JSC370 Final Report

Yongzhen Huang

4/18/2022

Introduction

We currently live in a special time in history. While we are still dealing with an ongoing global pandemic, the economy is also faced with high inflation. High inflation often leads to higher cost of living and decreased living standard.

This project aims to answer how inflation is tied to an economy and the economic growth, and how one may predict the future inflation given current state of the economy. In particular, the following questions will be answered

1. How does inflation affect the overall economy and economic growth (e.g. GDP)?
2. How does inflation affect the standard of living i.e. CPI?
3. Given the current state of the economy, how may one predict the future inflation?

Methods

Data Source

The main source of data is from Organisation for Economic Co-operation and Development (OECD). Data from OECD are downloaded as annual data in CSV files and includes the following information.

Category	Details
GDP	$USD millions, per capita
Population	# million persons, growth, % working age
Employment Rate	% of working age population
Inflation	annual %
Prices	housing (rent) and share prices
Interest Rate	government short-term interest rate
Household stats	Savings and spending

Stock data are mainly extracted from Alpha Vantage and Finnhub.

The API for Alpha Vantage and Finnhub are only available in Python and other non-R languages, and therefore the data is extracted in the code here

Data Cleaning

Overall, the main tools used for the project are R and Python.

Python is focused towards data collection from Finnhub and Alpha Vantage for market, mainly major indices (DJIA, NASAQ, and S&P500) with sector ETFs.

R is used for the data cleaning, analytics, and constructing interactive plots

- `tidylr` for piping
- `dplyr` for table manipulations
- `data.tables` for more table manipulations
- `ggplot` and `plotly` for visualization

Once data is read in from CSV to dataframe (one file per metric), they are filtered to contain only the information described above. Then they are all merged using `merge` based on country code and year. From here, any NA are dropped.

For the prediction problem, the target is next year's inflation number for each country. This data is first extracted from consolidated data to update the year and then merged back as target. Finally, the train and test data are split in 70% and 30%, respectively, in chronological order i.e. the test data occurs strictly after the training data by country.

Results

Predictive Models

Decision Tree

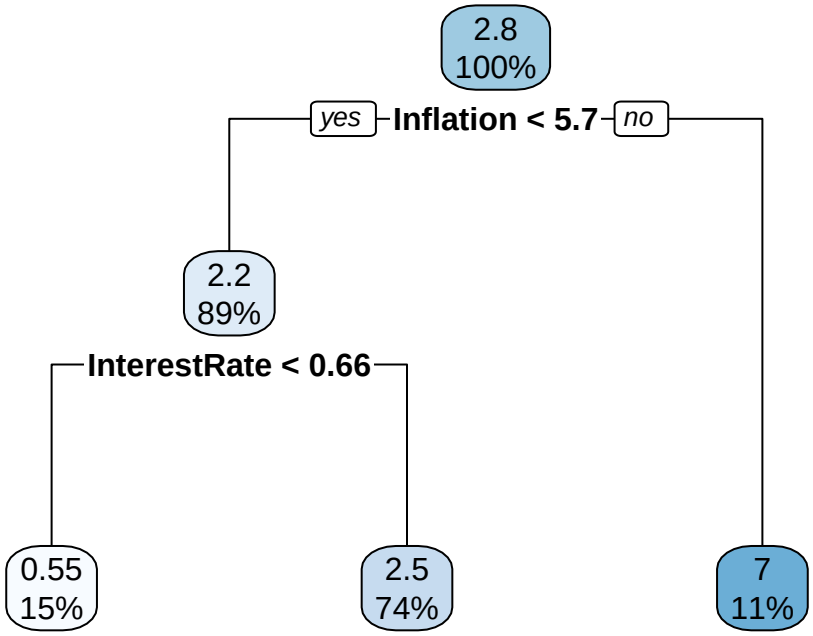


Figure 1: Pruned Decision Tree

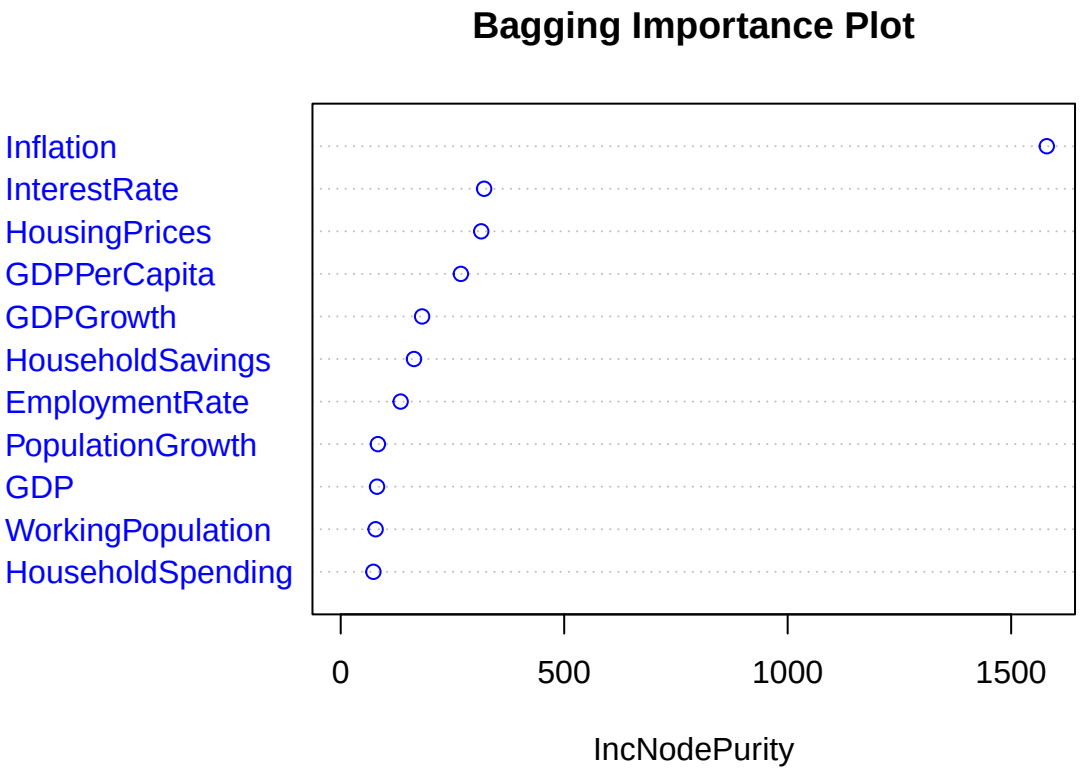


Figure 2: Bagging Variable Importance

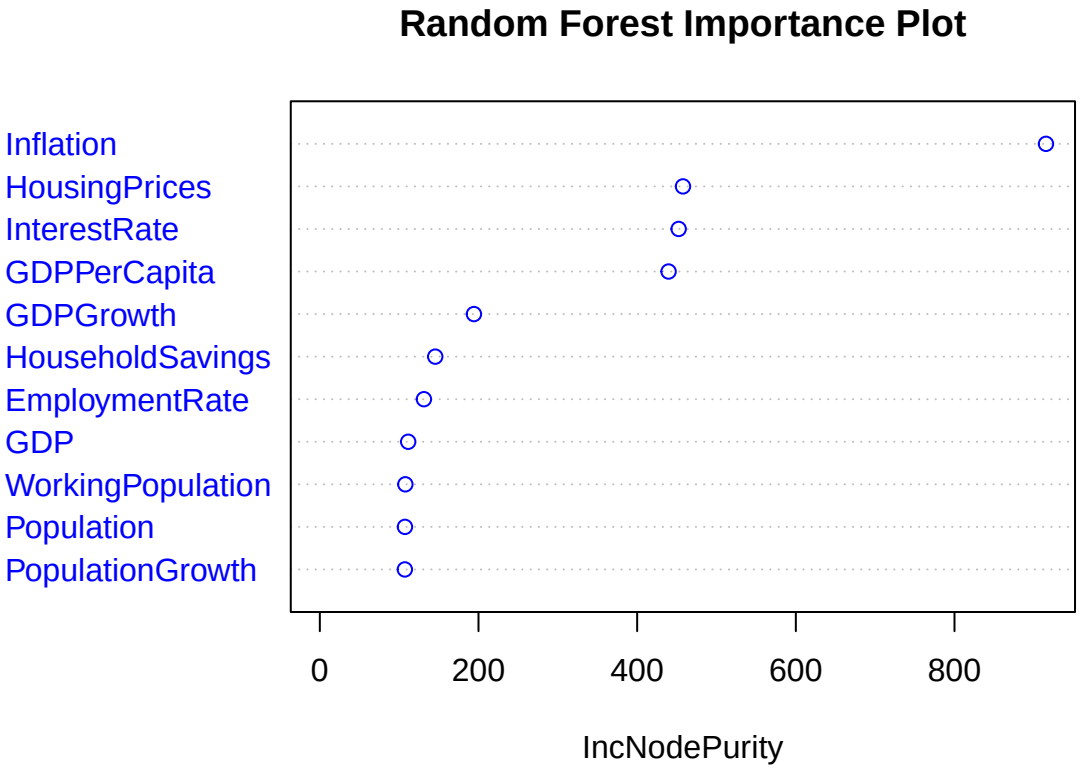


Figure 3: Random Forest Variable Importance

Bagging

Random Forest

GBM

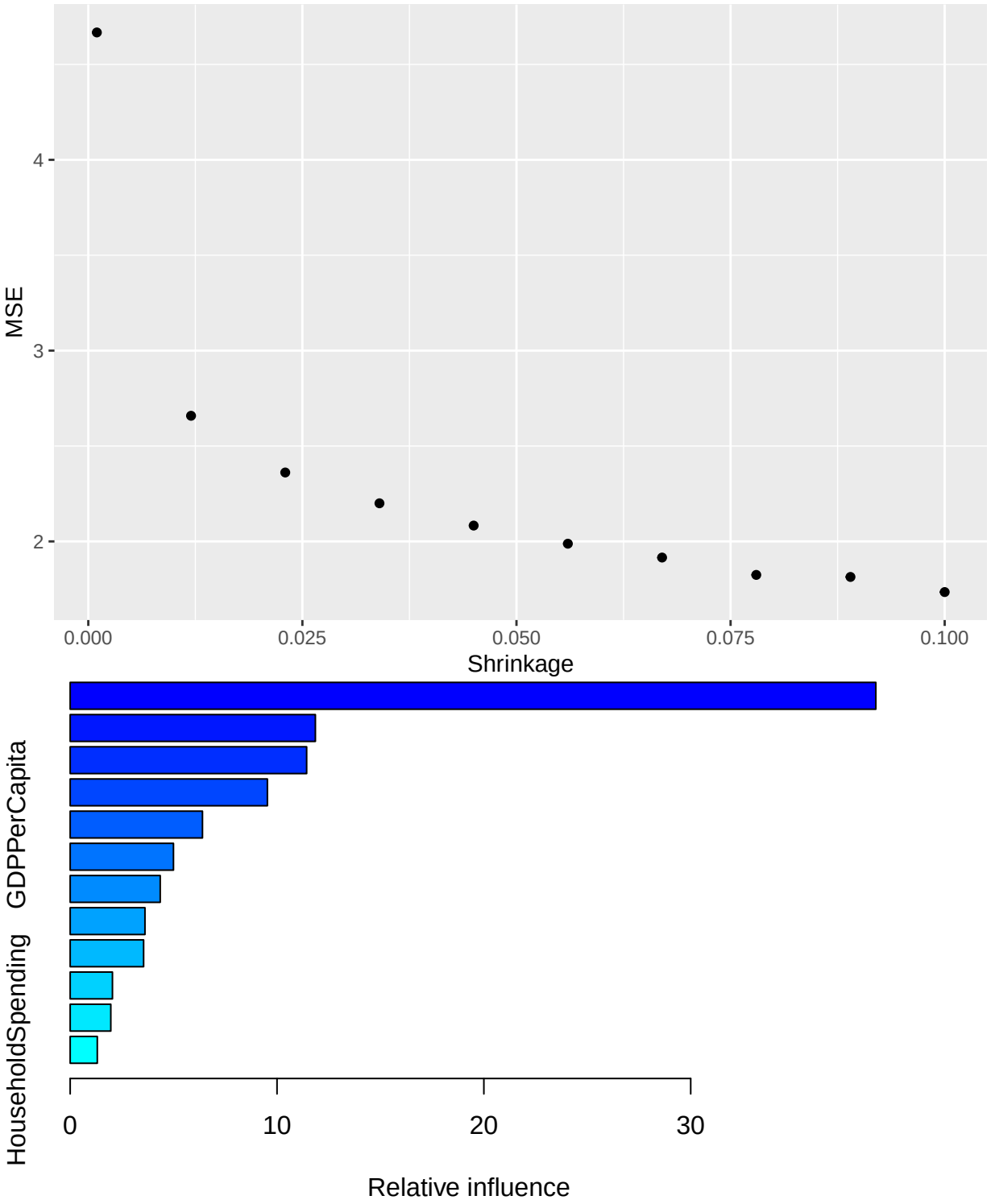


Table 2: Variable Importance

Feature	Relative.Influence
Inflation	38.950179
GDPGrowth	11.853939
InterestRate	11.432863
HousingPrices	9.534734
GDPPerCapita	6.395966
PopulationGrowth	4.993792
WorkingPopulation	4.353021
GDP	3.616138
HouseholdSavings	3.550575
EmploymentRate	2.043564
Population	1.964713
HouseholdSpending	1.310516

XGBoost

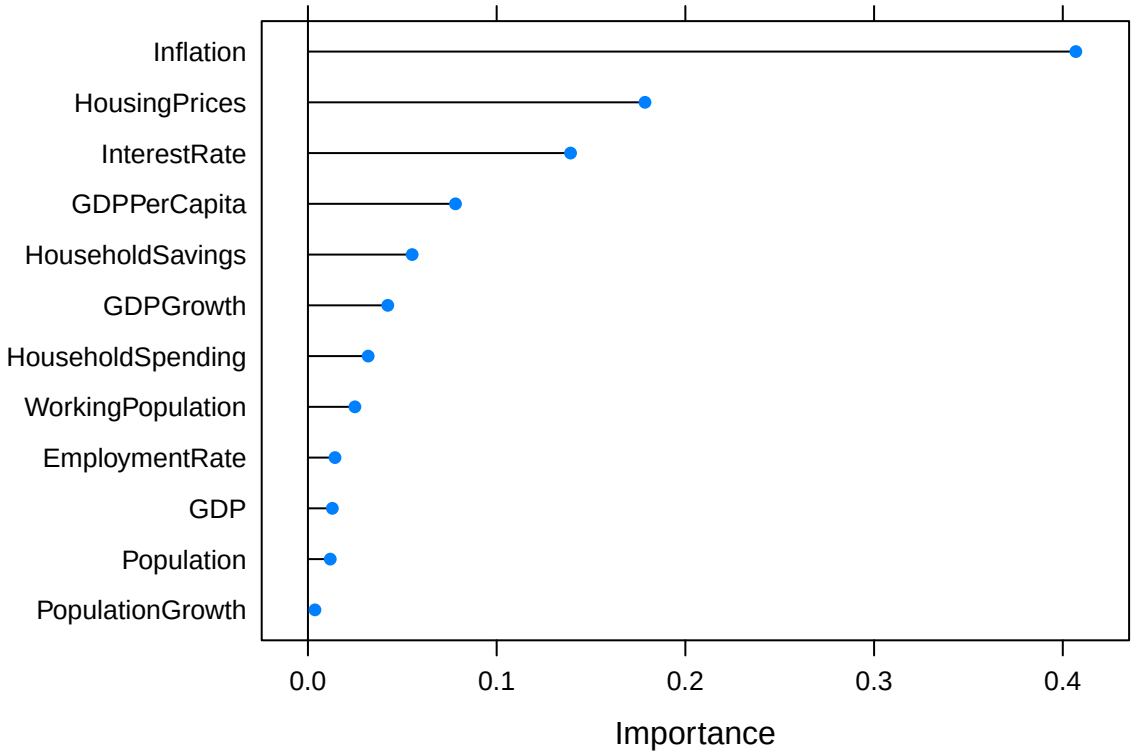


Figure 4: XGBoost Variable Importance

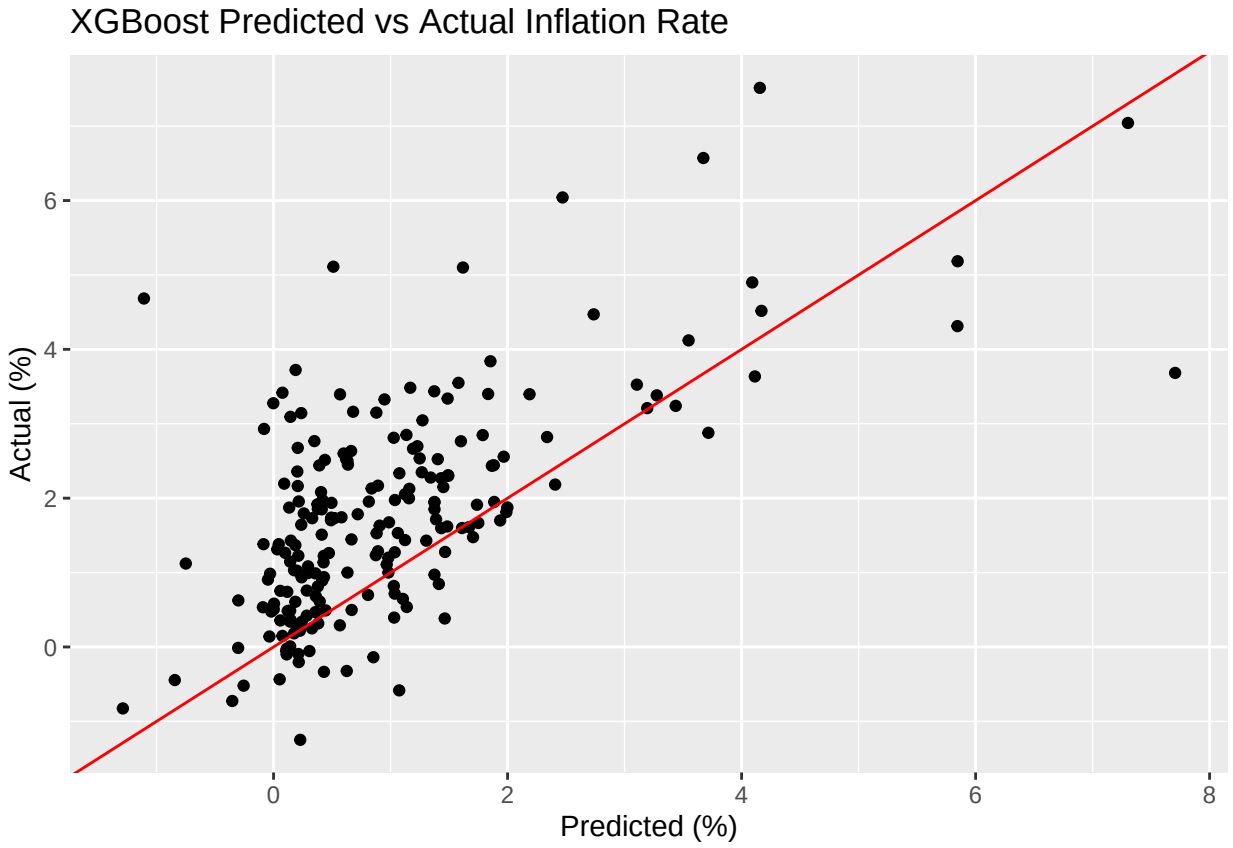
Model Performance

Table 3: Methods and RMSE

method	RMSE
Regression Tree	2.1236

method	RMSE
Bagging	2.1395
Random Forest	1.9907
Boosting	3.5656
XGBoost	1.9489

The following graph shows the performance of the XGBoost by plotting the predicted inflation rate against the actual inflation rate. The red line is the 45-degree line which indicates the case of perfect prediction where all points should lie on the line. From the graph, we can see there is a linear relationship between the predicted and actual inflation. The implication is that the XGBoost model is capable to predict future inflation to some extent. This is important because it can help policy makers to better understand the future state of economy and adjust policies accordingly.



Conclusions and Summary