Forecasting Trade Balance in Goods and Services for the United States

Agam Goyal
agoyal25@wisc.edu

Malkom Castellanos
mcastellano3@wisc.edu

May 3, 2022

1 Introduction

The economic variable that we chose to forecast is the U.S. International Trade Balance in Goods and Services. International Trade Balance in Goods and Services attempts to accurately measure the trade balance of the United States, which is the difference between imports and exports measured in the millions of US dollars ($).

This economic time series also captures what the top exports and imports in the United States were in a given month. However, for the purpose of the forecast, we will be looking at the value of the Trade Balance for the coming 12 months. To provide a more complete picture of the forecasts, we look at both point and interval forecasts.

2 Data and Metrics

The original publisher of the data we use is the Bureau of Economic Analysis (BEA). However, we obtained the data from the Federal Reserve Economic Data (FRED)[5] using the FRED code `BOPGSTB`.

For the purpose of comparing different models we use for our data, the metric we use is the Akaike Information Criterion (AIC) to choose the best forecasting model. Since we are just beginning to work out a model for the Trade Balance time series, we believe that all models we come up with are just approximations of the true model, and our goal here is to find the best approximation, rather than finding the true model[2]. Moreover, we do not have access to out-of-sample data. These are the reasons for us to choose the AIC criterion over others like the Bayesian Information Criterion (BIC).

3 Methods and Models[1]

To begin exploring the nature of the time series data, we look at the auto-correlation function and the partial auto-correlation function and plot correlograms for the same. We see that the auto-correlation correlogram shows a slow decay trend and the partial auto-correlation correlogram seems to show no fixed trend (see `Appendix A`), except for indicating 7 significant lags after which the subsequent lags become insignificant. These trends suggest that we should look at an Autoregressive (AR) model for our forecast[4].

Since the partial auto-correlation correlogram shows us that the first 7 lags would be important to look at for our AR model, we went ahead and ran the `ARIMA(7, 0, 0)`

---

[1]All code used to generate the results presented in the paper as well as the plots in the appendix can be found on `https://github.com/AGoyal0512/forecasting_project`

model on the time series data. On doing so, we saw that only the `p-values` for the first 4 lags were significant, and the AIC for this model was `6838.614` (See `Appendix B`). Updating the forecast model to just an autoregressive model of order 4 or `ARIMA(4, 0, 0)`, we got that all the four lags are significant, and more importantly, our AIC value was also lower at `6835.714`, which was the motivating factor for us to move forward with this model.

Before proceeding, we wanted to check for stationarity in our model, and see if we need any fixes to make the series stationary if it is non-stationary. For this, we used the Augmented Dickey-Fuller (ADF) test[1]. In this test, our null-hypothesis was that of assuming non-stationarity in the time series, while the alternative was the opposite, assuming stationary. We decided to use a `0.05` size-of-test for this hypothesis test.

When we ran the ADF test on our original time series, we got a `p-value` of `0.922304`, which means that we clearly fail to reject the null hypothesis for non-stationarity, which implies that there is a good chance that our series possesses non-stationarity, and that our series has a unit root. To fix this, we decided to use the method of differencing, and took a one-differenced time series of our original data, and then re-ran the ADF test on this series (See `Appendix C`). This time we obtained a `p-value` of almost `0.000`, which indicates that we have a strong evidence to reject the null hypothesis and conclude that this new series is stationary. So we now decided to integrate the one-differenced component in our ARIMA model, to get the `ARIMA(4, 1, 0)`. The lower AIC value of `6807.724` confirmed that we were thinking in the right direction.

Laslty, we saw that on running the `ARIMA(4, 1, 0)` model, the 4th lag became insignificant with a `p-value` of greater than `0.05`. So we decided to drop that lag and settle on an `ARIMA(3, 1, 0)` model, which had roughly a similar AIC value, but a lower BIC value[2], which was an indication that this model might be closer to what the true forecasting model for this time series looks like (See `Appendix B`).

4 Forecasts and Discussions

Having settled on the `ARIMA(3, 1, 0)` model, our next step was to check if the model actually works well on the data and whether it would be suitable for our out-of-sample forecasts. For this, we decided to split our time series into a testing set and a training set with a `2:1 train-test split`. We trained our model on the training set, and then used this model to predict values of the in-sample dates in the testing set.

On performing predictions on the test dataset, our Root Mean Squared Error[3] was `3851.806`, which is a very reasonable error given that our observations are in the order of tens of thousands to almost a hundred thousand. Moreover, on plotting our predictions on top of the test data, we see that the forecast model performs a good job (See `Appendix D`).

Our first out-of-sample point forecast for the month of March is a trade balance in goods and services of `-$ 89756.094707 million` for the United States.

---

[2]The reason we used BIC here is just for it to act as a sort of a tie-breaker, since the AIC values are very similar for being able to draw conclusions

The point forecasts for the one upcoming year out-of-sample, for each of the 12 months lie at a trade balance in goods and services of `-$ 89756.094707 million`, `-$ 89427.186116 million`, `-$ 91032.214411 million`, `-$ 90702.505979 million`, `-$ 90600.974737 million`, `-$ 90999.122135 million`, `-$ 90820.712238 million`, `-$ 90819.462587 million`, `-$ 90918.924467 million`, `-$ 90851.954543 million`, `-$ 90863.320926 million`, and `-$ -90886.540965 million`.

The `95%` interval forecasts along with the point forecasts for the 12 months are reported in a table in `Appendix D`. The plots showing these forecasts with an overall plot of a part of the original time series with the 12 out-of-sample predictions can also be found there. All observations are in the units of millions of United States $.

References

[1] *How to Check if Time Series Data is Stationary with Python: Machine Learning Mastery.* URL: `https://machinelearningmastery.com/time-series-data-stationary-python/`. (accessed: 05.02.2022).

[2] *Introduction to AIC — Akaike Information Criterion: Towards Data Science.* URL: `https://towardsdatascience.com/introduction-to-aic-akaike-information-criterion-9c9ba1c96ced`. (accessed: 05.02.2022).

[3] *RMSE: Root Mean Square Error - Statistics How To.* URL: `https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/`. (accessed: 05.02.2022).

[4] *Time Series Analysis: Identifying AR and MA using ACF and PACF Plots: Towards Data Science.* URL: `https://towardsdatascience.com/identifying-ar-and-ma-terms-using-acf-and-pacf-plots-in-time-series-forecasting-ccb9fd073db8`. (accessed: 05.02.2022).

[5] *U.S. Census Bureau and U.S. Bureau of Economic Analysis, Trade Balance: Goods and Services, Balance of Payments Basis [BOPGSTB].* URL: `https://fred.stlouisfed.org/series/BOPGSTB`. (accessed: 05.02.2022).

Appendix A

Appendix B

Appendix C

Appendix D