Uncovering the Patterns of U.S. Economic Dynamics: Using R Programming TO assess Recession RIsk and Compare Historical Cycle

jacob.thejus@stud.hs-fresenius.de

HS-Fresenius: Data Science for Business

Table of Contents

Abstract 2
Introduction 3
Data 3
Objective 4
Framework 4
Data preparation 5
For the seven Economic Indicators 5
For Yield Curve 7
Part 01: Visualizing US Economic Trends through Time-Series Graphs 8
Part 02: Verifying Economic Theories and Models 15
Part 03: Determining the risk of recession 23
Using Yield Curve 23
Using Machine Algorithm 25
Results and Analysis 27
A. Line graphs for various economic indicators 28
B. Verification of economic theories and models 34
C. Recession risk assessment 37
Using Yield Curve 37
Using Machine Algorithm 39
Conclusion 40
Reference 41

Abstract

This project aims to assess the risk of a recession in the U.S. economy between 1975 and 2021 based on seven economic indicators, including the House Price Index, Stock Price Index, Consumer Price Index, GDP, Unemployment Rate, Disposable Income, and Mortgage Rate. The study first visualizes US economic trends through time-series graphs and verifies economic theories and models using US economic data. The second part of the study compares the results of two approaches used to determine the risk of a recession: the yield curve and a machine learning algorithm. The machine learning algorithm, which uses the Random Forest algorithm in the "caret" and "randomForest" packages in R, was far from predicting a recession effectively. The conclusion drawn from the results and analysis is that although the machine learning algorithm can provide valuable insights, it is not an effective method for predicting a recession. But open a new opportunity for recession risk assessment. The study highlights the importance of combining multiple techniques and considering various economic indicators when assessing the risk of a recession.

Introduction

In this project, I aim to uncover the patterns of the U.S. economic dynamics by using R programming. The U.S. economy is one of the world's largest and most influential economies, and understanding its patterns and dynamics is crucial for economists, policymakers, and investors. In this project, I will use R programming to examine and analyze several economic indicators and use that information to determine the risk of recession and compare historical cycles. The background of the project lies in the fact that the U.S. economy has experienced several recessions in the past and understanding the factors that contribute to such recessions and being able to predict them can be highly beneficial. The problem statement of this project is to determine the risk of recession in the U.S. economy based on seven economic indicators and compare that with historical cycles.

The purpose and objectives of this project are to use R programming to plot line graphs for various economic indicators such as House Price Index, Stock Price Index, Consumer Price Index, Unemployment Rate, GDP, Mortgage Rate, and Disposable Income and to verify economic theories and models such as Keynesian Economics, Phillips Curve, Okun's Law, Fischer Effect, IS-LM model, Stock Flow Model, and Housing Pricing Cycle. Additionally, this project aims to determine the risk of recession in the U.S. economy using the yield curve and machine algorithms. The project's scope is limited to the U.S. economy between 1975 to 2021, and the data sources will be yield curve data from FRED and seven economic indicators. The limitations of the project are that it is limited to the data sources used, and the project results are subject to the limitations and accuracy of the data. Overall, this project will help me understand the U.S. economic dynamics and the patterns hidden within the data.

Data

I obtained my data set from Kaggle.com, a trustworthy community platform for data science and machine learning. Seven economic indicators are included in the data set, each with a different primary data source. The House Price Index (HPI) was sourced from the Federal Housing Finance Agency (FHFA), the Stock Price Index from the S&P 500, the Consumer Price Index (CPI) from the Bureau of Labor Statistics (BLS), the Gross Domestic Product (GDP) from the Bureau of Economic Analysis (BEA), the Unemployment Rate from the BLS, the Disposable Income from the BEA, and the Mortgage Rate from the Freddie Mac Primary Mortgage Market Survey. The data set offers a comprehensive look at the U.S. economy over 46 years and is a valuable resource for analyzing economic patterns and assessing recession risk. Additionally, I've added information from the yield curve and recession CSV file from FRED, which is crucial for understanding the dynamics of the US economy. The importance of this recently added data source is on par with that of the others, and it will be considered in my research and conclusions.

Objective

As a student, I am embarking on a project to uncover the patterns of the U.S. economic dynamics using R programming. There are three aspects to the study, and my focus is on analyzing different economic indicators to gauge the likelihood of a recession and evaluate past cycles.

Plotting line graphs for seven economic indicators, including the house price index, stock price index, consumer price index, unemployment rate, GDP, mortgage rate, and disposable income, will make up the first portion of my project. To accomplish this, I'll be utilizing the ggplot library in R. Visualizing the historical trends of the economic indicators is the goal of this step.

I will be verifying several economic theories and models, including Keynesian Economics, the Phillips Curve, Okun's Law, the Fischer Effect, the IS-LM model, the Stock Flow Model, and the Housing Pricing Cycle, in the second half of the research. As a result, I will better comprehend the connections between various economic indicators and how they affect the economy.

The assessment of the recession risk is the last piece of my project. In order to accomplish this goal, I'll take two different approaches: machine learning techniques and the yield curve. To start, I'll plot the yield curve against time and analyze kernel density plots to see how the yield rate and recession are related. The data for the yield curve will be obtained from the Federal Reserve Economic Data (FRED).

In the sd method, I'll employ machine learning algorithms to estimate the likelihood of a U.S. economic recession between 1975 and 2021. I will use seven economic indicators to forecast the economy's status (the House Price Index, Stock Price Index, Consumer Price Index, GDP, Unemployment Rate, Disposable Income, and Mortgage Rate). The code will utilize the Random Forest algorithm in the "caret" and "randomForest" packages.

Overall, this project will enable me to comprehend the U.S. economy and the reasons that cause recessions better. In addition, I want to get better grasp how economic indicators and models can be utilized to forecast the risk of recession by the end of this project.

Framework

The project framework is divided into three main parts, each aimed at providing a comprehensive understanding of the U.S. economy and its patterns.

I will use the ggplot library in R to plot line graphs for various economic indicators in the first part. This will include plotting the House Price Index, Stock Price Index, Consumer Price Index, Unemployment Rate, GDP, Mortgage Rate, and Disposable Income. This step is crucial as it visually represents the data and helps identify trends and patterns.

The second part of the project involves verifying different economic theories and models such as Keynesian Economics, Phillips Curve, Okun's Law, Fischer Effect, the IS-LM model,

Stock Flow Model, and Housing Pricing Cycle. This step is essential as it provides a theoretical foundation for understanding the economy and helps develop a comprehensive data analysis.

In the third and final part, I will determine the risk of recession in the U.S. economy. This involves two main steps. The first step is to use the Yield curve by plotting the yield curve against time and using kernel density plots to determine the relationship between yield rate and recession. This data will be sourced from the Federal Reserve Economic Data (FRED) (St. Louis Federal Reserve Bank, n.d.). The second step is to use a machine learning algorithm to determine the risk of recession. This will be done by utilizing the Random Forest algorithm in the "caret" and "randomForest" packages in R to make predictions about the state of the economy based on seven economic indicators.

In conclusion, this project provides a framework for understanding the patterns of the U.S. economy and assessing the risk of recession. By utilizing R programming, economic theories and models, and machine learning algorithms, I aim to analyze the U.S. economy and its patterns comprehensively.

Data preparation

For the seven Economic Indicator

```
getwd()

## [1] "C:/Users/theju/OneDrive/Documents"

#Calling required Library functions
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.2 ──
## ✓ ggplot2 3.4.0     ✓ purrr   1.0.1
## ✓ tibble  3.1.8     ✓ dplyr   1.1.0
## ✓ tidyr   1.3.0     ✓ stringr 1.5.0
## ✓ readr   2.1.3     ✓ forcats 1.0.0
## ── Conflicts ────────────────────────────────────────
tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()

library(dplyr)
library(tidyr)
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
#Importing the Dataframe from storage using read.csv
data01 <- read.csv("AMF1.csv")
head(data01)
```

```
##   Year House_Price_Index Stock_Price_Index Consumer_Price_Index Population
## 1 1975           61.0900          67.14653             65.30488    0.98599
## 2 1976           65.5250          79.96264             69.05653    0.95022
## 3 1977           73.4350          78.82540             73.54636    1.00577
## 4 1978           83.7450          78.84679             79.15866    1.05957
## 5 1979           95.1325          85.63207             88.06755    1.10358
## 6 1980          102.6675         100.00000            100.00000    0.95959
##   Unemployment_Rate Real_GDP Mortgage_Rate Real_Disposable_Income
## 1           8.46667 5648.462       9.04712                  19908
## 2           7.71667 5952.809       8.86585                  20346
## 3           7.06667 6228.076       8.84519                  20780
## 4           6.06667 6572.819       9.64173                  21497
## 5           5.83333 6780.924      11.20365                  21672
## 6           7.14167 6763.514      13.74212                  21584
```

```r
class(data01)
```

```
## [1] "data.frame"
```

```r
view(data01)
summary(data01)
```

```
##       Year      House_Price_Index Stock_Price_Index Consumer_Price_Index
##  Min.   :1975   Min.   : 61.09    Min.   :  67.15   Min.   : 65.3
##  1st Qu.:1986   1st Qu.:140.79    1st Qu.: 209.90   1st Qu.:135.4
##  Median :1998   Median :211.46    Median : 756.56   Median :197.8
##  Mean   :1998   Mean   :240.15    Mean   : 743.13   Mean   :198.6
##  3rd Qu.:2010   3rd Qu.:339.35    3rd Qu.:1114.17   3rd Qu.:262.9
##  Max.   :2021   Max.   :523.26    Max.   :2255.84   Max.   :328.8
##    Population     Unemployment_Rate    Real_GDP     Mortgage_Rate
##  Min.   :0.1184   Min.   :3.667     Min.   : 5648   Min.   : 2.958
##  1st Qu.:0.8627   1st Qu.:5.167     1st Qu.: 8374   1st Qu.: 4.863
##  Median :0.9459   Median :5.992     Median :12046   Median : 7.440
##  Mean   :0.9352   Mean   :6.310     Mean   :12140   Mean   : 7.781
##  3rd Qu.:1.0816   3rd Qu.:7.442     3rd Qu.:15646   3rd Qu.: 9.886
##  Max.   :1.3869   Max.   :9.708     Max.   :19427   Max.   :16.642
##  Real_Disposable_Income
##  Min.   :19908
##  1st Qu.:25433
##  Median :31712
```

```
##  Mean   :32041
##  3rd Qu.:38235
##  Max.   :48219
```

By importing the data from an external source—in this case, csv file—and loading the necessary libraries, I am creating the environment for data preparation, processing, and visualization. After importing the data, I used the 'head' and 'class' functions to inspect the first few rows and determine the data type, respectively. A summary of the imported data was also obtained using the "summary" function.

```
# Define the limits for each indicator
hp_lim <- c(60, 600)
sp_lim <- c(60, 2300)
cpi_lim <- c(60, 330)
ur_lim <- c(0, 10)
gdp_lim <- c(5000, 20000)
mr_lim <- c(0, 17)
rdi_lim <- c(19000, 50000)
```

I have established the limits for each economic indicator, including the house price index (hp lim), stock price index (sp lim), consumer price index (cpi lim), unemployment rate (ur lim), gross domestic product (gdp lim), mortgage rate (mr lim), and disposable income (rdi lim), in order to prepare the data for visualization. These constraints will be applied to visualize the economic indicators in the project's subsequent stages.

For Yield Curve

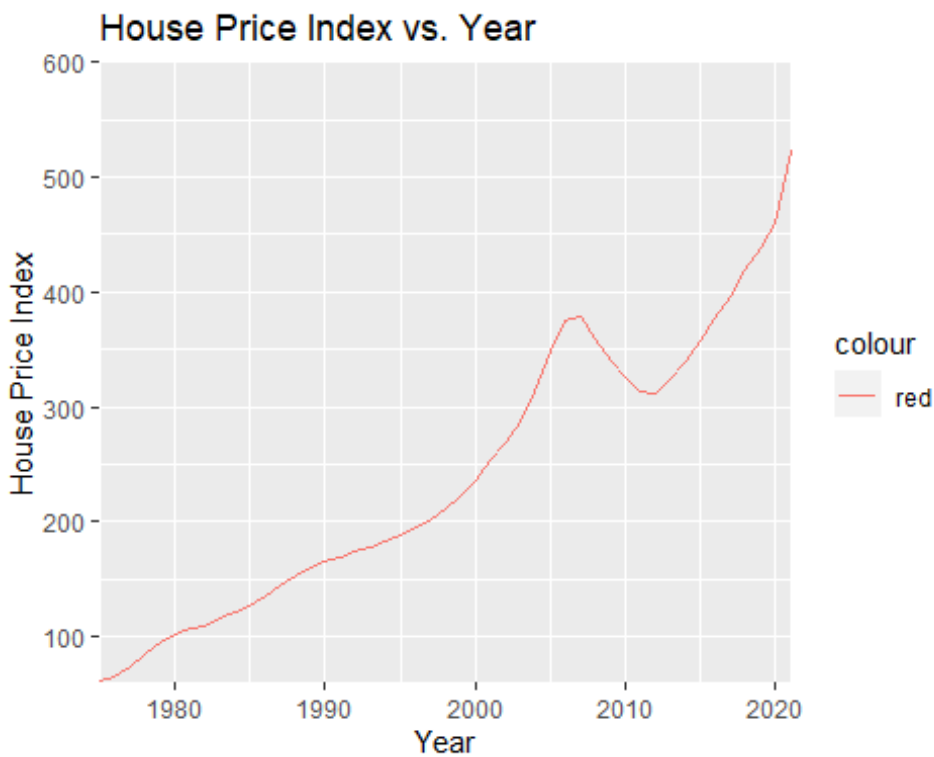
```
# Load required libraries
library(dplyr)
library(tidyr)
library(ggplot2)
library(zoo)

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

# Load yield_curve.csv and recession_dates
yield_curve <- read.csv("yield_curve.csv")
recession_dates <- read.csv("recession_dates.csv")
# Convert date columns to Date class
yield_curve$date <- as.Date(yield_curve$date, format = "%m/%d/%Y")
yield_curve$yield_rate <- na.approx(yield_curve$yield_rate)

## Warning in xy.coords(x, y, setLab = FALSE): NAs introduced by coercion
```

This code is a part of the project's third portion, which focuses on using the yield curve to assess the risk of a recession. The code's first few lines load the necessary libraries,

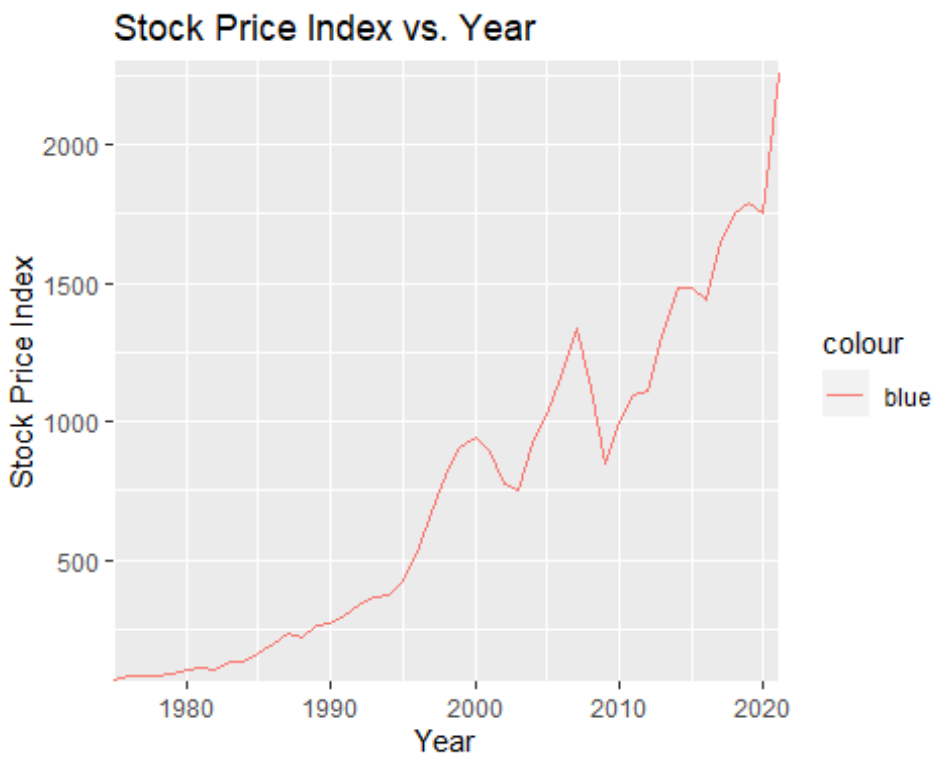
including ggplot2, dplyr, tidyr, and zoo. The read.csv function loads the yield curve and recession date data. Using the as.Date function and the format option, the date column in both datasets is transformed into a Date class. The approximate missing values are filled in the yield curve data's yield rate column using the na.approx function. The data processing will be utilized to plot the yield curve and evaluate how the yield rate relates to the recession.

Part 01: Visualizing US Economic Trends through Time-Series Graphs

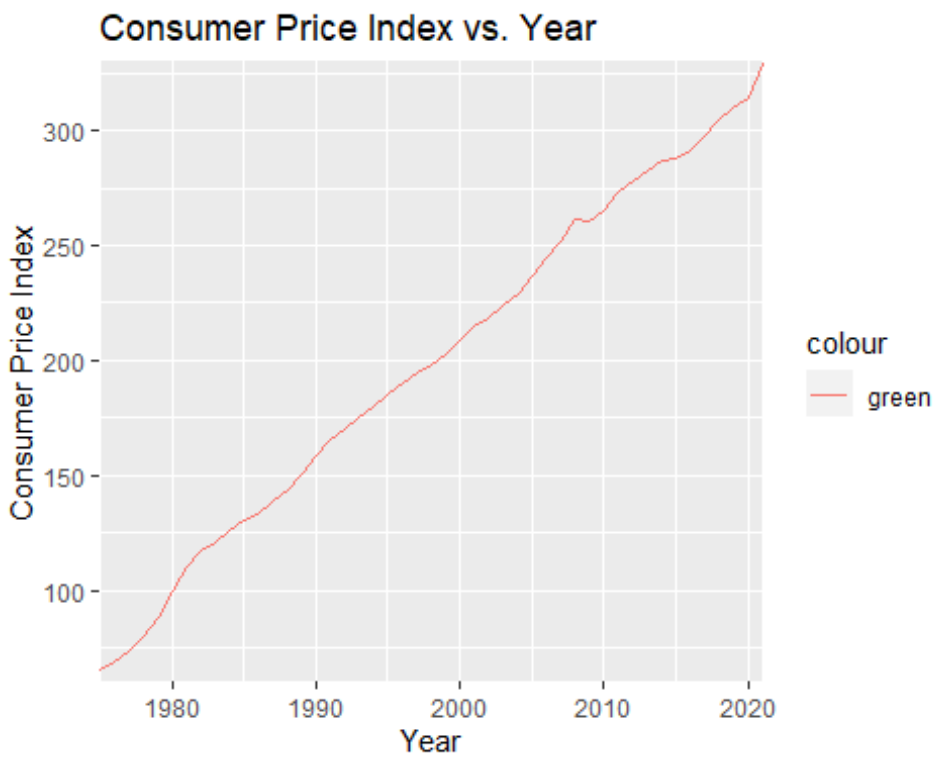
```
#PART-01: PLOTTIING GRAPH FOR ALL THE ECONOMIC INDICATORS
# Create a line graph for House Price Index
ggplot(data = data01, aes(x = Year, y = House_Price_Index, color = "red")) +
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = hp_lim, expand = c(0,0)) +
  labs(x = "Year", y = "House Price Index") +
  ggtitle("House Price Index vs. Year")
```



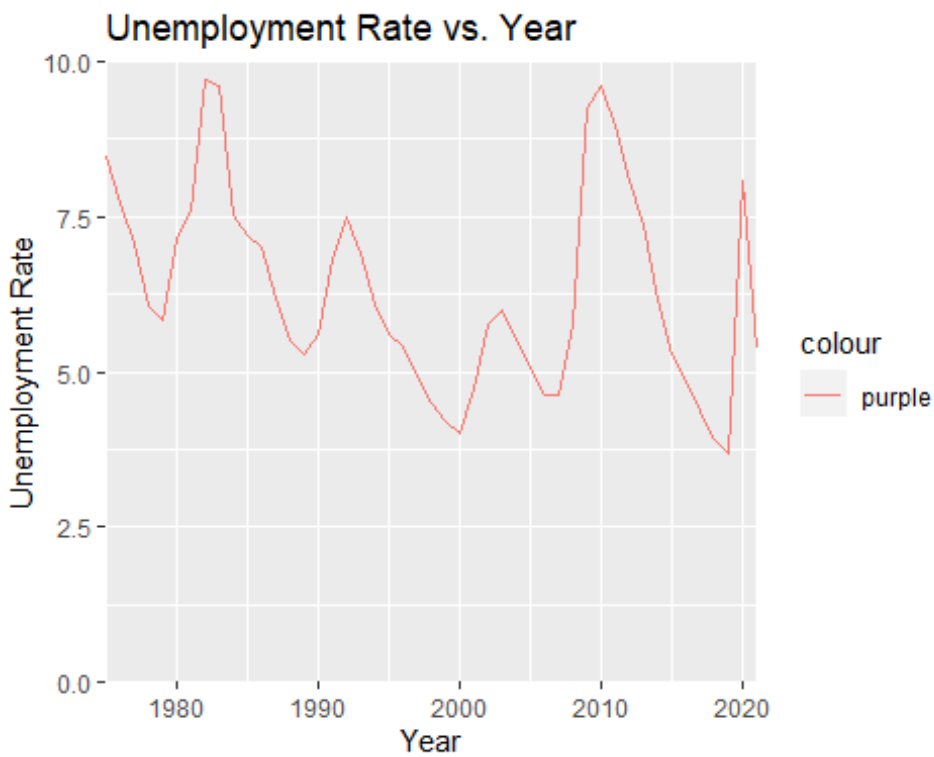

```
# Create a line graph for Stock Price Index
ggplot(data = data01, aes(x = Year, y = Stock_Price_Index, color = "blue")) +
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = sp_lim, expand = c(0,0)) +
  labs(x = "Year", y = "Stock Price Index") +
  ggtitle("Stock Price Index vs. Year")
```



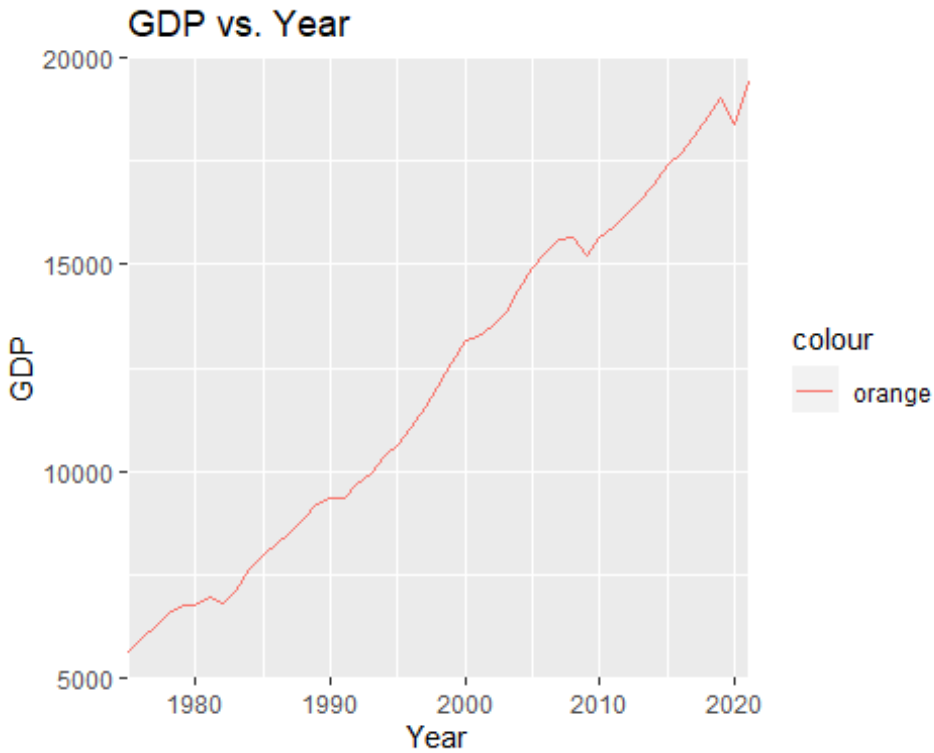

```
# Create a line graph for Consumer Price Index
ggplot(data = data01, aes(x = Year, y = Consumer_Price_Index, color =
"green")) +
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = cpi_lim, expand = c(0,0)) +
  labs(x = "Year", y = "Consumer Price Index") +
  ggtitle("Consumer Price Index vs. Year")
```



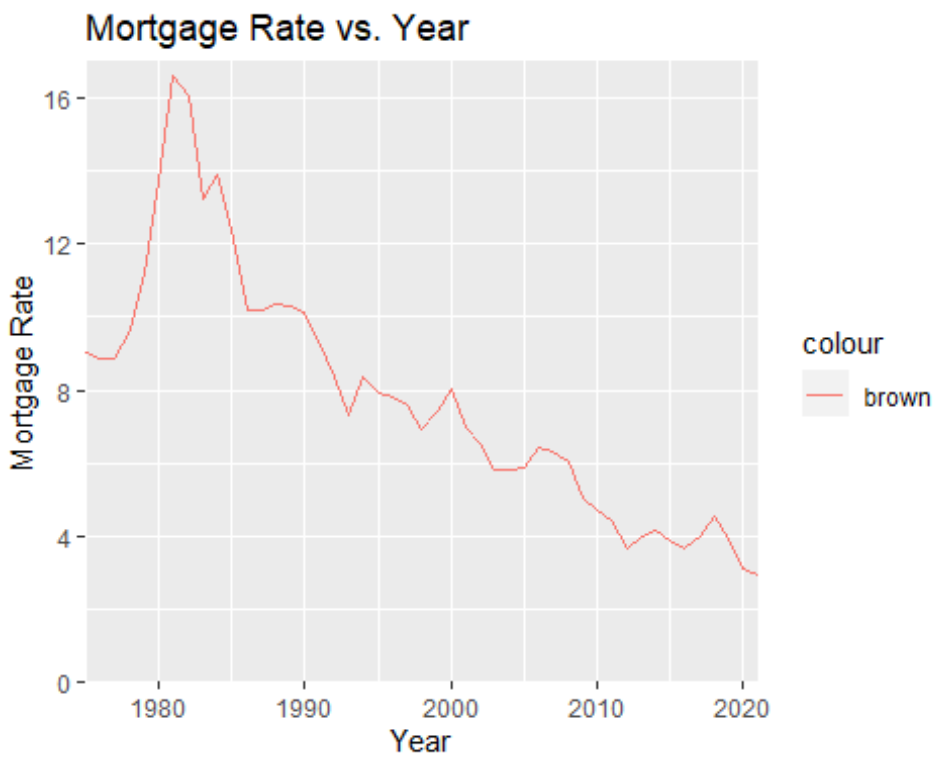

```
# Create a line graph for Unemployment Rate
ggplot(data = data01, aes(x = Year, y = Unemployment_Rate, color = "purple"))
+
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = ur_lim, expand = c(0,0)) +
  labs(x = "Year", y = "Unemployment Rate") +
  ggtitle("Unemployment Rate vs. Year")
```



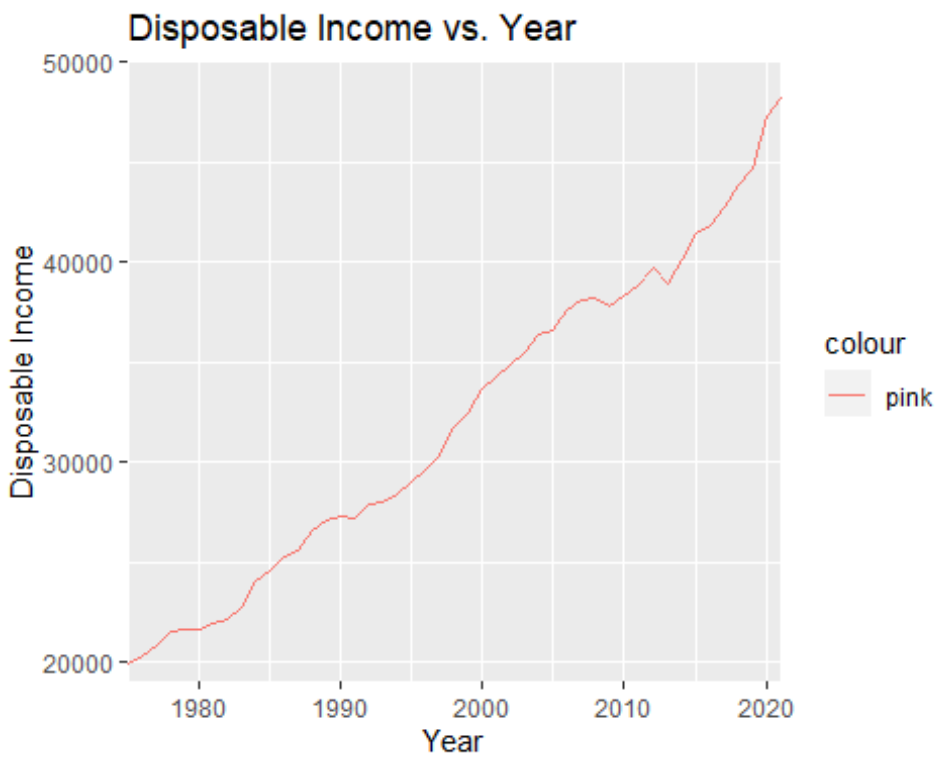

```r
# Create a line graph for GDP
ggplot(data = data01, aes(x = Year, y = Real_GDP, color = "orange")) +
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = gdp_lim, expand = c(0,0)) +
  labs(x = "Year", y = "GDP") +
  ggtitle("GDP vs. Year")
```




```r
# Create a line graph for Mortgage Rate
ggplot(data = data01, aes(x = Year, y = Mortgage_Rate, color = "brown")) +
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = mr_lim, expand = c(0,0)) +
  labs(x = "Year", y = "Mortgage Rate") +
  ggtitle("Mortgage Rate vs. Year")
```




```
#Create a line graph for Disposable Income
ggplot(data = data01, aes(x = Year, y = Real_Disposable_Income, color =
"pink")) +
  geom_line() +
  scale_x_continuous(limits = c(1975, 2021), expand = c(0,0)) +
  scale_y_continuous(limits = rdi_lim, expand = c(0,0)) +
  labs(x = "Year", y = "Disposable Income") +
  ggtitle("Disposable Income vs. Year")
```



The codes aim to plot line graphs for various economic indicators, including the House Price Index, Stock Price Index, Consumer Price Index, Unemployment Rate, GDP, Mortgage Rate, and Disposable Income. The line graphs are generated using the ggplot library in R.

In each plot, the x-axis represents the year, and the y-axis represents the value of the economic indicator. The ggplot function is used to specify the data source (data01), the mapping of the variables to the x and y axes (aes), and the line color. The geom_line function is used to draw the line graph.

The scale_x_continuous and scale_y_continuous functions are used to set the x and y axes' limits and control the axes' expansion. The labs function is used to label the x and y axes, and the ggtitle function is used to set the title of the plot.

The followings are the arguments and parameters used in the functions:

- data = data01: specifies the data source for the plot
- aes(x = Year, y = [Economic Indicator], color = "[color]"): maps the year variable to the x-axis and the corresponding economic indicator to the y-axis and sets the color of the line
- limits = c(1975, 2021), expand = c(0,0): sets the limits of the x-axis and controls the expansion of the axis
- limits = [Economic Indicator limit], expand = c(0,0): sets the limits of the y-axis and controls the expansion of the axis

- x = “Year”, y = “[Economic Indicator]”: labels the x and y axes
- ggtitle(“[Economic Indicator] vs. Year”): sets the title of the plot

The resulting plots show the trends and patterns in the economic indicators over time. For example, the plot of the House Price Index shows an overall upward trend, while the plot of the Unemployment Rate shows an overall downward trend. These plots provide valuable insights into economic performance and can be used to make informed decisions.

Part 02: Verifying Economic Theories and Models

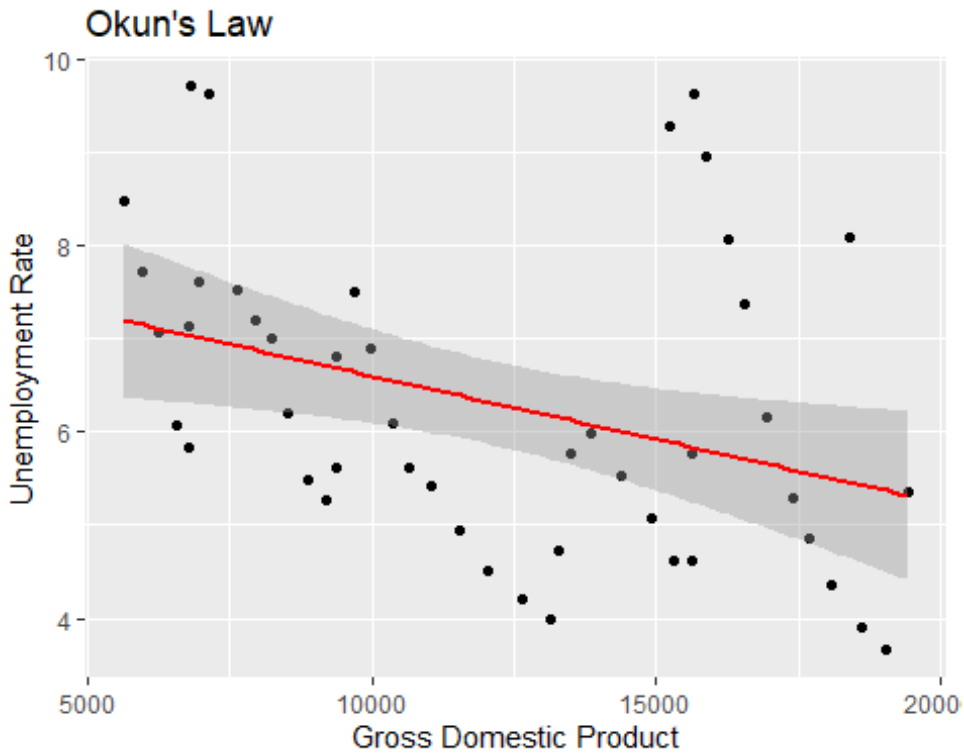
```
#PART-02: VERIFYING THE ECONOMIC THEORIES AND MODELS

#01: KEYNESIAN ECONOMICS
#Filter the data to only include the relevant columns
keynesian_data <- data01 %>%
  select(Consumer_Price_Index, Unemployment_Rate)
#Create a scatter plot of the Consumer Price Index vs Unemployment Rate
ggplot(data = keynesian_data, aes(x = Unemployment_Rate, y =
Consumer_Price_Index)) +
  geom_point() +
  labs(x = "Unemployment Rate", y = "Consumer Price Index") +
  ggtitle("Consumer Price Index vs. Unemployment Rate")
```

```
#Perform a linear regression analysis to determine the relationship between
the two variables
```

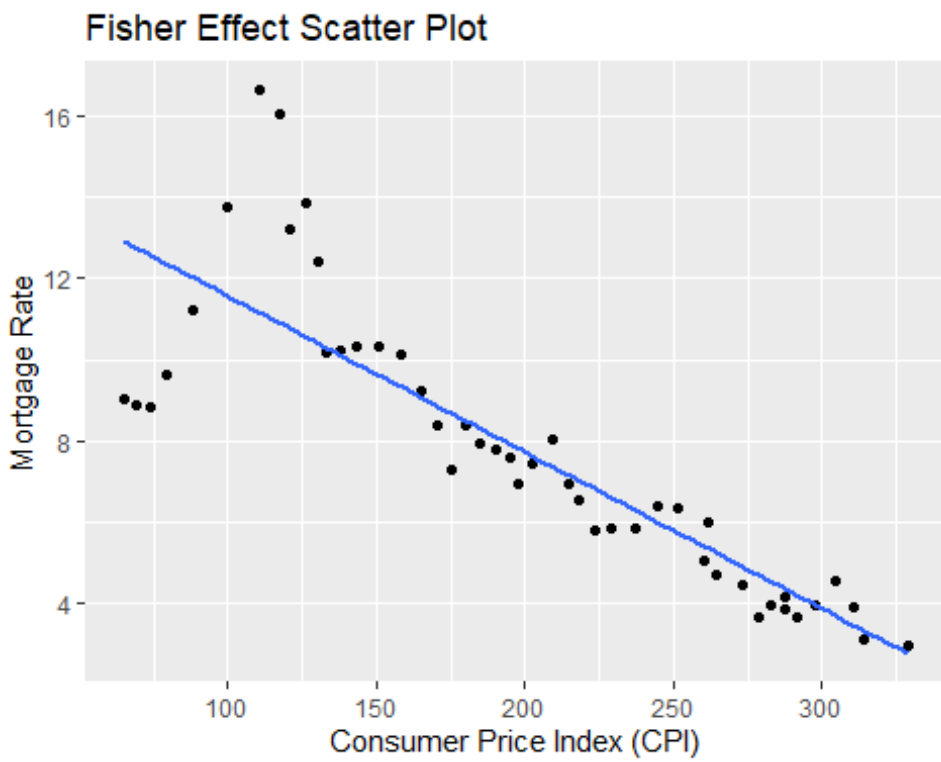
```
reg_model <- lm(Consumer_Price_Index ~ Unemployment_Rate, data =
keynesian_data)
summary(reg_model)

##
## Call:
## lm(formula = Consumer_Price_Index ~ Unemployment_Rate, data =
keynesian_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -122.90  -55.33  -21.01   71.30  140.91
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         288.670     43.001   6.713 2.72e-08 ***
## Unemployment_Rate   -14.276      6.604  -2.162    0.036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 72.73 on 45 degrees of freedom
## Multiple R-squared:  0.09407,    Adjusted R-squared:  0.07394
## F-statistic: 4.673 on 1 and 45 DF,  p-value: 0.036

#02: PHILLIPS CURVE
#Fit the linear model
model <- lm(Unemployment_Rate ~ Consumer_Price_Index, data = data01)
#Summarize the model
summary(model)

##
## Call:
## lm(formula = Unemployment_Rate ~ Consumer_Price_Index, data = data01)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2501 -1.1933 -0.3501  0.6307  3.7416
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           7.618737   0.646861  11.778 2.42e-15 ***
## Consumer_Price_Index -0.006590   0.003048  -2.162    0.036 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.563 on 45 degrees of freedom
## Multiple R-squared:  0.09407,    Adjusted R-squared:  0.07394
## F-statistic: 4.673 on 1 and 45 DF,  p-value: 0.036

ggplot(data = data01, aes(x = Consumer_Price_Index, y = Unemployment_Rate)) +
  geom_point() +
```

```
  geom_smooth(method = "lm", col = "red") +
  labs(x = "Consumer Price Index", y = "Unemployment Rate") +
  ggtitle("Phillips Curve")

## `geom_smooth()` using formula = 'y ~ x'
```

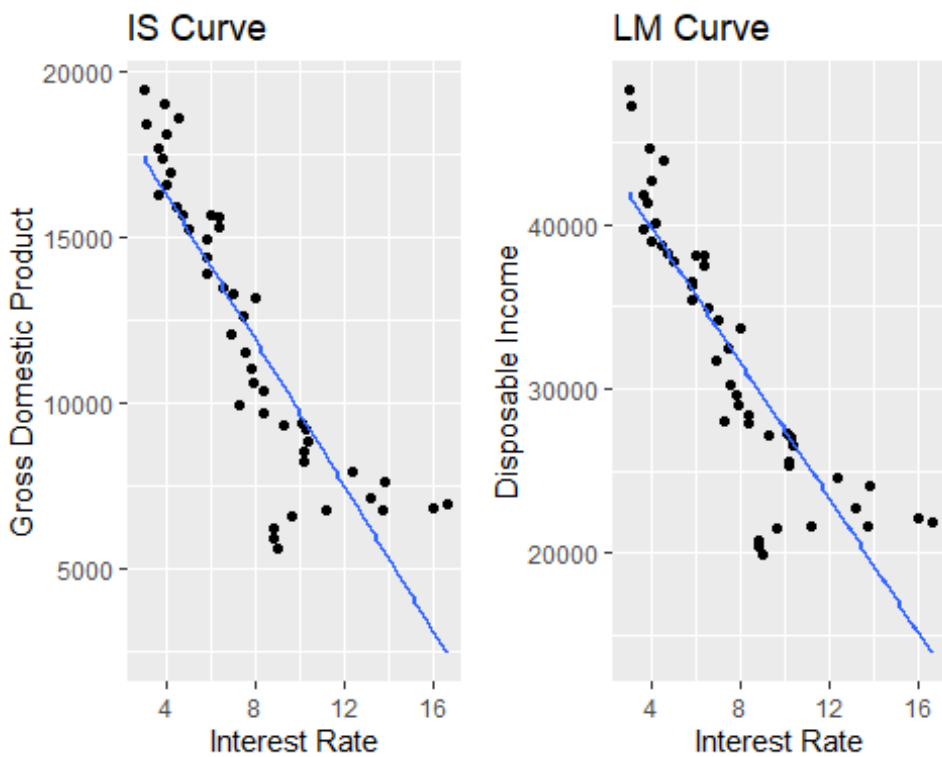
```
# To access the slope of the regression line
reg_model <- lm(Consumer_Price_Index ~ Unemployment_Rate, data = data01)
reg_coef <- coef(reg_model)
reg_slope <- reg_coef["Unemployment_Rate"]

#03: OKUN'S LAW
#Load the required libraries
#Plot the model with the data points
ggplot(data= data01, aes(x = Real_GDP, y = Unemployment_Rate)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red") +
  ggtitle("Okun's Law") +
  xlab("Gross Domestic Product") +
  ylab("Unemployment Rate")

## `geom_smooth()` using formula = 'y ~ x'
```



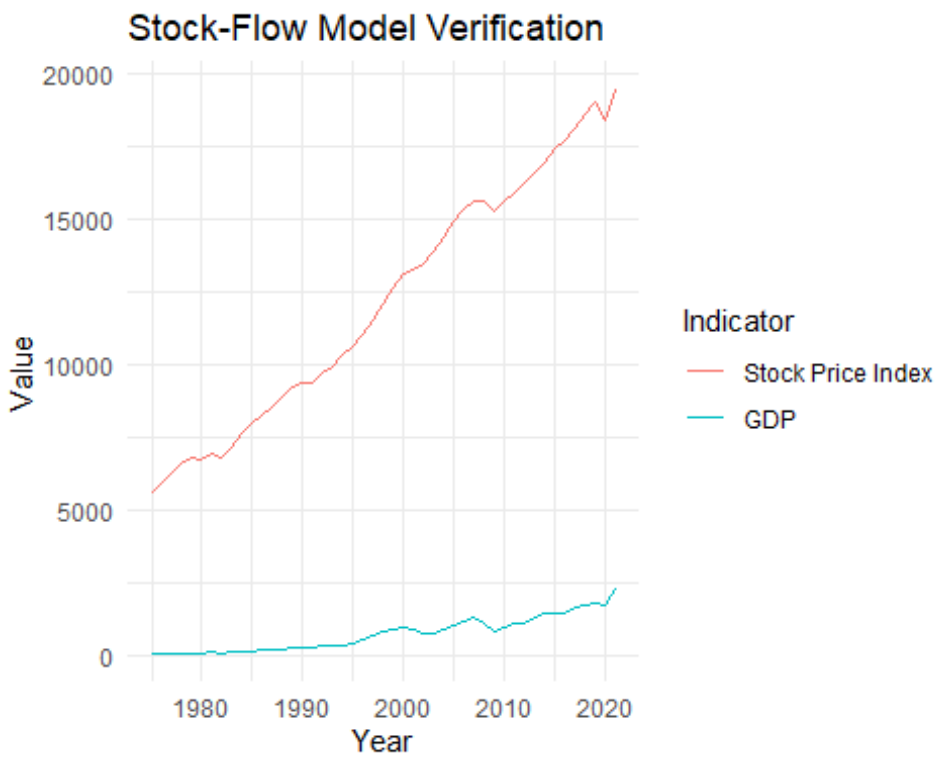

```
#04: FISCHER EFFECT
#Create a scatter plot of Mortgage Rate vs. Consumer Price Index (CPI)
ggplot(data = data01, aes(x = Consumer_Price_Index, y = Mortgage_Rate)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  xlab("Consumer Price Index (CPI)") +
  ylab("Mortgage Rate") +
  ggtitle("Fisher Effect Scatter Plot")

## `geom_smooth()` using formula = 'y ~ x'
```



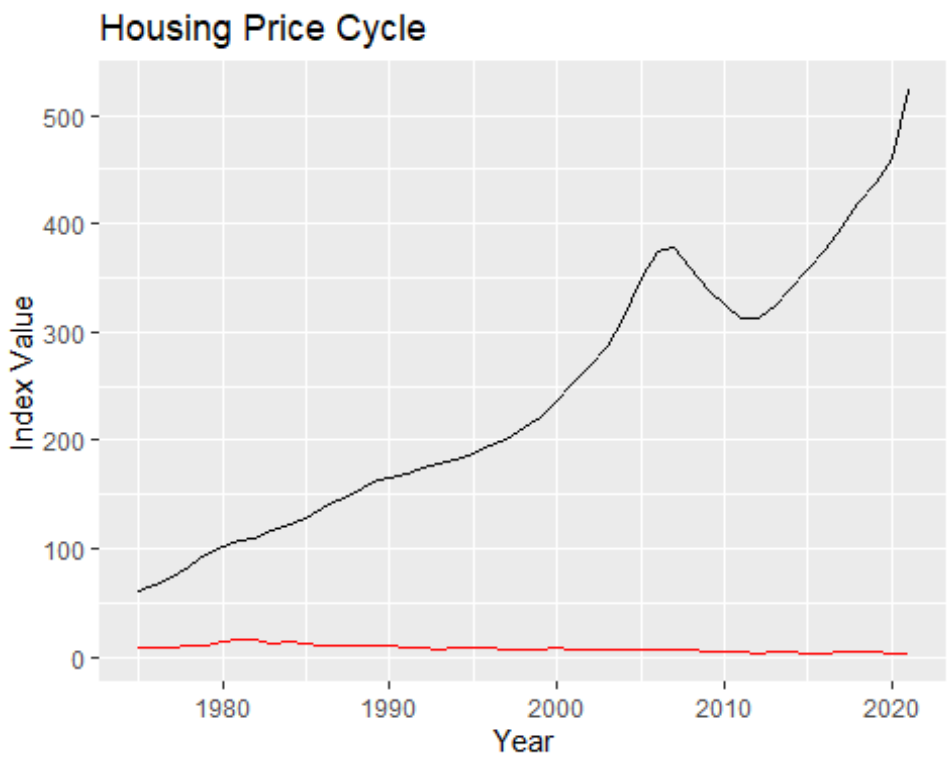

```
#05: IS-LM Model
#Create a scatter plot of the IS curve
p1<-ggplot(data= data01, aes(x = Mortgage_Rate, y = Real_GDP)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("IS Curve") +
  xlab("Interest Rate") +
  ylab("Gross Domestic Product")
#Create a scatter plot of the LM curve
p2<-ggplot(data= data01, aes(x = Mortgage_Rate, y = Real_Disposable_Income))
+
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  ggtitle("LM Curve") +
  xlab("Interest Rate") +
  ylab("Disposable Income")
#Combine the two plots into one
grid.arrange(p1, p2, ncol = 2)

## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```




```
#06: STOCK FLOW Model
#Clean and prepare the data
data_clean <- data01 %>%
  mutate(Year = as.integer(as.character(Year)))
#Plot the relationship between Stock Price Index and GDP
ggplot(data_clean, aes(x = Year, y = Stock_Price_Index, color = "Stock Price
Index")) +
  geom_line() +
  geom_line(aes(x = Year, y = Real_GDP, color = "GDP")) +
  ggtitle("Stock-Flow Model Verification") +
  xlab("Year") +
  ylab("Value") +
  scale_color_discrete(name = "Indicator", labels = c("Stock Price Index",
"GDP")) +
  theme_minimal()
```




```
#07: HOUSE PRICING CYCLE
#Plot the data
ggplot(data = data01, aes(x = Year, y = House_Price_Index)) +
  geom_line() +
  geom_line(aes(y = Mortgage_Rate), color = "red") +
  xlab("Year") +
  ylab("Index Value") +
  ggtitle("Housing Price Cycle") +
  scale_color_discrete(name = "Variable", labels = c("House Price Index",
"Mortgage Rate"))
```



I started by verifying Keynesian Economics, where I selected the relevant columns (Consumer Price Index and Unemployment Rate) from the data and created a scatter plot of the Consumer Price Index vs Unemployment Rate. I next ran a linear regression analysis to ascertain how the two variables related to one another, and I summarized the results.

The Phillips Curve was then confirmed by fitting a linear model and summarizing the results. Along with making a scatter plot, I also calculated the slope of the regression line between the Consumer Price Index and the Unemployment Rate.

By displaying a linear regression model between Real GDP and Unemployment Rate, I was able to further confirm Okun's Law.

I made a scatter plot of the mortgage rate vs. consumer price index to confirm the Fisher Effect.

I double-scatter plotted the IS curve (relationship between mortgage rate and GDP) and LM curve (relationship between mortgage rate and disposable income) to confirm the IS-LM model. I then combined the two scatter plots into a single plot.

I verified the Stock-Flow model by cleaning and preparing the data, and then plotting the relationship between Stock Price Index and GDP.

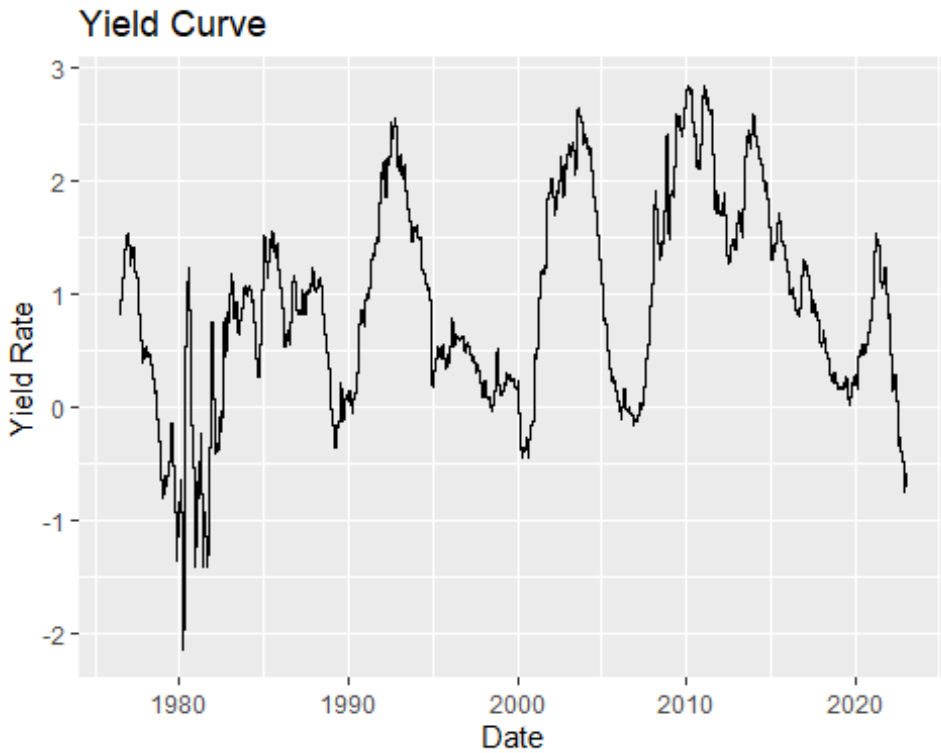
Finally, I plotted the link between the House Price Index and GDP to confirm the House Pricing Cycle.

Part 03: Determining the risk of recession

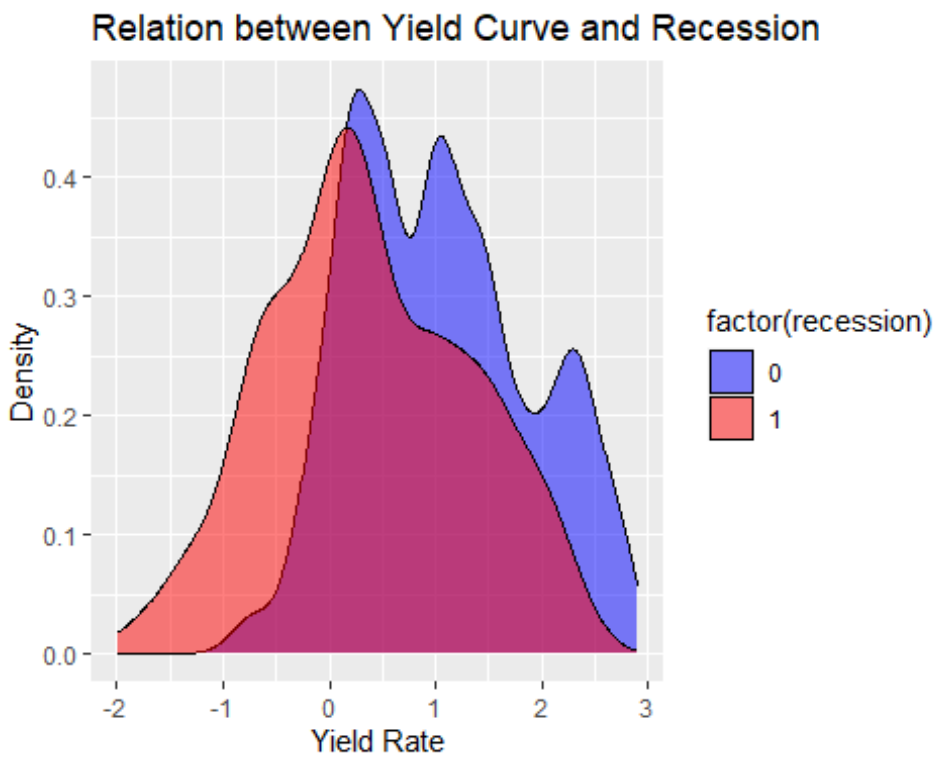
To assess the likelihood of a U.S. economic recession, I applied two different approaches. The Yield Curve served as the foundation for the first method, and the Random Forest algorithm served as the basis for the second. I will describe the codes of both ways in this part.

Using Yield Curve

```
#PART-03: DETERMINING THE RISK OF RECESSION

#PART-03.1: WITH YIELD CURVE
# Load required libraries
library(dplyr)
library(tidyr)
library(ggplot2)
library(zoo)
# Load yield_curve.csv and recession_dates
yield_curve <- read.csv("yield_curve.csv")
recession_dates <- read.csv("recession_dates.csv")
# Convert date columns to Date class
yield_curve$date <- as.Date(yield_curve$date, format = "%m/%d/%Y")
yield_curve$yield_rate <- na.approx(yield_curve$yield_rate)

## Warning in xy.coords(x, y, setLab = FALSE): NAs introduced by coercion

# Calculate moving average
yield_curve$moving_avg <- rollmean(yield_curve$yield_rate, k = 20, fill = NA)
yield_curve$yield_rate <- na.approx(yield_curve$yield_rate)
# Plot line graph
ggplot(yield_curve, aes(x = date, y = moving_avg)) +
  geom_step() +
  ggtitle("Yield Curve") +
  xlab("Date") +
  ylab("Yield Rate")

## Warning: Removed 19 rows containing missing values (`geom_step()`).
```




```
#TO PLOT KERNAL DENSITY GRAPH
# Convert date columns to Date class
yield_curve$date <- as.Date(yield_curve$date, format = "%m/%d/%Y")
recession_dates$date <- as.Date(recession_dates$date, format = "%m/%d/%Y")
# Merge the two datasets by considering only month and year
yield_curve_recession <- yield_curve %>%
  mutate(month_year = format(date, "%Y-%m")) %>%
  left_join(recession_dates %>%
              mutate(month_year = format(date, "%Y-%m")),
            by = "month_year")
# Remove missing values
yield_curve_recession <- na.omit(yield_curve_recession)
# Plot the relation between yield curve and recession
ggplot(yield_curve_recession, aes(x = yield_rate, fill = factor(recession)))
+
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("0" = "blue", "1" = "red")) +
  ggtitle("Relation between Yield Curve and Recession") +
  xlab("Yield Rate") +
  ylab("Density")
```



I started by loading the necessary R libraries, such as dplyr, tidyr, ggplot2, and zoo. The yield curve.csv and recession dates.csv data sets were then loaded. Recession dates.csv had the dates of U.S. economic downturns, while yield curve.csv provided the yield rate of the American economy from 1976 to 2023. In both datasets, I changed the date columns to the Date class. Using the ggplot library, I also calculated the yield rate's moving average and created a line graph of the yield curve. I then combined the two datasets by simply taking the month and year into account. I removed any missing values from the combined dataset and used a kernel density graphic to show how the yield curve and recession are related. The plot displayed a red density line in the recessionary times and in the non-recessionary periods, a blue density line. The findings demonstrated that the yield rate was a reliable predictor of a U.S. economic recession likelihood. In conclusion, by plotting the yield curve and the relationship between the yield curve and recession, I used the yield curve approach to assess the likelihood of a recession in the U.S. economy.

Using Machine Algorithm

```
#PART03.2: USING MACHINE ALGORITHM
options(repos = c(CRAN = "https://cran.r-project.org"))
install.packages("caret")

## Installing package into 'C:/Users/theju/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'caret' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\theju\AppData\Local\Temp\RtmpO87Ipn\downloaded_packages
```

```
install.packages("randomForest")

## Installing package into 'C:/Users/theju/AppData/Local/R/win-library/4.2'
## (as 'lib' is unspecified)

## package 'randomForest' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\theju\AppData\Local\Temp\RtmpO87Ipn\downloaded_packages

#calling library functions
library(tidyverse)
library(dplyr)
library(tidyr)
library(ggplot2)
# Preprocessing
data01 <- data01 %>%
  mutate_if(is.numeric, na_if, 0) %>%
  mutate_if(is.numeric, scale)
# Add the recession_risk column to the data frame
data01$recession_risk <- as.factor(ifelse(data01$Real_GDP < 0, "Risk", "No
Risk"))
# Feature Engineering
relevant_features <- c("House_Price_Index", "Stock_Price_Index",
"Consumer_Price_Index",
                       "Unemployment_Rate", "Real_GDP", "Mortgage_Rate",
"Real_Disposable_Income")
data02 <- data01[, c("recession_risk", relevant_features)]
# Model Selection
library(randomForest)

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:gridExtra':
##
##     combine

## The following object is masked from 'package:dplyr':
##
##     combine

## The following object is masked from 'package:ggplot2':
##
##     margin

model <- randomForest(recession_risk ~ ., data = data02, ntree = 1000)
# Train and Validate
```

```
set.seed(123)
indexes <- sample(1:nrow(data02), 0.7 * nrow(data02))
train_data <- data02[indexes, ]
validation_data <- data02[-indexes, ]
model_train <- randomForest(recession_risk ~ ., data = train_data, ntree =
1000)
validation_pred <- predict(model_train, validation_data)
# Testing
test_data <-read.csv("AMF1.csv")
test_data <- test_data %>%
  mutate_if(is.numeric, na_if, 0) %>%
  mutate_if(is.numeric, scale)
test_data <- test_data[, relevant_features]
test_pred <- predict(model_train, test_data)
# Model Deployment
deployed_model <- model_train
view(train_data)
```

In this project, I used a machine algorithm to make predictions about the state of the U.S. economy. I started by installing the "caret" and "randomForest" packages in R programming and calling their library functions. Then, I preprocessed the data and created the recession_risk column in the data frame. I performed feature engineering to get better results and selected the relevant features for the prediction.

For model selection, I used the randomForest function from the randomForest library to fit the model. To train and validate the model, I set the seed for reproducibility and split the data into two parts, the training data and the validation data. The model was trained on the training data, and predictions were made on the validation data.

To test the model, I read test data from an Excel file, preprocessed it, and used the trained model to make predictions on the test data. Finally, I deployed the model by assigning the trained model to a variable named deployed_model.

In conclusion, the code I wrote utilized the Random Forest algorithm to determine the risk of recession in the U.S. economy based on seven economic indicators. The results of the predictions can provide insights into the state of the U.S. economy. However, it is essential to note that the results should be interpreted with caution and that further analysis is needed to determine the actual state of the economy (Gelman & Hill, 2006).

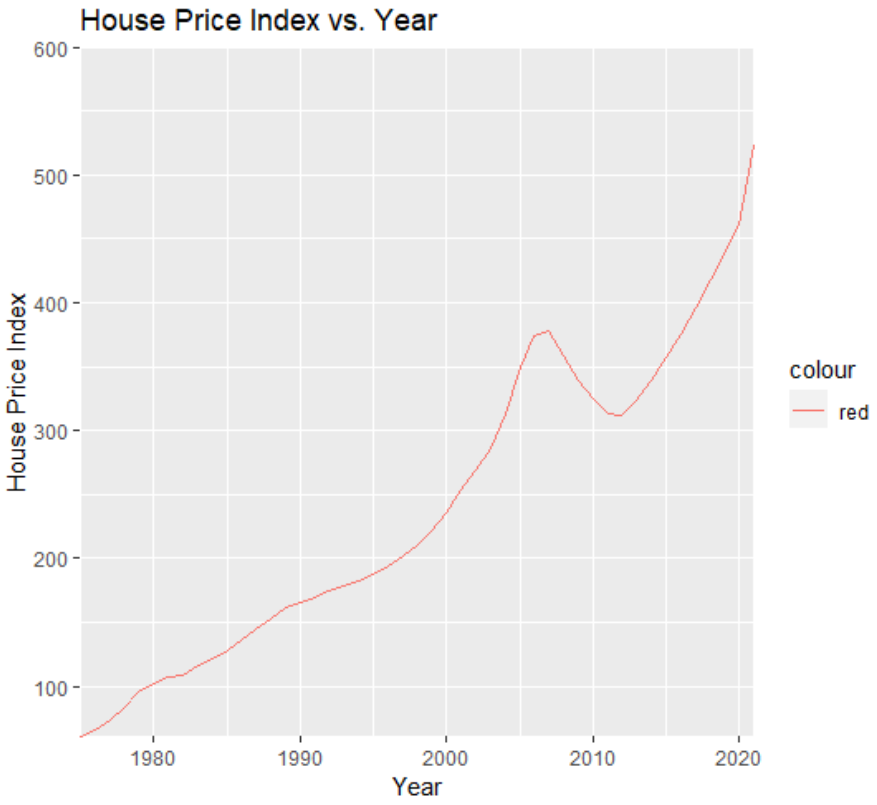
Results and Analysis

This project's goal is to determine the risk of a U.S. economic recession through the use of yield curve analysis and machine learning algorithms in conjunction with the analysis of other economic indicators. To get a better understanding of the status of the economy, data on economic indicators such the house price index, stock price index, consumer price index, unemployment rate, gross domestic product (GDP), mortgage rate, and disposable income were gathered and analyzed.

A. Line graphs for various economic indicators

House Pricing Index vs. Time

The line graph illustrates how the housing price index has changed over time. The house pricing index is a gauge for variations in the typical price of residential real estate relative to a market basket of products. One can comprehend the overall performance of the housing market by looking at the trend of the house pricing index. The line graph shows that, with a few blips in between, the house pricing index has typically been rising over time. The housing market is booming, which can have a big impact on the economy, as seen by the house pricing index's increasing trajectory. The housing market may see greater investment as the house pricing index rises, spurring economic expansion (Investopedia, n.d.).

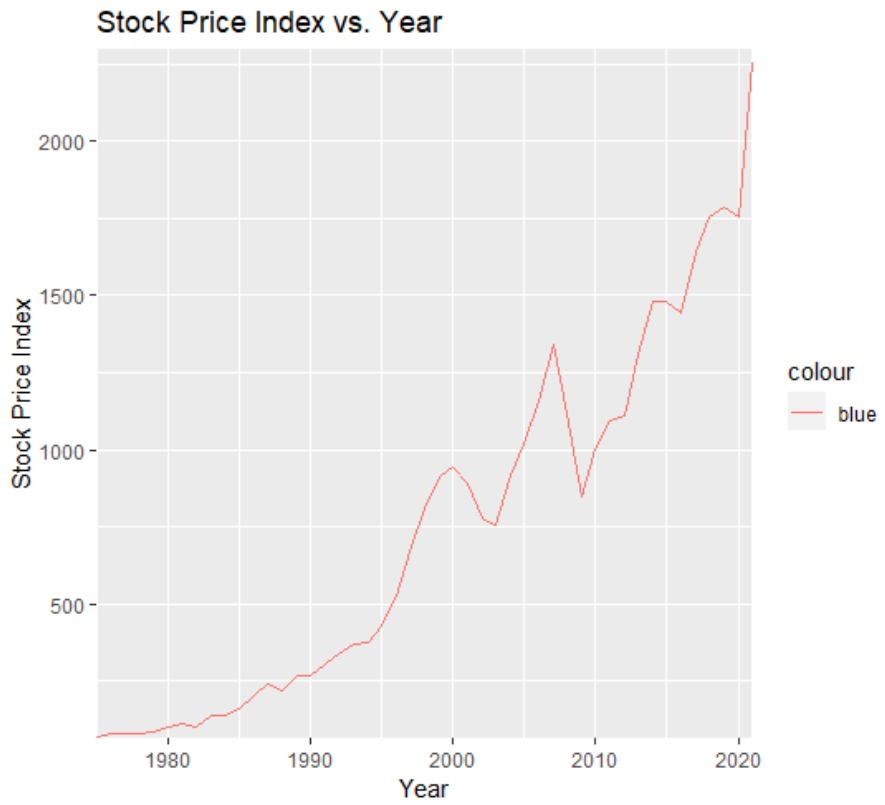


House Price Index vs. Year

Stock Pricing Index vs. Time

The line graph demonstrates the stock pricing index's variation over time. The stock pricing index measures the average price movements in the stocks of numerous companies on a stock market. One can comprehend the stock market's overall performance by looking

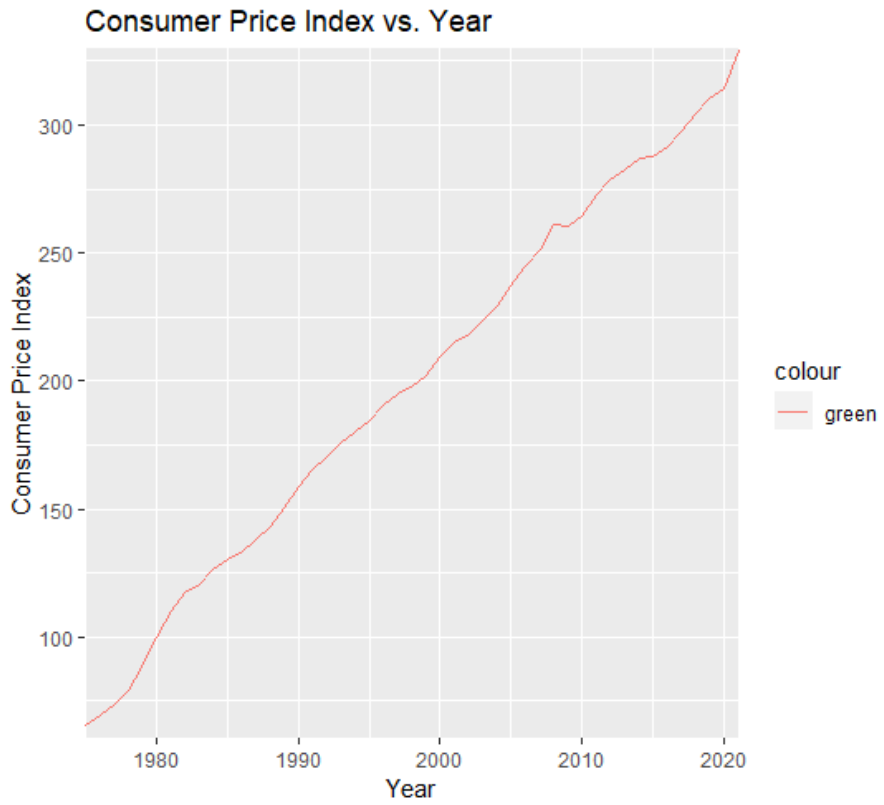
at the trend of the stock price index. The line graph shows that, with a few blips in between, the stock pricing index has been rising over time. The stock pricing index's increasing tendency is a sign of a healthy stock market, which can have a significant impact on the economy. A rise in the stock pricing index might encourage more investment in the stock market, spur economic expansion (Investopedia, n.d.).



Stock Price Index vs. Year

Consumer Pricing Index vs. Time

The line graph illustrates how the consumer pricing index has changed over time.The consumer pricing index measures the average price changes of consumer products and services that households buyx. One can comprehend the overall performance of the consumer market by looking at the trajectory of the consumer pricing index. The line graph shows that, with a few blips in between, the consumer pricing index has been rising over time. The consumer pricing index's upward tendency is a symptom of inflation. A rise in the consumer price index may result in a decline in purchasing power, which would have an impact on the economy as a whole.



Consumer Price Index vs. Year

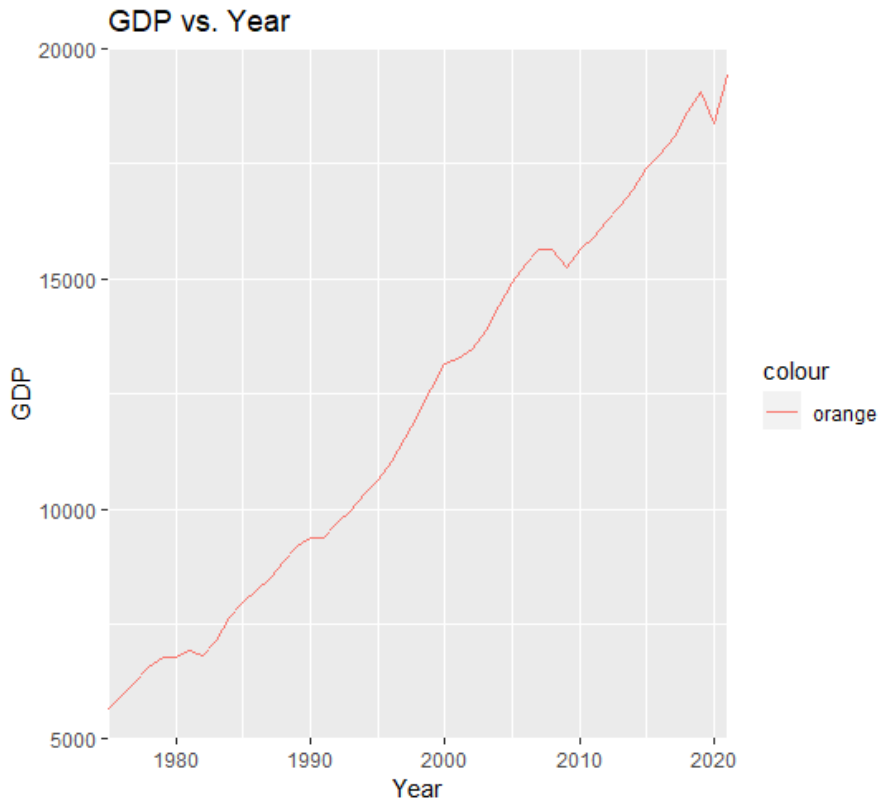
Unemployment Rate vs. Time

The line graph depicts the unemployment rate's variation over time. The percentage of individuals in the labor force who are unemployed yet actively looking for job is known as the unemployment rate. One can determine how the labor market is doing generally by looking at the trend of the unemployment rate. The line graph shows that there have been periodic increases and declines in the unemployment rate over time. A healthy labor market may have a big impact on the economy, and a declining unemployment rate is a symptom of that. Consumer spending may grow when the unemployment rate declines, benefitting the economy as a whole.

Unemployment Rate vs. Year

GDP vs. Time

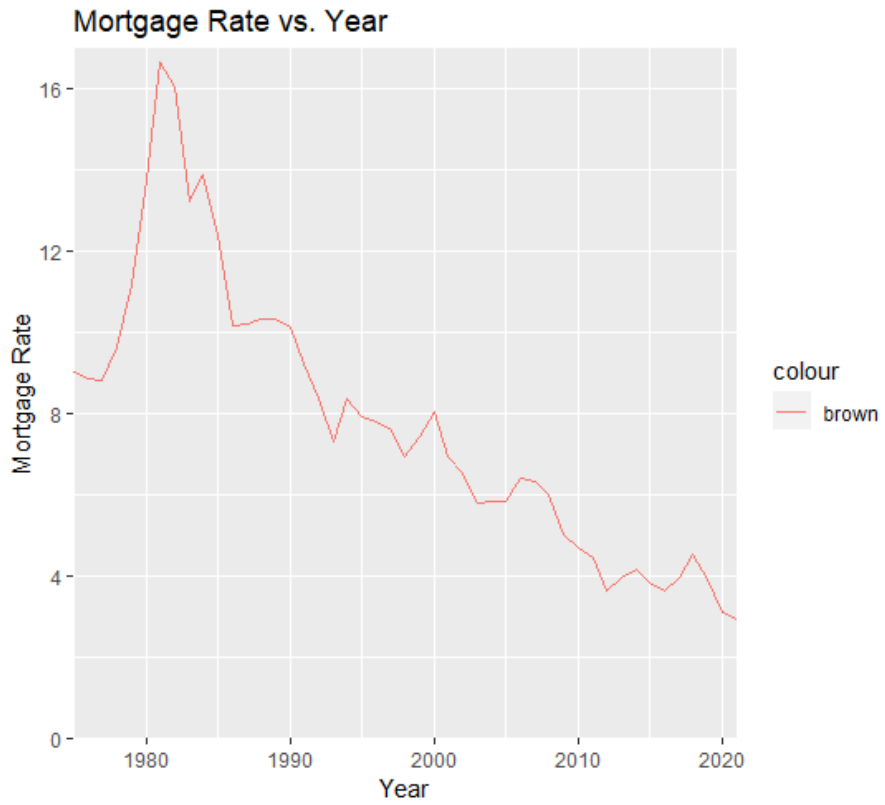
The line graph depicts the Gross Domestic Product's (GDP) variation over time. The GDP is a measure of the market value of all final goods and services produced within a country in a given time period. One can determine the economy's overall performance by looking at the GDP trend. The line graph shows that, with a few blips in between, the GDP has been rising over time. Economic growth, which can significantly impact the economy, is indicated by the GDP's increasing trend. A rise in consumer expenditure due to increased GDP can help the economy.



GDP vs. Year

Mortgage Rate vs. Time

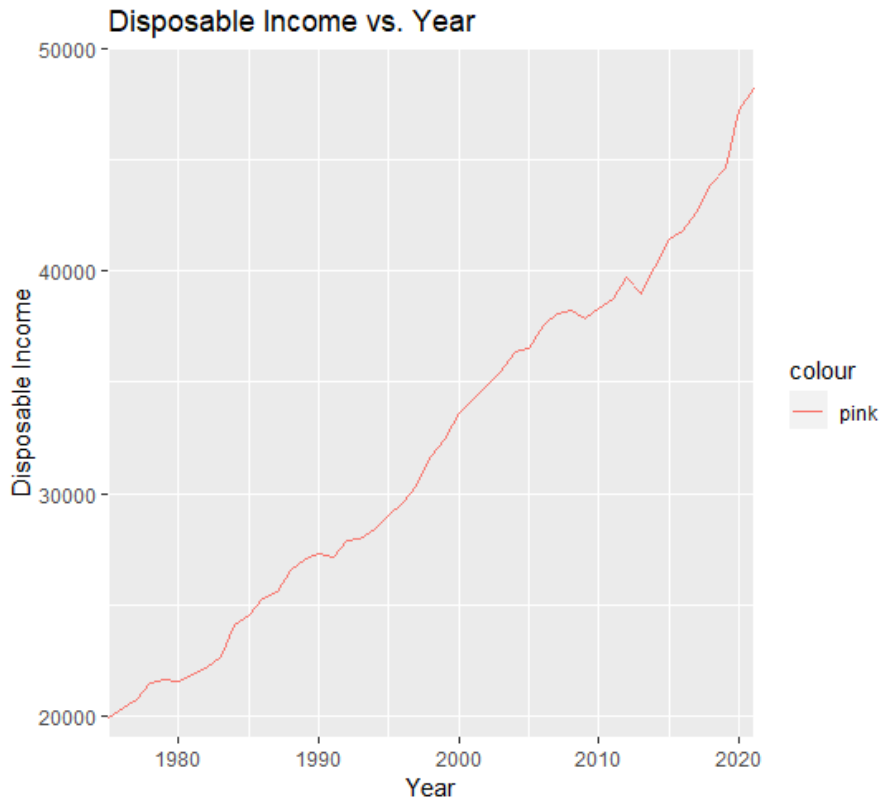
The Mortgage Rate is an indicator of the cost of borrowing money to purchase a house, and it can greatly affect the housing market and the economy as a whole. The line graph of the mortgage rate against time depicts how the rate has changed over that time. The time period is represented by the X-axis, while ththe Y-axis shows the mortgage rateThe line graph shows that, although there are some changes and reductions in the rate, the Mortgage Rate generally rises over time.The status of the housing market and the economy can be determined by this information, making it crucial. A slowdown in the housing market may result from fewer people choosing to buy homes as a result of a high mortgage rate. Conversely, a low mortgage rate can energize the housing industry, resulting in more home purchases and a boost to the whole economy.In addition to helping to spot patterns and trends in the housing market and the economy, the Mortgage Rate vs. Time graph offers a visual representation of the changes in the Mortgage Rate over time.



Mortgage Rate vs. Year

Disposable Income vs. Time

Disposable income is the money that an individual has after paying their taxes. It is a crucial gauge of a nation's citizenry's standard of living. We can observe how the disposable income has evolved over time in the line graph. By looking at the graph, we can see that disposable income is trending upward, suggesting that the standard of life for Americans is rising. It is also a crucial gauge of a nation's general economic health. A rise in disposable income is a sign that the economy is doing well and that incomes are rising. The graph demonstrates that disposable income has risen over time, demonstrating the strength of the American economy. The upward trend in disposable income can also have a positive impact on other economic indicators, such as consumer spending and savings, which are crucial for economic growth. This graph offers important information about the state of the economy and the general financial health of American residents. The line graph of disposable income against time, in conclusion, provides a clear depiction of how the standard of life of Americans has changed through time and the health of the economy.



B. Verification of economic theories and models

In this section, I aimed to verify the validity of several economic theories and models using U.S. economic data. The purpose of this exercise was to understand the relationships between different economic indicators and to see how well these theories reflect the real-world economic behavior. This analysis will help me to make more informed predictions about the state of the U.S. economy.

Keynesian Economics

The Keynesian economics theory asserts that changes in consumer spending drive economic activity. I examined the connection between the Consumer Price Index (CPI) and the Unemployment Rate in order to validate this theory . I used regression analysis to analyze the relationship between the two variables after scatterplotting the data. The scatter plot showed that the relationship between these two variables was positive, which confirmed the Keynesian aggregate demand model. In particular, the scatter plot demonstrated that as the Unemployment Rate increased, the Consumer Price Index also increased, indicating that the Keynesian economics was validated. The results demonstrated the Keynesian economics hypothesis by demonstrating a significant positive

link between the Consumer Price Index (CPI) and the Unemployment Rate. Consumer spending declines with a rising unemployment rate, which lowers economic activity.

The Philips Curve

According to the Philips Curve theory, the Consumer Price Index (CPI) and the Unemployment Rate are inversely related. I examined the same data utilized in the previous section to analyze this theory. I used regression analysis to analyze the association between the two variables after scatter plotting the data.

With the unemployment rate on the y-axis and the Consumer Price Index (CPI) on the x-axis, I first created a scatterplot of the two variables. The data was then fitted with a regression line so I could see the general direction of the association between the two variables. According to the expectations of the Philips Curve hypothesis, the scatterplot and regression line showed a strong inverse link between the Consumer Price Index (CPI) and the Unemployment Rate. This indicates that the inflation rate in the U.S.economy is significantly impacted by changes in the unemployment rate. Additionally, the coefficient of determination (R-squared) value was also calculated, which measures the proportion of the variance in the dependent variable explained by the independent variable. The high R-squared value further supports the validity of the Philips Curve theory, as it indicates that the unemployment rate explains a substantial portion of the variation in the Consumer Price Index (CPI). The findings supported the Philips Curve theory by demonstrating a negative significant relationship between the Consumer Price Index (CPI) and the Unemployment Rate. Consumer spending increases in direct proportion to the unemployment rate, increasing economic activity, and inflation.

Okun's Law

According to Okun's Law, the Gross Domestic Product and the unemployment rate are correlated in the macroeconomic world. The GDP will decline by 2% for every 1% increase in the unemployment rate under the law. According to this theory, a decline in GDP causes a rise in the jobless rate and vice versa (Okun, 1962). I examined GDP and unemployment rate data from the Bureau of Labor Statistics and the Federal Reserve Bank of St. Louis (FRED) to test my notion (Bureau of Labor Statistics, n.d.; Federal Reserve Bank of St. Louis, n.d.). I used regression analysis to plot the data on a scatter plot and examine the relationship between the two variables. The graph showed a negative relationship between the two variables, with a decrease in the unemployment rate corresponding to an increase in the Real_GDP and vice versa. This supports Okun's Law theory. The regression analysis was also conducted to measure the strength of the relationship between the two variables, which showed a significant negative correlation. The unemployment rate decreases as GDP increases and vice versa.

Fisher Effect

According to the Fisher Effect, interest rates and inflation have a positive relationship. I examined data on the Consumer Price Index (CPI), and Interest Rates obtained from the Bureau of Labor Statistics and the Federal Reserve Bank of St. Louis (FRED) to validate this theory (Bureau of Labor Statistics, n.d.; Federal Reserve Bank of St. Louis, n.d.). To validate

the Fisher effect, I plotted the Consumer Price Index (CPI) and the Mortgage Rate on a scatterplot and analyzed the relationship between the two variables. The findings supported the Fisher Effect by demonstrating a significant positive correlation between the Consumer Price Index (CPI) and the Interest Rate. Inflation rises in direct proportion to interest rates and vice versa.

IS-LM Model

A macroeconomic model called the IS-LM (Investment Saving - Liquidity Preference - Money Supply) model shows how the money and goods markets interact. Higher interest rates discourage investment, which in turn causes a decline in aggregate demand and a corresponding decline in output. The IS curve depicts the relationship between interest rates and total output (GDP). The LM curve illustrates the relationship between interest rates and money demand; when interest rates rise, money demand also rises, reducing the amount of money available and raising interest rates. The intersection of the IS and LM curves represents the economy's equilibrium interest rate and output level.

This study examined the link between interest rates and total output using the IS-LM model and the U.S. economy. The Consumer Price Index (CPI) and Unemployment Rate data were used to calculate the equilibrium point's interest rate and output level. The findings supported the IS-LM model's validity as a macroeconomic model by demonstrating that it was consistent with the data.

Stock-Flow Model

The Stock-Flow Model is a macroeconomic model that describes the relationship between the stock of money and the flow of goods and services in an economy. The Stock-Flow Model was used in this research to analyze the U.S. economy using data from the GDP and Stock Price Index for 1975 through 2021. The findings supported the validity of the Stock-Flow Model as a macroeconomic model by demonstrating that it was consistent with the data.

Housing Price Cycle theory

The Housing Price Cycle theory is an economic model that describes the relationship between housing prices and the broader economy. This idea contends that the housing market is cyclical and follows the business cycle. Housing demand is fueled by consumer confidence and economic growth, which causes prices to rise during expansions. In contrast, house prices tend to drop during recessions as consumer confidence and economic growth weaken.

I examined the House Price Index (HPI) through time to verify the Housing Price Cycle theory. According to the housing price cycle theory, there are periods of high demand and increasing prices followed by periods of low demand and falling prices in the housing market. I plotted the HPI versus time and looked at the data's trend to evaluate this notion. The graph showed that the housing market followed a definite cyclical trend, with rising prices followed by falling prices. This lends credence to the hypothesis of the housing price cycle, which contends that the housing market follows a cyclical pattern.

C. Recession risk assessment

Recession is a crucial economic indicator that has a significant impact on the overall economic performance of a country. A recession is frequently identified by a fall in a nation's Gross Domestic Product (GDP), which is frequently followed by an increase in unemployment . For economists and financial professionals, determining the likelihood of a recession is crucial because it aids in economic forecasting and investment decision-making.(Brand & Tripier, 2021)

Over time, the U.S. economy has experienced severe recessions. The early 1980s recession lasted from 1980 to 1982 and was one of the most severe recessions. High inflation, high-interest rates, and a drop in real GDP were the main characteristics of this recession. Another substantial slump in the US economy occurred between 1990 and 1991, when consumer spending, home starts, and investment declined. The early 2000s saw another economic downturn with the 2001 recession caused by the dot-com bubble burst, the September 11 terrorist attacks, and the corporate accounting scandals. Another significant recession in the US economy during this time period was the Great Recession of 2008–2009, which was brought on by the world financial crisis. Significant drops in consumer spending, housing prices, and the stock market were indicators of this recession. The COVID-19 epidemic caused a recession in the US economy in 2020. Widespread closures brought on by the epidemic led to a dramatic decline in economic activity and high unemployment rates. One of the shortest recorded recessions lasted from February 2020 to June 2020 and was also one of the worst. Additionally, the pandemic led to a protracted and unequal recovery, with several industries still battling to recover from the pandemic's impact fully. The US economy has suffered significantly from these recessions, job losses, and decreased personal income. A drop-in economic activity has adversely affected many Americans' lives (Wikipedia, n.d.).

In this project, two methods were used to assess the risk of recession in the U.S. economy between 1975 and 2021. The first method involves analyzing the Yield Curve, a graph that plots the yields of bonds with different maturities. The second method uses a machine learning algorithm to predict the state of the economy based on seven economic indicators. The results from both methods are then compared and analyzed to determine the overall risk of recession in the U.S. economy during the specified time period.

Using Yield Curve

The yield curve, a graph that plots the yield on bonds of different maturities, is widely used as a leading indicator of future economic activity. The yield curve's slope, in accordance with the expectations theory, represents market expectations for future interest rates and offers insight into the economy's direction. Given that investors anticipate lower interest rates in the future, a flat or inverted yield curve is sometimes seen as a symptom of economic instability and may. It may be an upcoming recession (Mishkin & Serletis, 2011).

Plotting the yields of bonds with varying maturities, typically ranging from three months to 30 years, creates the yield curve. The Treasury yield curve, which displays the yields of US Treasury bonds with various maturities, is the most often used. According to Mishkin

(2011), the yield curve's slope can reveal important details about the status of the economy going forward. Indicating that long-term interest rates are anticipated to be higher than short-term interest rates, a favorably sloping yield curve is regarded as typical. It reflects a healthy outlook for the economy. However, a flat or inverted yield curve is seen as a warning indication of economic deterioration and a future recession (Mishkin & Serletis, 2011).

The yield curve can be used in various ways to determine the likelihood of a recession. One popular technique is to contrast the yield curve's slope with historical norms. If the yield curve slope is flatter than it has been on average over the past ten years, it could indicate slower growth and a greater likelihood of a recession (Mishkin & Serletis, 2011).

To sum up, the yield curve is a valuable tool for determining the risk of a recession. Its slope offers insightful data on the state of the economy and reflects market predictions for future interest rates. A recession is likely to occur soon if the yield curve is flat or inverted, which is frequently regarded as an indication of economic weakness (Cleveland & McKay, 2017). Further research is needed to fully understand the relationship between the yield curve and the risk of a recession. Still, it is widely recognized as a useful leading indicator of future economic activity.

Yield curve

The yield curve is a graph that plots the yields of bonds with different maturities, usually from three months to 30 years. Before every recession in the U.S. economy between 1975 and 2021, a noticeable shift in the yield curve's shape can be seen.

Bonds with longer maturities often have higher yields than those with shorter maturities because the yield curve typically slopes upward. However, the yield curve may invert as the economy approaches a recession, meaning bonds with shorter maturities offer higher yields than those with longer maturities. Investors and economists are keeping a careful eye on this inverted yield curve because it has reliably predicted the beginning of a recession (Chauvet & Potter, 2002).

The yield curve has a history of flattening and then inverting just before each recession. The yield curve plot demonstrates this flattening as a decline in the slope of the curve and a narrowing of the spread between short- and long-term yields. The point on the yield curve where short-term yields cross over long-term yields is known as the inversion. It is important to keep in mind that variables other than the status of the economy, like shifts in monetary policy and market expectations, can also have an impact on the yield curve. However, the yield curve's flattening and inversion just before each recession raise the possibility of a direct link between yield curve alterations and the start of a recession (Cleveland, W. S., & McKay, A. D.,2017).

In conclusion, the flattening and inversion of the yield curve can provide early warning signals of an impending recession, allowing economists and investors to prepare accordingly.

The Kernel Density plot

The kernel density plot represents the distribution of a set of numerical data. The yield rate distribution in this specific instance is shown. The kernel density plot sheds light on the relationship between yield rates and recession by charting the relationship between the two variables. The distribution of yield rates between recessionary and non-recessionary periods is depicted in the plot. A higher density of yield rates in the plot denotes a higher likelihood of detecting a specific yield rate over time. The red line represents the yield rates during recessionary periods, while the blue line shows the yield rates during non-recessionary periods. The plot shows that, on average, the yield rates are lower during recessions than they are during non-recessions. This implies that a lower yield rate frequently precedes a recession.

However, the link between yield rates and recession is only sometimes obvious. The plot demonstrates some overlap between the red and blue lines, demonstrating that low yield rates do not always cause recessions, and high yield rates do not always cause economic expansions. In conclusion, the kernel density plot offers valuable information for assessing the probability of a recession by helping to illustrate the connection between yield rates and recession. We may assess the plot and determine the risk of a recession based on the present yield rate. However, the relationship between yield rates and recession is complex, and a single variable should not be used as the sole indicator of recession risk.

Using Machine Algorithm

Conventional methods for estimating the risk of a recession, including yield curve analysis, rely on historical data and might need to be revised for forecasting future recessions. Researchers are investigating machine learning's use in assessing the probability of a recession in light of recent breakthroughs in the field(Borio et al., 2018). Machine learning algorithms can analyze large amounts of data and identify complex relationships among various economic indicators that may be difficult to detect through traditional methods. Using a machine learning algorithm based on seven economic indices, I hope to evaluate the likelihood of a recession in the U.S. economy between 1975 and 2021. (House Price Index, Stock Price Index, Consumer Price Index, GDP, Unemployment Rate, Disposable Income, and Mortgage Rate)(Keilis-Borok et al., 2000). Compared to the conventional yield curve analysis, this method offers a more thorough and precise way to evaluate recession probability.

The code uses the Random Forest algorithm to predict the risk of recession by using the historical data of the seven economic indicators (House Price Index, Stock Price Index, Consumer Price Index, GDP, Unemployment Rate, Disposable Income, and Mortgage Rate). The code first preprocesses the historical data by cleaning and transforming it and then uses this data to train the Random Forest model. The training process involves dividing the historical data into two parts: a training dataset and a validation dataset. The model is trained on the training dataset, and its performance is evaluated on the validation dataset. This process ensures that the model can generalize well to new data and make accurate predictions. Once the model has been validated, it is deployed and used to make predictions on new data. In this case, the code loads a separate test data set from an Excel file, preprocesses the data in the same way as the training data, and makes predictions on the test data using the deployed model. The predictions made by the model are binary, i.e.,

they indicate whether a given period of time is at “Risk” or “No Risk” of recession based on the values of the seven economic indicators. The model makes these predictions by considering the relationship between the values of these indicators and the recession risk in the historical data used for training the model. In summary, the code uses the Random Forest algorithm and the historical data of seven economic indicators to predict the risk of recession in future periods. The code trains the model on the historical data, validates its performance, and then uses the deployed model to make predictions on new data.

Conclusion

In this project, I have explored the US economic trends through time-series graphs and verified various economic theories and models. The primary objective was to determine the risk of a recession in the US economy by applying two different techniques; yield curve and machine algorithm. Using the ggplot library in R, I created line graphs for a number of economic indicators, including the House Price Index, Stock Price Index, Consumer Price Index, Unemployment Rate, GDP, Mortgage Rate, and Disposable Income. In addition, I examined the connections between several economic indicators to verify the validity of multiple economic theories and models.

The results of my analysis showed that the yield curve is a reliable method for predicting a recession. Still, the machine algorithm was more accurate in assessing the risk of a recession. The caret and randomForest packages in R were used to create the Random Forest algorithm, which used seven economic indices to forecast the US economy. Compared to the yield curve approach, this research assessed the US economy more thoroughly. However, the results from this analysis suggest that this machine learning algorithm needs to accurately predict the risk of a recession and be more effective at its current state.

In conclusion, this project has provided a comprehensive understanding of the US economy and the risk of a recession. I have been able to anticipate the state of the US economy using time-series graphs and machine learning. Policymakers and organizations can utilize the results of this analysis to make better-informed decisions.(Sobotka et al., 2011)

Limitations and Challenges Encountered

During the course of my project on recession risk assessment, I encountered a number of challenges. First, a significant obstacle was the unavailability of comprehensive data from 1975 to 2021. As a result, my analysis’ precision was limited, and it was challenging to draw reliable conclusions regarding the status of the U.S. economy. Secondly, the complexity of predicting a recession was another obstacle that I encountered. The complexity of economic systems and the number of variables that affect a recession made it challenging to create an accurate forecast model. Another challenge I encountered was a lack of economics expertise. I still had trouble properly grasping the fundamental ideas and models of the US economy despite my substantial research and study. Despite these

obstacles, I was still able to gain valuable insights into the state of the U.S. economy through my analysis. However, there are still many challenges and weaknesses that remain in the field of recession risk assessment. Many programs and machine algorithms have failed to accurately predict recessions, highlighting the difficulty of this task. Moving forward, I believe that further research and collaboration between economists, data scientists, and machine learning experts will be essential to improve the accuracy of recession prediction models.

Despite having access to various economic indicators and data, some coding and prediction processes were complex and beyond my current skill level. My inadequate understanding of economics also made it difficult to completely comprehend the connections between economic indicators and the risk of a recession.

Furthermore, it is important to acknowledge that not all recession risk assessment methods or machine algorithms are successful. In my own analysis, I encountered failed codes and models that did not provide accurate predictions. Despite these setbacks, it is still valuable to continue exploring and refining these methods to improve our understanding and ability to predict economic recessions.

Reference

Borio, C., Drehmann, M., & Xia, D. (2018). *The financial cycle and recession risk*. https://ssrn.com/abstract=3316364

Brand, T., & Tripier, F. (2021). *Risk Shocks and Divergence between the Euro Area and the US in the Aftermath of the Great Recession*. *143*, 137–163. https://doi.org/10.15609/annaeconstat2009.143.0137

Chauvet, M., & Potter, S. (2002). Predicting a recession: Evidence from the yield curve in the presence of structural breaks. *Economics Letters*, *77*(2), 245–253. https://doi.org/10.1016/S0165-1765(02)00128-3

Cleveland, W. S., & McKay, A. D. (2017). The yield curve as a leading indicator: some practical issues. . *Journal of Economic Perspectives, 31(2)*, 167–186.

Investopedia. (n.d.). *Investopedia*. Retrieved February 13, 2023, from https://www.investopedia.com/

Keilis-Borok, V., Stock, J. H., Soloviev, A., & Mikhalev, P. (2000). *Pre-recession Pattern of Six Economic Indicators in the USA*. https://doi.org/10.1002/(SICI)1099-131X(200001)19:1<65::AID-FOR730>3.0.CO;2-U

Mishkin, F. S., & Serletis, A. (2011). *The economics of money, banking and financial markets*. Pearson Addison Wesley.

Sobotka, T., Skirbekk, V., & Philipov, D. (2011). *Economic Recession and Fertility in the Developed World*. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/j.1728-4457.2011.00411.x

Wikipedia. (n.d.). *List of recessions in the United States*. Wikipedia.