USD-YEN exchange rate analysis before and after a tumultuous week

Xiaojie Chong
The Ohio State University

ABSTRACT

In this paper, we analyze currency exchange rate of Japanese and U.S. I try to use different data analytic techniques to analyze the data. The dataset I use is currency exchange rate of YEN and USD between 1971-01-04 to 2019-05-10. The paper can be further divided into three parts. First part is about data preprocess. I will divide the whole dataset into different time periods to avoid steep points' influence and use splie and polynomial regression to process the data. In second part, I will slide the windows of different sizes to find the best window size for performing PCA. In the third part, I will give a case study about exchange rate changes before and after a tumultuous week. The result shows that before a tumultuous week, the exchange rate is under a regular behavior but after a tumultuous week, the exchange rate will become unpredictable.

1 INTRODUCTION

With the explosion of the information revolution, more and more data can be accessed by people through Internet. Faced with the data whose size is gigabyte or even terabyte, it is very difficult for people to find the rules simply by reading the data. In addition to data size, multi-dimensional data also prevent people from having an intuitive understanding. These characteristics are particularly evident in the financial field, and how to make uses of different tools to effectively analyze those data is of vital importance in the financial field. As a person who is learning to use MATLAB and python, I will use several common data analysis tools together in this paper.

In this project, I focus on currency exchange rate of Japanese and U.S. and show that the rates follow a "regular behavior" before the "catastrophe" and the rate dynamics becomes subject to great uncertainty afterward. After finishing the project I realize that the the economy is moving towards globalization, and the occurrence of a financial crisis will affect all kinds things in markets. In the current context, if we can learn how to deal with big data and use these data for speculation and judgment, it can help us a lot in market analysis. In the future, I will apply the concept of big data to everything.

Lightspots. This paper has some Lightspots to the study of currency exchange rate in the following aspects:

- The paper studies a problem of currency exchange rate between YEN and USD with a large dataset from 1971-01-04 to 2019-05-10.
- The paper use slide windows to find steep points in the whole dataset and divide the dataset into different time periods to make sure that the data in each time period is predictable and reasonable. The paper compares the performance of Spline and polynomial regression.

Table 1: Highest Absolute Ratio Changes.

Date	Ratio Change
1971-08-31	-0.048971596474
1973-02-13	-0.0906703948714
1974-01-07	0.0645538657875
1978-04-03	-0.0502414198095
1978-11-01	0.0362116991643
1998-06-17	-0.0442207566817
1998-10-07	-0.0547464735036
1998-10-08	-0.0413003145922
2008-10-06	-0.0425491679274
2008-10-24	-0.0508196721311
2009-03-19	-0.0431280587276
2016-06-24	-0.0371137905049

- The paper tries different window sizes in different time period to find the best window size for performing PCA such that the eigenvalues obtained consistently carry the most amount of information, which, in other words, means most valuable data is extracted from the original multi-dimensions data.
- The paper further give a case study of rate changes before and after a tumultuous week, the result shows the exchange rate is under a regular behavior but become unpredictable after that.

Roadmap. Section 2 presents the background of currency exchange rate problems. Section 3 explains how I use slide window, spline and polynomial regression to process the data. Section 5 describes how I normalize the input data and find the best window size for performing PCA such that the eigenvalues obtained carry the most valuable data. Section 5 . In Section 6, I will further analyze the data. Section 7 summarizes related work and reference. Section 8 concludes the paper.

2 BACKGROUND

In this section, I will mainly introduce two data analysis tools I used in the paper.

Principal Component Analysis. Principal Component analysis (PCA) [12] is a very powerful tools in data processing. The aim of PCA is to remove the reduction of a large scale of data and to find those independent factors. There are some other advanced PCA [8, 11, 13], which can be used in different cases. In this paper, I just use the original PCA and the experiment results show its promising performance in dimension reduction.

Linear Regression. Most of our daily data is discrete. However, discrete data is difficult to analyze and predict. At this time, we need linear regression to process discrete data. For Linear regression, there are many different models [5, 6, 9] that can be used to restore data relationships from discrete data. In this paper, I use two different models and compare their performance towards the currency exchange data.

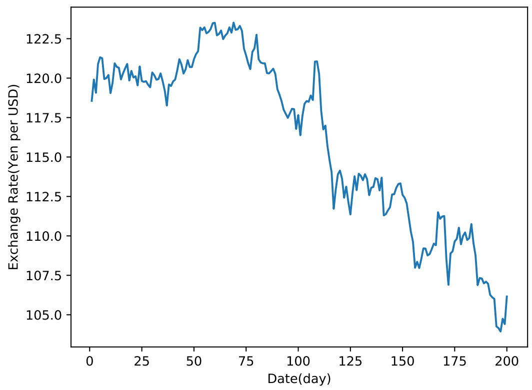

Figure 1: Raw data between 2015-08-24 to 2016-06-24.

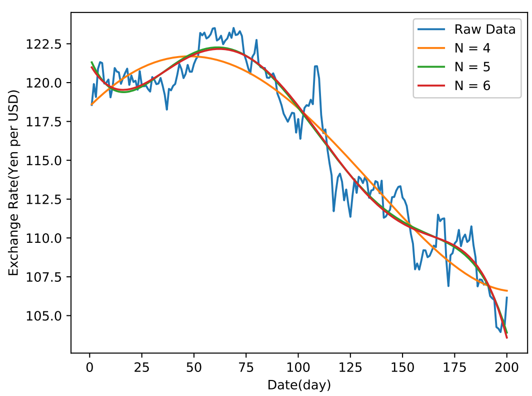

Figure 2: Polynomial regression with different degrees between 2015-08-24 to 2016-06-24.

3 DATA PREPROCESS

Part 1. Improve the quality of data. Find the best supplementary for the given daily rate data (the "best" in the sense of least square mean error for statistical moments such as mean and variance) using curve-fitting methods (e.g. use MATLAB).

In this part, we are going to find the best supplementary for the given daily rate data for the currency change rates between USD and YEN. The goal here is to find the minimum square mean error

3.1 Data Resource

The dataset we use is a exchange rate chart of USD and YEN, which we get from [3]. The chart include exchange rate from 1971-01-04 to 2019-05-10. We first use python[7] to read the CSV file, we use two lists to store the date and the exchange rate of that data. The two lists look like:

$$Date_{list} = [\text{“}1971-01-04\text{”}, \text{“}1971-01-05\text{”}, ..., \text{“}2019-05-10\text{”}]$$

$$Rate_{list} = [\text{“}357.73\text{”}, \text{“}357.81\text{”}, \text{“}357.86\text{”}, ..., \text{“}109.7190\text{”}, \text{“}109.945\text{”}]$$

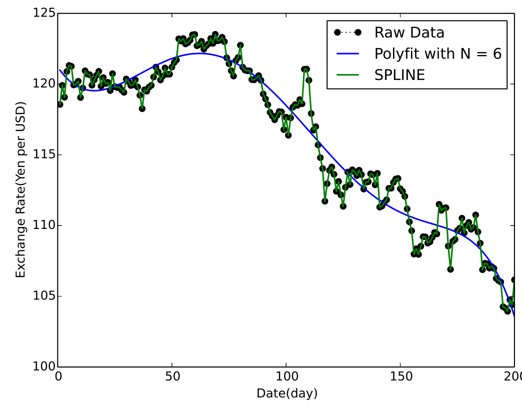

Figure 3: SPLINE versus polynomial regression.

We make sure $Date_{list}[x]$ means the xth date in the whole dataset and $Rate_{list}[x]$ means the exchange rate in the xth date.

Also, we create a dictionary and make the dictionary satisfy the rule below. The dictionary can help us find the exchange rate of certain date in a very quick way.

$$DIC(Date_{list}[x]) = Rate_{list}[x]$$

NOte there are total 12247 data in our dataset.

3.2 Partition the dataset

After getting the dataset, we then need to partition the dataset. The reason why we need to partition data into different sections is that there will be some sharp drop or raise points in the dataset. With those steep points, it is hard to do polynomial regression. Moreover, those steep points are usually caused by unstable markets or economic situation. It is meaningless to analyze the currency exchange rate when there is a Great Depression or financial crisis.

The way we decide those steep points is we compare the ratio of exchange rate changes. We create a list of absolute ratio change value:

$$Change_{list}[x] = ABS(Rate_{list}[x]/Rate_{list}[x-1] - 1)$$

, where x = 1, 2, 3, ..., 12247.

We soft the $Change_{list}$ to decrease order. Table 1 shows those date whose absolute ratio changes more than 3.5% in the history form 1971-01-04 to 2019-05-10. As we can see in Table 1, there are total 12 dates whose absolute ratio changes more than 3.5%.

In our experiment, we regard those dates whose dates whose absolute ratio changes more than 2.5% as steep points. There are total 72 steep points. Thus we can divide the whole dataset into 73 different **time periods**. Here we define **time period** as a continuous time where there is no steep points inside this time. The 72 steep points, the 73 time periods and the length of different time periods can be seen in the appendix.

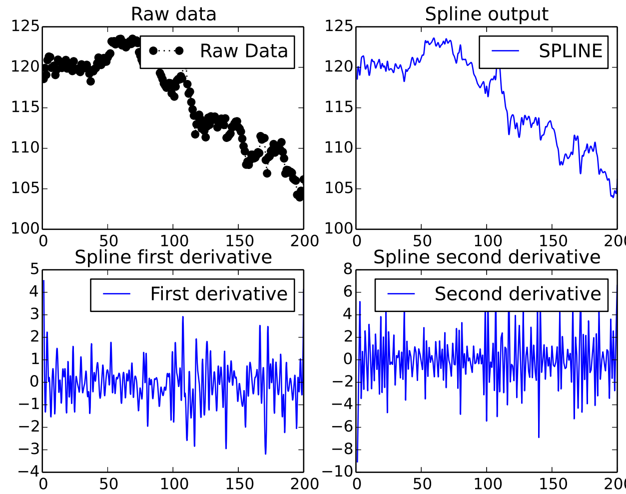

Figure 4: SPLINE output and first and second drivative.

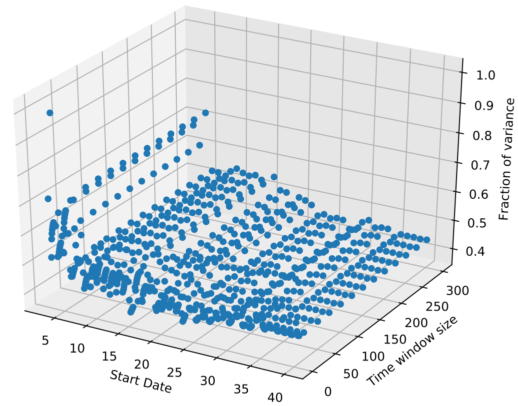

Figure 5: Fraction of variance of the first principal component in different T_{start} and W_{size} combination.

3.3 Spline and Polynomial Regression

In this part, we will do polynomial regression for different time periods. We only focus on time period whose length is longer than 50 days but less than 365 days. We take the time period between 2015-08-24 to 2016-06-24 (total 200 date) as an example in this section.

Figure 1 shows the exchange rates between 2015-08-24 to 2016-06-24 (total 200 date). Then we need to use polynomial regression to process the data.

We define $R(x)$ as the polynomial function, where x is the xth day in the time period and the value of $R(x)$ is the exchange rate (Yen per USD) of that day. We can see in Figure 2 that when the polynomial regression degree is equal or greater than 5, the curve will have a similar shape as the raw data. Here the polynomial regression's parameter when degree is equal to 6 is

$$\begin{aligned} R(x) = -5.162e^{-12}x^6 + 8.221e^{-10}x^5 + 4.598e6{-}07x^4 \\ -0.0001317x^3 + 0.01039x^2 - 0.2394x + 121.2 \end{aligned} \tag{1}$$

After patitioning the dataset and find the polynomial regression, we begin to to data analysis in next section.

4 DATA ANALYSIS USING PCA

Part 2. Determine the best time-window size for the analysis goal. Use data analytic techniques on time-windows of various size for the data to find the window sizes (i.e. time interval lengths) over which predictions of behavior can correctly be made (or with least error as mentioned above.)

4.1 Data normalization

The input matrix we used in data analysis is a three-dimensions matrix. We use M_{input} to denote the input data. M_{input} is a $N * 3$ matrix, where the first column is formatted by points found with PPVAL, the second column is formatted by points from the first derivative of the SPLINE, the third column is formatted by points from the second derivative.

Our goal in this section is to find the best window size for performing PCA such that the eigenvalues obtained consistently carry the most amount of information, which, in other words, means most valuable data is extracted from the M_{input}.

The first step for performing PCA is to normalize M_{input}. We use PREPSTD function, which will process the matrix so that the output matrix has a mean of zero and a standard deviation of one. We call the output matrix M_{STD}.

We then performing PCA toward M_{STD}. For the output of PCA, the total variance is the sum of the variances of all principal components. The **fraction of variance** R_{var_i} is the ratio of the **ith** principal component and the total variance. Here we have

$$\sum_{n=1}^{N} R_{var_i} = 1$$

where N is the number of principal components and in our example N = 3.

4.2 Goal function

Our goal is to find $MAX(R_{var_1})$ (the fraction of variance of the first principal component) in different time windows. So we assume we want to focus on a time period which ended in date T_{end}. The T_{end} might be a date when a tumultuous week happens or a crash happens. We want to adjust time windows W_{size} and the start point T_{start} of the time period in order to find the $MAX(R_{var_1})$. Note that each combination of $G(T_{start}, W_{size})$ will have a corresponding R_{var_1}.

Here we define a function called

$$G(Combination\ of\ (T_{start}, W_{size})) = Max(R_{var_1})$$

When given a certain time period, our goal then become to find the maximum value of $G(Combination\ of\ (T_{start}, W_{size}))$.

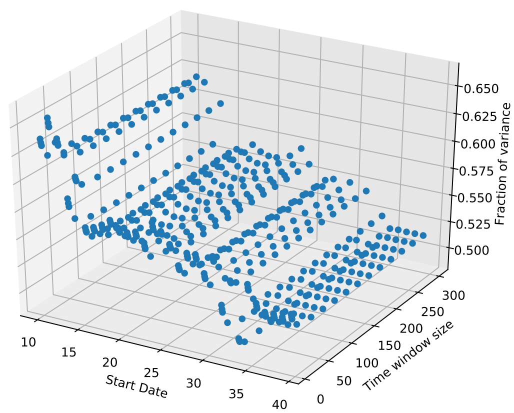

Figure 6: Before Catastrophe.

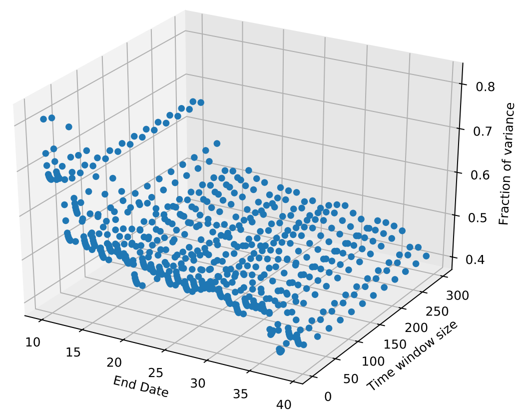

Figure 7: After Catastrophe.

4.3 Case Study

Our dataset in this section is consistent with the previous section, which contains date from 2015-08-24 to 2016-06-24 (total 200 dates). In this part, I will give a case study of how do I find the maximum value of our goal function. In Section 5, we will examine the Japanese/U.S. exchange rates before and after a tumultuous week (e.g. a crash) in the market.

In our case study, the T_{end} is 2016-06-24. The T_{start} varies from 2016-04-13 to 2016-06-22, which is 40 days to 2 days before T_{end}. The W_{size} varies from 300 per day to 1 per day.

We use Figure 5 to show the results. The X-axis is different T_{start} , the Y-axis is different W_{size} and the Z-axis is the corresponding fraction of variance of the first principal component. Using this way, I am able to find combination of W_{size} and T_{start} whose fraction of variance is higher than 0.95, which means most valuable data is extracted from the input data.

5 TUMULTUOUS WEEK'S INFLUENCE

In this section, I will examine the Japanese/U.S. exchange rates before and after a tumultuous week (e.g. a crash) in the market. We refer to the market crash as a "catastrophe". Moreover, I will test the hypothesis provided in project goals: The rates follow a "regular behavior" before the "catastrophe" and the rate dynamics becomes subject to great uncertainty afterward.

The tumultuous week I chose is the "catastrophe" occurred in 1997. On October 27, 1997, the Dow Jones Industrial Average dropped by 554 points or a little more than seven percent of its value. We will test in this section that the rates follow a "regular behavior" before October 27, 1997 and the rate dynamics becomes subject to great uncertainty afterward.

5.1 Before Catastrophe

Figure 6 show the rates changes before the catastrophe. The X-axis is different T_{start} , the Y-axis is different W_{size} and the Z-axis is the corresponding fraction of variance of the first principal component.

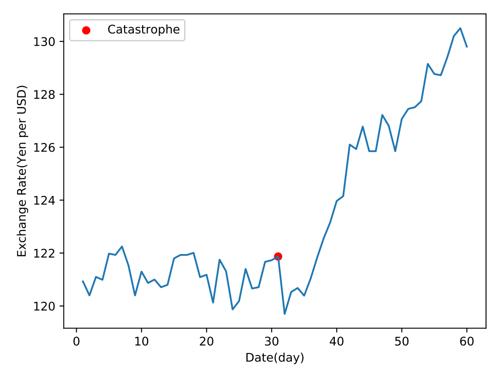

Figure 8: 30 days' rates before and after catastrophe.

5.2 After Catastrophe

Figure 6 show the rates changes after the catastrophe. However, there is some changes from the previous figures. Since here we analysis N days after the catastrophe, the X-axis is no longer T_{start} but T_{end}, the Y-axis is different W_{size} and the Z-axis is the corresponding fraction of variance of the first principal component.

When compared Figure 6 and Figure 7, we can see that after catastrophe, the distribution is more mess and the fraction of variance is higher, which means the data is more unpredictable. After around 30 days, the data become smooth again. We can also look at Figure 8. Figure 8 is the exchange rate changes before and after the catastrophe. Before the catastrophe, the exchange just fluctuate a little. I believe this is because the noise and uncertainty of market. However, after the catastrophe, we can see steep changes of the rates.

6 DISCUSSION

In this section, I further analyze more data and verify the hypothesis in different tumultuous weeks. All the data show those tumultuous weeks all follow that hypothesis and thus I believe the hypothesis is true.

7 RELATED WORK

In this section, I will talk about some related work about what I have done in this paper as well as some reference. Those related work and reference helped me a lot in writing the report.

There are many other works which also use some popular data analysis tools to analysis currency exchange rates. Most of their paper will focus on some economic events' influence on currency exchange rates. [10] demonstrates that the currency exchange rates are influenced by profit rates across international regulating capitals. [2] improves the predication by replacing the previous commonly given variance-smoothing weights to use Principal Components Analysis (PCA) instead. [4] proposes a new robust forecasting models for currency exchange rates prediction. The authors make use of PCA as well as some artificial neural network to analyze the exchange rate history and do a prediction. [1] gives a study of EUR/USD exchange rate. Moreover, the put forward a model about how to measure forecasting accuracy.

8 CONCLUSION

In this paper, we analyze currency exchange rate of Japanese and U.S. By doing the research, I become familiar with some data analytic techniques. By using those data analytic techniques, I am able to analyze large scale of data and verify different hypothesis in the dataset. In this project, I focus on currency exchange rate of Japanese and U.S. and show that the rates follow a "regular behavior" before the "catastrophe" and the rate dynamics becomes subject to great uncertainty afterward. I realize that the the economy is moving towards globalization, and the occurrence of a financial crisis will affect all kinds things in markets. In the current context, if we can learn how to deal with big data and use these data for speculation and judgment, it can help us a lot in market analysis. In the future, I will apply the concept of big data to everything.

REFERENCES

[1] Mauro Costantini, Jesus Crespo Cuaresma, and Jaroslava Hlouskova. 2016. Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate. *Journal of Forecasting* 35, 7 (2016), 652–668.

[2] Scott William Hegerty. 2013. Principal component measures of exchange market pressure: comparisons with variance-weighted measures. *Applied Financial Economics* 23, 18 (2013), 1483–1495.

[3] Macrotrends. 2019. Dollar-Yen-Exchange-Rate-Historical-Chart. https://www.macrotrends.net/2550/dollar-yen-exchange-rate-historical-chart. (2019).

[4] Babita Majhi, Minakhi Rout, Ritanjali Majhi, Ganapati Panda, and Peter J Fleming. 2012. New robust forecasting models for exchange rates prediction. *Expert Systems with Applications* 39, 16 (2012), 12658–12670.

[5] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. 2012. *Introduction to linear regression analysis.* Vol. 821. John Wiley & Sons.

[6] John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman. 1996. *Applied linear statistical models.* Vol. 4. Irwin Chicago.

[7] Python. 2019. Python. https://www.python.org/. (2019).

[8] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. 1997. Kernel principal component analysis. In *International conference on artificial neural networks.* Springer, 583–588.

[9] George AF Seber and Alan J Lee. 2012. *Linear regression analysis.* Vol. 329. John Wiley & Sons.

[10] Anwar M Shaikh. 1999. Real exchange rates and the international mobility of capital. (1999).

[11] Michael E Tipping and Christopher M Bishop. 1999. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61, 3 (1999), 611–622.

[12] Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. *Chemometrics and intelligent laboratory systems* 2, 1-3 (1987), 37–52.

[13] Hui Zou, Trevor Hastie, and Robert Tibshirani. 2006. Sparse principal component analysis. *Journal of computational and graphical statistics* 15, 2 (2006), 265–286.

Table 2: Full Steep Points (72).

Date	Ratio Change	Time Period Length (day)
1971-08-31	-0.048971596474	155
1972-06-23	-0.0263184030192	205
1973-02-13	-0.0906703948714	158
1974-01-07	0.0645538657875	224
1978-03-31	0.0338175113549	1060
1978-04-03	-0.0502414198095	1
1978-11-01	0.0362116991643	148
1979-05-07	-0.0302483069977	127
1979-12-10	-0.0314321398834	148
1980-04-23	-0.027	93
1985-09-23	-0.0341524364848	1358
1985-09-26	-0.0290877038343	3
1986-01-24	-0.0296956198961	80
1986-05-05	-0.0251912889935	70
1987-08-18	-0.0256803628602	325
1988-01-05	0.0342298288509	94
1988-01-15	0.0297482837529	8
1989-06-16	-0.0272022456891	357
1990-05-11	-0.0271974522293	227
1991-01-17	-0.0290098648155	171
1992-01-21	-0.033346410357	252
1993-06-18	0.0257611241218	357
1993-06-25	-0.0283752860412	5
1994-02-14	-0.0333426209715	160
1995-03-31	-0.0299341003016	283
1995-05-26	-0.0258671369782	40
1995-08-02	0.027027027027	46
1995-08-15	0.0327956414913	9
1995-09-21	-0.0305084745763	26
1997-05-09	-0.0251917642309	410
1997-08-08	-0.0261824324324	63
1997-12-17	-0.0295713303278	89
1998-06-17	-0.0442207566817	125
1998-06-22	0.0275707898659	3
1998-06-30	-0.0255777903044	6
1998-09-01	-0.0331440738112	45
1998-09-09	0.0329295987888	5
1998-10-06	-0.0251969674446	19
1998-10-07	-0.0547464735036	1
1998-10-08	-0.0413003145922	1
1998-10-20	0.0317141359427	7
1998-11-09	0.0273730312474	14
1999-01-12	0.0297712027933	42
1999-02-02	-0.0258058910418	14
1999-02-16	0.0328901329601	9
1999-09-09	-0.0269393639067	144
1999-09-13	-0.0257798165138	2
2000-03-01	-0.0253003276301	117
2000-03-31	-0.0265327395054	22
2002-03-07	-0.0288880397402	484
2005-07-21	-0.0260095221301	850
2005-12-14	-0.0274748147531	100
2007-08-16	-0.0269647580852	421
2008-03-17	-0.0331337325349	147
2008-10-06	-0.0425491679274	142
2008-10-08	-0.0251221896383	2
2008-10-14	0.0274537409493	3
2008-10-24	-0.0508196721311	8
2008-10-28	0.0257289879931	2
2009-01-21	-0.0264996119304	56
2009-03-19	-0.0431280587276	40
2010-09-15	0.0310656231186	377
2011-03-18	0.0304800609601	125
2011-08-04	0.0278392090542	97
2011-10-31	0.0297147385103	60
2013-04-04	0.0339931153184	357
2013-06-20	0.0310859063222	54
2014-10-31	0.0289149990821	343
2015-08-24	-0.0290721480632	202
2016-06-24	-0.0371137905049	200
2016-07-29	-0.0304930179538	25
2019-01-02	-0.026701551589	707