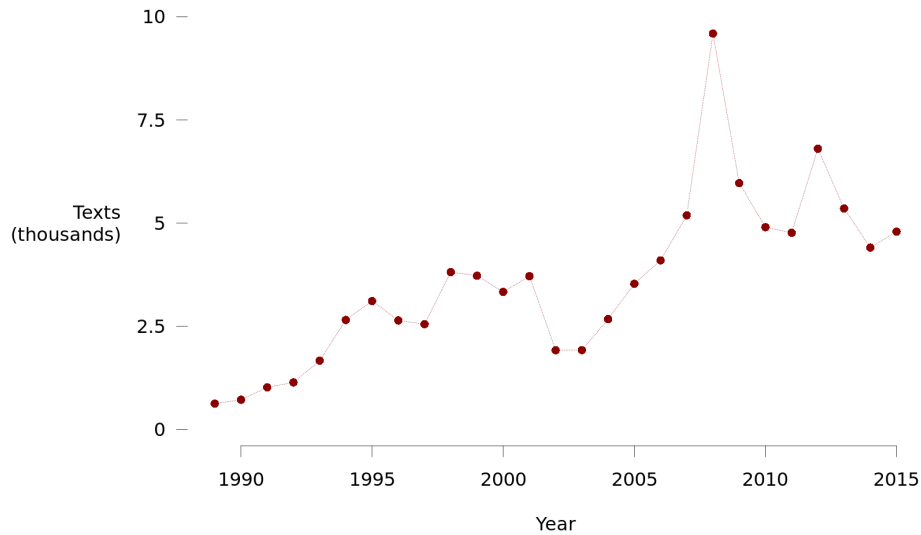The term'Beijing Consensus' was coined by UK Foreign Policy Centre Analyst Joshua Cooper Ramo in 2004 to refer to the Chinese alternative to the western neo-liberal model of economic growth ("the Washington Consensus"). According to Ramo, the Beijing Consensus features state- rather than market-driven investment policies aimed at "equitable, peaceful high-quality growth", and "defense of national borders and interests." To create the Beijing Consensus sub-corpus (hereafter, BC corpus), we selected all the documents from the larger corpus containing lowercased at least one of the two expressions "beijing consensus" and "china model". This procedure yielded a collection of 510 documents, of which 84 are from Australia, 63 from Canada, 256 from the UK, and 107 from the US. The oldest document was published on 19 May 1992 and the newest on 1 December 2015. Figure 8 shows the counts of the documents and Figure 9 of the words by year. An informal inspection of the data indicates that mid-year months are represented more than others. Table 1 shows this by country, indicating that this is the case for the texts from the UK. Figure 10 shows the ten most common words for each year in the BC corpus.

14

Figure 6: Top ten most frequent words (stems) in the HR corpus by year

A reliable interpretation of the data necessitates drilling down further into the stories themselves, but on the face of it, these figures suggest that the Beijing Consensus garnered far greater interest among journalists in the UK than in the other three English language mass media markets. This is at least partly due to the fact that the term itself was coined in the UK. Moreover, the greater dependence of the British economy on Chinese investment might account for why more articles

15

Figure 7: The most common words in the HR corpus by country and year

Figure 8: Document in the BC sub-corpus by publication year

about the Chinese economic model were published in the UK. In terms of overall trends, BC stories appear to rise from 2006 to 2010 across all countries as the term caught on in popularity. The peak of BC stories in the UK in 2015 (and to a lesser extent, in the the other three countries) is likely due to a spike in criticism of the Beijing Model in light of China's growing economic troubles during this time.

Tibet

For the Tibet sub-corpus, we included all documents from the larger corpus that contained at least one of the two expressions "tibet" and "dalai." This yielded 87,776 documents, the oldest of which
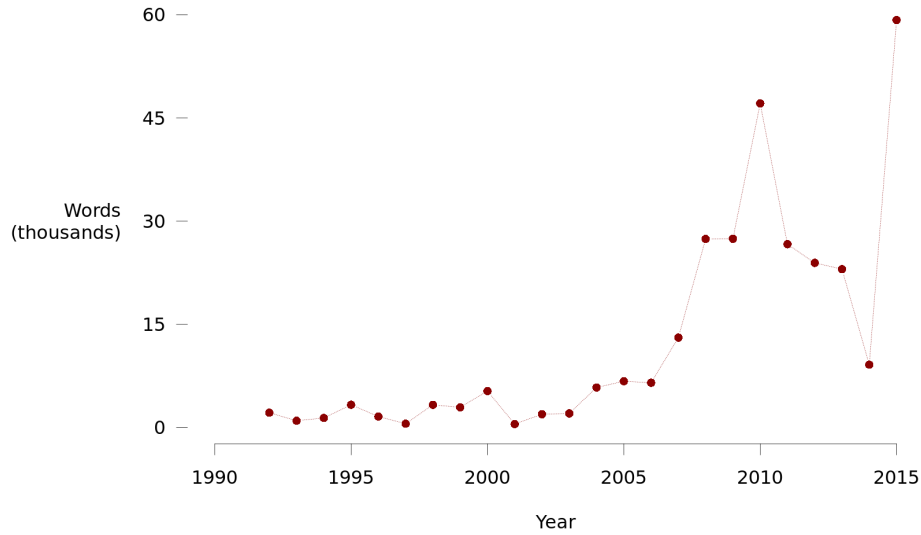
Figure 9: Words (stems) in the BC sub-corpus by publication year.

Table 1: Document counts in the BC corpus by month of the year and country

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	NA/s
AUS	8	8	9	3	12	7	6	7	6	10	6	2	0
CAN	0	8	5	8	6	7	6	8	6	3	5	1	0
UKG	11	12	17	29	20	26	29	25	18	24	16	14	15
USA	9	4	4	10	8	6	11	8	11	17	7	9	3

was published on 1 January 1989 and the newest on 31 December 2015. Of these documents, 11,743 were from Australia, 16,529 from Canada, 38,337 from the UK, and 21,167 from the US.

These results are largely self-explanatory in light of the March 2008 unrest in Tibet, which occurred in the run-up to the Beijing Olympics later that year. Possibly emboldened by the international spotlight due to the upcoming Olympics, Tibetan monks led peaceful protests against Beijing's policies of settling ever more Han Chinese in the Tibetan Autonomous Region (TAR), undermining the spirit of a long-standing agreement with Chinese authorities on Tibetan autonomy. The protests were held in and around Tibetan capital of Lhasa during the annual commemoration of Tibetan Uprising Day. After four days, these escalated to violent street fighting that pitted Tibetans against
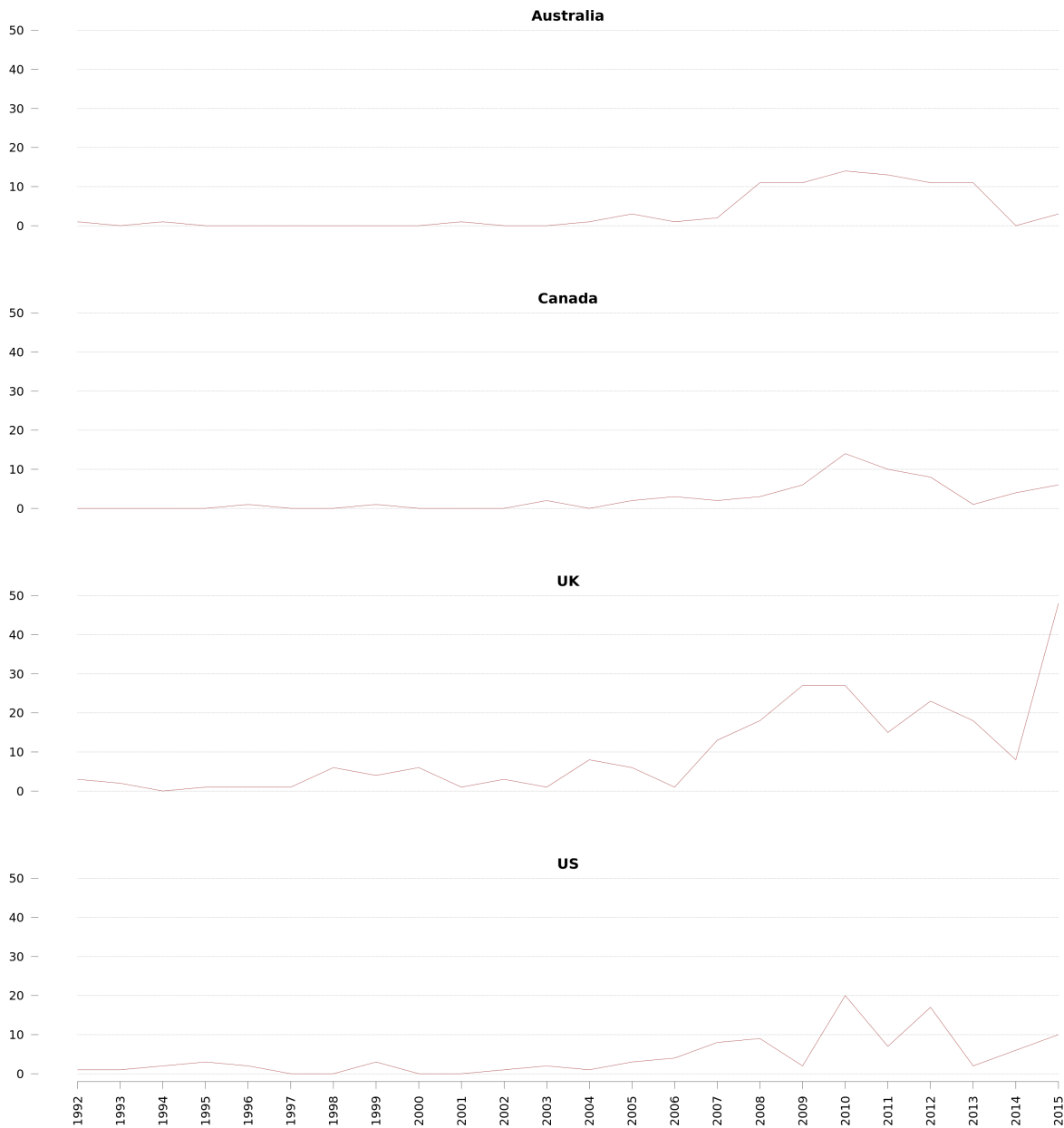
17

Figure 10: Documents in the BC sub-corpus by country and year of publication.

mostly Han Chinese. The Chinese authorities swiftly moved to quell the conflict, killing and detaining an unknown number of civilians (including monks) in the process. Journalists across all four media markets covered these events in Tibet, leading to international condemnation of China's repressive methods of societal management.

Figure 11: Ten most common words for each year in the BC corpus.

South China Sea

For the South China Sea (hereafter SCS) sub-corpus, we included all documents containing lower-cased at least one of the three expressions "south china sea," "spratly islands," and "paracel islands." This yielded 20,963 documents the oldest of which was published on 1 January 1989 and the newest
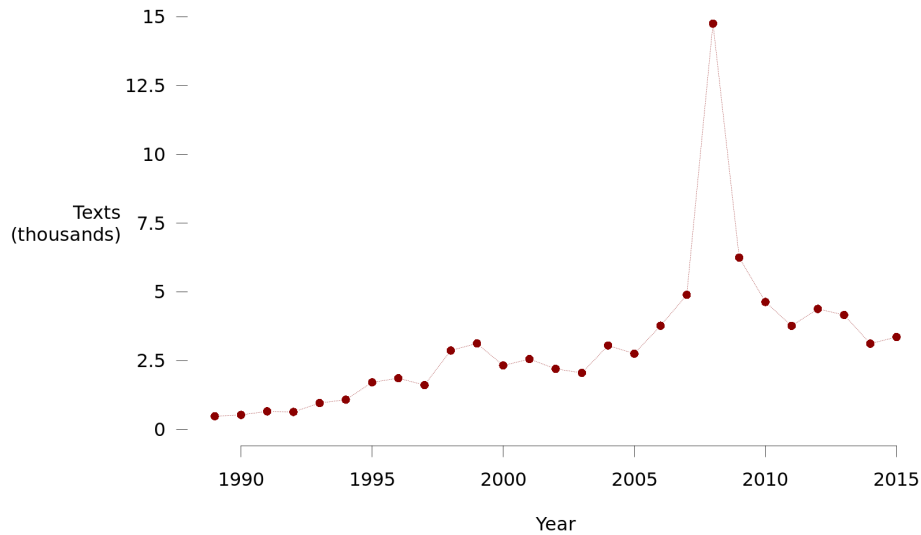
Figure 12: Documents in the "Tibet" sub-corpus by publication year.

on 31 December 2015. Of the documents, 2,994 were from Australia, 2,932 from Canada, 10,284 from the UK, and 4,733 from the US.

The trends in coverage of China with respect to the South China Seas serve as a broad indicator of international assessments of Chinese expansionist tendencies in the neighborhood, relating to the theme of "China as a threat." The 1995 coverage is likely associated with China's 1994-5 initial construction of oil drilling rigs in this area, which led to protests by the Philippine government. The 2001 coverage relates to the Hainan Island incident wherein a US Navy intelligence aircraft collided with a Chinese fighter jet, leading to a diplomatic incident between China and the US.

In 2014, we see an even greater uptick in the number of media stories about China and the South China Seas. In this year, Conflict had escalated between Vietnam and China when The China National Offshore Oil Corporation (CNOOC) began drilling for oil near the disputed Paracel Islands in the South China Sea, which are controlled by China but claimed by Vietnam. Demanding a three mile exclusion zone around the oil rig, Chinese vessels intimidated Vietnamese vessels in the area.

At the same time, China engaged in tense relations with the Philippines, whose patrols had apprehended a Chinese vessel near the disputed Spratly Islands on charges of smuggling hundreds of endangered turtles out of the area. Beijing swiftly condemned Manila's interference as politically
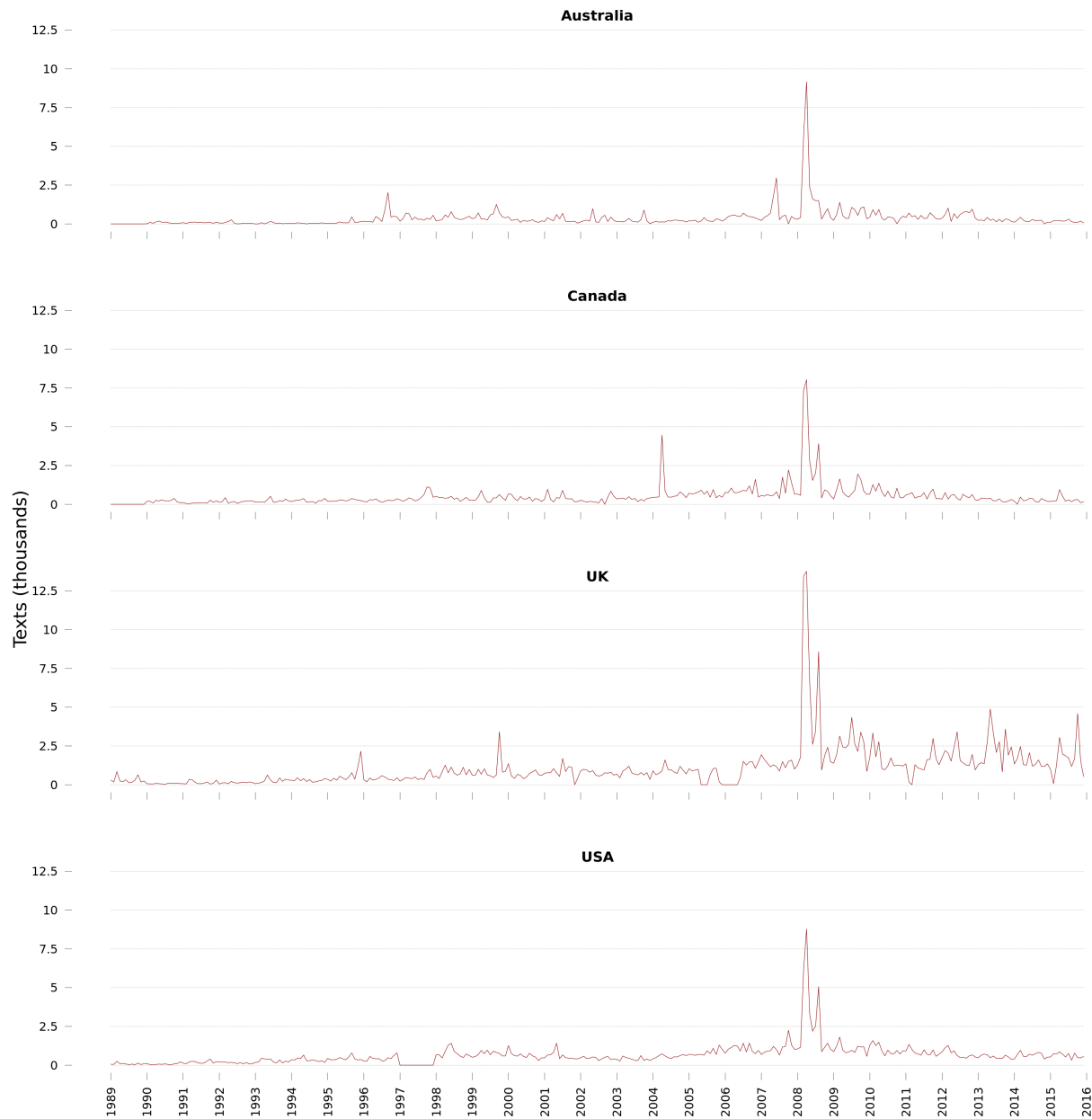
20

Figure 13: Documents in the "Tibet" sub-corpus by country and month of publication.

motivated. These events have left an impression in international media markets that the country could be pivoting to a more martial approach toward satisfying its national interests–Beijing reckons that up a third of its gas and oil reserves lie under the South China Sea.
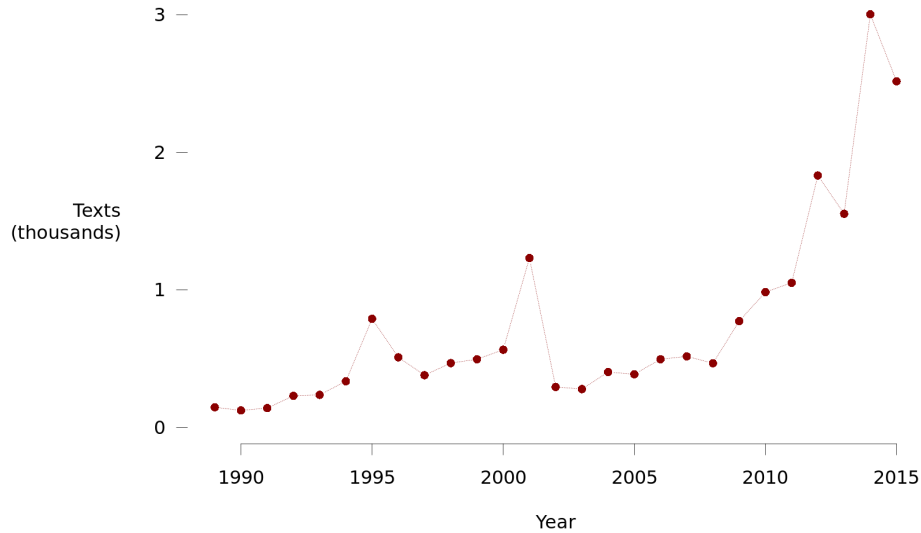
Figure 14: Documents in the SCS sub-corpus by publication year.

AIIB and BRICS banks

For the'banks' sub-corpus, we have selected all documents containing at least one of the two expressions'asian infrastructure and investment bank' and'brics bank.' This yielded 903 documents, the oldest of which was published on 1 April 2012 and the newest on 29 December 2015. The much narrower time window compared to the other sub-corpora is due to the fact that these topics are much newer. Of the documents, 219 were from Australia, 57 from Canada, 506 from the UK, and 120 from the US.

This smallish corpus offers little in the way of meaningful results thus parsed, but suggests that (as in other issue-areas) the UK has published a greater number of Chinese economic stories than the other three countries, indicating it has greater interest and sensitivity to economic events in that country. The uptick across all countries in early 2015 corresponds with Beijing's announcement that it was establishing the Asian Infrastructure Investment Bank as an alternative to the western-dominated IMF, thus offering other options to developing countries in need of financing. The second peak in late 2015 occurs only in British mass media, most likely in response to Xi's official visit to the UK.
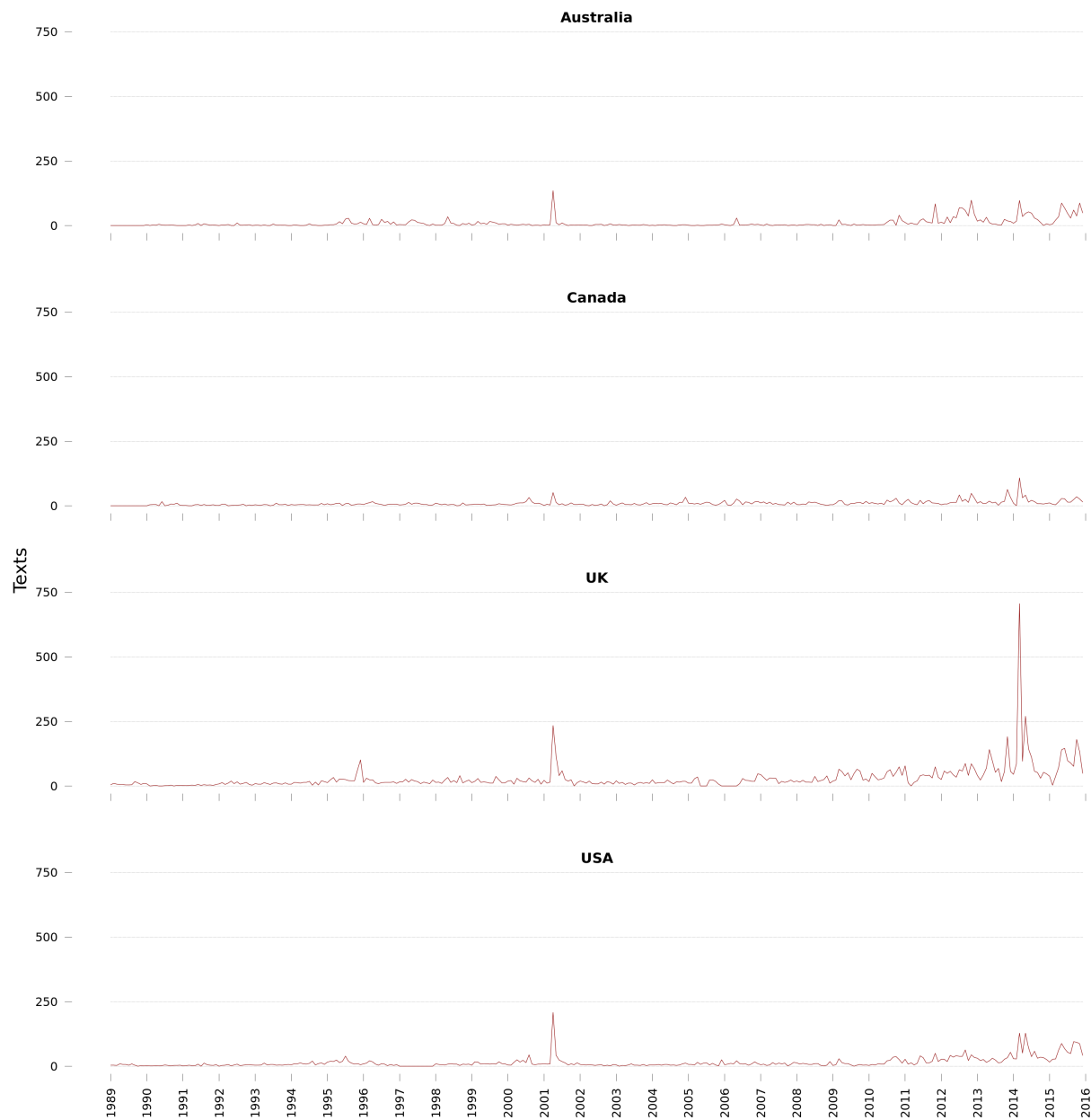
Figure 15: Documents in the SCS sub-corpus by publication year.

Conclusion

In this paper, we have introduced a corpus of 3.5 million English news texts on China from four English-speaking countries–Australia, Canada, UK, and USA–published between 1 January 1989 and 31 December 2015. Our main goal is to explore the corpus and identify document subsets that we

Figure 16: Texts in the "banks" sub-corpus by country and month of publication.

will investigate in depth. In this paper, five such sub-corpora were explored, each united by a theme: (1) human rights, (2) "Beijing Consensus", (3) Tibet, (4) South China Sea, (5) Asian Infrastructure Investment Bank and the New Development Bank. We have identified several patterns in cross-country and in-time variation and tentatively linked them with events and issues of substantive interest. Perhaps the most important of these patterns is the greater attention given to China by UK journalists, especially after 2000. Another interesting pattern is the sharp increase in coverage of China in 2008, related primarily to the Beijing Olympics as well as the crackdown in Tibet in the run-up to the Olympics. While Beijing had hoped that increased international attention to China during the 2008 Olympics would improve China's image abroad, these events show that soft power strategies can easily go sideways when unflattering features of the government are subjected to

24

increased media scrutiny.

In the present analysis, we have focused on patterns in the amount of coverage within our corpus and its five sub-corpora. In the next step, we will investigate the five sub-corpora, as well as additional smaller sub-corpora obtained the same way, to summarize patterns in the content of coverage. To do so, we will rely primarily on two techniques. First, we will use "topic" models to decompose the texts into substantively interpretable groupings ("topics"), and analyze the trends in the prevalences of these topics cross-country, cross-outlet, and in-time. Second, we will attempt to generate theme-specific dictionaries for "tonality"/"sentiment" of coverage of China using methods for supervised dictionary-building. Similarly to the topics, we will use the dictionaries to analyze the cross-country, cross-outlet, and in-time variation in the coverage of China.

These models will hopefully allow us to gauge different ways in which China is portrayed in different media markets. If we can then relate these patterns to variable public opinion of China across these countries, this will help us to arrive at a more complete picture of how China's soft power strategies have achieved variable success outside its borders.