Nate Jacobs

Data and Programming for Public Policy II

Final Project Writeup

**Research Question and Introduction**

Who is winning the US-China trade war? Ever since President Trump applied tariffs to hundreds of billions of dollars’ worth of Chinese imports beginning in 2018, the US and China have been locked in a cycle of escalating trade actions and hostility. My project seeks to understand how the trade war has affected each country’s overall economy by examining metrics like trade openness, import data, and export data.

The data I use cover the period 2017 to 2019, with the “treatment” being the launch of the trade war in 2018. This way, I can make direct comparisons of how the US and Chinese economies were performing before the trade war started and how they have performed since, both in isolation and relative to one another.

Trade data I used came from two main sources: The World Bank (imports and exports and foreign direct investment); and MIT’s Observatory of Economic Complexity (US-China bilateral trade flows). Altogether, I read 13 datasets into R Studio, and turned them into three tidy dataframes, for different types of analysis.

**Wrangling**

I collected two main types of data: data on bilateral trade flows between the US and China, and macroeconomic data, including trade data, on all countries. I had to download one csv per year of bilateral trade between the US and China, so after reading the data into R Studio I bound them together into one dataset using rbind().

On the worldwide macroeconomic data, I conducted extensive reshaping and tidying. For example, years were in their own columns, so I lengthened the dataset to put them in rows.

On all datasets, I standardized country names using the standardize.countrynames() function from the StandardizeText library. I also apply the clean\_names() function from the janitor library to standardize column names and eliminate spaces, which become troublesome when attempting analysis.

I then joined the two datasets containing macroeconomic and worldwide trade flow data into one, on the newly standardized country-year columns. I kept the data on US-China bilateral trade, originally composed of several different datasets, separate since it contains detail on goods and product categories that the macroeconomic dataset lacks.

I also compute trade openness, a measure of export and import values divided by GDP, for each country and add a column to that effect.

After tidying and merging, I turned the new world macroeconomic dataset (now called world\_trade\_final) into a panel dataset to prepare it for running models. I use the dummy\_cols() function to add dummies for year and country, to estimate time and entity fixed effects, to which I return below.

For purposes of analysis, I impute zeros for missing values. Leaving NAs in creates problems for plotting and modeling, and on examination of the distribution of NAs, I see that they mostly come from countries that lack robust macroeconomic data. (None were related to the US and China, the two countries I examine in the project).

**Plotting**

With the data prepared for analysis, I create a few plots to examine the effect of the trade war visually. The first is a choropleth showing inbound FDI in US dollars for most countries in the world. Its purpose is to give a snapshot of the state of world trade just before the trade war started. It shows, unsurprisingly, that the US is a top destination for FDI and China is a major destination for FDI. *(Note: A couple of countries do not appear on the choropleth—this is due to their absence from the source data. All countries present in the source data are accounted for, as proven by the standardization of country names across different datasets).*

I then visualize trade openness over time for the US and China in a bar chart. The chart clearly shows that trade openness for both countries decreased between 2017 and 2019, and slightly more for China. It also shows that compared to the world average, the US and China are not particularly trade-dependent economies.

Delving deeper into US-China bilateral trade, I visualize the change in US exports to China of the ten product categories that dominated US-to-China exports in 2017. The chart shows a major decrease in US exports of transportation goods to China from 2017-19, smaller decreases of other product categories, and a v-shaped trajectory for vegetable products.

Intrigued by this, I created an interactive plot that allows a user to toggle on product category and see a chart tracking US exports *and* imports to/from China over time, along with a table that shows the specific dollar figures of the value of trade in each category. Indeed, far fewer transportation goods were exchanged between the US and China in 2019 than in 2017, both in terms of US exports and US imports—a sign of the trade war’s broad effect.

**Text Analysis**

To gain a broader understanding of the effect of the trade war on the US and China beyond just economic data, I conducted text analysis on articles reporting on the trade war from two sources: The New York Times and the People’s Daily. The New York Times is considered the paper of record in the United States. The People’s Daily is considered one of the most authoritative organs of Chinese party-state media. Though the comparison is not exactly apples-to-apples—the NYT is a private-sector media outlet and the People’s Daily represents the viewpoint of the Chinese Communist Party—the papers’ respective statures in their countries and their robust coverage of the trade war make them appropriate for comparison.

Due to content acquisition limitations, I did not scrape articles directly from the papers’ websites. (I do not have a subscription to either paper). Instead, I searched for articles by time period and keyword using the UChicago library’s Factiva service and downloaded several articles as .txt files.

I chose two pair of articles from each paper, with one pair from around the beginning of the trade war in 2018 and another pair on US-China trade from 2021. Sentiment analysis reveals a surprising finding: both the NYT and the People’s Daily report high counts of positive words in all their reporting related to the trade war. However, the NYT also reports high (but lower) counts of negative words, while the People’s Daily is more uniformly positive. Perhaps this is a result of the NYT’s greater freedom to report on the negative effects of the trade war.

I then conduct dependency parsing to get a sense of the structure of sentences in the NYT and People’s Daily articles. This aspect of the project revealed no new interesting information within the confines of the project’s scope, but could form the basis for interesting future work on how different types of media report on major economic events.

Seeking to make my text analysis easily reproducible, I wrote functions to produce plots of sentiment analysis and dependency parsing.

**Analysis**

The panel dataset I created from tidying OEC and World Bank data allowed me to run a fixed-effects model estimating the effect of the trade war on the relative importance of trade in the economies of the United States and China. In other words, whose trade has actually been more affected by the trade war?

By interacting year and country, and controlling for a series of economic indicators and time-invariant entity fixed effects across other countries, I was able to generate coefficients for the effect of the treatment (trade war launch) on the trade openness of the US and China. The coefficients returned are negative for China and the US in 2018 and positive for both countries in 2019, which, if they were statistically significant, would provide further evidence of the damaging effect of the trade war on the role of trade in each economy. Neither is significant, however, so the model does not reveal any causal relationship.

**Conclusion, Limitations, and Directions for Future Research**

My project reveals mixed findings overall on who is winning the trade war. On the one hand, China’s trade openness appears to have been damaged more than the United States’s has on examination of visual evidence. However, a model controlling for economic indicators and the influence other countries may have on US-China trade does not lead to conclusive evidence. And the major newspapers of both countries report a surprising degree of positive sentiment, a puzzling finding worthy of further analysis.

Several limitations came into play. No data on specific tariffs was used; this was due to the difficulty of merging data on goods and services across the different Harmonized System (HS) levels. The model used likely suffers from weaknesses as well, including unobserved covariates that were not included. There are also other ways of measuring the effect of the trade war; I chose trade openness for the model because it is a direct measure of the influence of trade on a country’s economy, likely with fewer potential confounders than, say, GDP or unemployment.

Future research should attempt to address these shortcomings, and conduct more sophisticated Natural Language Processing, specifically dependency parsing, to examine further how media outlets report on adverse economic events.