China's GDP Forecasting Project

Yuan Tian

I chose the data of China's quarterly GDP from 1992 to 2019 from the website of Federal Reserve Economic Data repository (FRED), the original data is in Chinese currency and without seasonal adjusting.

Part I. Data preparation and visualization

First, I read in the data, changed the column name to "GDP" and converted it to quarterly time series data.

When visualized the GDP data, I noticed clearly upward trend and seasonality from the plot. So I decided to fit the trend both in linear and exponential methods, from the fitted lines, the GDP data showed an exponential trend. In this case, I made a log transformation to the original data.

Part II. Model Selection

We should ensure that the time series being analyzed is stationary before specifying a model. So I ran the ADF test with "trend" to check the stationary of the data. Based on the result of test statistics, all the three hypotheses (tau3, phi2, phi3) can be rejected at 5% level. So the data is stationary.

Then I checked the ACF and PACF plots, From the ACF plot, there are diminishing spikes last until 20. From the PACF plot, there is a significant spike at lag 1 and almost no significant spikes after that. It may suggest a non-seasonal AR (1) component.

Since the plot shows seasonality, I also ran a simple linear regression on trend and season, and the results showed significantly trend and seasonal components. In this case, although the ADF test shows there is no unit root, there can still be a trend component in the data.

So I took a seasonal difference to see the points' distribution, the plots turned out to be more stationary compared with the previous one. And with another difference (to capture the trend component), the plots improved further.

Consequently, I made the first assumption of my model, from the ACF and PACF plots, I recognized an AR (1) model. Besides, with the check of seasonality and trend, I also decided to add the (0,1,0)[4] term in the model to adjust for seasonal difference and an "xreg" term to capture the trend. My initial function is "order=c (1,0,0), seasonal=c (0,1,0), xreg=(1:T)"

Afterwards, I ran auto arima to see how it works, this function returns the model ARIMA (1,1,0)(0,1,0)[4]. It turned out that the auto arima uses first differencing to capture the trend. Although xreg in my model played similar role, the BIC and RMSE are both performed better in the auto arima model.

So I finally decide to choose auto arima model, which is ARIMA (1,1,0) (0,1,0)[4] that has both tend and seasonal components. The residual plot of the model showed no autocorrelations and the histogram plot is bell shaped, so the model passes the required checks.

As a result, my final model equation is:

$$\log(GDP)_t = 0.4685 \log(GDP)_{t-1} + \log(GDP)_{t-12} - 0.4685 \log(GDP)_{t-13} + e_t$$

Part III. Forecasting on the model

I used a train/validation structure to make the forecasting because my data contains seasonality and split data into two sets may be more useful in forecasting.

I split the original data into train and valid, then ran modeling on the training set.

I used ‘forecasting’ function to make predictions and checked the accuracy using the original valid data set.

From the forecasting plot, the predicted values fitted relatively well to the real value, it showed that the model forecasting performance is good.

Part IV. Model Comparison

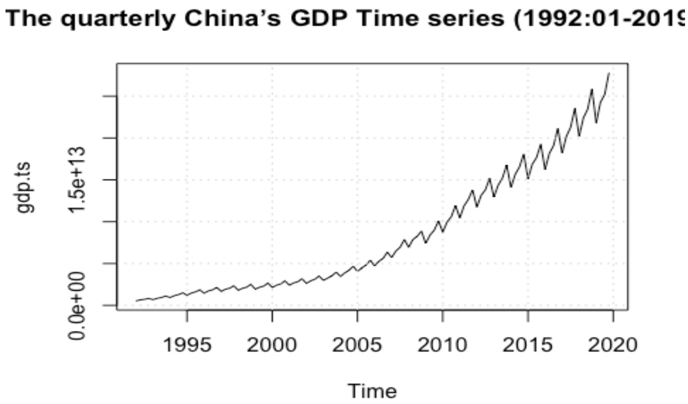
First, I compared my ARIMA model with Naïve model as the baseline model.

I ran the “dm” test both in train and valid data, set the alternative equals to “less”, both the p-values are almost 0, so I can reject the null in the two cases, which means my model is significantly better than the Naïve baseline model.

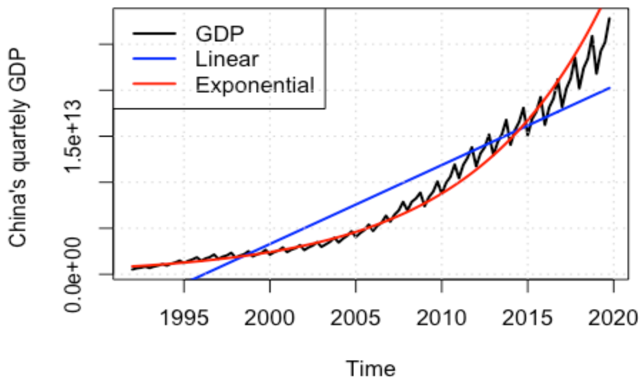
I also compared my model with a simple AR (1) model, from the “dm” test results on train and valid data, I can reject the null in the two cases, so my model is also better than the simple AR(1) model.

Part V. Visualizations

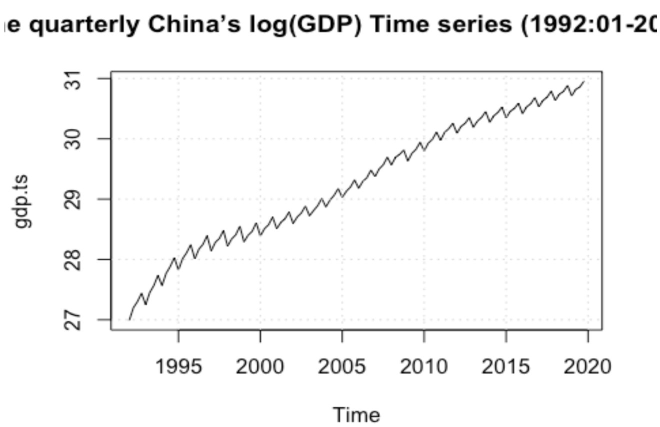
Plot1. Time Series of Original Data (China's GDP)



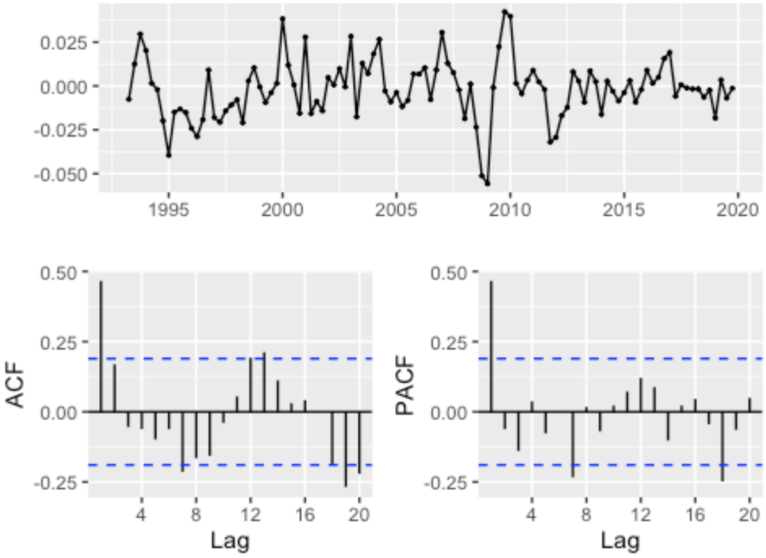
Plot2. Trend Fitted Plot of Original data



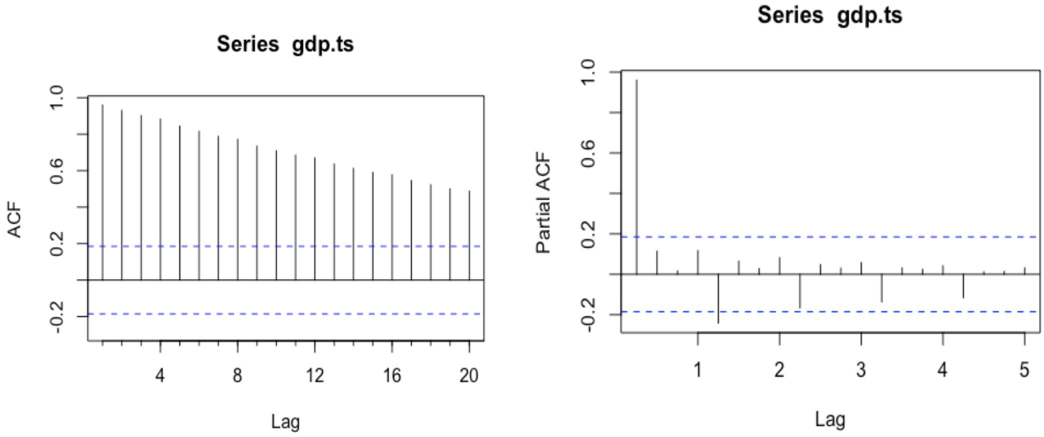
Plot3. Time Series of log Data (log GDP)



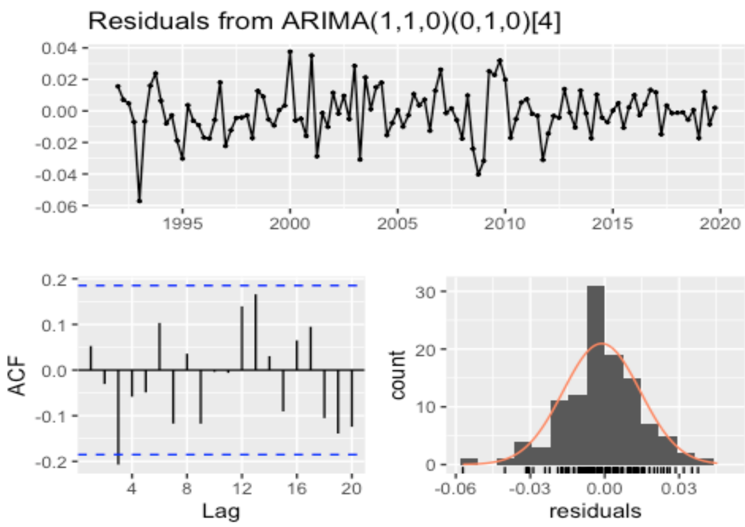
Plot4. Plots after Seasonal Differencing (diff(4))



Plot5. ACF and PACF Plots on log (GDP)



Plot6. Residual Plot of the Model



Plot7. Forecasting Plot