Exploring US Economic Forecasts and other Insights

**INTRODUCTION**

I am looking at several economic factors to see how well the United States of America is doing.  I am specifically looking at the factors of GDP, health expenditure, life expectancy, internet usage, unemployment rates and the increase in population. I am doing this as an exploratory analysis, but my main goal is to see how much money we will be spending on health and if it is going up or down, especially in relation to the other factors. I think that we can learn a lot about predicting the health expenditure and other variables I have chosen and can cover a wide range of clients, those being in the government, health field and even in communications.

**Data**

The data that I am looking at came from a data repository called Quandl. Quandl houses multiple time series data based off the major stock markets and how businesses are doing in a day to day fashion. They also deal with macro level time series like country economies, country health statistics and demographics of the different countries. The data I collected came from multiple data sets. The data sets in question came from a combination of the IMF Cross Country Macroeconomic Statistics(IMF) and the World Bank World Development Indicators database(WBWD).

Some of the main limitations of this data only shows the values of the variables, it does not go into a deep dive of why the numbers are what they do, as in no historical information just the pure numbers. So, the only thing I can predict with these is how high or low the values can go in the coming years.

I did quite a few alterations to my different data sets. The first thing I tried to do was import all the datasets I got into R and create a single table with all the variables but that turned out to not work as different variables had different lengths, so instead I decided to create six different data frames. Once I created the data frames I changed their value names to what those values were, as an example I changed the Value in GDP to USA\_GDP. Once I created the separate data frames I then reversed the order of the table so that the dates were chronologically accurate from earliest to present. Next I created an extended time series object in R by using the as.xts () function so I could then preform the proper analysis on the data, since I could not do the predictions with a data frame.

**Exploratory Analysis**

I explored the data with graphs of the time series to see how the values fluctuated all of them were in a positive direction and seemed to be steadily increasing. I did not notice any real dramatic shifts in values. I also tested the trend stationarity of my data with a kpss test. The test showed that there is evidence that it is not a trend stationarity with a p value less than 5%.

**DATA ANALYSIS**

The data analysis had to be changed for this project. I was thinking I was going to try out a regression model to predict the changes in either health expenditure or using health expenditure to see about change in life expectancy, but I decided with help from my mentor to change that idea and go with forecast predictions of my data. I like this approach better as it serves as more of an exploratory sense and, I can use it as a benchmark to see how well things are either improving or in some cases to see a potential decrease in values. I will be using ARIMA models as the base for my forecast predictions. Specifically, I will be using auto ARIMA as that way R can help me do the preprocessing for the ARIMA model better.

**DATA DEEP-DIVE**

The first data point that I want to talk about is the GDP of the USA in PPP (Purchasing Parity Power). The training data that I used was for the years 1980 to 2015. The test set was predictions from the world bank data set from 2016 to 2021. The GDP predictions from the IMF were within the confidence levels of my own forecast model. This I believe validates my model. Both data sets show that the GDP is increasing at a steady rate.

Here are my data predictions for 2016 to 2021:

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

37 57372.15 56589.50 58154.80 56175.19 58569.11

38 58620.99 57281.03 59960.95 56571.70 60670.28

39 59869.83 58144.01 61595.64 57230.42 62509.23

40 61118.67 59078.72 63158.61 57998.84 64238.49

41 62367.50 60055.72 64679.29 58831.94 65903.07

42 63616.34 61061.49 66171.20 59709.03 67523.66

And here are the predictions from the IMF:

2016-12-31 57293.79

2017-12-31 59407.47

2018-12-31 61667.78

2019-12-31 63835.42

2020-12-31 65874.33

2021-12-31 67938.17

The next data set I would like to look over is the data set of the population of people living in the USA by millions of people. The training data I have is from the years 1980 to 2015 and the test set is predictive numbers by the IMF for the years 2016 to 2021. My predictions were in line with that of the IMF as far as their numbers were within the confidence intervals of my predictions. The Population is also predicted to increase slightly over the next couple years, which seems appropriate as GDP as well is increasing. Here are my predictions for the Population of the USA:

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

37 324.1327 323.9593 324.3060 323.8676 324.3978

38 326.6691 326.2130 327.1253 325.9715 327.3668

39 329.2218 328.3843 330.0594 327.9409 330.5028

40 331.7878 330.5354 333.0402 329.8724 333.7032

41 334.3645 332.6861 336.0429 331.7976 336.9314

42 336.9500 334.8458 339.0543 333.7318 340.1682

These are the predictions by the IMF:

2016-12-31 323.978

2017-12-31 326.175

2018-12-31 328.386

2019-12-31 330.612

2020-12-31 332.854

2021-12-31 335.110

The next data prediction I would like to address is that of internet usage in the USA. The data comes from the WBWD. The years in the training data are from 1990 to 2010 and the test set are from 2011 to 2014. For this data set I wanted to see how my predictions turned out compared to that of the historical data that they have collected. My point predictions were not accurate but they were within the confidence interval bounds of the data.

Here is my forecast of the internet users in the USA:

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

22 73.60721 69.80956 77.40485 67.79921 79.41520

23 76.16338 69.25164 83.07512 65.59278 86.73398

24 79.05224 69.35311 88.75137 64.21871 93.88578

25 82.11433 69.95060 94.27805 63.51151 100.71714

Here is the WBWD data for those same years:

2011-12-31 69.72946

2012-12-31 79.30000

2013-12-31 84.20000

2014-12-31 87.36000

The next data forecast I would like to talk about is the data of health expenditure from the WBWD. The years in the training set are from 1995 to 2009. The test set is from 2010 to 2013. Just like the previous data there is no real predictions and only historical data so I would like to compare my prediction model to that of the data that the WBWD has collected. Unlike the data on internet users my prediction model was more in line with the data that the WBWD has.

Here are my own predictions:

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

16 8291.764 8237.889 8345.639 8209.369 8374.158

17 8529.077 8374.512 8683.642 8292.690 8765.463

18 8766.389 8484.389 9048.390 8335.107 9197.672

19 9003.702 8572.002 9435.403 8343.473 9663.931

Here is the WBWD data for the same years:

2010-12-31 8298.501

2011-12-31 8553.077

2012-12-31 8845.180

2013-12-31 9145.828

The last data set that I would like to talk about is my predictions on the unemployment rate in percent points, this comes from the IMF dataset. The years in the training data is from 1980 to 2015. The years in the test set are from 2016 to 2021. My predictions are more bleak then the IMF data. My predictions show a steady increase in unemployment rates whereas the IMF data shows a steady decrease. The IMF data does fall within the confidence intervals of my own prediction though.

Here are my own predictions:

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

37 5.452701 4.430267 6.475135 3.889023 7.016379

38 5.770294 4.079645 7.460942 3.184670 8.355917

39 5.982834 4.067051 7.898617 3.052897 8.912771

40 6.125071 4.116625 8.133516 3.053418 9.196723

41 6.220258 4.171671 8.268845 3.087215 9.353302

42 6.283960 4.217648 8.350272 3.123809 9.444111

43 6.326591 4.252390 8.400792 3.154374 9.498808

Here are the IMF predictions:

2016-12-31 4.895

2017-12-31 4.771

2018-12-31 4.729

2019-12-31 4.846

2020-12-31 5.049

2021-12-31 5.055

**CONCLUSION**

I think my predictions could have impact and help the government both the local and federal levels. They could use this information to prepare for increases in the population and unemployment rate. Medical professionals could look at this data and then determine the best way to care for people and how they will price treatments, since expenditure will increase. Even telecommunications can work with the data on internet usage to see coverage and if there is a way to spread that coverage or what ways to get more of the population on board with using the internet.

As I said earlier on I would like to look at ways to potentially fine tune these predictions and possibly find a different way to do these. I would like to see if these different variables correlate with each other and if there is a way to predict one based off the other. I would also want to look and see if I could find different data sources so I could compare the outputs of different data forecasts.