Import Competition and U.S. Sentiment Toward China

Rabah Arezki, Ha Nguyen, Duong Trung Le, Hieu Nguyen

Centre d'Études
et de Recherches
sur le Développement
International

CERDI WORKING PAPERS
2024/2

Import Competition and U.S. Sentiment Toward China

Rabah Arezki
Duong Trung Le
Ha Nguyen
Hieu Nguyen

To cite this working paper:

Arezki, A., Trung Le, D., Nguyen, H., Nguyen, H. (2024) "Import Competition and U.S. Sentiment Toward China". CERDI Working Papers, 2024/2, CERDI.

CERDI, Pôle tertiaire, 26 avenue Léon Blum, 63000 Clermont-Ferrand, France.

The authors

Rabah Arezki
Professor; Director of Research;
Université Clermont Auvergne; CNRS; IRD;
CERDI F-63000 Clermont Ferrand; France
Senior Fellow; FERDI; Harvard Kennedy School.
Email: rarezki.econ@gmail.com

Duong Trung Le
Economist; World Bank
Email: dle6@worldbank.org

Ha Nguyen;
Economist; Institute of Capacity Development;
International Monetary Fund
Email: hnguyen7@imf.org

Hieu Nguyen;
PhD candidate; Washington University in St. Louis.
Email: h.d.nguyen@wustl.edu

Acknoledgments

We thank JaeBin Ahn; Xuefei Bai; Andy Berg; Olivier Blanchard; Karim Barhoumi; Mai Dao; Jeffrey Frankel; Mercedes Garcia-Escribano; Gregoire Rota-Graziosi; Tarek Masoud; Tilla McAntony; Flavien Moreau; Rafael Machado Parente; Rick van der Ploeg; Dmitry Plotnikov; David Guio Rodriguez; Tarik Youssef and seminar participants at the IMF's Western Hemisphere Department on October 25; 2023 for helpful comments. The views expressed in this research are of the author(s) and do not necessarily represent the views of the IMF and the World Bank; their Executive Boards; or their management.

This work was supported by the LABEX IDGM+ (ANR-10-LABX-0014) within the program "Investissements d'Avenir" operated by the French National Research Agency (ANR).

CERDI Working Papers are available online at: https://tinyurl.com/2xwfzw8s

Director of Publication: Simone Bertoli
Editor: Catherine Araujo Bonjean
Publisher: Marie Dussol
ISSN: 2114 - 7957

Disclaimer:

Abstract

We empirically examine how import competition affects sentiment toward China in local communities in the United States using a news-based index for sentiment. Results are threefold. First; U.S. sentiment toward China peaked in 2007 before turning negative. Second; communities more exposed to import competition from China have experienced a greater deterioration in sentiment. Third; the trade-induced U.S. sentiment toward China is broad-based; encompassing political; military; and national security issues. These findings suggest that competition over trade may have important geopolitical implications through sentiment of local communities.

Keywords

Import competition; sentiment; fragmentation

JEL Codes

E24; F14; F16; J23; J31; L60; O47; R12; R23.

1. Introduction

The U.S.-China relationship is among the most important relationship in the 21st century. It does not only affect the two countries but also the rest of the world. The relationship is affected by the large imbalance of the U.S. with China,[1] which has very sizable spillovers over the global economy[2], the impact of which has become a topic of discussion by the International Monetary Fund.[3]

To understand views and sentiment of the U.S. toward China, this paper aims to quantify U.S. sentiment toward China and to examine if the trade imbalance influenced the sentiment. To the best of our knowledge, sentiment of local communities in the U.S. toward China has not been systematically available nor analyzed. In addition, this paper explores if trade-induced shocks affect sentiment toward China in local communities in the U.S. The findings will help inform strategic and policy discussions between the U.S. and China.

To do so, we use change in China-originated import penetration to U.S. commuting zones and an index of local news sentiment. In line with the pioneering work by Autor *et al.* (2013 a,b, 2014), we exploit cross-market variation in import exposure from initial differences in industry specialization and instrumenting for U.S. imports using changes in Chinese imports by other high-income countries. The increase in import from China, developed by Autor *et al.* (2019), is used to examine how the "China shock" affects sentiment of local U.S. communities for the period from 2000 to 2020. The first half of that period saw the steepest increase in U.S. imports from China following China's accession to WTO. China's share of world manufacturing exports surged from 4.8 percent in 2000 to 15.1 percent in 2010. To track sentiment in local communities, we provide a novel news-based sentiment index using computational text analysis methods. The method extracts sentiment from a database of news articles collected from the top 137 U.S. news media sources (ranked by circulation), totaling more than 420,000 news articles, spanning more than four decades from 1979 to 2020.

Our analysis yields several results. First, U.S. sentiment toward China peaked in 2007 before gradually turning negative. Second, local U.S. communities more exposed to import competition from China have experienced greater deterioration in sentiment. Third, the trade-induced negative effect on U.S.

[1] Data from the U.S. Census show that the total annual U.S.-China trade volume has increased over 560 percent since China's accession to the World Trade Organization (WTO) in 2021. The trade volume has increased from $116 billion in 2000 to $657 billion in 2021. The value of total imports from the U.S. from China reached $506 billion in 2021. That is over three times the value of exports from the U.S. to China for the same year.

[2] Together, the U.S. and China account for almost 40 percent of global GDP.

[3] See Gourinchas (2022), Georgieva (2023), IMF (2023) for discussions about the impacts of a global fragmentation.

sentiment toward China is broad-based, encompassing political, military, and national security issues. Together, these findings suggest that competition over trade may have important geopolitical implications through the channel of sentiments of local communities.

This paper is related to a strand of literature that examines the causal link between foreign competition and large and persistent decline in U.S. employment. Local labor markets with more-exposed industries experience greater job losses, greater declines in tax revenues and house prices, larger rises in male idleness and premature mortality (Autor *et al.*, 2013a; Feler and Senses, 2017, Autor *et al.*, 2019). Industries experiencing greater exposure to trade with China have seen higher exit of plants (Bernard *et al.*, 2006), larger contractions in employment (Pierce and Schott, 2016; Acemoglu *et al.*, 2016), and lower income for affected workers (Galle *et al.*, 2023; Caliendo *et al.*, 2019). Our paper complements the literature by exploring the consequence of the China shock on U.S. sentiment toward China. It shows that trade imbalance could affect broad-based sentiment toward China. The mechanisms from import penetration to sentiment could operate via the aforementioned negative labor and social effects on exposed U.S. localities. In addition, exposure to import competition affects local political outcomes as well. Autor *et al.* (2020) find that trade-impacted commuting zones or districts saw stronger ideological polarization in campaign contributions and a relative rise in the likelihood of electing a Republican to Congress.

Our paper also contributes to a rapidly expanding literature quantifying attitudes and opinions using news media sources and examining the predictive power of sentiment on economic activity (see Gentzkow *et al.* (2019) for a review). An example of the use of dictionary-based sentiment analysis in finance is Tetlock (2007), which considers changes in sentiment extracted from the Wall Street Journal's "*Abreast of the Market*" column as a predictor of market fluctuations. Born *et al.* (2014) find that central bank optimism, derived from Financial Stability Report and governors' speeches, increases equity prices and reduces volatility. Baker *et al.* (2016) develop a measure of economic policy uncertainty from the frequency of articles published in 10 leading U.S. newspapers that contain relevant terminologies. Shapiro *et al.* (2022) show that sentiment shocks have a positive effect on consumption, output, and real interest rates.[4] Our paper goes to the community level and inspects sentiment driven by structural exposure to trade patterns. Instead of examining the impacts of sentiment, our paper investigates the impact of import penetration on sentiment.

Our paper is most closely related to Lu *et al.* (2018) which examine the effect of the surge in Chinese imports on U.S. media slant against China using 147 U.S. local newspapers for the period running from

[4] Further applications of the text-as-data approach includes FOMC communication (Lucca and Trebbi, 2009; Hansen *et al.*, 2018), flu nowcasting (Ginsberg *et al.*, 2009), corruption in US cities (Saiz and Simonsohn, 2013), or economic reform discussions (Arezki *et al.*, 2020).

1998 to 2012. Lu *et al.* (2018) find that newspapers in U.S. counties facing greater exposure to Chinese imports report more negative news about China. The result is consistent with our findings. However, our paper differs from Lu *et al.* (2018) in at least three important ways. First, our paper is the first to provide systematic evidence of U.S. sentiment toward China using 420,000 news articles spanning four decades. Second, our paper constructs a text-based sentiment index utilizing the full text of newspaper articles. That text-based sentiment index contrasts with the much cruder media slant measure in Lu *et al.* (2018), based on arithmetic proportion of China-related newspapers with headlines containing keywords deemed to project a negative image of China. Third, our paper explores systematically non-trade topics. We find significant spillover effect of trade-induced shocks that encompasses several critical domains of public sentiment.

The remainder of the paper is organized as follows. Section 2 presents the data. Section 3 describes the empirical strategy. Section 4 presents the main results. Section 5 discusses robustness checks. Section 6 concludes.

2. Data and variable formation

2.1 News-Based Sentiment

We adopt a text-as-data approach to analyze the content of over 420,000 news articles extracted from the Dow Jones FACTIVA database via the Dow Jones "Data, News and Analytics" Platform[5]. In addition to the date of publication and the articles' content, Dow Jones provides several classifiers to navigate article searches. These classifiers include language code, subject code, country/region code (where the article is originated from), country discussed by the article, industry code, and publisher information. For our analysis purposes, we restrict the article search and data to cover all articles that included a discussion about China (from the classifier "country discussed by the article")[6] and originated from one of the top 137 U.S. news publishers (in terms of circulation). We obtain 420,642 news articles in total, from 1979 until June 2020. Appendix Table A-1 shows a tabulation of the number of China-related articles originated from the top 30 U.S. publishers in our database.

[5] Dow Jones FACTIVA is a global repository of news covering over one billion articles published in many different languages. See https://www.dowjones.com/professional/factiva/. Radio, an important source that shapes sentiment in some segments of the U.S. population, is not included.

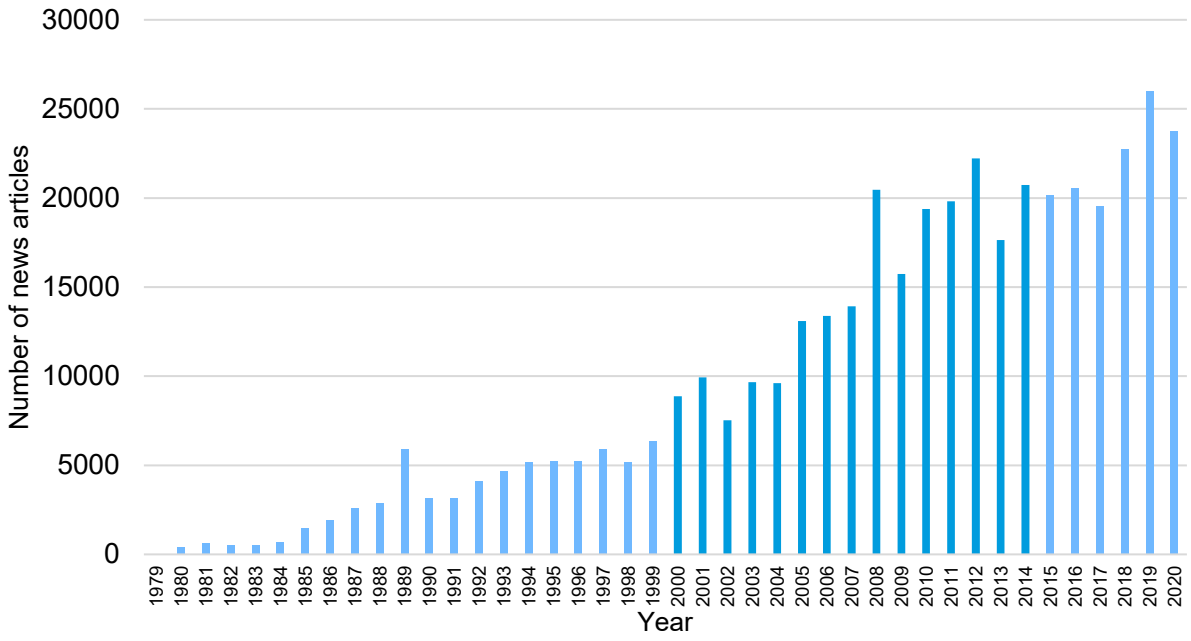
[6] Note that China alone or China and other countries can be discussed in one article. For example, if an article discusses the US-China political relationship, the United States and China are tagged as countries discussed by the article.

Figure 1 shows the distribution of news articles by year. The number of articles rises over time, reflecting the expansion of the news universe in FACTIVA and the rising importance of China in U.S. media. Our baseline analyses will focus on the period of 2000 to 2014. Our extension will consider the period of 1990 and 2000, albeit with fewer number of articles.

To examine the impact of import exposure across dimensions of public attitude, we use FACTIVA's classification in our main specification. All articles are classified by FACTIVA into one or more of a total of six broad groups, including *(i) Economics news*, *(ii) Commodity/Financial Markets News*, *(iii) Industry news*, *(iv) Politics/Social News*, *(v) General/Routine News*, and *(vi) Other News*.

Economic news contains macroeconomic news (for example, "*Trump victory freezes trade deals*"). *Commodity/ Financial Markets news* contains news about markets for trading in financial instruments or commodities *(e.g., "Europe, Asia Fall Further; Toronto Rises"). Industry news* contains all corporate and industrial news (e.g*., "CDC Games Hits Record Average Daily Revenue in China"*). *Politics/Social news* includes various topics: military and security, human rights and civic liberties, politics and international relations, arts and sciences, socioeconomic news and other social issues (e.g*., "Insurer Threatens to Sue Chinese Magazine Over Critical Article"*). *General/Routine news* consists of weather reports, natural disasters, sport news or coverages on personal and lifestyles (*e.g., "Earthquake Rocks China — Government Faces a Fresh Crisis as Thousands Are Killed"*). Articles that don't belong to any of these groups are marked as *Other news (e.g., "Business Education: On B-School Test, U.S. Can't Compete With Asia")*.

Figure 1: Number of news articles in the dataset.



Note: Data for 2020 are for the first six months.

As a robustness check, we employ a machine-learning method to classify news articles into semantic clusters. This approach—Vector Space Semantic methodology (Mikolov *et al.*, 2013)—generates word vector representations by mapping specific vocabulary items in high-dimensional space based on context probabilities (i.e., identifying words that tend to co-occur with a target word or term, and associated frequency). Appendix B describes the machine-learning-based classification approach in more details.

For each news article, we utilize a method popular in the literature to compute a value representing a sentiment index. The index is derived from taking the difference between the number of positive and negative words and scaling it by the sum of positive and negative words in the article.[7] By design, this index lies in the range $[-1, 1]$, where a higher value indicates more positive public sentiment toward China. A value of 0 indicates a neutral sentiment.[8]

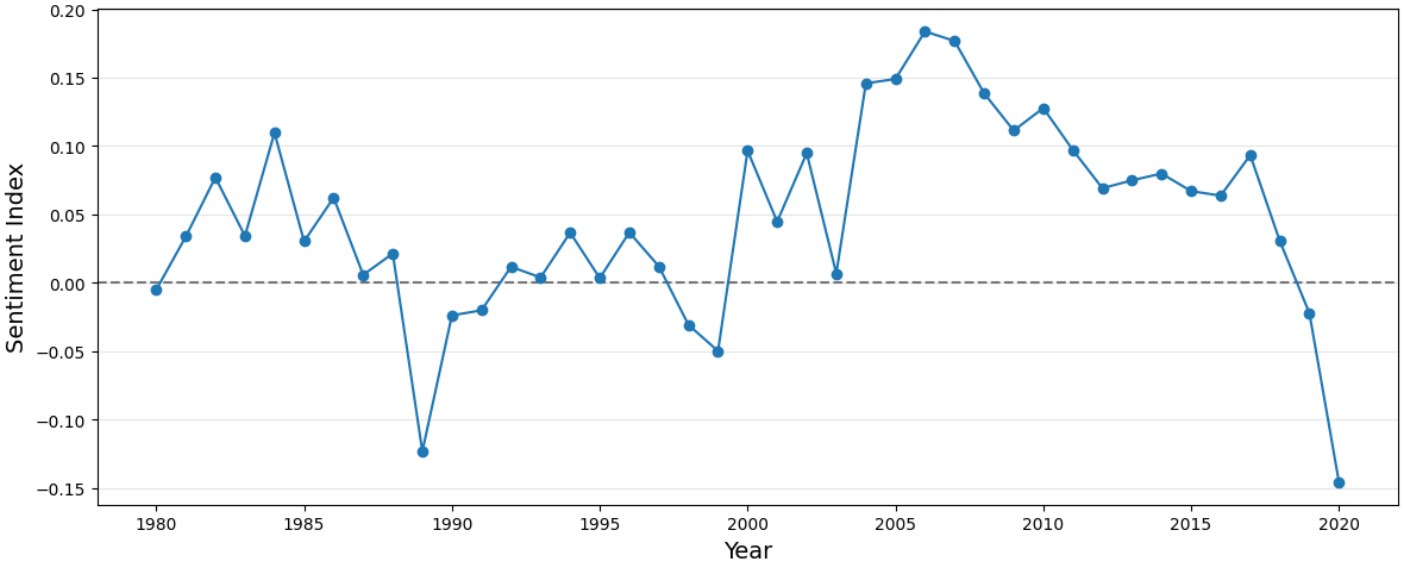
$$sentiment\ index = \frac{\#\ of\ positive\ words - \#\ of\ negative\ words}{\#\ of\ positive\ words + \#\ of\ negative\ words} \quad (1)$$

Figure 2 plots the average value of the index of sentiment computed from all U.S. news articles from all local and national sources referring to China between 1979 and June 2020. The U.S. sentiment toward China exhibits distinctive trends during three subperiods: China's pre-WTO period (1980–2001), China's WTO accession (2001–2017), and the contemporary period (2017–2020). Prior to China's accession to the WTO in 2001, the U.S. sentiment toward China was relatively neutral, with the annual average sentiment index fluctuating around zero for most years, except for a considerable dip in 1989 following the Tiananmen Square protests which led to the worsening of the bilateral relationship. China joining the WTO in 2001 was followed by over a decade-long growth surge and rising global importance. This period during which U.S.-China annual trade volume increased over three-fold coincided with positiveness in the U.S. media toward China. The sentiment peaked in 2007 and declined since then. The decline accelerated in 2017, coinciding with the start of the U.S.-China trade tensions. Continuous sharp drops brought the average annual U.S. sentiment index toward China to the negative territory for the first time since the late 1990s in 2019, followed by the decline to its lowest value during the entire sample recorded in the first half of 2020.

[7] This text-to-data method of deriving sentiment index is widely used in the literature. For examples, see Tetlock (2007) and Shapiro *et al.* (2022).

[8] We use the Python package polyglot (which builds on the work of Chen and Skiena (2014) to compute the sentiment value.

Figure 2: Aggregate sentiment index computed from U.S. news articles discussing China.



Note: The line plots the annual-average sentiment index value computed from all U.S. news articles in our sample (from both local and national sources). Data for 2020 is for the first six months.

China's import penetration had been shown to have heterogeneous effects across U.S. localities (Autor *et al.*, 2013a,b, 2014). To capture the spatial variation in sentiment across U.S. localities, we opt for the calculation of public sentiment index using just the content of news articles published by local news sources and omit articles from well-known national news sources. National sources include ABC News, CBS News, CNN, Fox News, MSNBC, NBC News, NY Times, and Wall Street Journal. These news sources cater to national audiences and are not necessarily representative of sentiment in individual local communities. Obviously, we do not refute the importance of these national news sources in shaping public sentiment (the aggregate sentiment index in Figure 2 includes them). They are removed in our regression analyses because we focus on local sentiment which is reflected by local news sources. After removing national news sources, we are left with 338,894 articles from 156 local new sources.[9]

Media-based sentiment, while innovative and increasingly used, has some caveats worth discussing. First, local news media can sometimes amplify, influence, or shape local public sentiment. While this is true, we argue that local news media cannot be too divergent from local public sentiment to maintain readership. Therefore, by and large, local media should be in sync with local sentiment. Second, there could be correlations among news sources across localities. For example, multiple news sources are owned by a common holding company at the national level and possibly at the local level. We still find

[9] The 156 local news sources in our data belong to 119 news publishers. Several publishers own multiple local news publishers. For example, MediaNews Group Inc. is a publisher headquartered in Denver, Colorado. Their publications belong to a total of 8 local news sources, including the Boston Herald headquartered in Massachusetts, The Orange County Register headquartered in California, and the Denver Post headquartered in Colorado. For our analysis, we strictly consider the headquarters of local news sources when identifying news exposure to a U.S. local community.

significant impact of import competition on the sentiment of exposed commuting zones, despite potential correlations of sentiment across commuting zones.

2.2 Trade exposure in U.S. localities

Commuting Zones

To examine the impact of trade exposure on public sentiment across U.S. localities, we follow the existing literature in defining geographic entities referred to as "commuting zones" (Tolbert *et al.*, 1996; Autor *et al.*, 2013a,b, 2014). Unlike the official administrative divisions (e.g., counties, cities, or towns), a commuting zone groups counties characterized by strong commuting patterns together with a total of 722 zones covering the entire mainland U.S. The use of commuting zones as appropriate units of analysis to study the impact of trade shocks on local U.S. labor market has been discussed in earlier studies. By design, a commuting zone better captures the overlap between employment and residential locations of an average individual, even if these locations might belong to different counties. Using commuting zones also helps circumvent issues of selective migration, for instance, migration from cities to suburbs. An analysis at the commuting-zone level, therefore, balances out the need to have a confined geographic unit that collectively captures the impact of trade shocks on labor market, public services, and local sentiment, while still maintaining appropriate spatial granularity.

Measure of exposure to China imports

Following Autor *et al.* (2019), we use change in import penetration from China to the US, $\Delta IP_{i\tau}^{CU}$, in commuting zones as a source of variation, where i is for commuting zone and τ is for the period.

$$\Delta IP_{i\tau}^{CU} = \sum_{j} \frac{L_{ji90}}{L_{i90}} \Delta IP_{j\tau}^{CU} \quad (2)$$

Equation (2) indicates that the change in import penetration from China to the U.S. in a commuting zone i is the weighted average of the change in import penetration of all industries j, $\Delta IP_{j\tau}^{CU}$, present in the commuting zone during period τ. The weight is the initial employment share of industry j in commuting zone i, $\frac{L_{ji90}}{L_{i90}}$. L_{ji90} is the initial (i.e., 1990) employment in commuting zone i in sector j, L_{i90} is the initial total employment in commuting zone i. Our paper uses the exposure from Chinese imports from Autor *et al.* (2019) for the periods between 2000 and 2014 (in the baseline regression), and between 1990 and 2010 (in the extension).

The change in import penetration for each industry j is calculated as the change in the value of U.S. imports from China in industry j during period τ, divided by initial U.S. absorption (which is industry shipment in the U.S. in 1991, Y_{j91}, plus net imports, $M_{j91} - X_{j91}$)

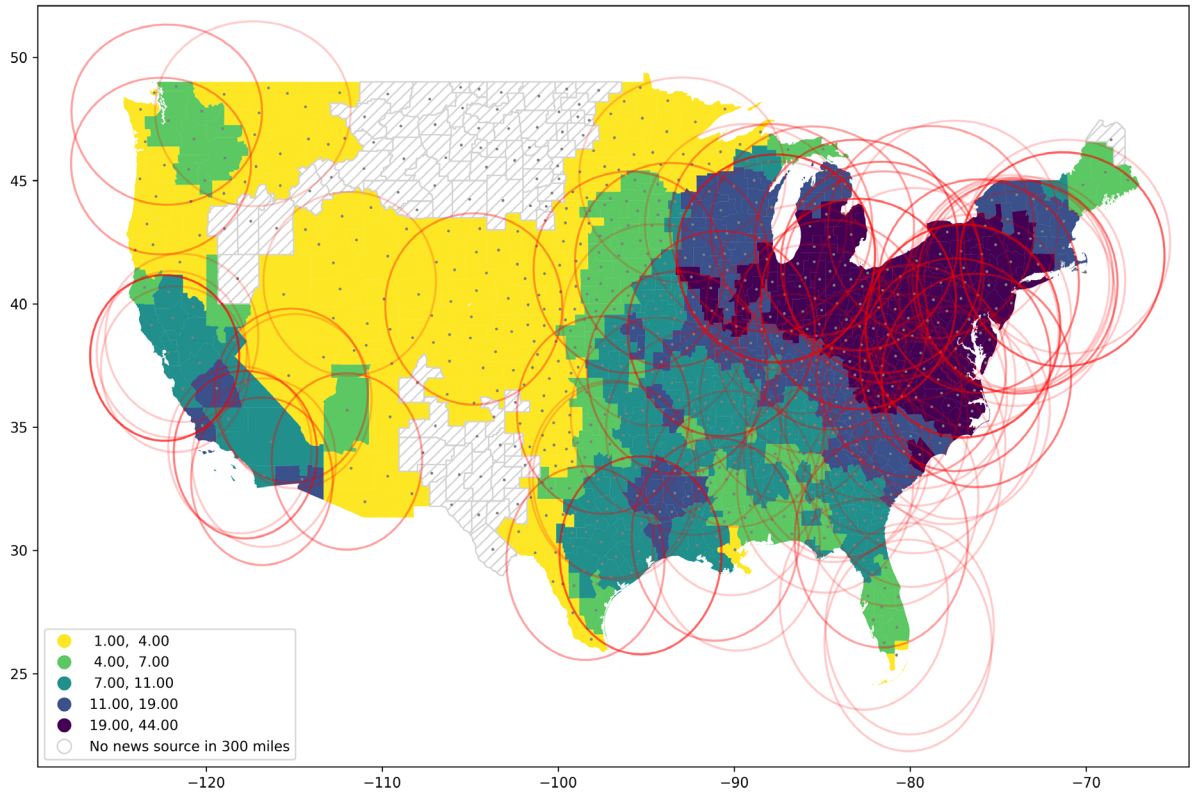
$$\Delta IP_{j\tau}^{CU} = \Delta M_{j\tau}^{CU} / (Y_{j91} + M_{j91} - X_{j91}) \quad (3)$$

2.3 Mapping news sources to commuting zones

We measure U.S. commuting zones' average sentiment by mapping locations of news sources to each commuting zone. To do so, we first calculate the coordinates of the local headquarters of each news source in our data (156 local news sources) and compute the geographical center of each commuting zone (722 zones). For each new sources, we generate a 300-mile spherical buffer and match the locations of all commuting zone centers that fall within this radius. The 300-mile radius is a cutoff to approximate feasible distance for daily distribution of articles' physical prints from the news source to the commuting zone (i.e., about 4 hours' drive). In the subsequent analysis, we provide robustness checks to the main result with varying the buffer radius. If multiple news sources are located within the 300-mile radius of a commuting zone's center, the commuting zone's sentiment is the weighted average of sentiment by each news source. The weight is based on the physical distance between the news source's location and the commuting zone's center to assign the importance of each source toward the computation of the commuting zone's sentiment:

$$\omega(news\ source_s, czone_i) = \frac{1/Distance\ (news\ source_s, czone_i)}{\sum_k 1/Distance\ (news\ source_k, czone_i)} \quad (3)$$

Figure 3: The distribution of news sources and the number of sources mapped to each commuting zone.



Note. Each red circle denotes a radius of 300 miles from a news source (156 in total). Each gray dot corresponds to a commuting zone's center. Color scale represents quantiles of commuting zones based on the number of corresponding sources.

Overall, 560 out of the total 722 U.S. commuting zones are matched with at least a news source falling within their 300-mile buffer, and thus is assigned a non-missing sentiment index value. Figure 3 shows the geographic distribution of the local news sources in our database and colors commuting zones by the number of local news sources by which they are influenced. Darker-shaded areas represent local communities influenced by more news sources. It is noted that our news database might not cover the universe of available news sources in the U.S., and in fact, might lack coverage of small local publishers in relatively remote areas (i.e., small publishers not belonging to the list of 137 publishers that we collected news articles from FACTIVA). Therefore, several commuting zones—mostly low-population localities—do not geographically locate inside the buffer radius of any news sources and are omitted from our analysis. To the extent that small local news sources do not often publish articles discussing China-related content[10], such omission arguably does not systematically affect our estimates.

3. Estimation Strategy

We estimate the effect of trade exposure on different dimensions of U.S. sentiment of local communities toward China as follows:

$$\Delta Y_{ig\tau} = \beta_0 + \beta_1 \Delta IP_{i\tau}^{CU} + X_i' \beta_2 + \alpha_g + \varepsilon_{ig\tau} \quad (4)$$

where $\Delta Y_{ig\tau}$ is the change in the average news-based sentiment index in commuting zone i in topic group g during time interval τ. Our focus is on the period between 2000 and 2014, following China's accession to WTO and fast rising import to the U.S. originating from China. The main explanatory variable is the corresponding import penetration from China to the U.S. in a commuting zone, $\Delta IP_{i\tau}^{CU}$, instrumented by China's exports to other advanced countries, $\Delta IP_{i\tau}^{CO}$. X_i' denotes locality controls capturing start-of-period demographic characteristics and labor force composition. Also, we control for semantic group dummies α_g for news articles. The standard errors $\varepsilon_{ig\tau}$ are clustered at the state level.

One concern in identifying a causal relationship between U.S. imports of Chinese goods and local sentiment across U.S. commuting zones is the potential existence of confounding factors. These factors, such as the (spatially) heterogeneous changes in U.S. income, could drive both sentiment and changes in U.S. imports. To isolate the exogenous variation in U.S. imports from China, we use the change in other high-income country imports from China, $\Delta IP_{i\tau}^{CO}$, as an instrument for U.S. changes in imports:

$$\Delta IP_{i\tau}^{CO} = \sum_j \frac{L_{ji80}}{L_{i80}} \Delta IP_{j\tau}^{CO} \quad (5)$$

[10] The bottom 10 news publishers in our 137-publisher database only record on average 2.5 articles with a China-related topic during the entire 1979–2020 period.

One element of the instrument in Equation (5) is the use of lagged employment levels by industry and region from prior decades, which helps mitigate potential contemporaneous adjustment in employment in response to Chinese trade. Another element of the instrument is the use of other high income countries' imports of Chinese goods, as opposed to U.S. imports of these goods, to account for the possibility that U.S. demand factors were simultaneously driving both the increase in Chinese imports and the changes in U.S. sentiment (the variable of interest). Our identification strategy relies on the assumption that the variation in $\Delta IP_{j\tau}^{CO}$ is driven by the increase in China's comparative export advantage in specific sectors and not by changes in U.S. demand.[11]

Table 1 presents summary statistics of the change in import penetration (from Autor *et al.*, 2019) across commuting zones with sentiment data. For the period of 2000 to 2014, a typical (median) commuting zone has the change in import penetration equals 0.998. To interpret this value, assuming workers in a commuting zone is entirely employed in an industry X, the change in import penetration of 0.998 implies the change in import values of industry X between 2000 and 2014 almost equals initial U.S. absorption.

Table 1: Summary statistics of change in import penetration

$\Delta IP_{i\tau}^{CU}$	N	Mean	Median	SD
$\tau = 2000 - 2014$	560	1.268	0.998	1.101
$\tau = 1990 - 2000$	431	1.003	0.7853	1.018

Note: Data are from Autor *et al.* (2019).

4. Main results

4.1 Baseline results

Table 2 reports the baseline results using our estimation strategy including the instrument variable described in Equation (5). Overall, we find a significant negative causal effect of China-import penetration on U.S. sentiment during the period going from 2000 to 2014. The estimated coefficients suggest over 0.03 unit decline in sentiment toward China in commuting zones experiencing one-unit larger trade exposure, which is about the median value of the change import penetration across

[11] Autor *et al.* (2013a) discuss some possible threats to identification: correlated product demand shocks in both the U.S. and other high-income countries, an increase in high-income country imports from China due to a negative productivity shock in the U.S., and common technological developments in the U.S. and other high-income countries that could drive the increase in imports from China. The authors dismissed the first threat to identification by estimating a modified gravity model designed to isolate changes in China's exports driven solely by supply and trade shocks. It is also arguable that forces internal to China are likely the dominant driver of Chinese export surge, which greatly outpaced that of other low- and middle-income nations during this period.

commuting zones between 2000 and 2014. Note that the aggregate sentiment ranges from about -0.15 to 0.2. Therefore, a 0.03-unit relative decline in sentiment represents a substantial impact.

Table 2: Impact of China-import competition on U.S. public sentiment (Baseline)

IV regression	(1)	(2)	(3)
	Δ Average sentiment (2000-2014)		
Δ Import penetration	-0.0309**	-0.0300**	-0.0324**
(2000-2014)	(0.0144)	(0.0133)	(0.0139)
Constant	YES	YES	YES
Observations	2,521	2,521	2,521
R-squared	0.052	0.059	0.061
Topic group FE	YES	YES	YES
Occupational controls	YES	YES	YES
Race controls	NO	YES	YES
Demographic controls	NO	NO	YES
Number of commuting zones	560	560	560

Note: Observation for 6 news-topic groups across 560 commuting zones. Dependent variable: change in average sentiment index for each group between 2000 and 2014. News sources are mapped to commuting zones based on physical proximity and volume of published articles. Occupational controls include start-of-period indices of employment in routine occupations, employment share of manufacturing, and of employment in offshorable occupations. Race controls are dummies representing start-of-period shares of commuting zone population that is Hispanic, Black, Asian, and other races. Demographic controls include start-of-period shares of foreign-born population, college educated, and female employment. The list of control variables follows Autor *et al.* (2019). Robust standard errors in parentheses clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Heterogeneity by topics

Having shown an aggregate negative effect of Chinese imports pressure on sentiment, we now explore the granularity of sentiment. A natural question is whether import penetration has a heterogeneous effect across different themes of the U.S.-China relationship. We rely on FACTIVA's classification of news articles by themes and generate 6 groups including *Economic news, Commodity/financial-markets news, Industry news, Political/social news, General/routine news*, and *Other news*. The heterogeneity in trade-induced effect on sentiment can be estimated with an interaction term between each of the themes and the Chinese imports indicator:

$$\Delta Y_{ig\tau} = \beta_0 + \beta_1 \Delta IP_{i\tau}^{CU} + X'_i \beta_2 + \beta_3 \Delta IP_{i\tau}^{CU} \times \alpha_g + \alpha_g + \varepsilon_{ig\tau} \quad (7)$$

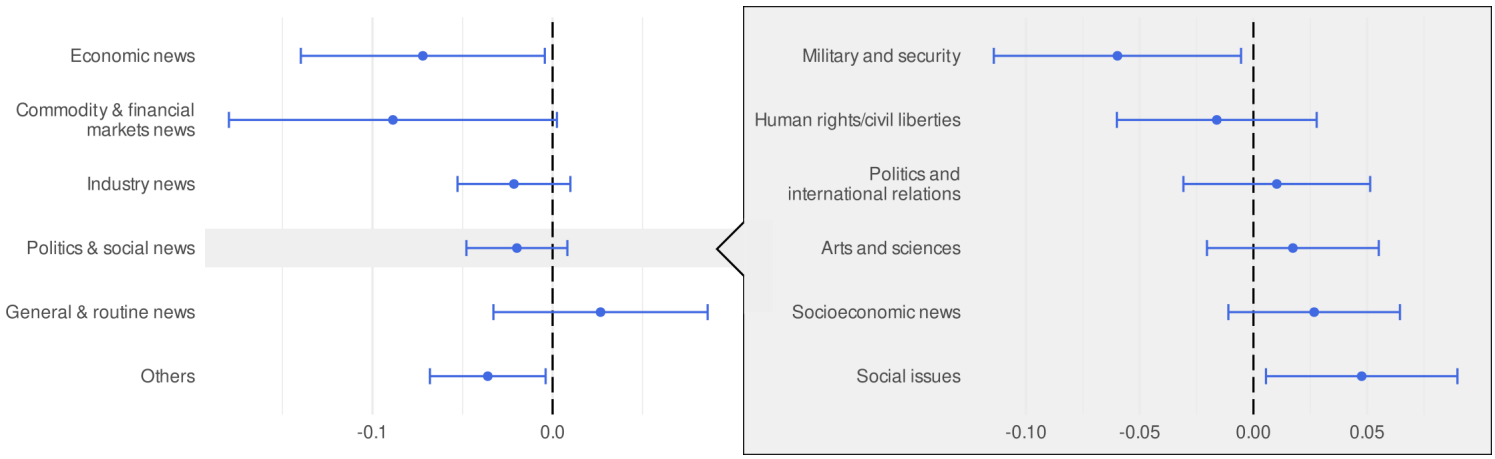
where the term α_g is an indicator for each thematic group. β_3 is the estimated coefficient corresponding to the interaction term $\Delta IP_{i\tau}^{CU} \times \alpha_g$ and illustrates the impact of Chinese imports on local sentiment about China across different topic groups.

Figure 4 plots the point estimates and 90-percent confidence intervals of the effect of Chinese imports on U.S. sentiment toward China for each of the six semantic themes ($\beta_1 + \beta_3$). By design, a negative estimate associated with a topic suggests that higher import competition from China reduces local U.S. sentiment toward China based on news articles categorized into that thematic news group. The estimates

in Panel A suggest that exposure to Chinese imports drives local U.S. sentiment about China downward on all thematic groups except for *General and routine news* (which discusses weather, natural disasters, sports, fashion...). The largest and most statistically significant effects are observed for articles grouped into *Economic news* and *Commodity & financial-market news*.

It is not surprising to observe a strong and negative effect on U.S. sentiment expressed through news articles on topics directly related to themes such as economics, financial markets, and industry. However, the fact that the marginal trade-induced effect is also negative for themes such as political & social news topics suggests that pressure from import competition could also have important implication across "non-economic" dimensions of sentiment.

Figure 4: Heterogeneous effects of China-import penetration across themes



Note: Panel A shows the 90% CI of the impact of import competition from China on the change of local sentiment. Panel B shows similar statistics but only for the sub-topics in political/social news. The dependent variable is the change in sentiment between 2000 and 2014. News sources are mapped to commuting zones based on physical proximity and volume of published articles. Occupational controls include start-of-period indices of employment in routine occupations, employment share of manufacturing, and of employment in offshorable occupations. Race controls are dummies representing start-of-period shares of commuting zone population that is Hispanic, Black, Asian, and other races. Demographic controls include start-of-period shares of foreign-born population, college educated, and female employment. The list of control variables follows Autor *et al.* (2019). Robust standard errors in parentheses clustered by state.

To further investigate the potential spillover effect on sentiment expressed in political and social news, we examine a more granular set of topics within these themes and estimate the effect of China trade-induced shock on sentiment. Among the available sub-topics, we find negative estimates for articles in *Military & national security*, and—albeit insignificant at conventional confidence intervals—*Human rights & civil liberties*. The negative trade-induced effect on sentiment expressed in articles discussing national security and human rights suggest trade-induced shocks could influence the U.S. awareness and attitude toward military and national security issues.

4.3 Persistent effects of import exposure on news sentiment

How persistent is the effect on sentiment? We examine this question by utilizing the news-based sentiment index calculated from news articles published in the years after U.S.-China trade competition bottomed off in the early 2010s. To analyze the effect of sentiment on an annual basis, we derive the outcome variable as the difference between the average sentiment for each year from 2014 to 2019 and the baseline sentiment level in 2000. The main explanatory variable is still the change in U.S. imports of Chinese goods between 2000 and 2014, from Autor *et al.* (2019).

The estimation reported in Table 3 suggests that trade competition continued to impact sentiment in the communities exposed to China imports in all the years after the 2000–2014 period. The negative impact on attitude reflected in U.S. news articles that featured China related issues was particularly strong during 2015–2017, and again in 2020.

Table 3: Persistent effect China imports penetration on U.S. public sentiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Δ Average sentiment (since 2000)						
VARIABLES	2014	2015	2016	2017	2018	2019	2020
Δ Import penetration (2000-2014)	-0.0324**	-0.0254*	-0.0241*	-0.0304**	-0.0174	-0.0143	-0.0325**
	(0.0139)	(0.0147)	(0.0139)	(0.0122)	(0.0134)	(0.0149)	(0.0151)
Constant	YES	YES	YES	YES	YES	YES	YES
Observations	2,521	2,682	2,617	2,475	2,317	2,504	2,598
R-squared	0.061	0.289	0.224	0.132	0.156	0.231	0.160
Topic group FE	YES	YES	YES	YES	YES	YES	YES
Occupational controls	YES	YES	YES	YES	YES	YES	YES
Race controls	YES	YES	YES	YES	YES	YES	YES
Demographic controls	YES	YES	YES	YES	YES	YES	YES
Number of commuting zones	560	560	560	526	560	528	560

Note: Observation for 6 topics groups across 560 commuting zones. Dependent variable: average sentiment indices for each semantic group in 2014, 2015, 2016, 2017, 2018, 2019 and 2020, all subtracted by average sentiment indices for each semantic group in 2000. News sources are mapped to commuting zones based on physical proximity and volume of published articles. Occupational controls include start-of-period indices of employment in routine occupations, employment share of manufacturing, and of employment in offshorable occupations. Race controls are dummies representing start-of-period shares of commuting zone population that is Hispanic, Black, Asian, and other races. Demographic controls include start-of-period shares of foreign-born population, college educated, and female employment. The list of control variables follows Autor *et al.* (2019). Robust standard errors in parentheses clustered at the state level. *** p<0.01, ** p<0.05, * p<0.1.

5. Extension and robustness checks

5.1 Results for the period from 1990–2000

In this extension, we estimate the impact of China-import penetration on U.S. public sentiment using changes in import competition in the decade between 1990 and 2000. This decade experienced the commencing of the China shock in 1992, when China expanded its "reform and opening" agenda to foster foreign trade and investment, and before the accession of China to the WTO in 2001.

Results on the effect of trade exposure on U.S. sentiment during 1990–2000 remain significant and negative. The regression coefficients, shown in Appendix Table A-2, suggest that the size of the effect could be larger than that of our baseline results shown in Table 2–covering the 2000–2014 period. This finding suggests commuting zones more exposed to rising imports from China saw a larger decline (or a smaller rise) in sentiment toward China during 1990-2000. Note that compared with our baseline regressions for the 2000–2014 period, our dataset contains fewer news articles between 1990–2000 and thus covers fewer commuting zones and topic groups. Between 1990–2000, our data contain 57,164 articles and cover 431 commuting zones. In comparison, our data for the 2000–2014 period contain 221,980 articles, covering 560 commuting zones.

5.2 Varying physical proximity between news sources and commuting zones

Next, we check the sensitivity of the causal relationship to our method of measuring the degree of news influence in each commuting zone. A concern is that 300 miles could be too far for a news source's influence. We re-estimate the effect using two alternative commuting zone's proximity buffer radius taking values of 100 and 200 miles (compared to the original 300-mile buffer). Doing this has pros and cons. On the upside, commuting zones' sentiment can be under stronger and cleaner influence of news sources. On the downside, the sample size shrinks.[12] Under more restricted sample sizes, the estimated results in Appendix Table A-3 show that the trade-induced effect on sentiment is stronger and statistically significant across all specifications, although the number of commuting zones with sentiment data is much smaller.

[12] There are 384 and 150 CZs matched with at least one nearby local news sources within 200-mile and 100-mile buffer radius, respectively. In comparison, there are 560 CZs matched under the baseline 300-mile radius buffer.

5.3 Machine-learning topic selection

Lastly, we check for the sensitivity of our estimates to methods in selecting thematic groupings of news articles. Our baseline specification uses the manual topic tags provided by FACTIVA to group news articles into six distinct themes. In this robustness check, we utilize a machine-learning method to systematically classify articles into topical groups instead (see Appendix B for details). One input parameter of the method is the number of semantic clusters. We allow the value of this input parameter to take three different values of six, eight, and ten topic groups. Appendix Table A-4 shows that estimations using these different number of thematic groups yields similar results to those obtained using our baseline specification.

6. Conclusion

This paper examines how import competition affect sentiment toward China in local communities in the United States using a news-based index for sentiment. We find that commuting zones which bear the brunt of the trade shock exhibit significantly more negative shifts in sentiment. This adverse effect is not confined to economic and financial news but permeates a range of topics such as military/security, and human rights. The negative effect on news-based sentiment persists over time and is robust to various checks. Our findings suggest that competition over trade may have important geopolitical implications through sentiment of local communities.

7. References

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price, "Import Competition and the Great US Employment Sag of the 2000s," *Journal of Labor Economics*, January 2016, *34* (S1), S141–S198.

Arezki, Rabah, Simeon Djankov, Ha Nguyen, and Ivan Yotzov, "Reform Chatter and Democracy," *World Bank Policy Research Working Paper 9319*, 2020.

Autor, David, David Dorn, and Gordon Hanson, "When Work Disappears: Manufacturing Decline and the Falling Marriage Market Value of Young Men," *American Economic Review: Insights*, September 2019, *1* (2), 161–178.

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013a, *103* (6), 2121–2168.

Autor, David H., David Dorn, and Gordon H. Hanson "The Geography of Trade and Technology Shocks in the United States," *American Economic Review*, 2013b, *103* (3), 220–225.

Autor, David H., David Dorn, Gordon H. Hanson and Jae Song, "Trade Adjustment: Worker-Level Evidence *," *The Quarterly Journal of Economics*, November 2014, *129* (4), 1799–1860.

Autor, David, David Dorn, Gordon Hanson, and Kaveh Majlesi "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." American Economic Review. 2020, 110(10): 3139-83.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, "Measuring Economic Policy Uncertainty*," *The Quarterly Journal of Economics*, November 2016, *131* (4), 1593–1636.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott, "Survival of the Best Fit: Exposure to Low-Wage Countries and the (Uneven) Growth of U.S. Manufacturing Plants," *Journal of International Economics*, January 2006, *68* (1), 219–237.

Born, Benjamin, Michael Ehrmann, and Marcel Fratzscher, "Central Bank Communication on Financial Stability," *The Economic Journal*, 2014, *124* (577), 701–734.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Chen, Yanqing and Steven Skiena, "Building Sentiment Lexicons for All Major Languages," in "Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2:

Short Papers)" Association for Computational Linguistics Baltimore, Maryland June 2014, pp. 383–389.

Feler, Leo and Mine Z. Senses, "Trade Shocks and the Provision of Local Public Goods," *American Economic Journal: Economic Policy*, November 2017, *9* (4), 101–143.

Galle, Simon, Andr´es Rodr´ıguez-Clare, and Moises Yi, "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade," *The Review of Economic Studies*, January 2023, *90* (1), 331–375.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, September 2019, *57* (3), 535–574.

Georgieva, Kristalina, "The Price of Fragmentation Why the Global Economy Isn't Ready for the Shocks Ahead", Foreign Affairs, 2023.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, February 2009, *457* 14 (7232), 1012–1014.

Gourinchas, Pierre-Olivier, "A more fragmented world will need the IMF more, not less", Finance and Development, June 2022.

International Monetary Fund, "World Economic Outlook, October 2023: Navigating Global Divergences", 2023.

Hansen, Stephen, Michael McMahon, and Andrea Prat, "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*," *The Quarterly Journal of Economics*, May 2018, *133* (2), 801–870.

Lu, Yi, Xiang Shao, and Zhigang Tao. "Exposure to Chinese imports and media slant: Evidence from 147 U.S. local newspapers over 1998–2012," *Journal of International Economics*. 2018, (114) 316–330.

Lucca, David and Francesco Trebbi, "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements," w15367, National Bureau of Economic Research, Cambridge, MA September 2009.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient Estimation of Word Representations in Vector Space," in "International Conference on Learning Representations" arXiv 2013.

Pierce, Justin R. and Peter K. Schott, "The Surprisingly Swift Decline of US Manufacturing Employment," *American Economic Review*, July 2016, *106* (7), 1632–1662.

Rehurek, Radim and Petr Sojka, "Gensim–Python Framework for Vector Space Modelling," *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 2011, *3* (2).

Saiz, Albert and Uri Simonsohn, "Proxying For Unobservable Variables With Internet Document-Frequency," *Journal of the European Economic Association*, February 2013, *11* (1), 137–165.

Shapiro, Adam Hale, Moritz Sudhof, and Daniel J. Wilson, "Measuring News Sentiment," *Journal of Econometrics*, June 2022, *228* (2), 221–243.

Tetlock, Paul C., "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *The Journal of Finance*, June 2007, *62* (3), 1139–1168.

Tolbert, Charles M., Molly Sizer, Charles M. Tolbert, and Molly Sizer, "U.S. Commuting Zones and Labor Market Areas: A 1990 Update," Economic Research Service Staff Paper 9614, United States Department of Agriculture 1996.

8. Appendix

Appendix A: Tables

Appendix Table A1: News articles about China by top U.S publishers in FACTIVA database, 1970-2020

Publisher name	Number of articles about China
Dow Jones & Company, Inc.	85178
Business Wire	32953
The New York Times Company	29657
Business Wire, Inc.	24731
Hearst Communications, Inc.	23211
Washington Post	21047
New York Times Digital (Full Text)	16582
Dow Jones & Company	13395
Seattle Times	12509
Investor's Business Daily	11892
News World Communications, Inc.	10409
Gannett Co., Inc. - Newspaper Division	7191
St. Louis Post-Dispatch	6123
CQ-Roll Call, Inc.	6067
National Public Radio, Inc.	5450
NewsBank Inc.	4914
MediaNews Group Inc.	4628
Knight Ridder Digital	4374
Journal Communications Inc.	3034
Media General, Inc.	3007
Grand Rapids Press	2597
Hollinger	2410
Deseret News Publishing Co.	2323
N.Y.P. Holdings, Inc.	2308
Advance Publications, Inc.	2184
Boston Herald Library	2105
Hearst Newspapers	1980
Gannett Co., Inc. / Newspaper Division	1929
Union-Tribune Publishing Company	1873

Appendix Table A-2: Impact of Import Competition from China on U.S. Sentiment (1990-2000)

IV regression	(1)	(2)	(3)
	Δ Average sentiment (1990-2000)		
Δ Import penetration	-0.0499*	-0.0485**	-0.0511***
(1990-2000)	(0.0259)	(0.0204)	(0.0196)
Constant	YES	YES	YES
Observations	1,271	1,271	1,271
R-squared	0.151	0.158	0.160
Topic group FE	YES	YES	YES
Occupational controls	YES	YES	YES
Race controls	NO	YES	YES
Demographic controls	NO	NO	YES
Number of commuting zones	431	431	431

Note: Observation for 6 topics groups across 431 commuting zones. Dependent variable: change in average sentiment index for each group between 1990 and 2000. News sources are mapped to commuting zones based on physical proximity and volume of published articles. Occupational controls include start-of-period indices of employment in routine occupations, employment share of manufacturing, and of employment in offshorable occupations. Race controls are dummies representing start-of-period shares of commuting zone population that is Hispanic, Black, Asian, and other races. Demographic controls include start-of-period shares of foreign-born population, college educated, and female employment. The list of control variables follows Autor *et al.* (2019). Robust standard errors in parentheses clustered by state. *** p < 0.01, ** p < 0.05, * p < 0.1.

Appendix Table A-3: Robustness check with varying buffer zones

IV regression	(1)	(2)	(3)	(4)	(5)	(6)
	Δ Average sentiment (2000-2014)					
Δ Import penetration	-0.109***	-0.101**	-0.0965**	-0.0738***	-0.0674***	-0.0797***
(2000-2014)	(0.0441)	(0.0424)	(0.0414)	(0.0235)	(0.0217)	(0.0225)
Constant	YES	YES	YES	YES	YES	YES
Observations	634	634	634	1,608	1,608	1,608
R-squared	0.074	0.097	0.112	0.051	0.069	0.071
Proximity buffer radius	100-mile	100-mile	100-mile	200-mile	200-mile	200-mile
Topic group FE	YES	YES	YES	YES	YES	YES
Occupational controls	YES	YES	YES	YES	YES	YES
Race controls	NO	YES	YES	NO	YES	YES
Demographic controls	NO	NO	YES	NO	NO	YES
Number of commuting zones	150	150	150	384	384	384

Note*:* Observation for 6 topic groups across 150 commuting zones for model (1–3) and 384 commuting zones for model (4–6). Dependent variable: change in average sentiment index for each group between 2000 and 2014. News sources are mapped to commuting zones based on physical proximity and volume of published articles. Occupational controls include start-of-period indices of employment in routine occupations, employment share of manufacturing, and of employment in offshorable occupations. Race controls are dummies representing start-of-period shares of commuting zone population that is Hispanic, Black, Asian, and other races. Demographic controls include start-of-period shares of foreign-born population, college educated, and female employment. The list of control variables follows Autor *et al.* (2019). Robust standard errors in parentheses clustered at the state level. *** p < 0.01, ** p < 0.05, * p < 0.1.

Appendix Table A-4: Robustness check with machine-learning topics selections

	Δ Sentiment (2000-2014)		
	6 groups	8 groups	10 groups
	(1)	(2)	(3)
Δ Import penetration	-0.0343***	-0.0326**	-0.0214*
	(0.0118)	(0.0131)	(0.0117)
Constant	YES	YES	YES
Observations	2,768	3,507	4,445
R-squared	0.156	0.112	0.102
Controls	YES	YES	YES
Topic FE	YES	YES	YES
Number of CZs	560	560	560

Note: Observation for 6,8, and 10 semantic groups across 560 CZs (in 2000-2014). Dependent variable: change in average sentiment index for each semantic group between 2000-2014. News sources are mapped to commuting zones based on physical proximity and volume of published articles. Occupational controls include start-of-period indices of employment in routine occupations, employment share of manufacturing, and of employment in offshorable occupations. Race controls are dummies representing start-of-period shares of commuting zone population that is Hispanic, Black, Asian, and other races. Demographic controls include start-of-period shares of foreign-born population, college educated, and female employment. The list of control variables follows Autor *et al.* (2019). Robust standard errors in parentheses clustered on state. *** p<0.01, ** p<0.05, * p<0.1.

Appendix B: Topic analysis of news articles

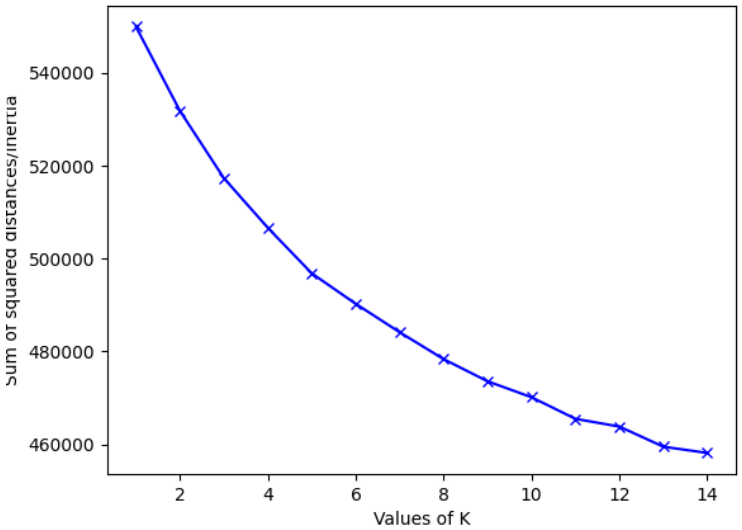
We created word vector representations (Mikolov *et al.*, 2013) to represent words through vectors relying on Vector Space Semantics methodology.[13] This is done by mapping specific vocabulary items in high-dimensional space based on context probabilities (i.e., identifying words that tend to co-occur with a target word or term, and how often).

Vector space representations have been shown to efficiently summarize the thematic relationships between words in a corpus, and enable to measure thematic relatedness between any two given words. Word-vector models can computationally determine "semantic clusters" containing words that belong together.

We used a pre-trained model from Google that includes word vectors for a vocabulary of 3 million words and phrases that they trained on roughly 100 billion words from a Google News dataset. Typical vector dimensionality used in implementations is between 100 and 300. In our implementation, the vector dimensionality is 300. Vector representations of words were computed using the package gensim (Rehurek and Sojka, 2011) in Python.

Using this model, each article's title in our dataset is mapped into the 300 dimensions of vector representation. We then proceed to use the K-means algorithm to cluster these titles into semantic clusters. We implemented with different number of clusters (*K*). Higher number of clusters means that clusters are better fitted (lower sum of squared distance to clusters' centroids) and display more nuanced distinction in topics that articles represent. However, the more clusters we used, the more fragmented our data becomes, and inference of clusters' themes are increasingly difficult.

Appendix Figure B.1: Sum of squared distances to each cluster centers by the number of clusters (K)



[13] Specifically, we use the word2vec family of vector space models.

Appendix Figure B.2: Word clouds for each topic clusters

a) Cluster 1 b) Cluster 2

c) Cluster 3 d) Cluster 4

e) Cluster 5 f) Cluster 6

g) Cluster 7 h) Cluster 8

Note: The clusters show the most popular words used by local US media for each cluster (i.e., topic).