Team 1

Kishore Vasan, Lucy Eun, Ji Yun Martel, Sean Yeager

Contents

Background

On July 4th of 2017, North Korea performed its first publicly announced flight test of its new ICBM weapon system which included a miniaturized nuclear warhead capable of fitting inside the long-range missile system. The timing of the event on July 4th, as well as the subsequent statements by North Korea that it was now "a full-fledged nuclear power that has been possessed of the most powerful intercontinental ballistic rocket capable of hitting any part of the world" lead to a great deal of worry and attention on this issue.

On August 8th, 2017, President Trump began to use antagonistic language over Twitter toward the North Korean government, saying now famously that they would "be met with fire, fury, and frankly power, the likes of which the world has never seen before." What followed were increasingly escalating tensions as the rhetoric between the two leaders continued over Twitter over the coming days.

South Korea is a major United States ally in Northeast Asia and its capital, Seoul, is within conventional weapons range of North Korean artillery. Seoul is home to over 25 million people, and is believed to be an immediate target of North Korean aggressive actions should violence begin. The vulnerability of South Korea in general and Seoul in particular is widely regarded as the reason that military interventions have not yet been undertaken against the North Korean regime, and thus, they represent a primary stakeholder in the escalating tensions of the political rhetoric.

With such close relationships between South Korea and the United States as economic and political allies, the exchange rate between the two nations is very important to their ongoing interactions. The international exchange rates are the subject of market speculation, similar to the public stock exchange, and like the public stock exchange, fluctuate with the confidence that investors have in those enterprises. Foreign currency exchanges are similarly affected by the confidence of the stability of that country, and in return, has a very real impact on the liquidity of funds and the available capital within that country.

The sensitivity of foreign exchanges in general to turbulent public statements in general is of interest to political and economic science. The specific example of Donald Trump's public engagement on Twitter with the state leader of North Korea is the subject of our analysis. The primary question we would like to answer is: **"To what extent do the president's tweets have an effect on the currency exchange rate with South Korea?"**

Data Exploration: Currency Exchange Data

Data Collection:

We collected the dataset from the Board of Governors of the Federal Reserve System. They release daily rates of exchange of major currencies against the U.S. dollar. It was possible to select the date range and the country, so we collected the currency data for South Korea from 8/1/2016 to the most current update, which was 10/27/2017.

Major Features:

	DATE	DEXKOUS
0	2016-08-01	1106.1500
1	2016-08-02	1108.3500
2	2016-08-03	1116.2100
3	2016-08-04	1111.9900
4	2016-08-05	1114.5400

The dataset only has two features (columns): date and currency. Date is in the format of yyyy-mm-dd, and the currency column has the currency Won value equivalent to one dollar for the day. For example, the first row tells us that on August 1st, 2016, 1106.15 Won was equivalent to one US dollar. Since the dataset is simple, little data cleaning was required. We changed the column name from “DEXKOUS” to “currency”. Some of the rows have no currency value because the corresponding date was a holiday. In this case, we opted not to remove rows, since we do want to keep the date increasing by one day.

Observations:

There are total 325 dates from 8/1/2016 to 10/27/2017. Out of 325 dates, 13 dates have no currency value because it was Sunday or holiday. For each month, the average currency rate is shown in the table below. The last month, which is October 2017’s average currency rate was 1132.005(Won). Out of this period, the highest currency was 1212.22 (Won) on 12/28/2016, and the lowest currency was 1090.03 (Won) on 9/7/2016

Date	8, 2016	9, 2016	10, 2016	11, 2016	12, 2016	1, 2017	2, 2017
Currency(Won)	1110.018	1108.36	1128.166	1162.708	1183.07	1179.107	1140.487

Date	3, 2017	4, 2017	5, 2017	6, 2017	7, 2017	8, 2017	9, 2017

Currency(Won)	1133.862	1134.176	1125.136	1130.846	1131.746	1130.27	1131.599

This graph is South Korea Currency from 8/1/2016 to 10/27/2017. As shown, the peak during this time period was above 1200(Won). This happened in December, even though the x-axis' mark is not showing because the texts are too long. The lowest time was towards beginning, where it was less than 1100(Won). Because we are more interested in the changes than the currency value of the day, having a table of delta values will be more relevant.

Data Exploration: Twitter Data

Data Collection:

We were able to get the latest **3227** tweets by Donald Trump using the Twitter API. The oldest tweet goes back to **'2016-08-22 01:19:06'** till **'2017-10-20 05:12:44'.** We are not considering Retweets(RT) which comes out to about 353 tweets(~10% of the dataset).

Data Cleaning:

Example of an unclean tweet : "b'It was great to have Governor @RicardoRossello of #PuertoRico\\xf0\\x9f\\x87\\xb5\\xf0\\x9f\\x87\\xb7with us at the @WhiteHouse today. We are with you!\\xe2\\x80\\xa6 https://t.co/Q2NhV2MAXD'"

We removed all the emoticons in the tweets which was decoded as “\\xe2”,”\\xa6” etc. We also removed the starting two indexes of the tweets that was “ b’ “ and the ending single quote. After cleaning, the tweet became: 'It was great to have Governor @RicardoRossello of #PuertoRico with us at the @WhiteHouse today. We are with you! https://t.co/Q2NhV2MAXD'

Dataset Features:

	id	created_at	text
0	921242760685998080	2017-10-20 05:12:44	b'Great news on the 2018 budget @SenateMajLdr ...
1	921209530628956161	2017-10-20 03:00:42	b'Big ratings getter @seanhannity and Apprenti...
2	921207772233990144	2017-10-20 02:53:42	b'The Fake News is going crazy with wacky Cong...
3	921189980843663360	2017-10-20 01:43:00	b'Keep up the GREAT work. I am with you 100%!\...
4	921113816053755904	2017-10-19 20:40:21	b'It was great to have Governor @RicardoRossel...

The dataset contains 3 columns Id, created_at and text. The first column is a tweet id created by Twitter, the second column is that date-time of the tweet and the third column is the tweet itself.

The first column is of type integer, and the other two columns are of type string.

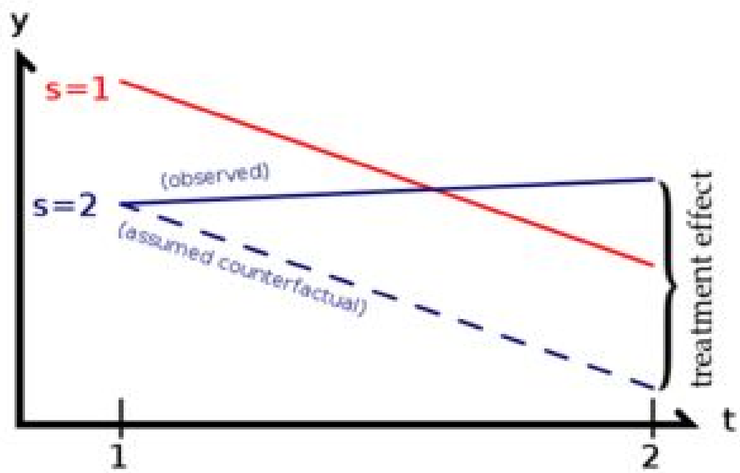
Observations:

Tweets that had mention of Korea, Kim Jong Un, the rocket tests, or any other mention of the political difficulties were hand selected. The dates for tweets that used aggressive or threatening language, even indirect threats, were used for evaluation in our model. The total number of dates selected was 56.

Data Analysis: Econometrics

Difference in Differences Method:

In order to mimic a experimental research design using observational data (in this case the currency exchange rates), we study the differential effect on a control group versus a treatment group.



As you can see from this graph, given a base line s = 1 at t = 0 we expect a trend in difference. Once the treatment takes place, at t = 1, we expect it to follow the assumed counterfactual line, but instead it follows a different line at s = 2. The **difference in differences** here is the treatment effect.

In this case, we will model all the days that Trump did not tweet to get an expected s = 1 line. Now, given the treatment(Trump Tweeting), we observe the bold blue line as opposed to the counterfactual(dotted blue line). The difference between the bold blue line and dotted blue line will act as the treatment effect.

Taking motivation from Card and Krueger (1994):

Card and Krueger hoped to see the effect of increase in minimum wage on the employment in the fast food industry in New Jersey. They noted that, observing a change in employment in New Jersey only, before and after the treatment, would fail to control for omitted variables, such as weather and macroeconomic conditions of the region. So, they used the employment in fast food industry in Pennsylvania as a control group. By including Pennsylvania as a control in a difference-in-differences model, any bias caused by variables common to New Jersey and Pennsylvania is implicitly controlled for, even when these variables are unobserved. Assuming that New Jersey and Pennsylvania have parallel trends over time, Pennsylvania's change in employment can be interpreted as the change New Jersey would have experienced, had they not increased the minimum wage, and vice versa. (Card and Kreuger)

This same example can be modeled in our case. The days Trump didn't tweet about North Korea can be considered as the control group and the days Trump tweeted about North Korea as the treatment group. While this doesn't act as straightforward like the example above as the currency exchange is much more fragile than the fast food industry, this is a step to take to model the observational data in an experimental design.

Modeling this method:

We can run regression analysis using Ordinary Least Squares(OLS) model to get the results.

$$y = \beta_0 + \beta_1 T + \beta_2 S + \beta_3 (T \cdot S) + \varepsilon$$

Here, T is a dummy variable set to 1 if the observation is from the second time period group, else set to 0 if it is from the first time period group. S is also a dummy variable set to 1 if it is from the treatment group, else 0 if it is from the control group. T*S is a composite variable indicating if T=S=1.

The beta terms can be transformed into the values given below.

$$\hat{\beta}_0 = (y \mid T = 0,\ S = 0)$$

$$\hat{\beta}_1 = (y \mid T = 1,\ S = 0) - (y \mid T = 0,\ S = 0)$$

$$\hat{\beta}_2 = (y \mid T = 0,\ S = 1) - (y \mid T = 0,\ S = 0)$$

$$\begin{aligned}\hat{\beta}_3 = &\big[(y \mid T = 1,\ S = 1) - (y \mid T = 0,\ S = 1)\big] \\ &- \big[(y \mid T = 1,\ S = 0) - (y \mid T = 0,\ S = 0)\big],\end{aligned}$$

In simple terms without the intimidating math:

DID =E[YA−YB|D=1]−E[YA−YB|D=0]

Where, the first term is the expected difference between the treatment and control group. Here, YA is the expected Y from the second time period A and YB is the expected Y from the first time period A. and D indicated if the treatment or interference(in this case Trump's tweet) has happened or not. And hence, DID explains the difference in the interference.

Assumptions:

Basic OLS assumptions -

- The linear regression model is "linear in parameters.".
- There is a random sampling of observations.
- The conditional mean should be zero.
- There is no multi-collinearity (or perfect collinearity).
- Spherical errors: There is homoscedasticity and no autocorrelation.
- Error terms should be normally distributed.
  ("Key Assumptions of OLS: Econometrics Review")

Common trend assumption - The DID econometric model requires a parallel line assumption. This means that the model requires a control group that has as similar and parallel a pattern as possible to the treatment set as possible. It is very difficult to find a truly parallel process in real life examples, and this is expected by the model. The degree to which parallelism can be observed is proportional to the accuracy of the DID model. (Bertrand, Duflo, and Mullainathan)

To establish a control set, we chose the Canadian - South Korean Won exchange rate.

This exchange rate’s trend pattern closely matches the trend pattern we see for the US - South Korean exchange rate. This is within acceptable bounds to facilitate our use of the DID method for our analysis. (Chabé-Ferret)

Results:

The DID regression summary is shown below. The interaction coefficient shows the Difference in Difference estimate. Since the coefficient of interaction is positive 0.4405, the result shows the effect being significant at 100% with the treatment having a positive effect. The adjusted r-squared value, which shows the goodness of a fit, is .9947, which is very high. Also the p-value is less than 2.2e-16.

```
> reg1 = lm(value ~ time + treatment + interaction, data = alldf)
> summary(reg1)

Call:
lm(formula = value ~ time + treatment + interaction, data = alldf)

Residuals:
    Min      1Q  Median      3Q     Max 
-56.454 -24.688   7.606  21.530  32.806 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 887.9178     6.6369 133.785   <2e-16 ***
time          1.8701     9.3860   0.199    0.843    
treatment   695.7442     9.3860  74.126   <2e-16 ***
interaction   0.4405    13.2738   0.033    0.974    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.7 on 56 degrees of freedom
Multiple R-squared:  0.9949,	Adjusted R-squared:  0.9947 
F-statistic:  3665 on 3 and 56 DF,  p-value: < 2.2e-16
```

With the high significance and adjusted r-squared value as well as a p-value lower than 0.05 allow us to conclude that **there is a positive effect that Trump’s posts have on South Korean currency rate.** However, due to the nature of our statistical model, there are some considerations to include along side this declaration that mitigate the conclusiveness of this statement.

As discussed in the previous section, the DID model requires a control group that exists in the same space of time as the treatment group in order to form its “natural experiment". Due to the nature of the observation, there is no "other US to South Korea" exchange rate to consider alongside our treatment exchange rate. Canada's exchange rate is used as a substitution for the United States, but it is not itself, the United States. While Canada and the United States are together in the North American Free Trade Agreement (NAFTA), and are important political and economic partners, they are subject to many, many different factors. Our model cannot account for all of these missing variables, and the results that we demonstrate must be understood in this context. We cannot make the claim that Canada would otherwise be identically affected by circumstance and thus cannot say it is a "perfect" control.

Additionally, the most important assumption behind DID model is the parallel trend assumption, and our control and treatment groups have somewhat parallel trends. However, to what extent they have to be parallel to be fair to use DID model is unclear, and the degree to which they are not perfect parallels is the degree to which our model does not account for all variables. Both of these considerations in the selection of the Canada to South Korean exchange rate should moderate the claim that we make above stating a causal link between Trump's tweets and the changes in exchange rate that we observed.

The second significant concern is this test's small degree of freedom. The degree of freedom is only 56, despite starting with 3227 tweets and more than a year worth of daily currency exchange rates. While the event itself spanned several weeks of political and media attention, the actual number of specific tweets about North Korea were relatively slim by comparison. As well, some of the dates that represent the tweets as our instrumental variable fell on holidays when currency data was not recorded, and could not be considered fully.

Work Breakdown

Each member of the team contributed to the project in their own way. The work breakdown structure is as follows:

Lucy Eun collected the currency exchange data from the IMF, and performed the cleaning and study of that dataset. She performed parts of the statistical analysis for the results, helping to formulate it and understand it for the team. She also wrote significant portions of the analysis for this paper.

Jiyun Martel performed initial t-tests during the data analysis, although they ultimately proved to be an inadequate analysis technique and were not included in these results. As well, she searched out and reviewed previous work on the topic. This essential portion of the project helped to come up with the DID method that formed the most important part of the analysis. Jiyun's personal experiences also inspired the initial research question regarding South Korean currency exchange rates and their sensitivity to the political atmosphere.

Kishore Vasan collected the data from the Twitter set, and performed the cleaning from the API, as well as performed filter searches to locate relevant tweets. Kishore also wrote the econometrics section of this paper, and performed the linear regression models.

Sean Yeager hand-sorted the tweets from the set and classified them by relevance. He helped to implement the DID model in our data, and acted as editor and primary writer for the final paper displaying our findings.

Further Goals

Given our work so far, we show a promising correlation, and because of the nature of the DID test's nature as a "natural experiment" can make some causal claims. However, more could be done to make a more convincing claim and to suggest and test the further implications of such claims.

We also want to judge the positivity/negativity of Trump's posts by using Natural Language Processing. Right now, we can filter Trump's posts by its topical content, but our analysis makes no effort to decide whether the post is positive or negative in any official categorical capacity. Our research question, whether the government official's posts have a relationship with global currency exchange, can be studied in the context of positive posts and negative posts.

This same research could be further expanded by studying the exchange rates of different countries vs Trump's Tweets. After thoroughly studying data about Korea, we can move onto other countries such as Russia, Germany, or any other countries and study in the same manner we studied with Korean currency rate.

It would also be possible, with further data and analysis, to cross-tabulate between these expressions of exchange rate differences and actual dollar losses or gains and other possible specific economic impacts that such changes could suggest. With more data we could estimate actual potential losses in terms of loans issued or denied, productivity shortfalls, and other econometrics. This might put the research into more substantial terms in order to make specific policy recommendations in reaction to this understanding.

References

1. Card, David, Krueger, Alan B. “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania”. The American Economic Review, Vol. 84, No. 4. 1994, pp. 772-793.
2. “Key Assumptions of OLS: Econometrics Review”,*Albert,* 25 Nov. 2016, https://www.albert.io/blog/key-assumptions-of-ols-econometrics-review/. Accessed 17 Nov. 2017.
3. Waldinger, Fabian. “Lecture 3 : Difference in Differences”, https://www2.warwick.ac.uk/fac/soc/economics/staff/ffwaldinger/teaching/ec9a8/slides/lecture_3_-_did.pdf. Accessed 17 Nov. 2017.
4. Bertrand, Marianne, Duflo, Esther, and Mullainathan, Sendhil. “How Much Should We Trust Differences-In-Differences Estimates?“. *The Quarterly Journal of Economics*, Vol 119, Issue 1, 1 Feb. 2004, pp. 249–275.
5. Chabé-Ferret, Sylvain. “Why Does Difference In Difference Matching Work?”. conference.iza.org/conference_files/ESSLE2014/10424.pdf. Accessed 17 Nov. 2017.