My project is a study of the real GDP of USA and India. It primarily looks into individual GDP forecasts followed by a comparison between the two countries to see how the production has changed in terms of goods and services between an established global market (USA) and an emerging market (India).

All my data was collected from the Federal Reserve Bank of St. Louis Economic Data (FRED). Y data is an annual time series. The limitation I faced while collecting data was that the real GDP for India was only an annual time series. The range if the quarterly data was from 2004, so I disregarded the annual one. In order to compare all the data, I preferred to keep them at the same level. Therefore, I used annual data for both countries, for all the variables used in this study.

I also chose a range of 1961 to 2019 for USA data and 1968 to 2019 for India based on data availability for all the variables and to eliminate the covid shock from the final forecast.

The graph below shows the trends followed by both countries in GDP. I chose adjusted to PPP data because it allows for a much more accurate scale of comparison between the two countries. The data is logged to reduce variation in the data. As we can see from the plot, the difference gap between the two lines is decreasing, which is mostly from the fast increase in India’s production compared to USA, which has a slower trajectory. You can also see that there was a dip for both countries during covid but there seems to be a recovery as well. Whether or not the GDP bounced back in either case is not clear from the plot.

Chart, scatter chart

Description automatically generated

**Diving into diagnostics for USA**

**Real GDP for USA looks like a unit-root process from the simple plot against time.**

**A picture containing text, diagram, plot, line

Description automatically generated**

Chart

Description automatically generated

From the plot above, we can clearly see that the real GDP data for the USA is a unit-root process. After running a kpss test, the unit root was confirmed.

* P-value < 0.05 (0.01), data is a unit-root process

Therefore, the data will be differenced once to eliminate the unit root and used in the ARIMA model. The plot below shoes that after differencing, Real GDP look stationary.

Chart, line chart

Description automatically generated

Chart

Description automatically generated with low confidence

The ACF plot of the differenced Real GDP confirms that the data is stationary now and does not require any further differencing. Just a note: from the plot it looks like an ARIMA(1,1,1) would be the best model to forecast the data. The ARIMA estimation will be discussed in a later part of the report.

**Forecast Model estimations:**

1. **Exponential Smoothing Model- Holt’s Damped**

**A picture containing text, plot, diagram, screenshot

Description automatically generated**

1. **ARIMA**

For the ARIMA model estimation, I tested the white noise process of all the estimated models. (\*\*An ARMA model does not work for the real GDP data since the data is a unit-root process and needs to be differenced.)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ARIMA Models | ARIMA(1,1,0) | ARIMA(1,1,1) | ARIMA(2,1,0) | ARIMA(0,1,1) |
| P-value | 0.63 | 0.756 | 0.654 | 0.779 |

From the ACF plot of the differenced data, and the white noise test values, I decided to use ARIMA(1,1,0) and ARIMA(1,1,1).

|  |  |  |
| --- | --- | --- |
| ARIMA Models | ARIMA(1,1,0) | ARIMA(1,1,1) |
| AIC | -261 | -260 |
| BICc | -256 | -252 |
| RMSE (In-sample) | 0.0185 | 0.0184 |
| RMSE (Out of Sample) | 0.0219 | 0.0237 |

The report for the two models:

|  |  |  |
| --- | --- | --- |
| ARIMA Models | ARIMA(1,1,0) | ARIMA(1,1,1) |
| AIC | -261 | -260 |
| BICc | -256 | -252 |

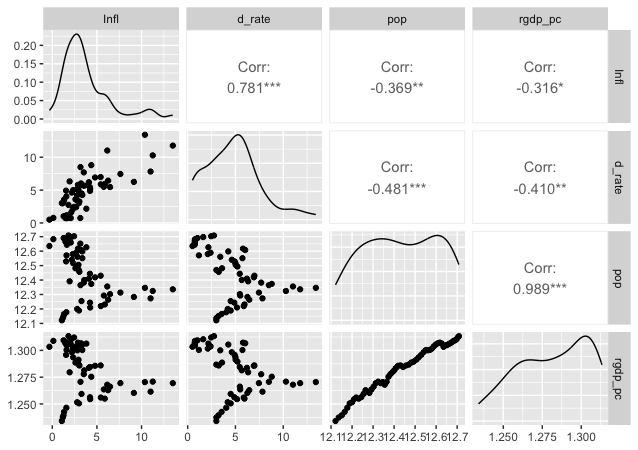
Based on the RMSE of both models, ARIMA(1,1,0) performed a little better. The in-sample forecast plot below is based on that model.

**A picture containing text, plot, diagram, screenshot

Description automatically generated**

1. **VAR**

To estimate VAR model, I first made a correlation matrix to select the right secondary variable from a list consisting of the discount rate, population and inflation. From the plot below, I chose Discount rate as the 2nd variable as it’s correlation with the variable does not cause multicollinearity with the primary variables.



An AICc and BIC models were created to choose the best model. Since the BIC model poses a stronger penalty on the parameters, often times the models vary. For this dataset, both the models gave the same VAR(2) model. The ACF plots for both the AICc and BIC model are shown below.

A picture containing text, screenshot, line, parallel

Description automatically generated

We see no significant autocorrelation between the residuals for either model. I ran a white noise test to further confirm the better model and the P-values for the BIC model performed slightly better than the AICc for both the variables.

I finally picked the VAR(2) BIC model to forecast the series.

A picture containing diagram, plot, line, text

Description automatically generated

The table below shows the comparisons between all the models using the various statistical measures:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Exp. Smoothing Damped-Holt's (Additive) | Exp. Smoothing Holt's (Additive) | ARIMA | VAR |
| Estimated Model | Holt's additive | Holt's damped add | (1,1,0) | VAR(2) BIC |
| BIC | -188 | -183.22 | -261 | -73.1 |
| AICc | -198 | -191.91 | -256 | -88.5 |
| RMSE (In-sample) | 0.0186 | 0.0193 | 0.0185 | 0.0172(RGDP) |
| RMSE (Out of Sample) | 0.0205 | 0.0145 | 0.0219 | 0.00913(RGDP) |

Finally picking the VAR model to forecast for RGDP for the next 10 years. As expected, real GDP increases when there is a decrease in the discount rate. Of course, the actual data of real GDP and discount rate are not remotely close to this plot below, because from 2020 there was a negative shock to the economy due to COVID-19 which my study is not capable of measuring.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated with medium confidence

**Diving into diagnostics for India**

**The Real GDP of INDIA looks like a unit-root process from the simple plot against time.**

**A picture containing text, plot, diagram, line

Description automatically generated**

**A picture containing plot, line, diagram, screenshot

Description automatically generated**

The P-value for the kpss test for the real GDP of India is 0.01.

Therefore, the ACF plot and the kpss test confirms that it is in fact a unit-root process.

A picture containing text, font, diagram, plot

Description automatically generated

After Differencing the data, it looks to be stationary.

A picture containing text, diagram, line, plot

Description automatically generated

The ACF plot confirms that the data is stationary now.

**Forecast Model estimations:**

1. **Exponential Smoothing Model- Holt’s**

Holt’s additive performed a little better than the Damped model for Indian data.

A picture containing text, plot, screenshot, line

Description automatically generated

**ARIMA**

|  |  |  |  |
| --- | --- | --- | --- |
| ARIMA Models | ARIMA(1,1,0) | ARIMA(2,1,0) | ARIMA(0,2,1) |
| P-value | 0.623 | 0.52 | 0.913 |
|  |  |  |  |
| ARIMA Models | ARIMA(1,1,0) | ARIMA(2,1,0) | ARIMA(0,2,1) |
| AICc | -190 | -188 | -186 |
| BIC | -185 | -182 | -183 |
| RMSE (In-sample) | 0.0274 | 0.0273 | 0.0272 |
| RMSE (Out of Sample) | 0.0603 | 0.0585 | 0.0307 |

From the white noise test statistics and the RMSE for the sample ARIMA models, it looks like the ARIMA(0,2,1) which was auto-generated by R is the best model. I went ahead and forecasted the series using the auto-generated model. There were a few things I found challenging in this ARIMA forecast. The data was white noise process after differencing it once, which made predicting the appropriate model difficult. Therefore I am unsure how to select the right ARIMA model under these circumstances. One thing to note from the plot, the forecast looks very similar if not the same to the ETS Holt’s additive model forecast.

A picture containing text, plot, screenshot, diagram

Description automatically generated

**VAR**

A screenshot of a graph

Description automatically generated with low confidence

From the correlation matrix, I chose inflation as the second variable for India data as it shows a significant correlation, and is not high enough to cause multicollinearity.

After fitting the data to the VAR model, the AICc process gave me a VAR(2), whereas the BIC gave me a VAR(3) model. The ACF plot and the white noise test for both models are below.

A picture containing text, screenshot, line, parallel

Description automatically generated

White noise test:

AICc model for Real GDP:

p-value = 0.445

BIC model for Real GDP:  
p-value = 0.265

With all the stats in mind, I went ahead and chose the VAR(2) model to forecast the series.

A picture containing plot, line, text, diagram

Description automatically generated

Some remarks about the forecast. The model did pretty well for the Real GDP but inflation prediction was quite off. There must be more done to this study for India, which is beyond my capacity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Exp. Smoothing Damped-Holt's (Additive) | Exp. Smoothing Holt's (Additive) | ARIMA | VAR |
| Estimated Model | Holt's additive | Holt's damped add | ARIMA(0,2,1) | VAR(2) |
| BIC | -137 | -128 | -186 | 103 |
| AICc | -145 | -137 | -183 | 91.5 |
| RMSE (In-sample) | 0.0282 | 0.0291 | 0.0272 | 0.0231 |
| RMSE (Out of Sample) | 0.0472 | 0.0289 | 0.0307 | 0.0507 |

Based on the table above, compared to all the models, the holt’s performed the best.Forecasting the series finally using holt’s method.

A picture containing text, screenshot, plot, diagram

Description automatically generated

Of course, the actual data of real GDP and discount rate are not remotely close to this plot below, because from 2020 there was a negative shock to the economy due to COVID-19 which my study is not capable of measuring.

From my study, it seems like both real GDP of India and USA were predicted to improve from 2019 going forward, which was a hopeful future for both nations. But unfortunately, COVID-19 shook both the economies and the actual Real GDP for both countries took a downturn.

Limitations of this study: Because the data is annual, the predictions are bound to be far less accurate then they could have been if it was month and/or quarterly.