

# Economic Recession Forecasts Using Machine Learning Models Based on the Evidence from the COVID-19 Pandemic

Yuhuan Huang<sup>1</sup>, Erik S. Yan<sup>2</sup>

<sup>1</sup>Department of Finance, Southern University of Science and Technology, Shenzhen, China

<sup>2</sup>Carey Business School, Johns Hopkins, Baltimore, USA

Email: yuhuanlucky@163.com, kavanster@g.ucla.edu

---

**How to cite this paper:** Huang, Y. H., & Yan, E. S. (2023). Economic Recession Forecasts Using Machine Learning Models Based on the Evidence from the COVID-19 Pandemic. *Modern Economy*, 14, 899-922.  
<https://doi.org/10.4236/me.2023.147049>

**Received:** May 10, 2023

**Accepted:** July 14, 2023

**Published:** July 17, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This paper focuses on the use of machine learning models to forecast economic recessions caused by incidents such as the COVID-19 pandemic. Relevant economic variables are selected to fit into the VAR, SVR, Random Forest, and LSTM models. The study examines the cases of the US and Italy, analyzing how the models predict the Euro crisis, 2008 Financial Crisis, and the economic recession induced by COVID-19. Evaluations and comparisons among these models and cases are made to determine appropriate models. Additionally, an analysis based on US 2020 mobility data is applied to demonstrate the difference in economic activities between normal and crisis times.

## Keywords

Economic Recession, COVID-19, Machine Learning Applications, US, Italy

---

## 1. Introduction

The COVID-19 pandemic has triggered worldwide economic downturns. With businesses shutting down, financial market depressed, unemployment rates surging, and personal consumption declining, the economic decline raised many social problems. Given the current situation with the pandemic, it is meaningful to understand the economic impact of the incidents like the pandemic and predict future recession scenarios. An economic recession can be attributed to the economic cycle or some other common reasons, such as oversupply, bursts of economic bubble, and incidents like wars and pandemic. We can identify an economic recession by observing successive decreases in GDP growth rate. While a financial-market-induced crisis like the 2008-2009 subprime crisis would typically

exhibits warning signs prior to its onset, an economic decline triggered by events like COVID-19 is more likely to be a sudden blow. Since macroeconomic forecasts would potentially be utilized for policy making and planning, it is imperative that we find a sound and feasible way to identify a possible economic recession. Machine learning has been exerting its great advantages in various fields including environment, medicine and healthcare, transportation, finance as well as economy, assisting people in making predictions or decisions. The researcher hopes to explore the effectiveness of machine learning as a method for forecasting economic recessions in this study (Connaughton, 2010; Liu & Tang, 2022).

There has been some recent research concerning forecasts of economic recessions caused by incidents like COVID-19. According to Ludvigson et al., incidents like COVID-19 are multi-period and have massive influence worldwide, which differs from conventional economic shocks. According to their study, they use data of US over forty years and construct a costly disaster (CD) time series to measure costs of the incident and analyze its dynamic impacts. Another study focuses on the underlying assumptions of the forecasting problem. They synthesize the current shock from typical economic shocks in history based on some elements. Kuyo et al. uses natural language processing model (NLP) to do sentiment analysis. Machine learning models including Naïve Bayes and N-grams are applied on social medial data. Chetty et al. build a public database at a granular and high frequency level by themselves, which helps them obtain observations more precisely. The study conducted by Baker et al. mainly focuses on the uncertainty triggered by the incident. Levanon utilizes Markov Switching model to calculate the probabilities of economic recessions. In the study conducted by Liu et al., big data analysis is used to predict government economic situation. They design a system for the early warning of the government economic situation. The researchers consider pertinent economic factors as well as responses of policy makers facing these situations (Ludvigson, Ma, & Ng, 2020; Primiceri & Tambalotti, 2020; Kuyo, Mwalili, & Okang'o, 2021; Chetty et al., 2020; Baker et al., 2020; Levanon, 2011; Liu & Tang, 2022).

In a broader context, researchers are committed to using machine learning and data analysis methods for predicting overall future economic or GDP growth. Nosratabadi et al. summarize the applications of advanced machine learning and deep learning models in Economics-related fields. According to their study, models including SVR, Naive Bayesian and C4.5 Decision Tree Classifiers, BP Neural Network, Deep Neural Network, Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS) are some of the classic machine learning methods used in economics-relevant fields. More cases of using hybrid machine learning and deep learning methods are also summarized. Another group of study utilizes data-mining techniques such as non-parametric classification and hierarchical clustering to examine the effects of COVID-19 on the economies of G20 countries. In the study conducted by Malladi, machine learning method including linear SVM (LSVM) and KNN-weighted are used to make

forecasts mainly in financial markets. The researchers discover that these machine learning algorithms can predict stock market crashes up to two months in advance. In another study, Huang combines PCA and BP Neural Network to conduct economic data analysis and intelligent predictions. Chu et al. compare ten forecasting models in out-of-sample predictions of GDP. The models examined in the study include Mixed Data Sampling (MIDAS), Ridge Regression, Lasso Regression, Random Forest, XGBoost, LSTM, Deep NN, and so on. Their findings suggest that tree-based models such as XGBoost and Random Forest may perform better with a larger number of predictors, whereas LSTM may perform better with fewer numbers of predictors. Meanwhile, Deep NN performs fairly between these two cases (Nosratabadi, Mosavi, Duan, Ghamisi, Filip, Band, Reuter, Gama, & Gandomi, 2020; Taylan, Alkabaa, & Yilmaz, 2022; Malladi, 2022; Huang, 2022; Chu & Qureshi, 2022).

Therefore, this study would like to implement different machine learning models on economic recession forecasts and evaluate their performances. The models would be established based on historical data of relevant economic indicators and be applied to different economies.

## 2. Methodologies

The economy is a complex system made up of various economic factors, which has mutual influence on each other. The primary task of this study is to use models built on economic factors to forecast future economic recessions. Machine learning models including Support Vector Regression (SVR), Random Forest (RF), and Long-short Term Memory Model (LSTM) are used in this study as prediction models. The traditional time series model Vector Autoregression Model (VAR) is also applied as a benchmark for comparisons. GDP is selected as the research variable that reflects whether an economy is undergoing a recession: According to research, economic recessions can be indicated by two consecutive quarters of GDP growth rate<sup>1</sup>. In this study, the logarithmic differential form (log-diff) of GDP is used as an approximation of the GDP growth rate. The input variables of the models are selected economic factors, which are good indicators of economic activities and might also experience shocks during recessions. Then the researcher fits the time series data of these variables into the model for prediction and evaluates the performance of the model (Huang, 2022).

### 2.1. Models

In this study, the researcher would like to firstly establish a basic benchmark using the traditional time series model VAR. Moreover, machine learning models SVR, Random Forest, and LSTM are also applied.

**VAR** model is the multivariate form extended from the Autoregressive Model (AR). An AR model describes how a time series  $r_t$  can be regressed to its lag orders. For a multivariate time-series vector  $X_p$ , where  $a_t$  is the white noise vector, the VAR (p) can be expressed as:

<sup>1</sup>Referenced from: <https://www.santander.com/>.

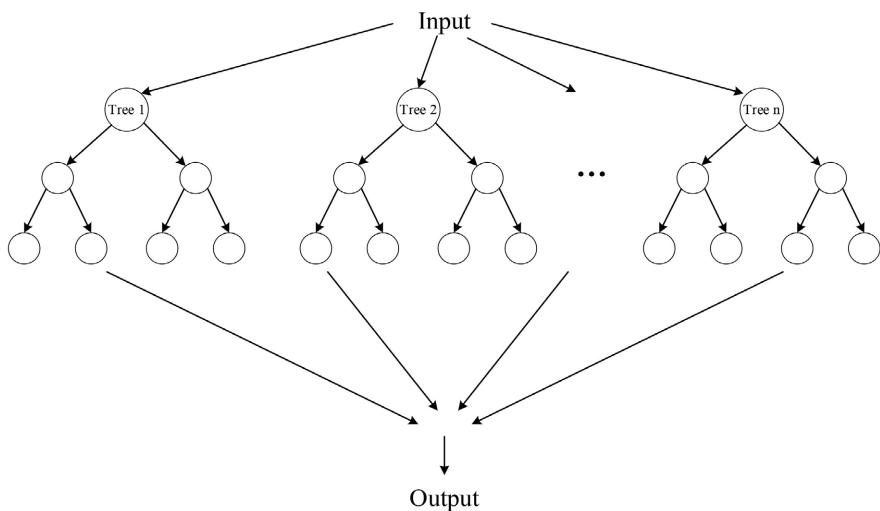
$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + a_t$$

It is worth noting that before building the VAR model, we should ensure that all components of  $X_t$  are stationary, which can be tested by ADF tests. After building the VAR model, Granger Causality tests can be applied to check whether each component has granger causality on other components (Tsay, 2005).

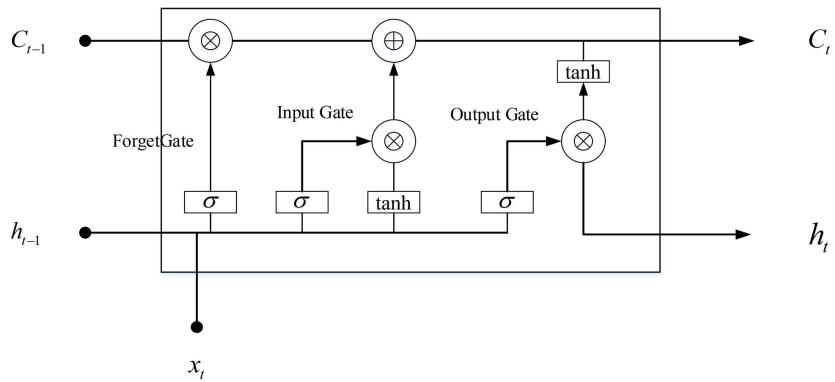
The basic idea of SVR model is that we ensure the point farthest from the hyperplane has a distance  $d$  smaller than the tolerant deviation  $\varepsilon$ , which can be expressed as a constraint optimization. For the support vector regressor, minimizing distance  $d$  can be transformed to the goal of minimizing  $\frac{1}{2}\|\omega\|^2$ . For a linear function  $y = \omega x + b$ , the constraint becomes:  $|y_i - (\omega x_i + b)| \leq \varepsilon, \forall i$ . Improvements including introducing slack variables  $\xi_i$  to the problem or using kernels for mapping are proposed by researchers later (Smola & Schölkopf, 2004; Vapnik, 1999).

**Random Forest** is a bagging (bootstrap aggregating) ensemble learning model made up of multiple decision trees, where each decision tree is an estimator of the ensemble learning. In ensemble learning, predictions of each estimator are integrated with other estimators to form the output. Bootstrapping is a sampling-with-replacement method: by randomly selecting  $n$  samples each time for  $k$  times, we can get  $k$  datasets with a sample size of  $n$ . **Figure 1** shows the basis structure of a random forest (Chu & Qureshi, 2022; Li et al., 2013).

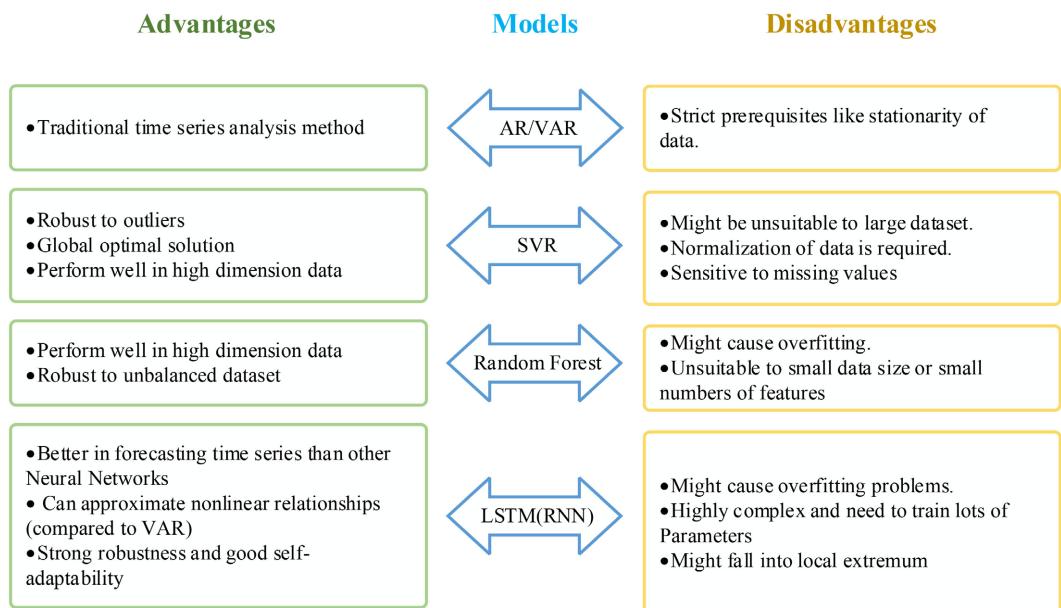
As a type of recurrent neural network (RNN) model, **LSTM** is a machine learning method that is adept at processing time series data. However, a traditional RNN model may cause gradient vanishing problem, making the model unable to transfer long term memories. LSTM has been proposed as an improvement to traditional RNN, since it can store long-term memory by adding a memory unit and using gates to control memory storing and discarding. The three gates of a LSTM model are input gate, output gate and forget gate. Besides a hidden state  $h_t$ , there is also a memory cell state  $C_t$  in the LSTM model. The



**Figure 1.** Basic structure of constructing a random forest.



**Figure 2.** Basic structure of constructing a LSTM model.



**Figure 3.** Advantages and disadvantages of different models.

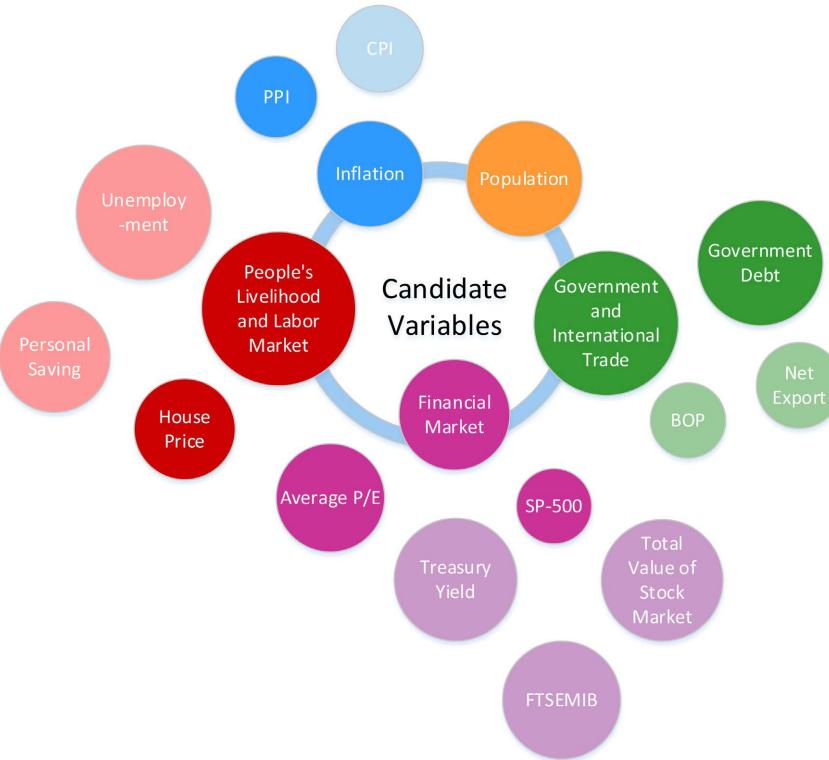
state of cell  $C_t$  is updated with new information  $x_t$  and information passed from  $h_{t-1}$ . The model calculates the information and decides the activation of the gates. **Figure 2** shows the basic structure of a LSTM cell. **Figure 3** shows the comparisons of pros and cons among the four models (Nosratabadi, Mosavi, Duan, Ghamisi, Filip, Band, Reuter, Gama, & Gandomi, 2020; Staudemeyer & Morris, 2019).

## 2.2. Candidate Variables for Models

In order to make accurate forecasts, it is crucial to use input variables which are indicative of economic performance. In the study conducted by Taylan et al., they use economic factors such as GDP, government spendings and productions, and medical data of the pandemic including new cases, vaccinated population, available hospital bed number. In Levanon's study, he examined three groups of indicators: economic activities, including factors like unemployment and

part-time working due to economic reasons; sentimental indicators, including factors like NFIB optimism index and change in consumer comfortable index posted by media; financial indicators, including interest rate swap and LIBOR-Treasury spread (Taylan, Alkabaa, & Yilmaz, 2022; Levanon, 2011).

In this study, economic and financial factors are desired to be the input of the four models by the researcher. These variables are derived from different aspects, which are summarized by the researcher as People's Livelihood and Labor Market, Financial Market, International Trade, Inflation and Population. **Figure 4** shows the specific variables that are included in this study. By incorporating these variables into the models, we can gain a more comprehensive understanding of the economic landscape and make more accurate predictions about future GDP trends. Other non-economic factors, such as people's mobility data, and COVID-19 related situation, are discussed in the following part as a supplement of the economic forecasting models.



**Figure 4.** Candidate economic variables for models.

### 3. Study Based on Google COVID-19 Community Mobility Reports of US in 2020

This study uses Google community mobility data and relevant materials of US in 2020 to gain insights of the impact of public health emergency events like the COVID-19 pandemic on people's daily activities. The level of people's daily activities is also a good indicator of economic situation. Through this research, we may understand how changes in people's mobility may be used to forecast eco-

nomic recessions in a shorter period.

### 3.1. Backgrounds of the Pandemic in US, 2020

US has gone through waves of outbreaks in 2020. **Figure 5** shows new cases of inflections of US in 2020 based on data provided by WHO<sup>2</sup>. As is shown in the figure, the first massive outbreak in US happened in March 2020. The second outbreak happened in early summer and gradually declined. This may be due to stricter prevention measures. However, the situation worsened again in the fourth quarter of 2020 and continued a rising trend until early January 2021<sup>3</sup>.

The COVID-19 pandemic caused an unprecedented fall in economic activities in US. According to Federal Reserve's Monetary Policy Report, over the first half of 2020, the level of US GDP fell a cumulative ten percent and unemployment rate even rose to a post-War high level. The inflation decreased, and the personal consumption expenditures (PCE) also decreased drastically. Financial market was also negatively impacted. The heightened uncertainty and weak demand led many businesses to delay investment plans. In addition, loans became unavailable for firms and consumer borrowings declined since the spending stagnated. Fortunately, as the situation gradually became under control, economic activities recovered gradually, and unemployment rate fell. Although the recover trend slowed down in the late autumn and early winter 2020 with the worsening of the pandemic, the economy resumed. **Figure 6** describes how the sudden hit by the pandemic influences important economic variables GDP, Unemployment Rate, PCE, and SP-500 index. The x-axis stands for dates while the y-axis represents the values of these variables. Start from the first quarter of 2020, there was a sudden decline of GDP and PCE and a dramatic increase in unemployment rate. The stock index also experienced a drop. One piece of good news is that GDP and PCE recovered to previous level by the third quarter of 2020, and the stock price continues to rise. The data are retrieved from FRED<sup>4</sup> ([Board of Governors of the Federal Reserve System, 2020, 2021](#)).

The US government introduced public laws to tackle with health, economic, and people's livelihood issues brought by the pandemic duly. In March 2020, when the nation was firstly struck by the pandemic, CORONAVIRUS PREPAREDNESS AND RESPONSE SUPPLEMENTAL APPROPRIATIONS ACT<sup>5</sup> and FAMILIES FIRST CORONAVIRUS RESPONSE ACT<sup>6</sup> were announced for additional support for CDC, Public Health & Social Services Emergency Fund, and other relevant institutions and funds. The acts aided diagnostics and therapeutics of the virus, aided vaccine development, and provided social assistance like food and

<sup>2</sup>Data Source: WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: <https://covid19.who.int>.

<sup>3</sup>Source: International Monetary Fund. Policy Responses to COVID-19.

<https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19#U>.

<sup>4</sup>Data Source: <https://fred.stlouisfed.org/series/>.

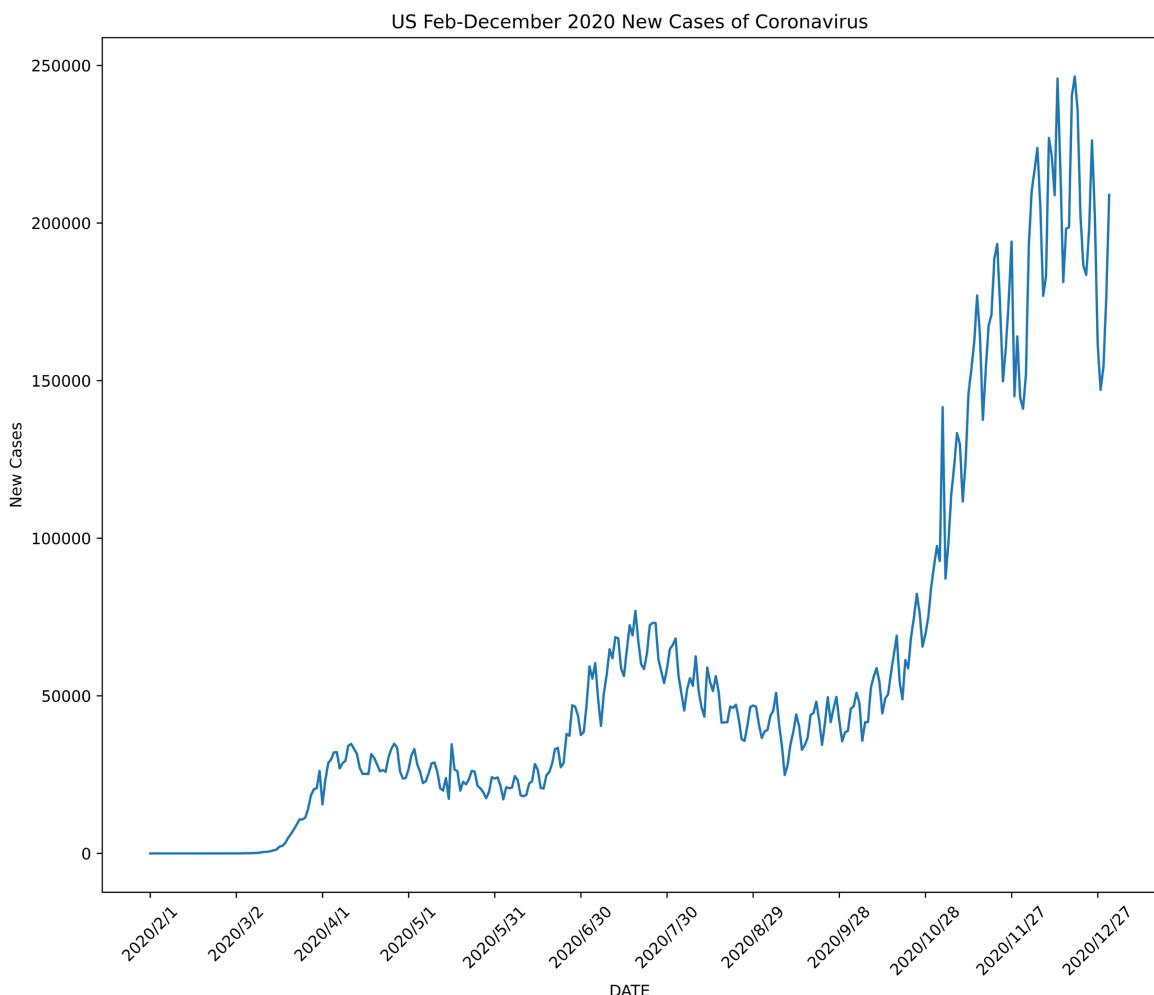
<sup>5</sup>Source: <https://www.govinfo.gov/content/pkg/PLAW-116publ123/pdf/PLAW-116publ123.pdf>.

<sup>6</sup>Source: <https://www.govinfo.gov/content/pkg/PLAW-116publ127/pdf/PLAW-116publ127.pdf>.

insurances. CARES Act<sup>7</sup> intensively provided details of protection for laborers, assistance or compensation for families and businesses. It supported American's health care system and fought against the disease. More acts including those aim at extending the supports and stabilizing financial market were announced throughout the battles with the pandemic. **Figure 7** presents the timeline of some of the laws announced by the US government during the pandemic.

### 3.2. Community Mobility Data

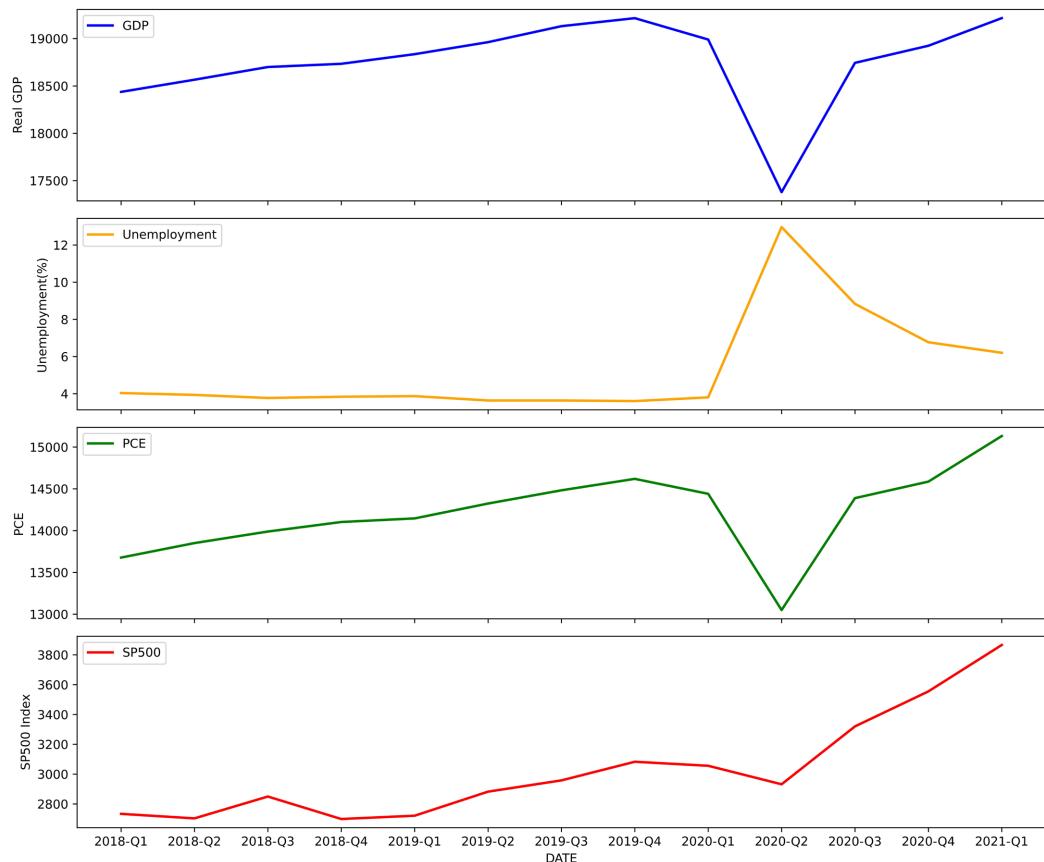
Google COVID-19 Community Report<sup>8</sup> shows changes in people's visits to different categories of places. The data include items of Retail and Recreation, Grocery and Pharmacy, Park, Transit Station, Workplaces, and Residential. The data are collected from 2020 February to 2020 December daily, and are all observed in the form of percentage change from baseline, showing changes of these factors over time.



**Figure 5.** New cases of inflections in the US from Feb, 2020 to December, 2020.

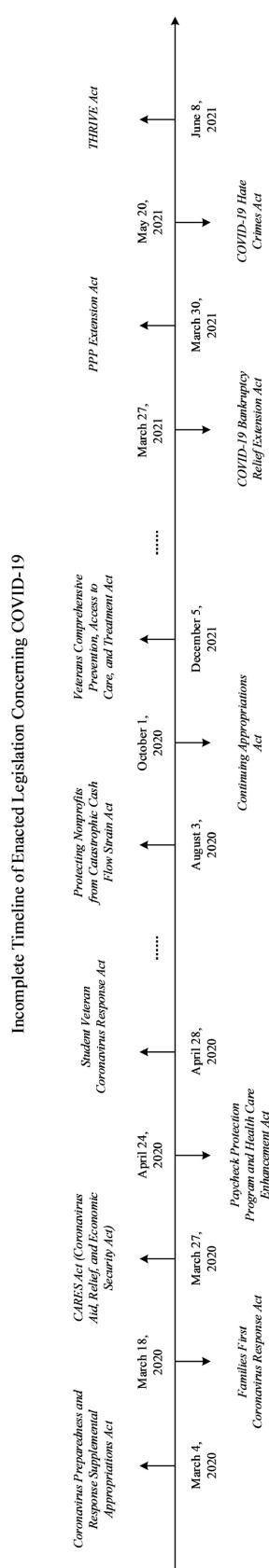
<sup>7</sup>Source: <https://www.govinfo.gov/content/pkg/BILLS-116hr748enr/pdf/BILLS-116hr748enr.pdf>.

<sup>8</sup>Data Source: <https://www.google.com/covid19/mobility/>.



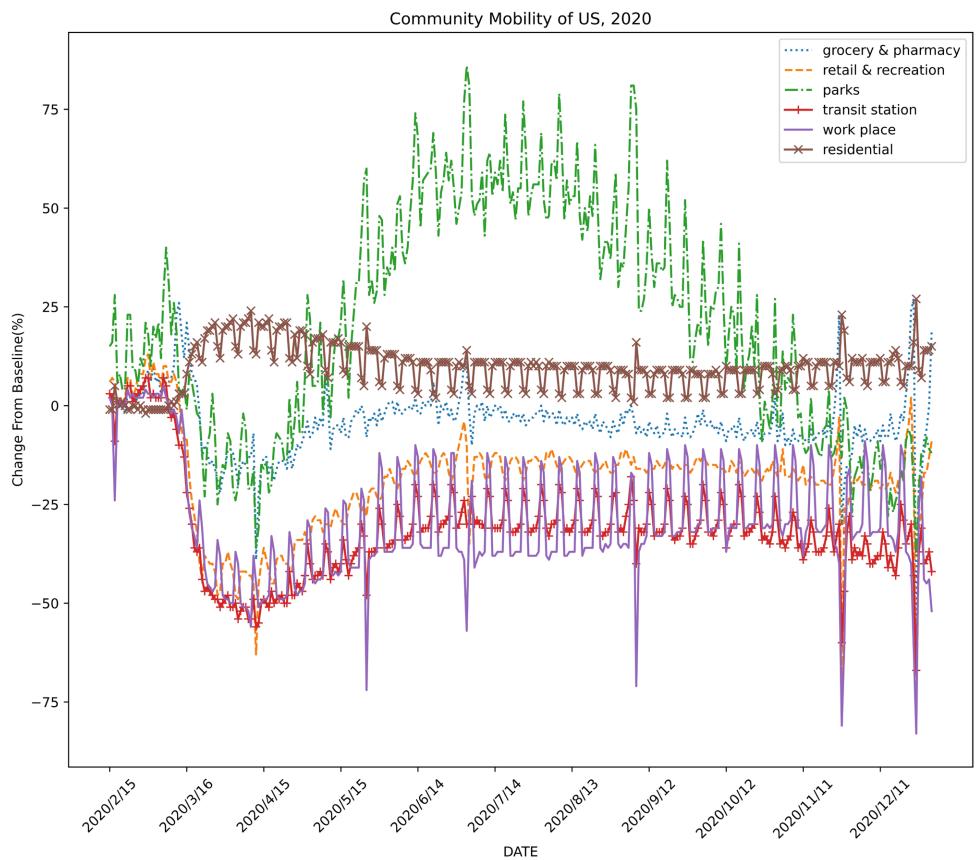
**Figure 6.** Major economic indexes before and during the pandemic.

**Figure 8** depicts the trend graph of the data, the x-axis stands for the date, while the y-axis stands for the values of the observed variables. According to the graph, except for “residential change”, other factors went through a fierce drop in March, when US first combated with COVID-19, and “residential change” increased to a positive rate then. Since “residential” measures the duration people stay in residences, it is explicable that instead of visiting other places, people spent more time isolating at home. After the drop, factors except for “residential change” began to recover since the early April. This might be attributed to acts passed by the government since early March, and the nation built up a defense system to better react to COVID-19. The recovery lasted about three months. Compared to the previous level, “parks change” soon increased drastically about two to four times. This might be caused by restricted indoor activities, and people turned to parks to soothe their discontents caused by COVID-19. Other factors that had dropped, including “transit station change”, “retail and recreation change”, “grocery and pharmacy change”, and “workplace change” remained at a negative rate and did not recover to the previous positive level in 2020. What’s more, the rates fluctuated fiercely twice in November and December, which agreed with the new waves of infections in the second half of year 2020. It took less and less time to dissipate each shock, probably because people were more experienced in dealing with the problems.



Note: The graph only presents an incomplete collection of laws passed in US regarding COVID-19

**Figure 7.** US laws reacting to COVID-19 pandemic (part).



**Figure 8.** Visualization of patterns of community mobility variables.

#### 4. Empirical Study of US Data

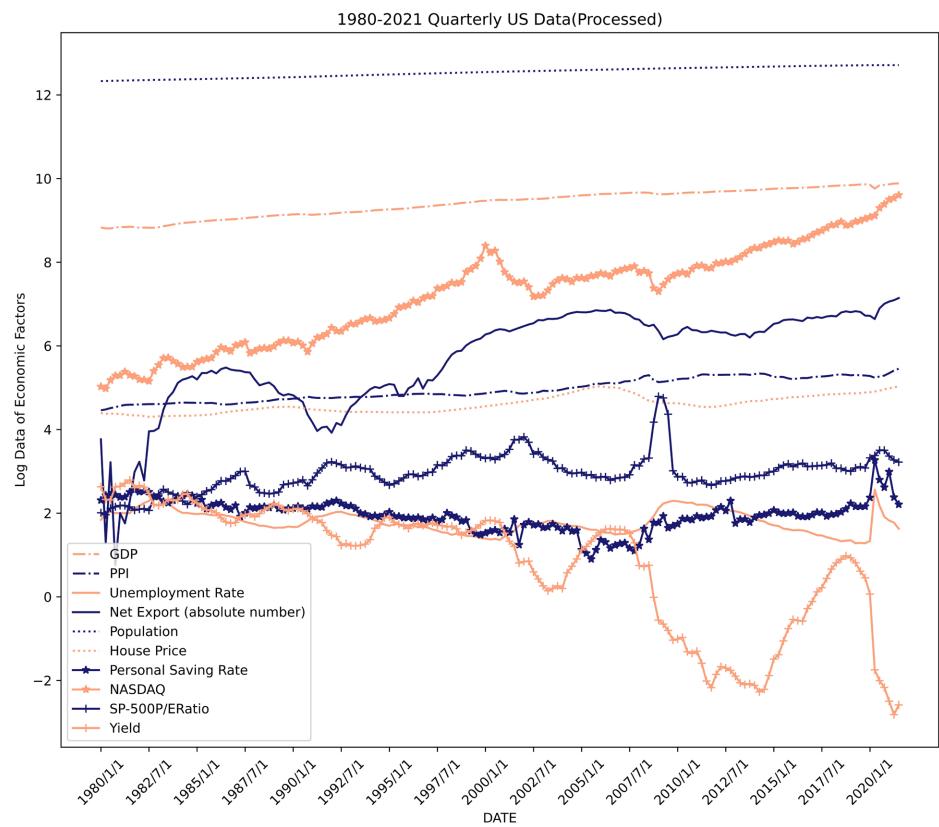
The study would like to investigate the performance of machine learning models for predicting recessions in different countries. To obtain representative cases, the models will first be applied to US data since the US is an influential economy. Subsequently, the four models will be applied to Italy data, which serves as an example from the Eurozone.

##### 4.1. Data

All the data are retrieved from FRED, Federal Reserve Bank of St. Louis<sup>9</sup>, except for the data of SP-500 P/E Ratio, which are retrieved from Macrotrends<sup>10</sup>. The data are collected on a quarterly basis, from the first quarter of 1980(1980-Q1) to the third quarter of 2021 (2021-Q3). The original data are transformed to log differences. The researcher uses the absolute number of the Net Export data, since the original data are always negative. **Figure 9** shows how these economic indicators change over the forty years, with the y-axis representing the logarithmic form of the processed data. **Table 1** provides an overview of the original data.

<sup>9</sup>Data Source: <https://fred.stlouisfed.org/series/>.

<sup>10</sup>Data Source: <https://www.macrotrends.net/>.



**Figure 9.** Economic data (log) of GDP and other indicators.

**Table 1.** Variables used in the study.

| Variable Name    | Notes   |
|------------------|---|
| GDP              | Real GDP, Units: Billions of Chained 2012 Dollars, Seasonally Adjusted, Data Source: FRED   |
| PPI              | Producer Price Index by Commodity: All Commodities, Index 1982 = 100, Not Seasonally Adjusted, Data Source: FRED  |
| House Price      | Real Residential Property Prices for United States, Index 2010 = 100, Not Seasonally Adjusted, Data Source: FRED  |
| Population       | Units: Thousands, Data Source: FRED   |
| Net Export       | Real Net Exports of Goods and Services, Units: Billions of chained 2012 Dollars, Seasonally Adjusted, Data Source: FRED                                       |
| Unemployment     | Unemployment Rate, Units: Percent, Seasonally Adjusted, Data Source: FRED   |
| Personal Saving  | Personal Saving Rate, Units: Percent, Seasonally Adjusted, Data Source: FRED  |
| Treasury Yield   | Market Yield on US Treasury Securities at 1-Year Constant Maturity, Quoted on an Investment Basis, Units: Percent, Not Seasonally adjusted, Data Source: FRED |
| SP-500 P/E Ratio | SP-500 P/E Ratio, Data Source: Macrotrends  |
| NASDAQ Index     | NASDAQ Composite Index, Close Price, Index Feb 5, 1971 = 100, Not Seasonally Adjusted, Data Source: FRED  |

## 4.2. Model Performances

Quarterly data spanning from 1980 to 2019-Q4 are used to construct models. The data are used in log difference forms. Instead of a random train-test split, the researcher uses a sequential split on the dataset with a 0.9:0.1 ratio to divide a validation set for the time series data. To construct the VAR model, ADF tests are applied to drop the factors with non-stationary time series. Factors “Population”, “House Price” and “Government Debt” are excluded from the model based on the ADF tests. All factors, except for “Treasury Yield”, are used to construct the SVR model. Additionally, the data are normalized when they are used to build the SVR and LSTM. To make optimal performance of the machine learning models, the researcher uses simple alterations in parameters and Grid Search to tune the parameters. The parameters used in the models are provided in the appendix.

The study would like to find out how the four models VAR(2), SVR, RF, and LSTM perform in the predictions. To evaluate their performance, the researcher uses two common metrics, MSE and  $R^2$ , which provide a general idea of how closely the predictions match the actual values. The metrics are calculated based on log differential data. On the training set, Random Forest outperforms the other three models with a large  $R^2$  of about 0.77 and a small MSE about  $1.18e^{-6}$ . Among the remaining models, LSTM performs better than VAR and SVR with a  $R^2$  of about 0.59. However, it turns out that none of the models performs well in future predictions since  $R^2$  on testing set are negative. Despite of this, the traditional time series model VAR(2) shows relatively better performance among the remaining models (**Table 2**).

The DM test is widely used to compare the performance of different models on time series data. The researcher applies the DM test on all training and testing set predictions. The results given by **Table 3** reveal that RF outperforms VAR, which is evidenced by a negative DM-statistic and a p-value smaller than 0.05. On the other hand, the performance of SVR and LSTM is comparable to that of VAR and RF. Since training set predictions make up of a major proportion of total predictions, the results from the DM test are consistent with the results given by the evaluation metrics.

## 4.3. Model Predictions

Among all the predictions, the primary interest of the study lies in examining forecasts of economic recessions. To this end, the study firstly makes in-sample predictions of log GDP during the US financial crisis from 2008 to 2009. **Figure 10** shows the predictions given out by the four models and the true GDP. Also, the study presents the out-of-sample predictions from 2020-Q1 to 2021-Q2, when the country was hit by the COVID-19 pandemic. The predictions are compared with the true log GDP and among all the other models, which are shown in **Figure 11**. **Table 4** provides the out-of-sample predictions of log GDP using the four models.

**Table 2.** Model evaluation metrics.

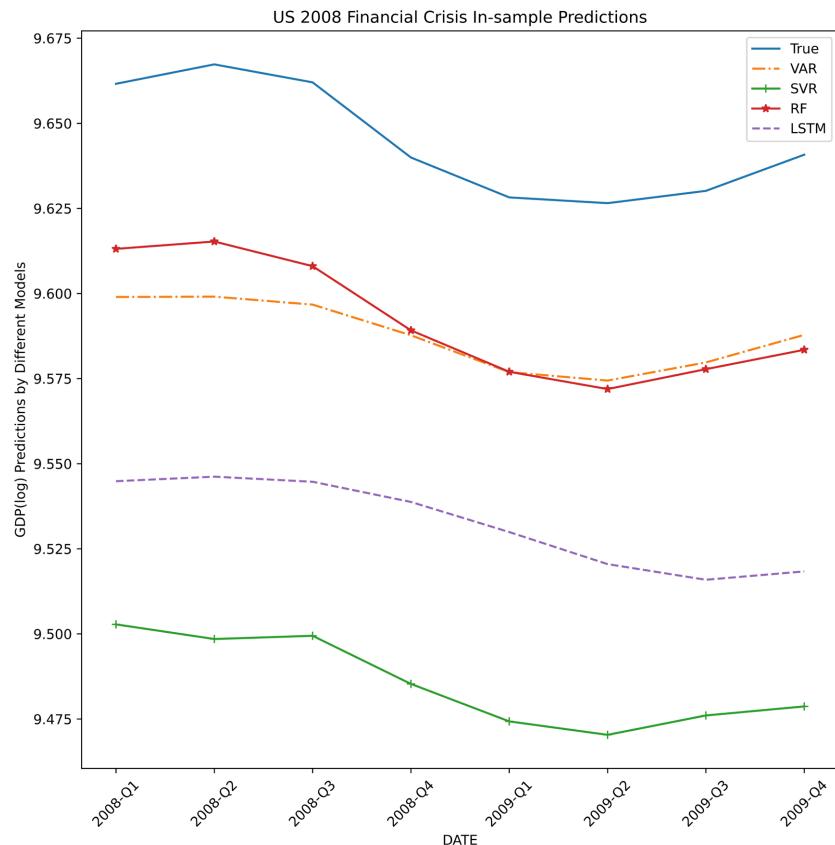
|                        | <b>VAR</b>              | <b>SVR</b>              | <b>RF</b>               | <b>LSTM</b>             |
|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| MSE(Train)             | 3.433146e <sup>-5</sup> | 3.078222e <sup>-5</sup> | 1.186559e <sup>-6</sup> | 2.274719e <sup>-5</sup> |
| R <sup>2</sup> (Train) | 0.350172                | 0.417352                | 0.775407                | 0.593098                |
| MSE(Test)              | 5.607217e <sup>-6</sup> | 1.110079e <sup>-5</sup> | 8.340478e <sup>-6</sup> | 4.244879e <sup>-5</sup> |
| R <sup>2</sup> (Test)  | -0.336931               | -1.646767               | -0.988624               | -0.956968               |

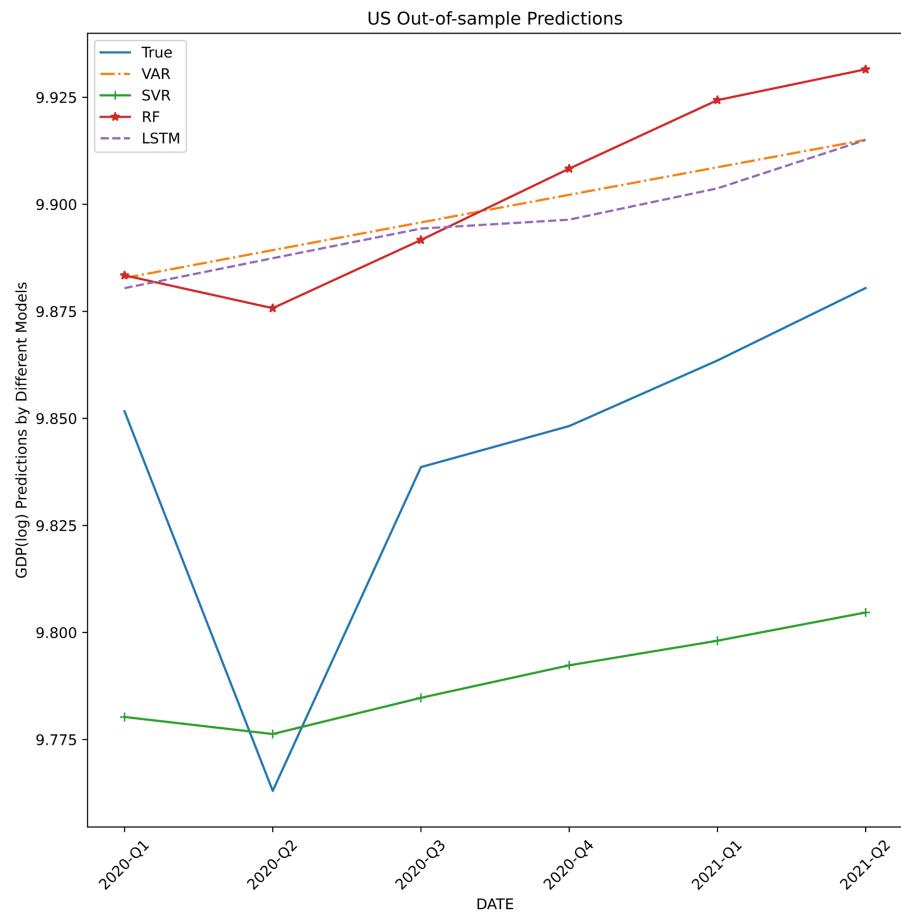
**Table 3.** Model DM tests based on VAR.

|          | <b>DM-statistic</b> | <b>p-value</b>          |
|----------|---------------------|-------------------------|
| SVR-VAR  | 0.031157            | 0.512406                |
| RF-VAR   | -5.346800           | 1.689180e <sup>-7</sup> |
| LSTM-VAR | -0.059851           | 0.476177                |

**Table 4.** 2020-2021 out-of-sample Log GDP Predictions.

|         | <b>True</b> | <b>VAR</b> | <b>SVR</b> | <b>RF</b> | <b>LSTM</b> |
|---------|-------------|------------|------------|-----------|-------------|
| 2020-Q1 | 9.851661    | 9.882786   | 9.780258   | 9.883398  | 9.880417    |
| 2020-Q2 | 9.763001    | 9.889290   | 9.776273   | 9.875734  | 9.887403    |
| 2020-Q3 | 9.838614    | 9.895765   | 9.784733   | 9.891658  | 9.894327    |
| 2020-Q4 | 9.848200    | 9.902215   | 9.792325   | 9.908325  | 9.896402    |
| 2021-Q1 | 9.863510    | 9.908642   | 9.798066   | 9.924319  | 9.903719    |
| 2021-Q2 | 9.880436    | 9.915050   | 9.804663   | 9.931520  | 9.915029    |

**Figure 10.** In-sample predictions during the financial crisis (2008-2009).



**Figure 11.** Out-of-sample predictions during the pandemic (2020-2021).

In both in-sample and out-of-sample predictions, it is evident that the predictions can reflect a rough trend of the data. Nevertheless, during the COVID-19 pandemic, the out-of-sample predictions fall short as all models fail to reflect the violent decline caused by the sudden shock of the pandemic outbreak, and the time series-based model VAR(2) and LSTM are ineffective in showing signs of shocks, as demonstrated by the graph. This could be attributed to the fact that all the models are trained on mostly regular data, making them relatively “obtuse” in this extreme situation. Additionally, the time-series-based model VAR and LSTM are heavily reliant on historical data, which restricts their ability to predict the sudden decline, since the economy was relatively stable and growing steadily before the COVID-19 shock.

## 5. Empirical Study of Italy Data

In order to assess the robustness of the prediction models, the analysis is extended to include the economy of Italy, a member of the Euro area. By applying the same models used in the US case, we can compare model performances in different economies and identify the similarities and differences in both model applications.

## 5.1. Data

The data for the study of Italy are retrieved from Eurostat<sup>11</sup>, Bank of Italy Statistical Database<sup>12</sup>, Wind Database<sup>13</sup>, OECD<sup>14</sup>, and FRED, Federal Reserve Bank of St. Louis<sup>15</sup>. **Table 5** shows the variables that are used in this study.

In order to make reasonable comparisons, resembling variables are chosen for the studies of the two cases. The data of US are available for a longer period of forty years, whereas the data of Italy are from 1995 (1995-Q1) to 2021(2021-Q3) on a quarterly basis. The original data are processed before they are fit into the models. It is worth noticing that the value of variable BOP is converted to BOP + 30,000, and the value of variable Treasury Yield is converted to Treasury Yield + 2. The data are further transformed to log differences, and missing values are filled to ensure completeness of the dataset. **Figure 12** shows how these economic indicators change over these years. The y-axis stands for the logarithmic form of the processed data. **Table 5** provides an overview of the original data.

**Table 5.** Variables for the study.

| Variables       | Notes   |
|-----------------|---|
| GDP             | Real GDP, Units: Millions of Chained 2010 Euros, Seasonally Adjusted, Data Source: Eurostat   |
| PPI             | Domestic Producer Prices Index: Manufacturing for Italy, Index 2015 = 100, Not Seasonally Adjusted, Data Source: Eurostat               |
| House Price     | OCED Real House Price Index, Index 2015 = 100, Seasonally Adjusted, Data Source: OECD, retrieved from retrieved from FRED               |
| Population      | Units: Thousands, Not Seasonally Adjusted, Data Source: Eurostat  |
| BOP             | Balance of Payment BMP6 (Current Account), Units: Euros, Not Seasonally Adjusted, Sum over component months, Data Source: Bank of Italy |
| Unemployment    | Harmonized Unemployment Rate, Units: percent, Seasonally Adjusted, Data Source: OECD  |
| Government Debt | General Government Gross Debt, Units: Euros, Data Source: Bank of Italy   |
| FTSEMIB         | FTSEMIB Close Price, Units: Euros, Data Source: Wind  |
| Yield           | Gross yield of benchmark CTZ, Units: Euros, Type of security: Zero coupon bonds with maturity within 1 year, Data Source: Bank of Italy |

<sup>11</sup>Data Source:

[https://ec.europa.eu/eurostat/databrowser/explore/all/economy?lang=en&subtheme=prc&display=list&sort=category&extractionId=PRC\\_HICP\\_MIDX\\_custom\\_3378783](https://ec.europa.eu/eurostat/databrowser/explore/all/economy?lang=en&subtheme=prc&display=list&sort=category&extractionId=PRC_HICP_MIDX_custom_3378783).

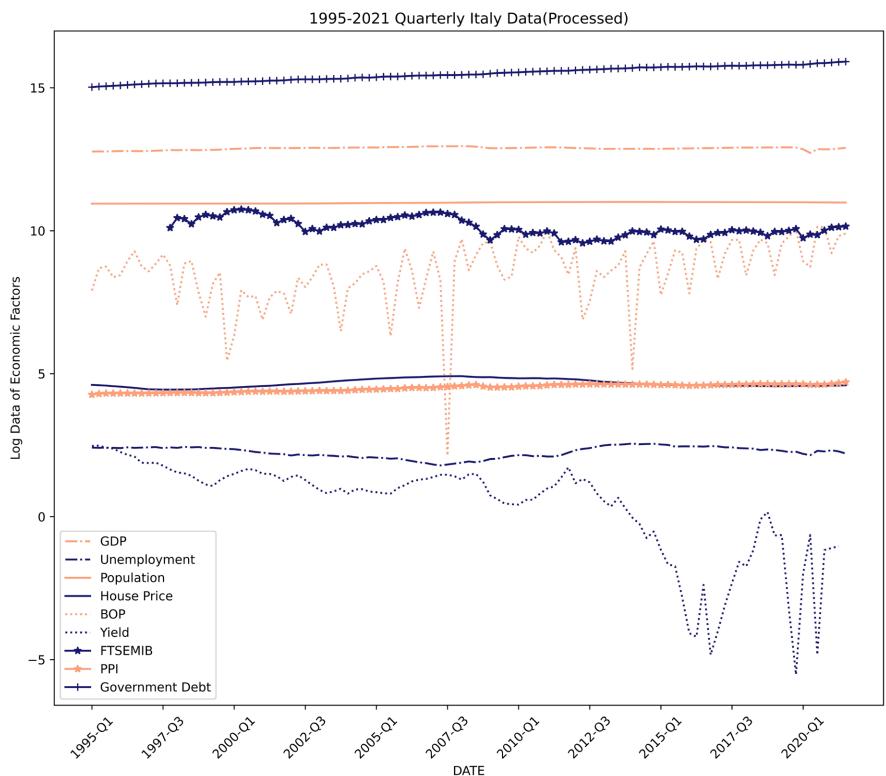
<sup>12</sup>Data Source:

[https://infostat.bancaditalia.it/inquiry/home?spyglass/taxo:CUBESET=&ITEMSELEZ=&OPEN=true&ep:LC=EN&COMM=BANKITALIA&ENV=LIVE&CTX=DIFF&IDX=1&/view: CUBEIDS="](https://infostat.bancaditalia.it/inquiry/home?spyglass/taxo:CUBESET=&ITEMSELEZ=&OPEN=true&ep:LC=EN&COMM=BANKITALIA&ENV=LIVE&CTX=DIFF&IDX=1&/view: CUBEIDS=)

<sup>13</sup>Wind Database is a commercial financial information database that is widely used in China and Worldwide. The data of FTSEMIB close price are retrieved from Wind

<sup>14</sup>Data Source: <https://data.oecd.org/>.

<sup>15</sup>Data Source: <https://fred.stlouisfed.org/series/>.



**Figure 12.** Economic data (log) of GDP and other indicators.

## 5.2. Model Performances

Quarterly data spanning from 1995-Q4 to 2019-Q4 are used to build models and make in-sample predictions. The researcher uses a sequential split on dataset with a ratio of 0.9:0.1 instead of a random train-test split. The hyper-parameters of models are displayed in the following table. To construct the VAR model, ADF tests are applied to drop the factors with non-stationary time series. Factors “Population”, “House Price” and “Government Debt” are excluded from the model according to the ADF test. All factors, except for “Treasury Yield”, are used to construct the SVR model. Additionally, the data are normalized when they are used to build the SVR and LSTM.

The evaluation metrics MSE and  $R^2$ , as well as the DM-test, reveal similar results in both the US and Italy cases. According to the evaluation metrics results displayed in **Table 6**, among the four models, Random Forest outperforms the others on the training set, while VAR(2) is the only model with a positive  $R^2$  on the testing set. The DM test results shown in **Table 7** indicate that the RF model outperforms the VAR, whereas no significant difference is observed between the other models, which aligns with the results concluded from the training set. The comparisons of performance among models in the case of Italy are consistent with those in the case of the US, suggesting that in the time-series predictions, the machine learning models, even the LSTM, which is based on time series, should not be considered superior to traditional VAR in sequence time series predictions.

**Table 6.** Model evaluation metrics.

|                        | <b>VAR</b>              | <b>SVR</b>              | <b>RF</b>               | <b>LSTM</b>             |
|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| MSE(Train)             | 3.068620e <sup>-5</sup> | 2.163874e <sup>-5</sup> | 7.412359e <sup>-6</sup> | 2.206200e <sup>-5</sup> |
| R <sup>2</sup> (Train) | 0.419433                | 0.590606                | 0.859762                | 0.605355                |
| MSE(Test)              | 1.071295e <sup>-6</sup> | 3.435431e <sup>-6</sup> | 1.800640e <sup>-5</sup> | 5.029789e <sup>-6</sup> |
| R <sup>2</sup> (Test)  | 0.161082                | -1.690241               | -0.410057               | -1.318826               |

**Table 7.** Model DM tests based on VAR.

|          | <b>DM-statistic</b> | <b>p-value</b>          |
|----------|---------------------|-------------------------|
| SVR-VAR  | -0.720420           | 0.236556                |
| RF-VAR   | -4.223023           | 2.853033e <sup>-5</sup> |
| LSTM-VAR | -1.143467           | 0.127921                |

### 5.3. Model Predictions

Like the US case, the researcher makes in-sample predictions of log GDP during the Euro Crisis and presents out-of-sample predictions from 2020-Q1 to 2021-Q2, which cover the period when the pandemic took place. **Figure 13** and **Figure 14** show respectively the in-sample predictions and out-of-sample predictions generated by the models. **Table 8** provides the out-of-sample predictions of log GDP using the four models.

The performance of model prediction resembles that of the US case. The results in both cases indicate that the models can show the trend of how the economy changes, although predictions of the four models deviate from the true values. However, in the case of Italy, the models can't signal the drastic decline in 2020-Q2 in the out-of-sample prediction clearly.

## 6. Summaries and Limitations of Models

It has been shown that Random Forest stands out from the four models in the training set, where the rest of models have no significant difference, and that VAR is relatively better than the other models in the testing test (using the same training and testing set, which is sequentially split from the time-series). Since both the training and testing sets are composed of economic data from “stable” periods (except for the 2008 financial crisis period, which is also in the training set), therefore, it can be inferred that the Random Forest model is a favorable option for economic forecasting purposes. In addition, we may not tell significant improvement from traditional time series models VAR to machine learning models like the SVR or LSTM according to the results of the study. However, the traditional time series model VAR does have limitations since VAR relies on several previous steps before the prediction, therefore it might not be flexible concerning a sudden shock while machine learning models (except for LSTM) are not constrained by time period. Another major limitation for a VAR(p) model is that the forward forecast would be replaced with the mean value after p steps. This would also trouble the predictions for long term predictions without new data added.

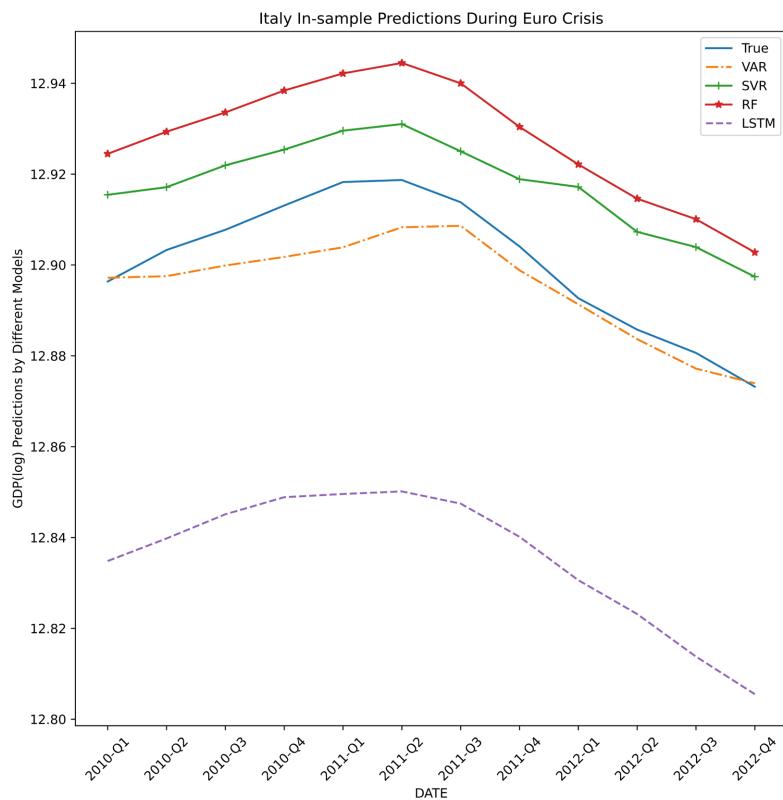


Figure 13. In-sample predictions during the Euro Crisis (2010-2012).

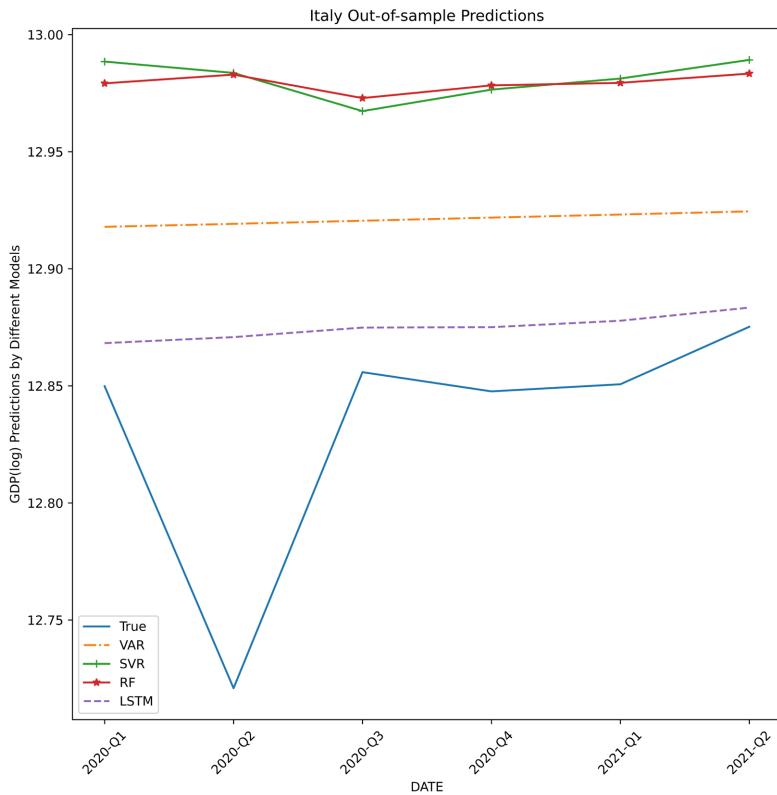


Figure 14. Out-of-sample predictions during the pandemic (2020-2021).

**Table 8.** 2020-2021 out-of-sample Log GDP Predictions.

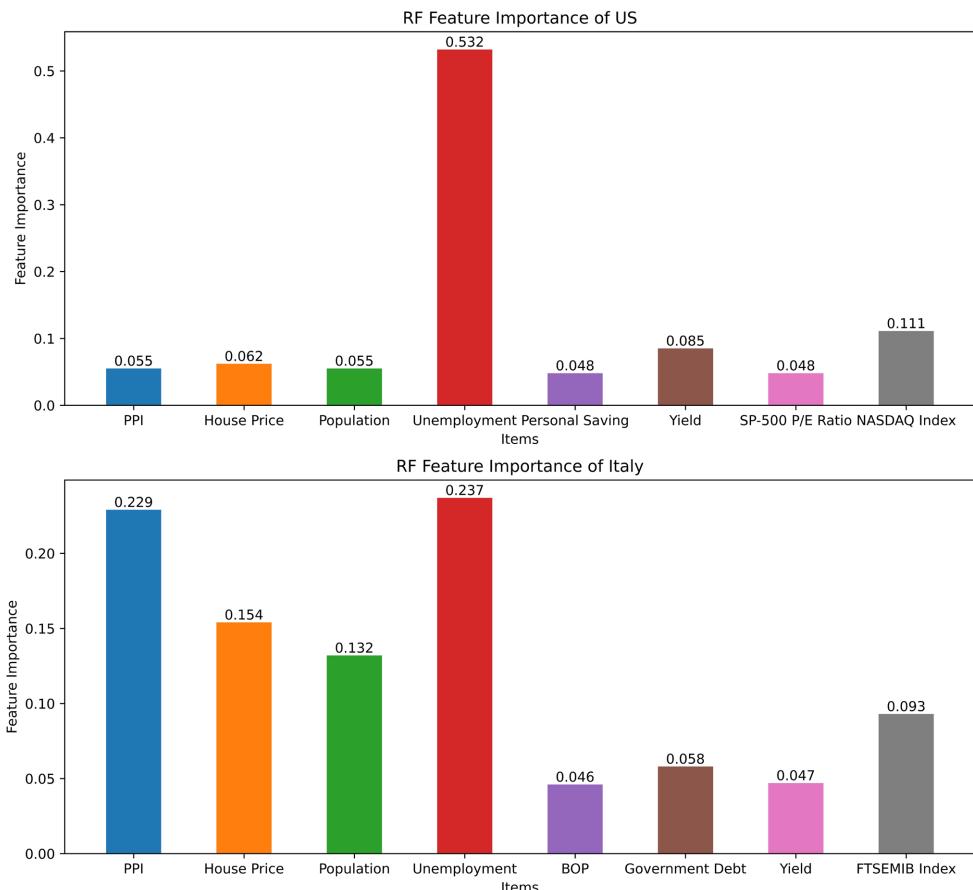
|         | <b>True</b> | <b>VAR</b> | <b>SVR</b> | <b>RF</b> | <b>LSTM</b> |
|---------|-------------|------------|------------|-----------|-------------|
| 2020-Q1 | 12.849854   | 12.917918  | 12.988490  | 12.979188 | 12.868252   |
| 2020-Q2 | 12.720877   | 12.919184  | 12.983636  | 12.982914 | 12.870817   |
| 2020-Q3 | 12.855849   | 12.920517  | 12.967327  | 12.972894 | 12.874887   |
| 2020-Q4 | 12.847644   | 12.921848  | 12.976512  | 12.978322 | 12.875054   |
| 2021-Q1 | 12.850678   | 12.923166  | 12.981234  | 12.979402 | 12.877815   |
| 2021-Q2 | 12.875225   | 12.924512  | 12.989196  | 12.983350 | 12.883411   |

Another drawback of these models is that the fitting degrees of predictions, especially on the testing set, are not satisfactory. In both the US and Italy cases, according to the evaluation metrics, the four models can approximate relatively well on the training set, but they indeed give out imprecise predictions on the testing set. The failure of testing data predictions is probably caused by the simple sequential split of testing data: For time-series-based models VAR and LSTM model, only a sequential split of the time-series data is reasonable and valid. Since the testing data are the last ten percent of the quarterly time series data, the models are built using outdated data, at best a few years before the testing data, making our models outdated for the constant changing economy. Unlike VAR and LSTM model, ordinary machine learning models SVR and RF allow for a random split in the data set, therefore the researcher tries to use a random split. As is shown in **Table 9**, for the machine learning models RF and SVR, the models using a random split training data with the same train-test ratio exhibit a better performance on the testing set. We can conclude that the models are quite unstable. It also raises a concern that our models should be always updated with the latest data, since an economy undergoes structural changes over time. In order to have more accurate predictions, more intricate enhancements of the basic models are suggested.

In speak of the economic sense provided in the study, the researcher would like to see the importance of each economic variable in building the model. **Figure 15** displays the feature importance given out by the Random Forest Models. From the above results we've known that the RF model performs predictions on real GDP growth almost comparably well in both countries. For specific factors, in the US case unemployment rate has a major importance on predicting real GDP, and NASDAQ Index also has a relatively high feature importance. Whereas in the Italy case, factors including Unemployment, PPI, House Price, and Population all have relatively high feature importance. According to Okun's Law, an empirical relationship between real GDP and Unemployment Rate has been proposed: A 1% increase in real GDP matches to a 0.5% decline in the unemployment rate. It can be explained by the idea that a deviation of unemployment rate from the natural rate of unemployment results from not fully optimizing production factors (real GDP is considered as the national production). Based on the results of this study, it can be concluded that unemployment has a significant impact on real GDP growth. Another limitation is that the study

could benefit from a more comprehensive selection of variables for the models, and incorporating a greater number of economic features may make machine learning models more advantageous (Okun, 1962).

When comparing the quality of predictions between two countries, although the Italian model was trained on a shorter period of data, we should not assume the predictions made by the Italian models are inferior to those of the US models. In fact, the evaluation metrics of both cases suggest that dropping “older” data may result in better performance for the models. Therefore, it’s possible that a shorter period of data may have contributed to its favorable performance.



**Figure 15.** Feature importance given out by the random forest models.

**Table 9.** Evaluation metrics for the US data using different set splits.

|                        | SVR(s) <sup>16</sup>    | SVR(r) <sup>17</sup>    | RF(s)                   | RF(r)                   |
|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| MSE (Train)            | 3.078222e <sup>-5</sup> | 3.024604e <sup>-5</sup> | 1.186559e <sup>-6</sup> | 7.147493e <sup>-6</sup> |
| R <sup>2</sup> (Train) | 0.417352                | 0.377378                | 0.775407                | 0.861080                |
| MSE (Test)             | 1.110079e <sup>-5</sup> | 2.855674e <sup>-5</sup> | 8.340478e <sup>-6</sup> | 1.969927e <sup>-5</sup> |
| R <sup>2</sup> (Test)  | -1.646767               | 0.328100                | -0.988624               | 0.386786                |

<sup>16</sup>“s” stands for sequential train test split.

<sup>17</sup>“r” stands for random train test split.

## 7. Conclusion

In conclusion, predicting future economic trend is always challenging work. This study sets out to use machine learning models to forecast economic recessions. The study has selected some relevant economic variables and fitted their historic data into models including VAR, SVR, Random Forest, and LSTM models. By comparing the performance of these models in two different countries, it is found that machine learning models do not necessarily give out better predictions than traditional VAR model, even though VAR model has more constraints with the time series. Limitations have been put forward and hopefully, more advanced research concerning the topic would be implemented by other researchers soon.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). *Covid-Induced Economic Uncertainty*. Technical Report, National Bureau of Economic Research.  
<https://doi.org/10.3386/w26983>
- Board of Governors of the Federal Reserve System (2020, June 12). *Monetary Policy Report*.  
[https://www.federalreserve.gov/monetarypolicy/files/20200612\\_mprfullreport.pdf](https://www.federalreserve.gov/monetarypolicy/files/20200612_mprfullreport.pdf)
- Board of Governors of the Federal Reserve System (2021, February 19). *Monetary Policy Report*.  
[https://www.federalreserve.gov/monetarypolicy/files/20210219\\_mprfullreport.pdf](https://www.federalreserve.gov/monetarypolicy/files/20210219_mprfullreport.pdf)
- Chetty, R., Friedman, J. N., Hendren, N., Stepner, M. et al. (2020). *The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data*. Technical Report, National Bureau of Economic Research.  
<https://doi.org/10.3386/w27431>
- Chu, B., & Qureshi, S. (2022). Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast US GDP Growth. *Computational Economics*.  
<https://doi.org/10.1007/s10614-022-10312-z>
- Connaughton, J. E. (2010). Local Economic Impact of the Great Recession of 2008/2009. *Review of Regional Studies*, 40, 1-4. <https://doi.org/10.52324/001c.8157>
- Huang, Z. (2022). Research on Economic Data Analysis and Intelligent Prediction Based on BP Neural Network. In *Proceedings of the 3rd Asia-Pacific Conference on Image Processing, Electronics and Computers* (pp. 878-882). The Association for Computing Machinery. <https://doi.org/10.1145/3544109.3544372>
- Kuyo, M., Mwalili, S., & Okang'o, E. (2021). Machine Learning Approaches for Classifying the Distribution of Covid-19 Sentiments. *Open Journal of Statistics*, 11, 620-632.  
<https://scirp.org/Journal/paperinformation.aspx?paperid=112294>  
<https://doi.org/10.4236/ojs.2021.115037>
- Levanon, G. (2011). Forecasting Recessions and Slow-Down Probabilities with Markov Switching Probabilities as Right-Hand-Side Variables. *Business Economics*, 46, 99-110.  
<https://doi.org/10.1057/be.2011.8>

- Li, X. H. et al. (2013). Using “Random Forest” for Classification and Regression. *Chinese Journal of Applied Entomology*, 50, 1190-1197.
- Liu, Y. S., & Tang, A. Y. (2022). Prediction Method of Government Economic Situation Based on Big Data Analysis. *Digital Government: Research and Practice*, 3, 1-16. <https://doi.org/10.1145/3563042>
- Ludvigson, S. C., Ma, S., & Ng, S. (2020). *COVID-19 and the Macroeconomic Effects of Costly Disasters*. Technical Report, National Bureau of Economic Research. <https://doi.org/10.3386/w26987>
- Malladi, R. K. (2022). Application of Supervised Machine Learning Techniques to Forecast the COVID-19 US Recession and Stock Market Crash. *Computational Economics*. <https://doi.org/10.1007/s10614-022-10333-8>
- Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S. S., Reuter, U., Gama, J., & Gandomi, A. H. (2020). Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics*, 8, Article No. 1799. <https://doi.org/10.3390/math8101799>
- Okun, A. M. (1962). *Potential GNP: Its Measurement and Significance*. Cowless Foundation.
- Primiceri, G. E., & Tambalotti, A. (2020). *Macroeconomic Forecasting in the Time of COVID-19* (pp. 1-23). Manuscript, Northwestern University.
- Smola, A. J., & Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing*, 14, 199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Staudemeyer, R. C., & Morris, E. R. (2019). *Understanding LSTM—A Tutorial into Long Short-Term Memory Recurrent Neural Networks*.
- Taylan, O., Alkabaa, A. S., & Yilmaz, M. T. (2022). Impact of COVID-19 on G20 Countries: Analysis of Economic Recession Using Data Mining Approaches. *Financial Innovation*, 8, Article No. 81. <https://doi.org/10.1186/s40854-022-00385-y>
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons. <https://doi.org/10.1002/0471746193>
- Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4757-3264-1>

## Appendices

### Hyper-Parameters Used in the Models

**Table A1** and **Table A2** shows the parameters that have been tuned. The remaining parameters that have not been adjusted are currently using default values within the functions.

**Table A1.** US model parameters.

| Model | Hyper-Parameter  |
|-------|--|
| VAR   | lag = 2  |
| SVR   | kernel = "sigmoid", gamma = 0.001, C = 25  |
| RF    | n_estimators = 20, max depth = 5   |
| LSTM  | epochs = 300, batch size = 10, optimizer = "adam", n_past = 12 Layers:<br>LSTM1(80), LSTM2(32), Dropout(0.5), Dense(1) |

**Table A2.** Italy model parameters.

| Model | Hyper-parameters   |
|-------|--|
| VAR   | lag = 2  |
| SVR   | kernel = "rbf", gamma = 0.1, C = 120   |
| RF    | n estimators = 30, max depth = 7   |
| LSTM  | epochs = 500, batch size = 10, optimizer = "adam", n past = 7<br>Layers: LSTM1(128), LSTM2(64), Dropout(0.5), Dense(1) |