

DSCI 550 Final Project Report

Member Name: Zhuoyou Shen, Ni Huang, Fanglin Ma

Title: Interpretation and Prediction of US Inflation Rate during Big Event

1 Project Background

Recently, we have found that the prices of many goods are going up dramatically, like cars, fuel, electricity and so on. Many economic newsletters like Morning Brew explained this situation by the increase of inflation rate. Learning the inflation rate can help us better understand the current economic and social situation. There was previous research investigating the relationship between inflation rate, exchange rate, interest rate and stock price[1]. However, we hope to pay more attention to various factors instead of only macroeconomic variables. So, after reading more papers and articles, we found that inflation rate may be related to factors like interest rate, unemployment rate, import and export trade value, GDP, and house price.

2 Problem Definition

The most concerning problem is which features have significant relationships with the inflation rate and how to predict future inflation rate based on those significant features, especially during the time of some big events like the great recession and the pandemic.

3 Description of Dataset

All of the datasets' quality are good, there are no null values, or any outliers in the datasets, and the details of each dataset are described in Table 1.

Table 1. Table of Datasets Description

Factor	Description	Attributes	Quantity	Source
Inflation Rate	Web script. Monthly US inflation rate from 2000 to 2022.	Date by month; Inflation rate	264 x 2	Inflation Rate 2000-2022
Interest Rate	Web script. It contains a history lookup, monthly average, rates table, and annual average.	Date by year and month and a fixed-30 year interest rate	10 x 13	Interest Rate 2012-2022
Unemployment Rate	Manually download. Civilian Unemployment rate, seasonally adjusted. It contains data from 2012 to 2022. Also, the consumer index.	Date by year and months. 10 years' unemployment rate	10 x 12	Unemployment rate; CPI

Trade Value	API or Manually download. Monthly import/export trade value information of different countries from 2010 to 2020.	Date by month; Trade Flow; Reporter; Partner; Commodity code; Trade Value	Over 100000 x 6	UN Comtrade
GDP	Manually download. Quarterly GDP value information of the US from 1951 to 2021	Date by quarter; GDP	284 x 2	IMF Data
House Price	Manually download. The dataset provides monthly US house prices.	Region names; Date by month.	908 x 269	Zillow Housing Data

4 Project Plan and Methods

We would like to predict the inflation rate based on the most significant features by various regression models. So, we plan to conduct the project as the following steps:

1. Collect data of all the factors by web scripting, API or manually downloading them.
2. Conduct preprocessing work including removing the null values or outliers, data rescaling and merging datasets of different factors.
3. Select features by hypothesis testing with t-statistics. Suppose that there is no relationship between the inflation rate and each feature, then use t-statistic and p-value to decide whether to accept or reject the null hypothesis.
4. Use linear and polynomial regression models to predict the inflation rate based on each feature solely.
5. Take the inflation rate as the response variable, different combinations of factors as the independent variables, and then utilize linear and polynomial regression models to predict the inflation rate.
6. Use R-Square to evaluate the regression models and compare the accuracy to find the best model.
7. Explore the reason for the abnormal inflation rate and how the government reacts during the time of some big events like the great recession and the pandemic.

5 Experiment

In this section, we would like to describe the whole process about our experiment and analysis.

5.1 Data Collection

First, we scraped different websites by using Python libraries like requests and BeautifulSoup to get inflation rate and exchange rate datasets. Then, We extracted the trade value dataset through API with Python requests and json libraries. We also manually downloaded the GDP value, house pricing and unemployment rate dataset from the websites we mentioned before.

5.2 Data Preprocessing

Next, we preprocessed data with techniques like rescaling and integration. For example, we have a monthly inflation rate dataset while a quarterly GDP dataset, so we had to rescale inflation rate data by calculating the average quarter inflation rate. Besides, we did data integration to merge the feature datasets with the inflation rate dataset respectively by implementing SQL queries in Python with libraries like sqlite3 or directly using csv and pandas libraries in Python.

5.3 Feature Selection

Then, we use Python statsmodels library to construct simple linear regression of the inflation rate and each other factor respectively, and identify the significance according to the t-statistic and p-value.

The following is an example of inflation rate and house price:

We first make the null and alternative assumptions about their relationship:

H0: There is no relationship between the inflation rate and the house price.

H1: There is a significant relationship between the inflation rate and the house price.

```

=====
                        OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.026
Model:                  OLS    Adj. R-squared:       0.022
Method:                 Least Squares    F-statistic:      6.880
Date:                   Sun, 27 Mar 2022    Prob (F-statistic): 0.00923
Time:                   13:04:02    Log-Likelihood:    -455.93
No. Observations:       264    AIC:              915.9
Df Residuals:           262    BIC:              923.0
Df Model:                1
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
const                1.1390        0.430        2.646      0.009        0.292        1.987
x1                   5.679e-06      2.17e-06        2.623      0.009      1.42e-06      9.94e-06
=====
Omnibus:                1.841    Durbin-Watson:      0.114
Prob(Omnibus):          0.398    Jarque-Bera (JB):    1.687
Skew:                   -0.005    Prob(JB):            0.430
Kurtosis:               3.392    Cond. No.            1.02e+06
=====

```

Figure 1. Figure of OLS Regression Results

From Figure 1, the OLS model results show that the p-value is 0.009, which is less than 0.05, the level of significance in our project. so we reject the null hypothesis and conclude the evidence that

there is a significant relationship between inflation rate and house price. We summarized the regression results of all the factors into Table 2.

Table 2. Table of Hypothesis Test Conclusion

Factors	t-statistic	p-value	Whether significant(p-value<=0.05)
Unemployment Rate	2.602	0.004	Significant
Import value	1.565	0.1	Not significant
Export value	2.613	0.01	Significant
GDP	-0.579	0.564	Not significant
House Price	2.623	0.009	Significant
Exchange Rate	2.589	0.05	Significant

5.4 Prediction

Finally, after selecting the significant features, we use various models to fit the features and predict the inflation rate with python library scikit-learn.

We first conduct linear and polynomial regression with each significant feature solely by using sklearn classes like LinearRegression and PolynomialFeatures. We use the function train_test_split to split the dataset into the training set and the testing set. We run each model in different degrees and set the attributes of PolynomialFeatures to default which are interaction_only=False and include_bias=False. We use R-squared value to evaluate the performance of the models, and we summarize the best results of each feature to Table 3.

Table 3. Summary Results of Regression Models

Feature	Polynomial Degree	R-squared value
House Price	3	0.41
Export Value	3	0.29
Unemployment Rate	4	0.42
Interest Rate	2	0.22

From Table 3, we can see that the cubic regression model with the factor house price and the quadratic regression model with unemployment rate have the best prediction results. However, these

two models are still not strongly fitting the data. Thus, we also use multiple regression models to predict the inflation rate. We use the same method but we set the attributes of PolynomialFeatures to `interaction_only=False` and `include_bias=True` this time. Table 4 shows the best result of different feature combinations:

Table 4. Results of Regression with Different Feature Combinations

Features	Polynomial Degree	R-squared value
House price, Export value	1	0.30
House price, Unemployment rate	3	0.61
House price, Interest rate	4	0.56
Export value, Unemployment rate	2	0.28
Export value, Interest rate	1	0.32
Unemployment rate, Interest rate	4	0.56
House price, Export value, Unemployment rate	1	0.29
Export value, Unemployment rate, Interest rate	1	0.31
House price, Unemployment rate, Interest rate	3	0.75
House price, Export value, Interest rate	1	0.31
House price, Export value, Unemployment rate, Interest rate	1	0.32

We can see that the best model is the cubic regression model with the features house price, unemployment rate and interest rate, of which the R-squared value achieves 0.75. The prediction result is still not perfect and it is reasonable, since the inflation rate is a complicated index depending on many factors. The features we chose were just part of the influential factors. Besides, as the time goes by, there were many significant events in the United States, so the inflation rate could also be affected by the big events like pandemics, financial crisis, responses of the government, etc.

6 Observation

As we mentioned before, some big events like financial crises and pandemics can have a dramatic influence on the inflation rate, so we conducted further research about the reason for the abnormal inflation rate and how the government reacts during these special events.

6.1 Specific Reasons for Abnormal Inflation Rate during Big Events

According to Figure 2, we found two dramatic decreases. Thus, we will mainly discuss those two time periods: Nov. 2014 to Sept. 2015 and Jan. 2020 to Dec. 2021.

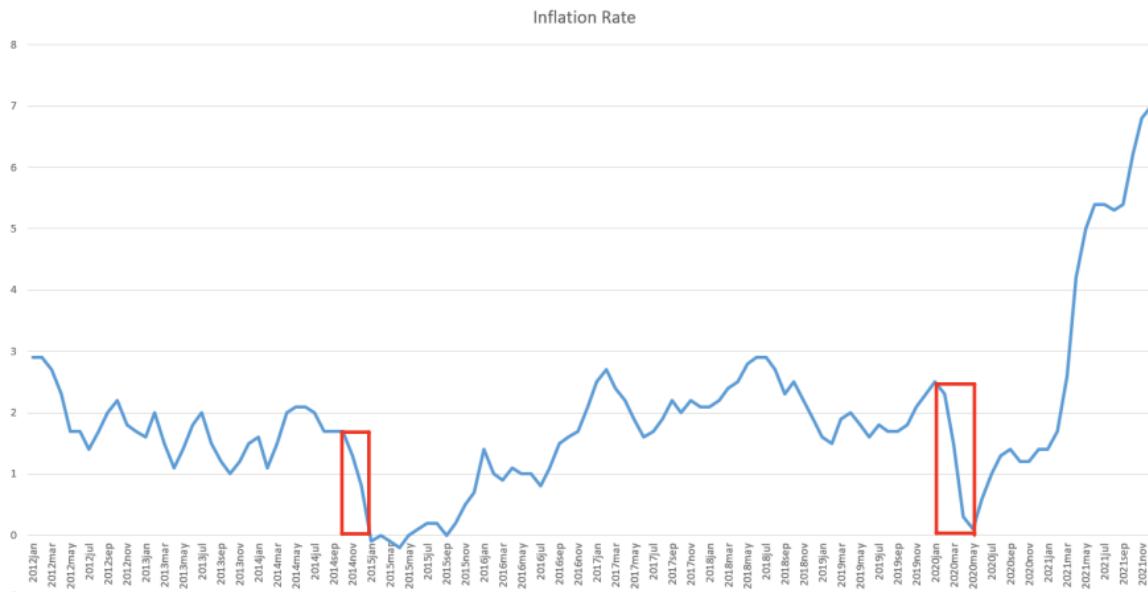


Figure 2. Figure of 10 Years' Inflation Rate

6.1.1 Period 1: Nov. 2014 to Sept. 2015

According to Table 2, inflation rate is significantly related to the Unemployment rate, Export value, House price, and Exchange rate. Thus, we thought that these factors could lead to the huge fall in inflation rate. In order to test it, we extract the data from each time period and conduct the linear regression models with each feature. The result is that there is no significant relationship between these factors and the inflation rate during this time period. To find the main influential factor we did some research and found out that in 2014 the states experienced the Commodity Price Shocks. Also, in that year, the crude oil price fell sharply. Thus, we take it into consideration and conduct a quadratic model between crude oil prices and the inflation rate. The R-square is 0.69578 which is above 0.5 and it is acceptable. Thus, the oil price is also one major factor contributing to the inflation rate during this time. Therefore, in 2014, the major factor that caused the inflation rate to decrease is crude oil price instead of the factors we considered before.

6.1.2 Period 2: Jan. 2020 to May. 2020

In Jan. 2020, the coronavirus started to influence the world economy, at this time, the inflation rate was 2.5%, while in May 2020, the inflation rate became 0.1%. To study the reason behind this, we

conduct the linear regression models for that specific five months period with the significant features we selected. As a result, we found out that the unemployment rate is highly correlated with the inflation rate as the R-squared value is 0.8997. Then we conducted the quadratic regression model with the unemployment rate that has a R-squared value of 0.9773. Since the R-squared value of both models is very close to 1, we conclude that the unemployment rate is the main reason for the decreasing of the inflation rate from January to May in 2020.

6.2 Presidents' Responses to the Abnormal Inflation Rate

When the inflation rate is too high or too low, economic conditions and people's daily life tend to be influenced dramatically. So, the president and the ruling party will take some steps to deal with the situation.

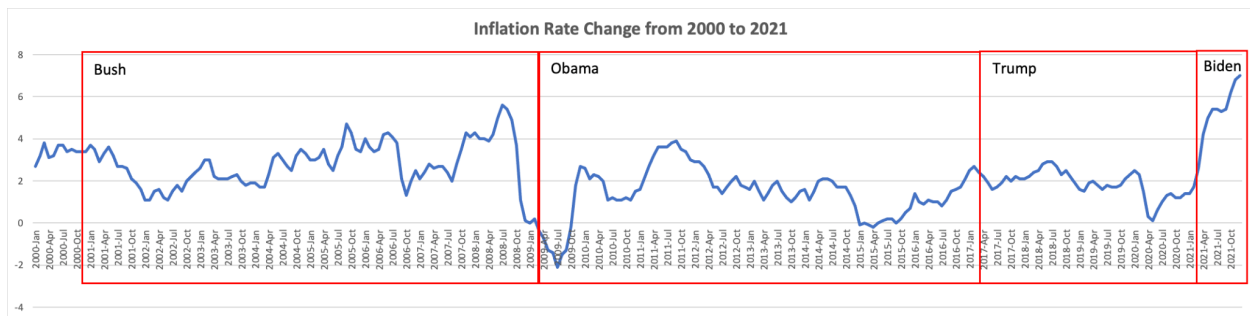


Figure 3. Figure of 22 Years' Inflation Rate

The first dramatic decrease in inflation rate took place from Jul. 2008 to Jul. 2009, which was related to the financial crisis. This time also witnessed the change of the president. President Bush's response to the financial crisis includes bailing out those damaged financial institutions, providing bank debt guarantees[2] to improve the confidence of the banking system and so on. Presidents' Bush paid much attention to saving banks, which was beneficial to the mortgage rate, interest rate and some other related factors. These policies could influence the inflation rate indirectly by having a direct effect on the factors we researched before. Besides, President Obama continued bailing out companies and banks, and also increased private sector jobs[3] to help improve the unemployment rate, stimulate consumptions and in turn make inflation rate increase to the normal.

The second significant increase in inflation rate happened after the eruption of Covid-19, during which time President Biden took office. In his plans to control the high inflation rate, he mentioned that he hoped car factories could produce vehicles in the US instead of depending on the foreign supply

chains[4]. In this way, we could see that he hoped to reduce import trade value and raise wages in car industries so as to decrease the contributions of car industries to the high inflation rate.

7 Conclusion

For this project, we conducted models based on significant economic factors to predict the inflation rate. Then, we also made further explorations of reasons for inflation rate change during specific times and investigated presidents' responses. We could conclude our findings as below:

1. The best prediction model of the inflation rate is the cubic model based on house price, unemployment rate and interest rate, which has an R-squared value of 0.75.
2. Inflation rate and unemployment have a stable and inverse relationship.
3. The higher the inflation rate, the more interest rates are likely to rise.
4. The main reason for the moderate performance is that the inflation rate is a complicated index, which has much unexplained variance and cannot be depicted with only a few factors. For instance, Demand-supply relationship could also be considered because high demand will cause the rise in the prices, which will also lead to a rise in the inflation rate.
5. The contributions of economic factors to the abnormal inflation rate are different between long run and short specific time, which teaches us when a general model achieved from long run datasets behaves terribly on a short time dataset, we need to analyze based on the specific situations.

References:

- [1] Mohseninasab, Elaheh. "Investigating the Relationship between Inflation Rates, Exchange Rates, Interest Rates and Stock Prices (Case Study Mapna and Takinco Companies)."
- [2] Wikipedia contributors. (2022, March 10). Economic policy of the George W. Bush administration. https://en.wikipedia.org/wiki/Economic_policy_of_the_George_W._Bush_administration
- [3] Wikipedia contributors. (2022, April 10). Economic policy of the Barack Obama administration. https://en.wikipedia.org/wiki/Economic_policy_of_the_Barack_Obama_administration
- [4] Anneken Tappe. CNN Business. (2022, March 2). Here's how President Biden wants to fight inflation. <https://www.cnn.com/2022/03/01/economy/us-inflation-biden-state-of-the-union/index.html>