

**Final Project**

By

Rylie Austin (8210704)

Gabe Bach (8179614)

Thomas Detlor (8319090)

Genevieve Guay (8821581)

Tina Samaha (8610185)

Ayoub Sheik-Ahmed (7370221)

Work presented to Professor Okhaide Akhigbe

Predictive Analytics - ADM4307

Fall 2018

Monday November 26th, 2018

Table of Contents

[Introduction 4](#_Toc530943158)

[More on Imports and Exports 4](#_Toc530943159)

[Economic Prediction of Imports and Exports 5](#_Toc530943160)

[Data preparation 7](#_Toc530943161)

[Model Fitting and Selection 8](#_Toc530943162)

[Simple Exponential Smoothing 8](#_Toc530943163)

[Holt Winters Exponential Smoothing and Additive Model 8](#_Toc530943164)

[ETS 8](#_Toc530943165)

[ARIMA 8](#_Toc530943166)

[Linear Regression 8](#_Toc530943167)

[Combination of Models 9](#_Toc530943168)

[Selection Method 9](#_Toc530943169)

[Difference in Export Value Between US and China 9](#_Toc530943170)

[Net Balance of the US 11](#_Toc530943171)

[Net Balance of China 12](#_Toc530943172)

[Interpretation of the Forecasting 13](#_Toc530943173)

[Difference in Export Value Between US and China 13](#_Toc530943174)

[Refitting for Difference in Export Value Between US and China 13](#_Toc530943175)

[Conclusion 14](#_Toc530943176)

[Appendix 1 16](#_Toc530943177)

[Appendix 2 16](#_Toc530943178)

[Appendix 3 16](#_Toc530943179)

[Appendix 4 17](#_Toc530943180)

[Appendix 5 18](#_Toc530943181)

[Simple Exponential Smoothing 18](#_Toc530943182)

[Holt Winters 19](#_Toc530943183)

[ETS 21](#_Toc530943184)

[Auto-ARIMA 22](#_Toc530943185)

[Linear Regression 23](#_Toc530943186)

[Combinations 24](#_Toc530943187)

[Forecast Hybrid with Equal Weights 24](#_Toc530943188)

[Forecast Hybrid with Sample Weights 25](#_Toc530943189)

[Appendix 6 26](#_Toc530943190)

[Appendix 7 26](#_Toc530943191)

[Appendix 8 27](#_Toc530943192)

[Simple Exponential Smoothing 27](#_Toc530943193)

[Holt Winters 28](#_Toc530943194)

[ETS 29](#_Toc530943195)

[ARIMA 31](#_Toc530943196)

[Linear Regression 32](#_Toc530943197)

[Combinations 33](#_Toc530943198)

[Forecast Hybrid with Equal Weights 33](#_Toc530943199)

[Forecast Hybrid with In-Sample Weights 34](#_Toc530943200)

[Appendix 9 35](#_Toc530943201)

[Appendix 10 35](#_Toc530943202)

[Appendix 11 36](#_Toc530943203)

[Appendix 12 38](#_Toc530943204)

[Appendix 13 39](#_Toc530943205)

[Appendix 14 40](#_Toc530943206)

[Personal Ethics Statement Concerning Telfer School Assignments 42](#_Toc530943207)

# Introduction

In the last few years, headlines about trade-war and the economic race between the US and China have become a topic of interest. In terms of economic prosperity and dominance, it’s clear that the 20th century belonged to America. Nevertheless, China’s importance in the 21st century remains to be seen. In fact, China surpassed the US economy in total value of exports quite sometime ago, however, the net balance of each economy has fluctuated over time, and the per-capita figures still put China far behind. While an all-out trade war appears to be off the table for now, China’s rise is destined to pose ongoing challenges for Donald Trump and his “America First” ethos. This being said, the purpose of this report is to make a comparison between the imports and exports of the Chinese and US economies, in terms of the value of total exports, and the net balance of each economy.

# More on Imports and Exports

The strength of a nation’s economy is largely reliant on a country’s imports and exports, and the balance of these in relation to one another. Both imports and exports are widely used economic indicators and are key in determining many other economic factors. Imports and exports are most commonly used to help calculate a nation’s gross domestic product through the calculation of net exports, which are the total exports minus the total imports. Imports are the summation of the value of all the goods and services that are brought into a country from abroad (in this case USD). Whereas exports are the summation of the value of all the goods and services that are sold in foreign countries. The value of a nation’s imports and exports are directly related to a country’s trade policies. As much as imports and exports impact the gross domestic product, they also directly affect a nation’s economic performance. The balance of a nation’s net exports can alter the valuation of their currency, which is a determining factor in economic prosperity. The United States and China are second and fourth in the world for total imports respectively, as well as being China being first in total exports, and United States being third (CIA, 2018) (Appendix 1). These are crucial factors in why both the United States and China have been experiencing rapid growth in their economy.

The purpose of imports and exports is to act as an economic indicator for the valuation of currency and the effectiveness of a nation’s trade policies. Exports also make up a large portion of gross domestic product, with exports being 11.89% of gross domestic product for the United States, and 19.76% of gross domestic product for China (World Bank Group, 2018). Imports and exports are not directly related to one another, but directly influence the balance of trade (or net exports) of a country. If a country imports more goods and services than it exports, the country will be a trade deficit. Being in a trade deficit will not only decrease gross domestic product, it will also affect the value of the currency. If a country is in a trade deficit, there is little demand for a country’s goods and services, meaning there is little demand for their currency, this ultimately leads to a devaluation of currency. Likewise the opposite is true when a country enters a trade surplus, the valuation of a country’s currency will increase, as there is high demand for the currency.

Furthermore, when you examine recent events in global trade you can begin to see the groundwork for our prediction to come into fruition. Recently, the United States and China have begun placing tariffs on many goods exported and imported into each country (Martin, 2018). The United States initially placed many tariffs on Chinese goods, with China doing the same in retaliation (Martin, 2018). This has caused a major shift in global trade and has decreased United States exports to China drastically. The impacts of these tariffs have most been felt by United States soybean farmers as China implemented a 25% tariff on all U.S. soybeans (Weinraub, 2018). China imports approximately 60% of all soybean production in the U.S., and with these new tariffs have essentially stopped all soybean imports from the U.S (Martin, 2018). This has resulted in a huge loss of exports for the United States. However, there is the potential for these tariffs to become increasingly drastic. If these Chinese tariffs on U.S. goods are not lifted, the United States has threatened to enact tariffs on $500 billion of Chinese goods coming to the U.S (Martin, 2018). If this were to come into action, it would have serious implications on the United States economy. This would certainly mean a sharp decrease of all trade with China, in total causing a massive drop in U.S. imports and exports. As long as this trade dispute continues, it further supports the results of our prediction.

However, in comparison to the trade dispute between China and the United States, there are other recent economic events that challenge the results of our prediction. Just as the United States are imposing new tariffs on certain goods, they have also repealed some trade policies to make the exporting of goods more appealing. In particular, the United States weapons industry has seen great growth as a result of looser restrictions on the sale of weapons (Reuters, 2018). United States arms sales to foreign governments grew by 13% last year totaling $192 billion worth of exports (Reuters, 2018). This could indicate the start of a period of immense growth of the value of exports in the United States. Furthermore, we see this trend continue when we examine the U.S. dairy industry and its exports. Recently, the United States has seen a large growth in the sale of dairy products internationally, with exports up a total of 10% from 2017 (Levitt, 2018). The value of milk exports from the U.S. totaled $4.25 billion (Levitt, 2018). A steady amount of growth in an industry as large, and as steady as the dairy industry bodes incredibly well for the growth of exports coming from the United States. China has also faced serious implications in reference to their trade dispute with the United States. The growth in Chinese exports decreased to just under 10% from the previous total of 12% (Shane, 2018). This trend is expected to continue as well, as tariffs imposed by the United States increase. As we know, the goal of any forecast is to not be exactly correct, but to have an estimation of what is going to happen in the future. These events indicate that even though our prediction was made on statistical information, it is not guaranteed to come true.

# Economic Prediction of Imports and Exports

Forecasting of imports and exports is most useful for Forex traders, the government, and economists. Forecasting these metrics are extremely important for these parties as it allows them to predict the exchange rate, value of currency, and gross domestic product more accurately. When these parties attempt to forecast exports, they are typically focused on events abroad, as exports are dependent on events outside of the domestic economy. Traditionally, when forecasting exports, experts used econometric models to predict export relationships (Coccari, 1978). In these models, these parties will typically look at the export relationship the country has with each country that it exports to, and more importantly the changes in income and price (Coccari, 1978). To examine these export relationships, experts would typically look at but are not limited to the following factors; aggregate foreign demand, domestic demand, grants and loans, foreign investment, research and development, and the total exports of a given country (Coccari, 1978) (Appendix 2).

This method was effective but it was very time consuming, so experts began using time series models to start forecasting exports of a country. In these time series models, experts only require few data sets to make an accurate prediction, most importantly these models only require either yearly, monthly, and quarterly export data, but improve with the combination of these data sets. The time series models that these experts typically selected are; single exponential smoothing, Holt’s, Winter’s, adaptive response rate, linear trend, and classical decomposition; however, Box-Jenkins are typically not evaluated (Mahmoud, Motwani, 1989). These models would be run over multiple periods to ensure the accuracy of the models. Multiple models for the forecast would be run, and then one chosen based on minimal values of MAPE, MPE, RMSE, or a combination of the three (Mahmoud, Motwani, 1989). Overall time series models can provide as accurate or more accurate forecasts of a country’s exports when compared to econometric models (Mahmoud, Motwani, 1989) ; this is why the most commonly chosen method of forecasting exports is through time series models.

As important as forecasting exports with these models is, the proper way to forecast export growth is by forecasting the difference in exports. This is the case because the error reflected in any single prediction intervals is amplified when making comparisons between two or more prediction intervals. In this case to make any inference on the two series, statistical significance in the difference must be evaluated using pairwise testing. In addition, two models must be developed for the separate series before comparison. It is, therefore, advantageous to derive difference between the corresponding points in time, creating a new time series and fitting a model to this new time series. In addition, for many ARIMA models to work, they require stationary data which is required for sample statistics to accurately describe the data at all time points. One very common way of making data stationary is by differencing the data. It is more likely that the difference of two series with seasonality will carry less of a seasonality component itself, creating another opportunity for increased efficiency in the methodology. Overall, forecasting with the differenced data reduces the complexity of comparing two metrics and allows those conducting the forecast to be able to more accurately describe the correlation between exports.

Again, forecasting projected imports is a key function of many economists and governments. Whereas forecasting exports relies on events happening abroad, imports are directly related to the domestic economy. Therefore, import growth rate and GDP are the factors that forecasters are most interested in. As exports are more variable due to the presence of foreign events, forecasts of imports tend to be more accurate. When forecasting the imports of a country, the International Monetary Fund chooses to use a formal econometric model (Genberg, Martinez, Salemi, 2014). As these types of models require large amounts of data and resources to complete, many of these experts chose other time series models, such as an autoregressive ARIMA model, an autoregressive distributed lag model or a vector auto-regression model (Keck, Raubold, Truppia, 2009). These methods are effective the forecasting of imports. As we can tell, the preferred method of forecasting imports and exports by some experts is still to use complex econometric models that rely on large amounts of data and expert opinion. However, using time series models to predict these economic indicators performs just as well in comparison to econometric models, with much less resource requirement for forecasters.

As anticipated, there will be some limitations associated with not only the calculation imports and exports, but with the forecasts of these metrics as well. One major limitation that is noted in the forecasting of imports and exports is the assumption that trade policies are unchanging, when realistically; trade policies can be very volatile. If an imported good has a high tax imposed on it, there will be less of it imported due to the increased cost. Another limitation is that import data is much more reliable that then export data, as exports are calculated by the assessment of tariffs by customs authorities (Keck, Raubold, Truppia, 2009). It is also much more challenging to forecast export data, as it requires information on global demand and global prices (Keck, Raubold, Truppia, 2009).

# Data preparation

Data on the value of US and Chinese imports and exports was sourced from the Federal Reserve Bank of St. Louis. The three metrics we wanted to asses were; the net balance of US imports and exports, the net balance of Chinese imports and exports, and the difference in total value of US and Chinese exports. In order to achieve this, we used four data series; total value of US exports from 1960 – 2018, total value of US imports from 1960 – 2018, total value of Chinese exports from 1992 – 2018, total value of Chinese exports from 1992 – 2018.

For the net balance of the US, the value of imports were provided in monthly intervals while the value of exports were provided in quarterly intervals. For this reason, the imports data set was converted into quarterly data points and then the difference between them was derived. This new time series was used in the model selection phase.

For the net balance of China, the value of imports and exports were both provided in monthly intervals, so the difference between them was derived and this new time series was used in the model selection phase.

For the difference in value of exports between US and China, the value of exports for the US was provided in quarterly intervals starting in 1960 while the value of exports for China was provided in monthly intervals starting in 1992. First, the data preceding 1992 in the US export time series was omitted. Next, the Chinese export time series was converted into a quarterly series. Finally, the difference between the US and Chinese value of exports was derived by subtracting the value of US exports from that of the Chinese exports. Although this resulted in initial negative vale, many of the most recent time series is positive which is more visually appealing and fits well with the various forecasting methods tested. This new time series was used in the model selection phase.

# Model Fitting and Selection

To be able to make a comparison between the imports and exports of the Chinese and US economies and to anticipate China’s importance in the economy of the 21st century in terms of value of total exports and the net balance of each economy, multiple forecasting methods are fitted on these factors to, then, select the best one and make the most accurate forecast possible. Among these methods, the time series will be fitted with simple exponential smoothing model, Holt Winters, ETS, ARIMA, linear regression and a combination of models, more specifically Hybrid.

## Simple Exponential Smoothing

This model uses a weighted average of the past available data as a forecast value. As the most recent observations are often the most relevant and accurate to the forecasting, the SES method will attribute the a more important weight to the most recent values and will reduce it exponentially as the values are further in the past using a smoothing constant α. It is best used when the data have no or little trend.

## Holt Winters Exponential Smoothing and Additive Model

Holt Winters model involves the use of three smoothing parameters: a simple, a trend and a seasonality smoothing parameter. Therefore, this model is appropriate when both trend and seasonality component are present.

## ETS

This model uses an estimation of the probability of the parameters used in Holt Winters based on the data to optimize the forecast.

## ARIMA

An Auto-Regressive Integrated Moving Average (ARIMA) model incorporates differencing, autoregression, and moving averages into a single model. The model has 3 main parameters; p, d, and q. P is obtained from the autoregressive component of the Model. Auto Regression forecasts variables based on a linear combination of past values of the variable, so, p refers to the order of past values. A moving average model uses past forecast errors where q refers to the order of the past forecast errors. Finally, differencing refers to the act of taking the difference of current and previous values, so, d refers to the order of differences taken on the data. To account for seasonality, the model is formed by including additional seasonal terms; P, D, and Q, also having m represent the number of observations per year. Where the terms are like the lowercase versions in the non-seasonal model. The best ARIMA model for the data would be the model with the best combination of these variables, given the characteristics of our data. The first step was to ensure our data was stationary.

## Linear Regression

The basic concept for time series regression models is that we forecast the time series of interest, based on its linear relationship with another time series. The other time series, that we expected to have a linear relationship to our main dataset, were the trend and seasonality of the data. By doing a time series decomposition it is easier to determine visually if there are strong trend and seasonal components to the data. Therefore, we built our regression model where our predictor variables were the trend and seasonal component and where the dataset was the explained variable.

## Combination of Models

This method consists in combining a number of forecasting methods to create the best combination. The forecasting of the factors of the US and Chinese economy will be done through the Hybrid forecast, which is an automatic method that had a pre-determined set of methods that were combined. In fact, it can be used through in two ways; using equal or in-sample weights. Evidently, equal weights make each method have the same impact on the forecast, whereas in-sample weights calculate the impact based upon the in-sample errors. The combination that will be used throughout this report is the Forecast Hybrid function in R which combines various automatic forecasting methods to create the best forecast using weights. It can be done using equal weights, where all methods have the same impact in the model and it can be done using in-sample weights, where the impact of each method is calculated based upon the in-sample errors.

## Selection Method

The most appropriate way to select the best fit to forecast is, generally, by examining the residuals. This can be done by quantifying the size of the deviations between the model and the data using a number of different calculations such as the mean absolute percentage deviation (MAPD), the mean squared deviation (MSD), R2 and many others.

Although AICc can be used when selecting the best ARIMA model, and R­2-Adj can be used to select the best regression model; when selecting what the best type of model is for a time series, the common metric becomes root mean squared error (RMSE), which is calculated as the root of the sum of the squared errors divided by (V-1), where V is the number of observations in the validation set. Therefore, in this report, the model with the lowest RMSE will refer to the most accurate forecasting technique.

# Difference in Export Value Between US and China

Upon examination of the difference in Export Value between US and China time series (Appendix 3), a couple things should be immediately noticed; China begins to overtake the US in terms of total value of exports around 2001, there is a clear upwards trend and seasonality in the data. This is confirmed in the decomposition of the time series (Appendix 4). Considering this, multiple forecasting methods were tested on this data series to select the most accurate function such as simple exponential smoothing, Holt’s damped, Holt Winter’s, ETS and ARIMA, for all of which the model fits, forecasts, accuracy reports and residual summaries were developed (Appendix 5).

In fact, the RMSE for each method is outlined as follows:

|  |  |
| --- | --- |
| **Method** | **RMSE** |
| Simple Exponential Smoothing | 30143558217 |
| Holt Winters | 13626892760 |
| ETS | 13905720587 |
| ARIMA | 12410416857 |
| Linear Regression | 45186618088 |
| Hybrid Equal Weights | 14541428216 |
| Hybrid in-sample Weights | 13213412830 |

Analyzing more closely, some models seem to be a better fit than others. For example, knowing the simple exponential smoothing model cannot adjust to trend or seasonality, it is obvious that this will not be an ideal method for the value of total exports, since China’s exports seem are highly seasonal and seem to be rising significantly and consistently compared to the US’s.

On the other hand, the Holt-Winters works best when there is heavy seasonality and trend and, as shown in the fit, this model seems to be following closely the data set. When analyzing the residuals, there seems to be no pattern, the RMSE value is significantly lower than as seen in the first method and the histogram of the residuals is slightly skewed to the left but close to normal, which indicates the mean is close to 0. ). Also, the autocorrelation function (ACF) plot shows the degree of correlation between the current value and each of the past values. In this case, there is no sign of a significant correlation.

Similar results can be seen in the ETS for the additive for errors, additive for trend, and additive for seasonality. However, the RMSE has a value of 13905720587, which is higher than Holt-Winters and is, therefore, not the best forecasting method. In addition, as the ACF plot shows two lags (2 and 15) reach the minimum allowance threshold, suggesting a correlation between the in-sample residuals and predicted values provided by our model, a Ljung-Box test was conducted to test the significance of the correlation at a 5% level. As the p value found is lower than 0.05 with a value of 0.0014, the correlation does exist, which is not ideal.

Furthermore, we used R to auto-fit an ARIMA model to the original time series which is designed to deliver the best possible ARIMA. The auto.arima function proposed an ARIMA(2,1,0)(1,1,2)[4]. As mentioned before, ARIMA models are best examined using the AICc, however, since this is not a comparable measure with all the other models, its RMSE value of 12410416857, proves this method to be the best fitting model for the data set.

As for the linear regression, the model was used with trend and seasonality as input variables; it proves to be the least fit model for this data set with a value of almost seven times higher than the others. First, the RMSE proves to be extremely high with a value of 71361421859, which is almost 7 times higher than the other methods. Also, when examining the residual plots, there is an evident cyclic pattern and heavy seasonality, which means it was not modelled properly, whereas the residual histogram is skewed to the left and, therefore, the mean is not 0. Finally, the ACF plot shows heavy correlation with a number of lags, so this model is not a match with the data set.

Examining the Hybrid Forecast, the in-sample weights had an RMSE value ranking it as the third more appropriate model for the dataset, whereas the equal weights were a bit higher and, therefore, less useful for the total exports value of the two economies. More specifically, the equal weight model shows normally distributed residuals per the histogram, however, the graph reflects seasonality and the ACF plot shows a heavy correlation between the model and past values. On the other hand, the in-sample weights hybrid forecast shows approximately the same results: there is seasonality in the residuals, they seem to be distributed close to normal so the mean is close to 0 and the ACF plot shows the correlation are mostly in between the maximum and minimum allowance threshold. In both cases, the RMSE still does not reflect the best model.

In conclusion, as can be interpreted from the table above, the ARIMA model best fit the difference of total exports value between China and the US factor; it can be assumed that this ARIMA model is the most appropriate for the data because the auto.arima function was used in R without approximation or stepwise. The best model for this time series was therefore determined to be ARIMA(2,1,0)(1,1,2)[4].

# Net Balance of the US

Upon examination of the net balance of the US time series (Appendix 6), a couple things should be immediately noticed; there is a clear upward trend and seasonality, which means the US’s exports are generally higher than the imports, and are still growing. There is also a noticeable and significant leap in the net balance around 2008 and grows back until, approximately, 2011-2012; this is confirmed in the decomposition of the time series (Appendix 7). Considering this, multiple forecasting methods were tested on this data series to select the most accurate function such as simple exponential smoothing, Holt’s damped, Holt Winter’s, ETS and ARIMA, for all of which the model fits, forecasts, accuracy reports and residual summaries were developed (Appendix 8).

Starting with the simple exponential smoothing, the model seems to be following extremely closely the data set. However, looking at the forecast, the average seems to stagnate at the last known value of the data set. Also, knowing this model works best when there is little or no trend, it is easy to assume that the results will not be very useful for this data. Looking further into the residuals; the histogram reveals that the residuals are distributed normally, but the residual plot reveals heavy seasonality and the ACF plot shows a significantly high correlation with most of the lag values. In addition, the RMSE value classifies this model as one of the least fit for the US net balance of exports.

On the other hand, Holt Winters model seems to be fitting the model better. First, the residuals seem to be distributed normally, the ACF plot shows more significant spikes at lag values 1,3 and 24, which proves correlation between past and model values. The RMSE stands at 13678280868, but the residuals plot shows a seasonal pattern.

As for the ETS model, it appears to be very fitting to the US net balance of exports: the residuals show a slight seasonal pattern, the ACF reveal a slight correlation in lag value 1 and the residuals histogram reveals normally distributed residual values. With an RMSE of 12908734529, this model is one of the best fit for the data set.

Furthermore, we used R to auto-fit an ARIMA model to the original time series, which proposed ARIMA(3,0,0)(0,1,1)[4]. With no pattern in the residuals, a normal distribution in the histogram and almost no correlation in the ACF plot, this model seems to be the best fit for the data. Although being the better-fit, the Ljung-Box test still shows that, with a p-value of 0.476, which is lower than 0,05, the correlation is slight, but existant.

Oppositely, linear regression is not at all a good fit for the data set. With a visible cyclic pattern for the residuals, a very abnormal distribution and an extremely high correlation for all lag values observed, this model has an RMSE of almost 7 times the value of other methods.

Finally, the hybrid forecast method is also interesting to look at. As the for equal weights, the RMSE is a bit higher, the correlation reflected in the ACF plot is remarkable at a number of lag values, there is a slight seasonality in residuals, which are normally distributed. On the other hand, in-sample weights show very similar results with a lesser correlation and lower RMSE.

In short, the RMSE for each method is outlined as follows:

|  |  |
| --- | --- |
| **Method** | **RMSE** |
| Simple Exponential Smoothing | 17270596142 |
| Holt Winter’s | 13678280868 |
| ETS | 12908734529 |
| ARIMA | 10748135007 |
| Linear Regression | 71361421859 |
| Hybrid Equal Weights | 14342191027 |
| Hybrid in-sample Weights | 13353263923 |

As can be interpreted from the table above, the ARIMA model best fit the net balance of US exports; it can be assumed that this ARIMA model is the most appropriate for the data because the auto.arima function was used in R without approximation or stepwise. The best model for this time series was therefore determined to be ARIMA(3,0,0)(0,1,1)[4].

# Net Balance of China

In fact, the RMSE for each method is outlined as follows:

|  |  |
| --- | --- |
| **Method** | **RMSE** |
| Simple Exponential Smoothing |  |
| Holt Winter’s |  |
| ETS |  |
| ARIMA |  |
| Linear Regression |  |
| Hybrid Equal Weights |  |
| Hybrid in-sample Weights |  |

# Interpretation of the Forecasting

## Difference in Export Value Between US and China

Based on the ARIMA(2,1,0)(1,1,2)[4] forecast of the differences between US and Chinese exports. It can be predicted that the Chinese value of exports will continue to outgrow that of the US at declining rate, but only when looking at the point estimates. As outlined in the graph below, when looking at the 95% prediction interval, the same conclusion cannot be drawn. This is because within the same level of statistical significance, we see the gap growing, staying steady, and decreasing.

One way to help interpret the information would be make inferences based on current events, whether a favourable, consistent, or unfavourable climate would be more likely. However, due to the nature of this report, we decided to go back to the drawing board and develop a model with better predictive ability.

## Refitting for Difference in Export Value Between US and China

The time series models tested in the initial model selection process allow for the inclusion of information from past observations of the time series, but not for the inclusion of other information that may also be relevant. For example, the effects of GDP, trade sanctions, economic recessions, or other external variables. This time, using dynamic regression with ARIMA errors we allowed the errors from regression to contain autocorrelation. The idea here is to use other related economic indicators to help predict the difference in the value of exports between the China and US

Data preparation:

In order to use dynamic regression; all the values in the series of interest must be positive, because the difference between the economies was negative up until the point that China surpassed the US and the largest value where the US was ahead of China is 136105000000.00 USD, a new Difference in Export Value Between US and China series was made by adding this value to every value in the original time series (Appendix 9). This number can simply be subtracted once the forecast is made.

In dynamic regression, the input variables must be stationary and we must know the values for the input variables for the time period that we are forecasting for the series of interest. It follows that input variables must be differenced until their values become stationary, it’s also important to note that the form of the relationship between predictors and the variable of interest should be maintained by differencing all input series the same number of times. Luckily for us, the auto.arima() function in R can take care of all differencing for specified predictor series.

Predictor Variable: GDP of US and China

First we used the GDP yearly GDP in each country (Appendix 10). To do so, we needed to determine the values of our predictors over the forecast horizon, so we used the auto.arima() function to create a forecast of our Chinese GDP and US GDP time series to do so (Appendix 11) . Because the series of interest is given in quarterly time periods and GDP is given yearly, the GDP level was simply repeated 4 times for each year (Appendix 12). This is appropriate as the effect of the current year’s GDP will have varying levels of correlation per quarter in either case.

Predictor Variable: Total Share Prices for all Shares in the US: Index 2015

After testing the dynamic model with projected GDP from both countries, the total of all share prices of US was added to see if an improved level of accuracy could be reached. The total US share price data was provided as an index of the corresponding 2015 value on a monthly basis, so the average price for each quarter was first calculated in order to fit with the time series of interest(Appendix 13). Next we determined the values of this predictor over the forecast horizon again, using the auto.arima() function to create a forecast (Appendix 14).

Predictor Variable: US and Chinese Populations

After testing the dynamic model with projected GDP from both countries and the total of all share prices of US, the US and Chinese population figures (Appendix 15) were added to see if an improved level of accuracy could be reached. Auto.arima() was used to predict the populations for both countries over the forecast horizon (Appendix 16). Because the series of interest is given in quarterly time periods and population levels are given yearly, the populations were simply repeated 4 times for each year (Appendix 17). This is appropriate as the effect of the current year’s populations will have varying levels of correlation per quarter in either case.

Model Fit:

After creating a dynamic regression that included 5 predictors, it was time to select the best subset. The subset combinations are outlined in Appendix 18, and the measures of forecast error are outlined in Appendix 19. After running through 19 combinations, combination 14 was determined to have the best performance with an RMSE of 119840000000000 and AICc of 4998.61. This model out-performs the original ARIMA model and was used to perform the same analysis as above. The model’s forecast, summary, and residuals can be found in Appendix 20.

Analysis:

Based on Regression with ARIMA(1,1,1)(0,1,1)[4] errors (model 14 in Appendix 19) forecast of the differences between US and Chinese exports. It can be predicted that the Chinese value of exports will continue to outgrow that of the US. As outlined in the graph below, when looking at the 95% prediction interval, the same conclusion can be drawn, although, a slight portion of the low 95% interval seem to be decreasing, it is only a very little bit and most of the interval has a positive trend.

We are also interested in the year over year rate of difference in the value of exports between the US and Chinese economies, based on Regression with ARIMA(1,1,1)(0,1,1)[4] errors, China will continue to grow the gap between it’s exports and that of the US at a fairly steady rate. This is true for all level is the confidence interval. This can be concluded from the graph below.

# Conclusion

Answer the question addressed in the introduction

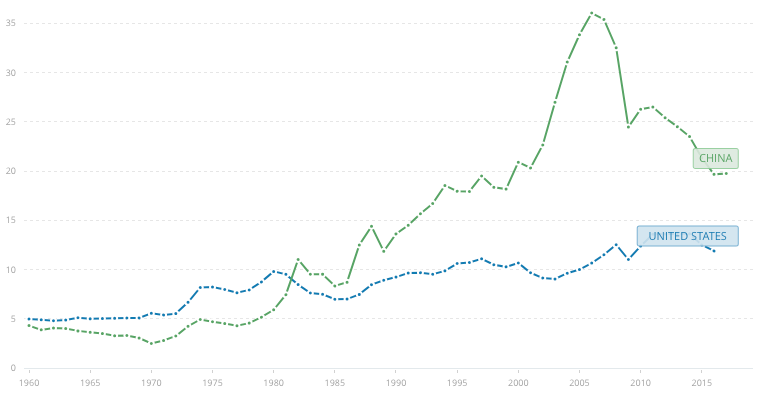
Who do we anticipate having the biggest importance on the economy in the 21st century

Consider and incorporate all the quantitative (forecasting) and qualitative factors

This being said, the purpose of this report is to make a comparison between the imports and exports of the Chinese and US economies, in terms of the value of total exports, and the net balance of each economy.

Appendix

# Appendix 1



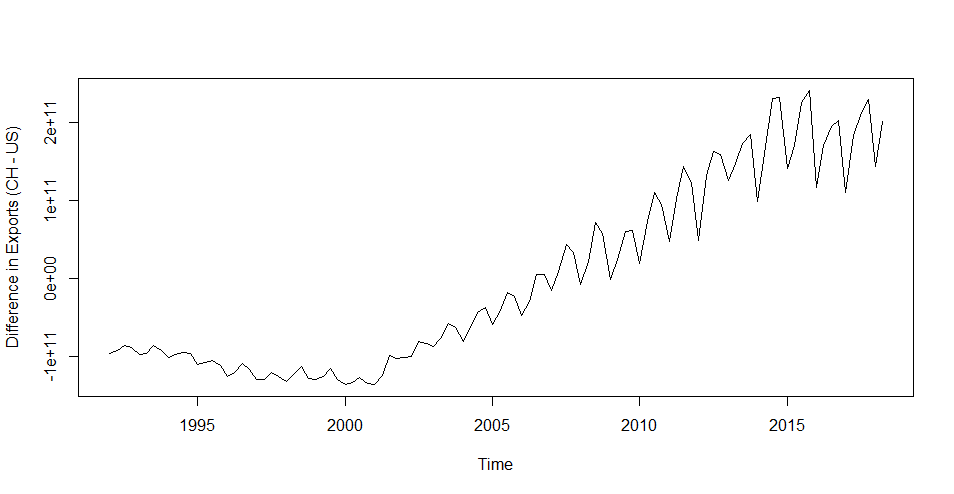
# Appendix 2

(Coccari, 1978)

* Aggregate Foreign Demand
  + Specified in the form of a combined index of industrial production for the given country
* Domestic Demand (GNP)
  + A variable representing business conditions in the given country measured by GNP
* Grants and Loans
  + Consists of government grants and long-term and short-term bank loans to other countries
* Foreign Investment
  + Direct capital outflows to other countries
* Research and Development
  + Research and development expenditure
* Exports
  + Total exports of a given country

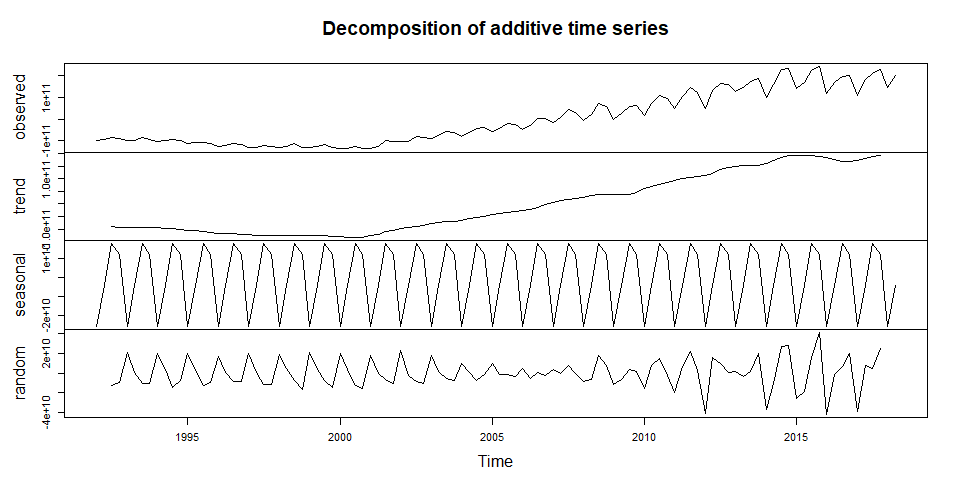
# Appendix 3

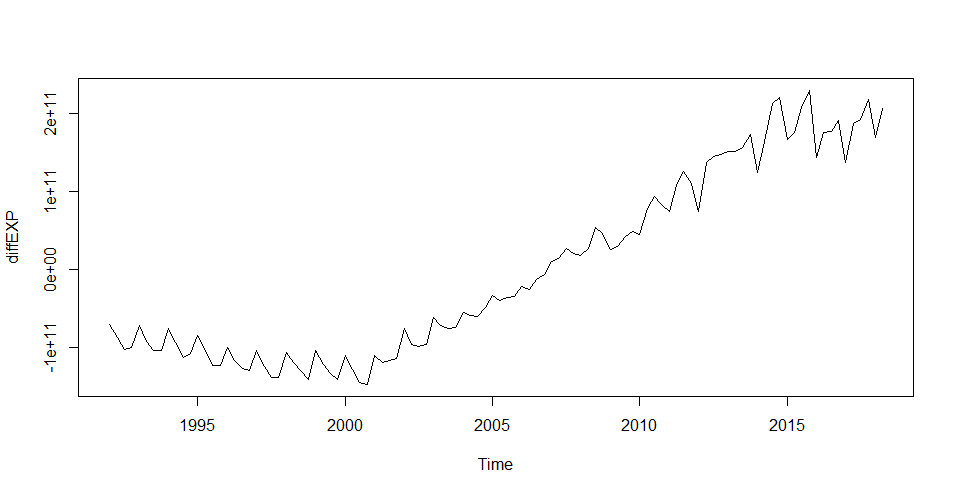
Difference in Export Value between US and China – Time Series



# Appendix 4

Difference in Export Value between US and China – Time Series Decomposition

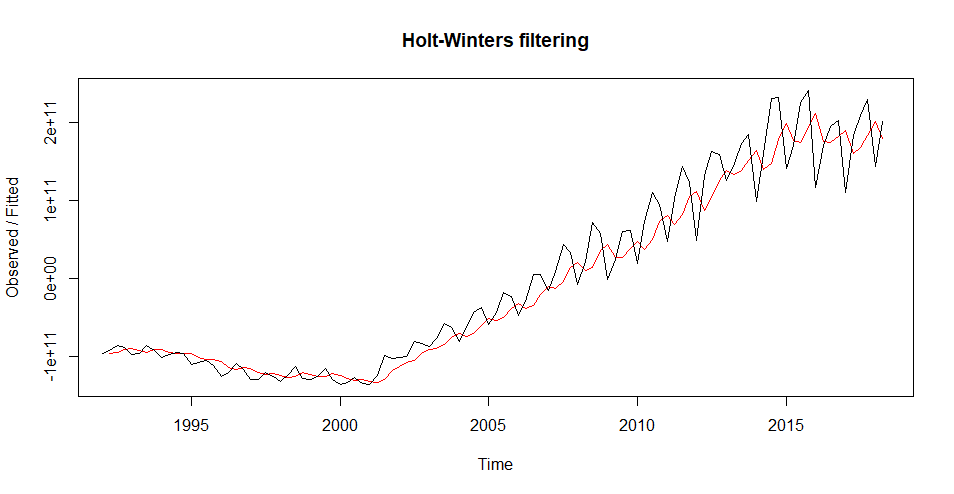


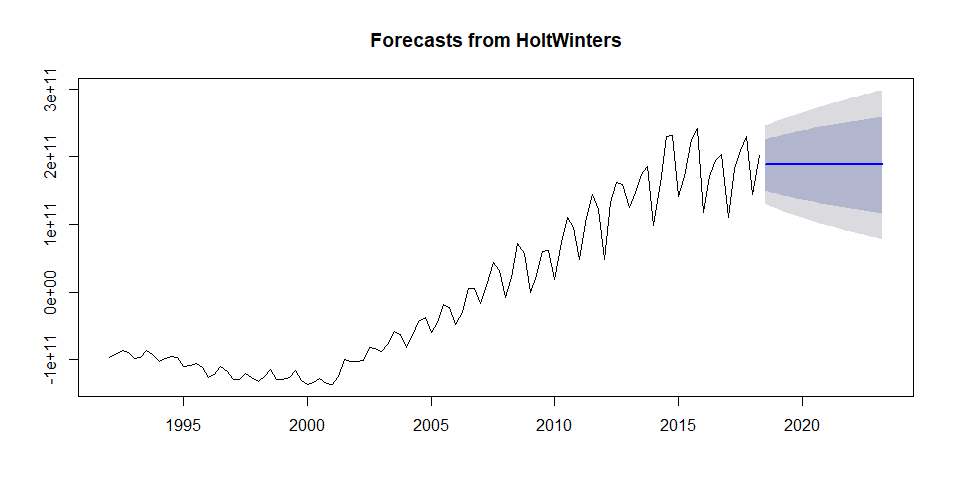


# Appendix 5

Difference in Export Value between US and China – Fitting the best model

## Simple Exponential Smoothing

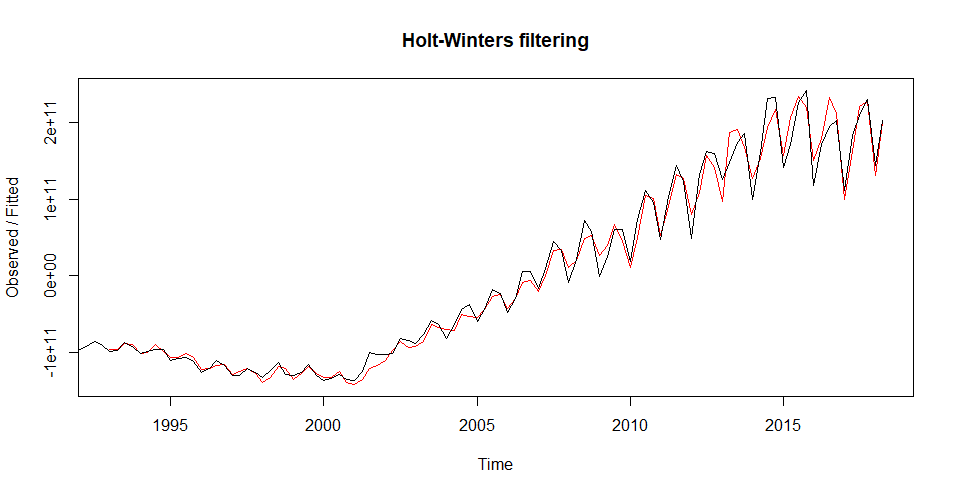


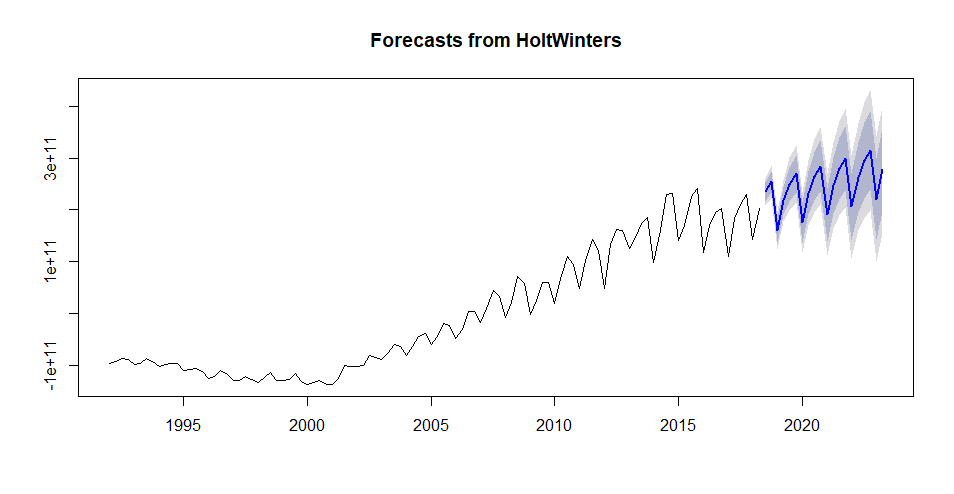


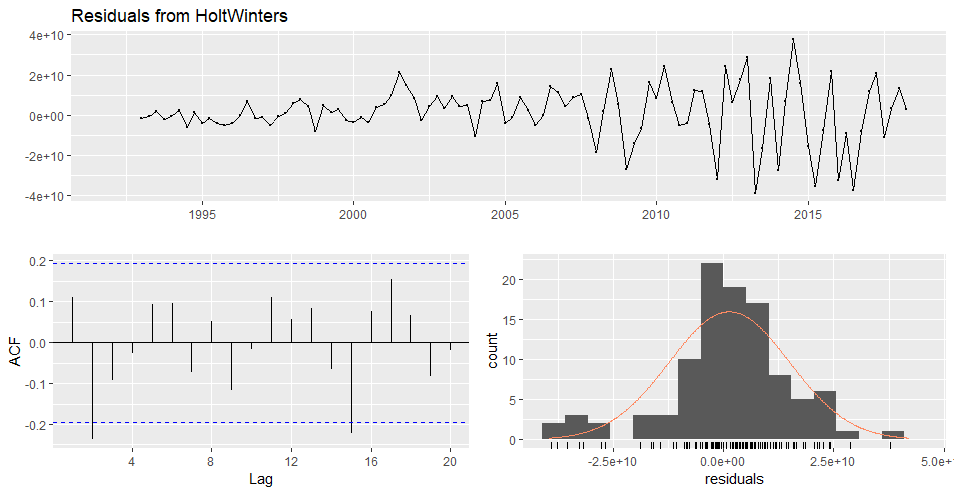
ME RMSE MAE MPE MAPE MASE ACF1

7258086819 30143558217 22145559849 71.05167 97.34416 1.250091 0.01066441

## Holt Winters



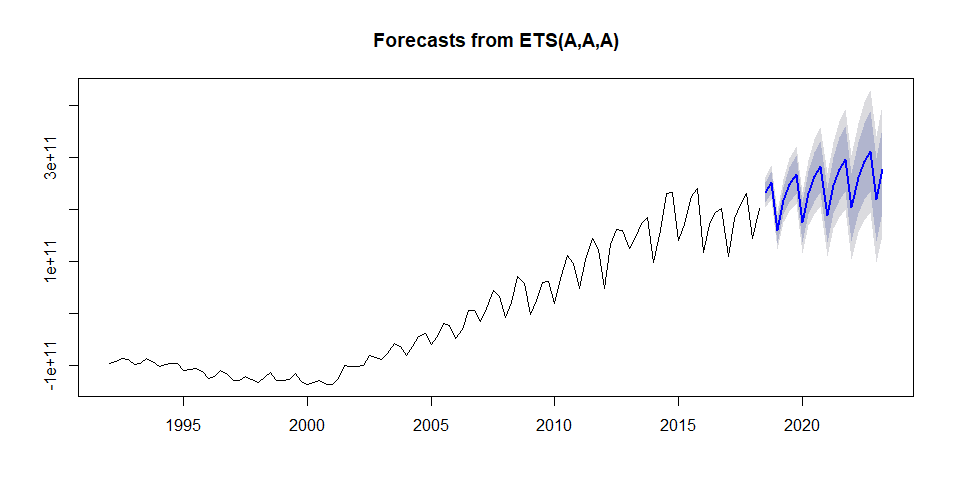


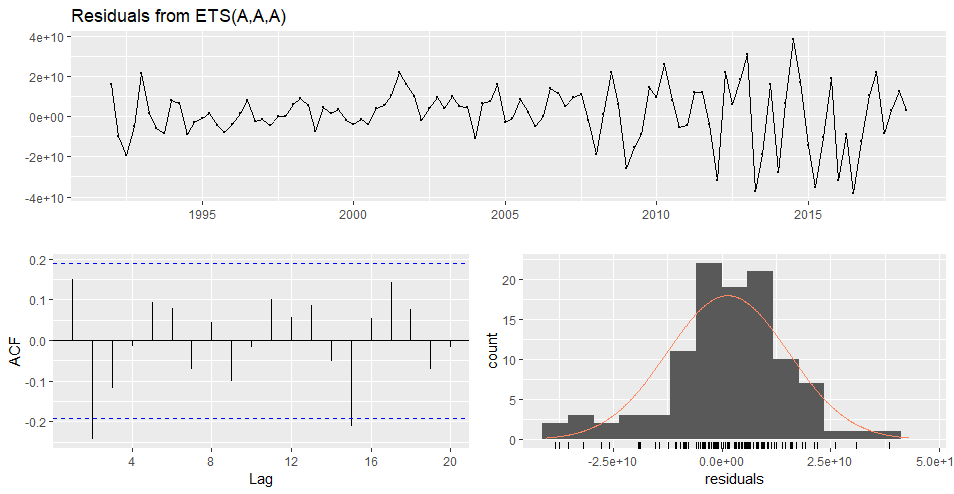


ME RMSE MAE MPE MAPE MASE ACF1

1271107037 13626892760 9878626029 40.10261 52.16816 0.5576368 0.1109597

## ETS





Ljung-Box test

data: Residuals from ETS(A,A,A)

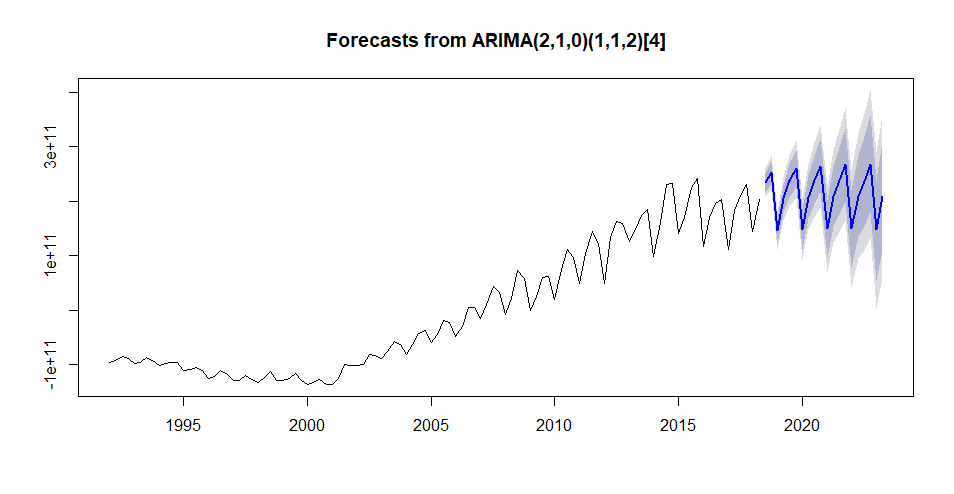
Q\* = 15.526, df = 3, p-value = 0.001418

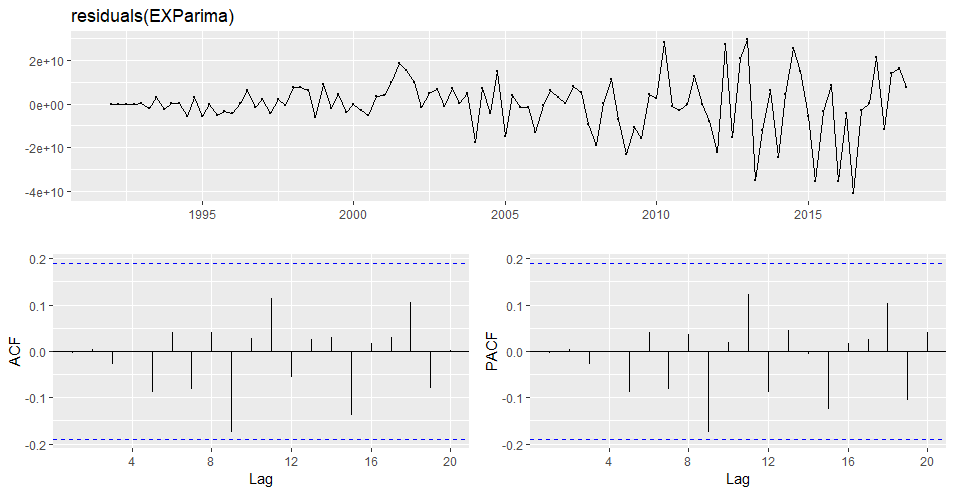
Model df: 8. Total lags used: 11

ME RMSE MAE MPE MAPE MASE ACF1

1287289347 13905720587 10488954049 37.13743 50.04595 0.5920891 0.1491436

## Auto-ARIMA





ARIMA(2,1,0)(1,1,2)[4]

Coefficients:

ar1 ar2 sar1 sma1 sma2

-0.1884 -0.3479 0.8749 -1.5627 0.7047

s.e. 0.0939 0.1078 0.0807 0.1389 0.1295

sigma^2 estimated as 1.701e+20: log likelihood=-2495.46

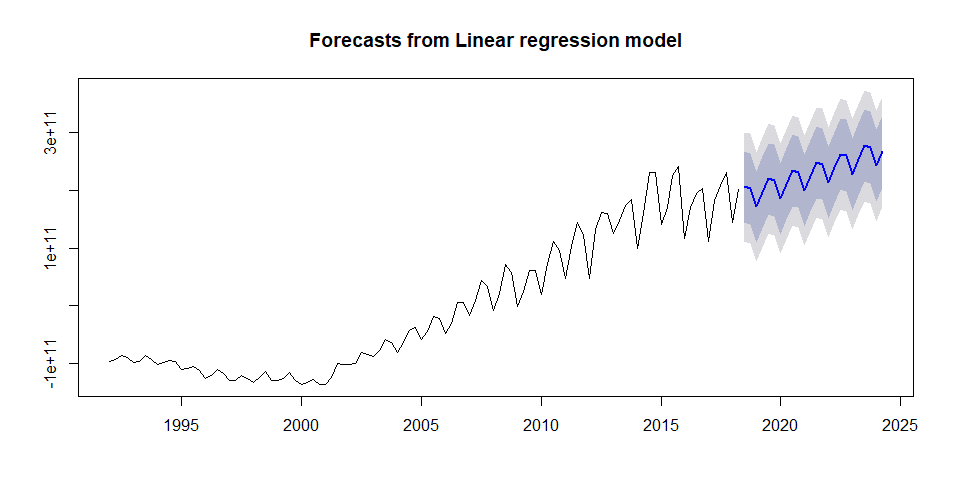
AIC=5002.91 AICc=5003.81 BIC=5018.6

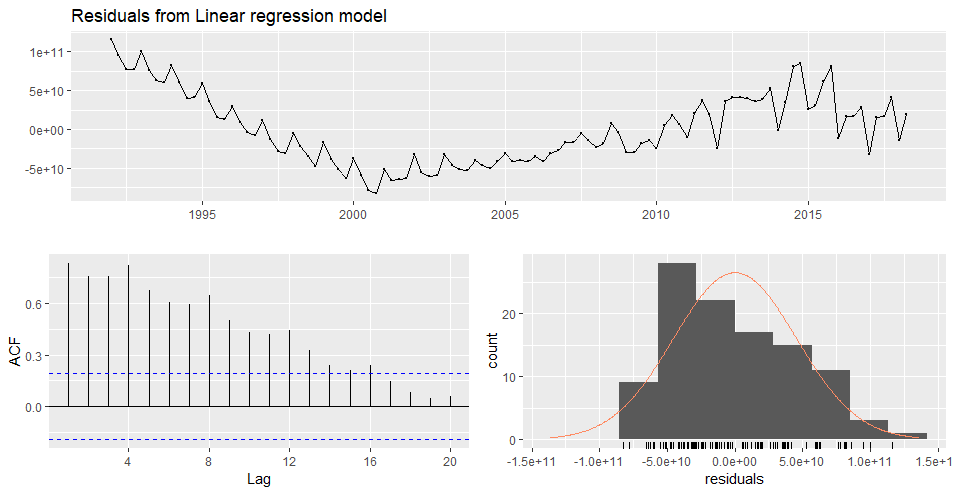
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set -134431860 12410416857 8488877917 31.40808 41.1162 0.4791871 -0.003596246

## Linear Regression





Residuals:

Min 1Q Median 3Q Max

-8.270e+10 -3.502e+10 -8.638e+09 3.549e+10 1.163e+11

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.161e+11 1.183e+10 -18.262 < 2e-16 \*\*\*

trend 3.557e+09 1.470e+08 24.206 < 2e-16 \*\*\*

season2 2.145e+10 1.260e+10 1.702 0.09181 .

season3 4.188e+10 1.272e+10 3.293 0.00137 \*\*

season4 3.519e+10 1.272e+10 2.766 0.00675 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.629e+10 on 101 degrees of freedom

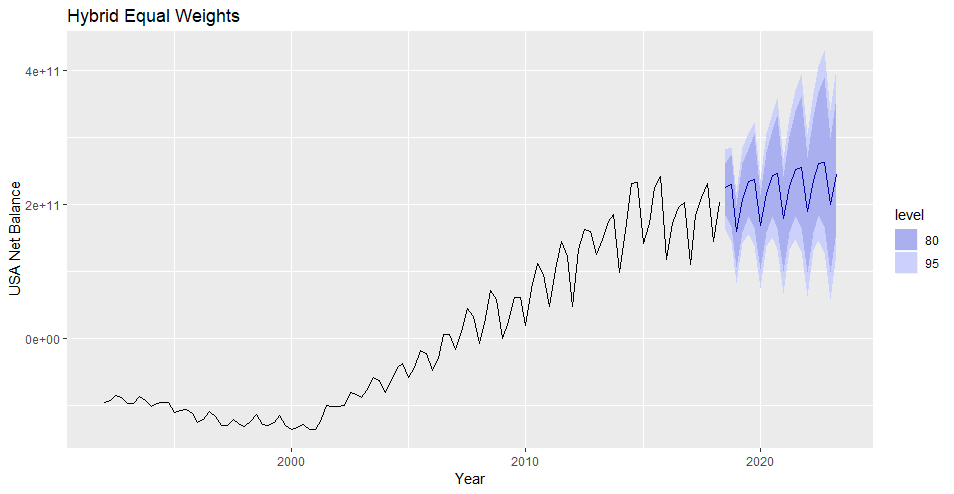
Multiple R-squared: 0.8558, Adjusted R-squared: 0.8501

F-statistic: 149.9 on 4 and 101 DF, p-value: < 2.2e-16

## Combinations

Combination of Auto-ARIMA, ETS, Thetam Model, Netar and STLM and TBATS

### Forecast Hybrid with Equal Weights

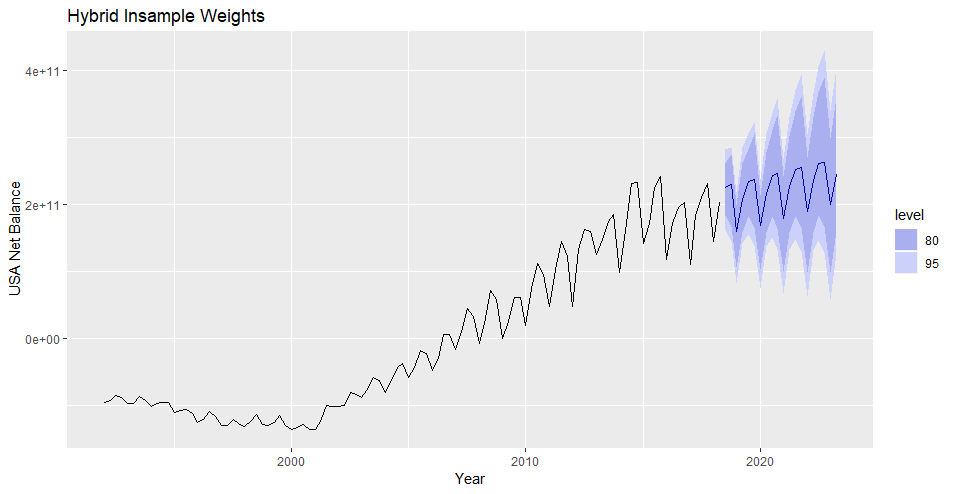




ME RMSE MAE MPE MAPE MASE ACF1

Training set 2293222577 14541428216 10548042676 39.3316 52.78025 0.5954246 0.09357328

### Forecast Hybrid with Sample Weights



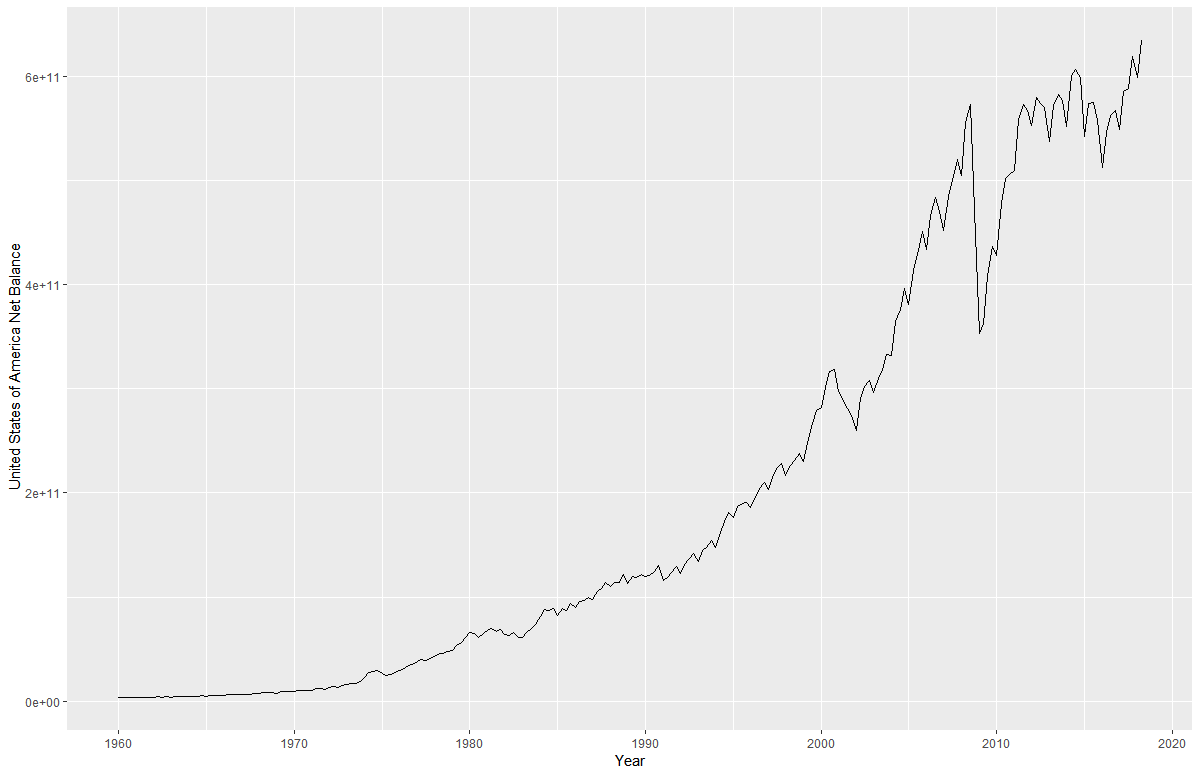


ME RMSE MAE MPE MAPE MASE ACF1

Training set 1542903196 9358251811 34.96899 46.82792 0.5282623 0.1073805

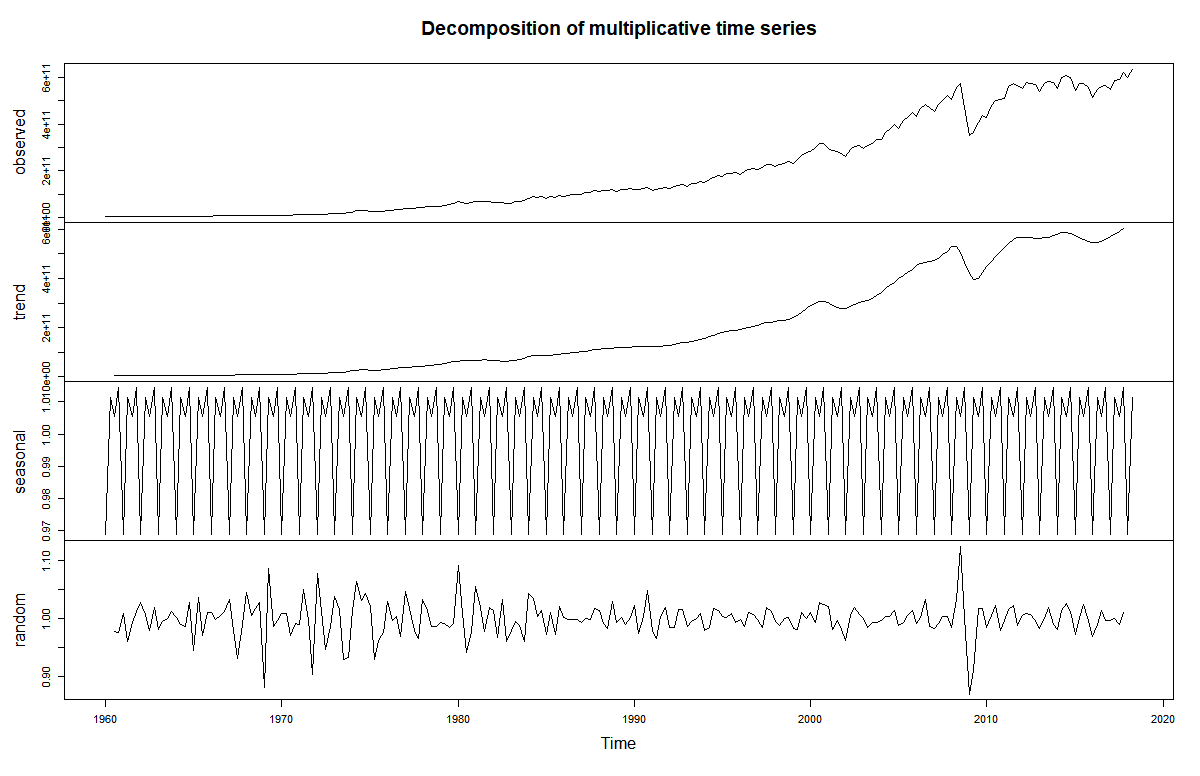
# Appendix 6

U.S. quarterly net balance 1960-2018 – Time Series



# Appendix 7

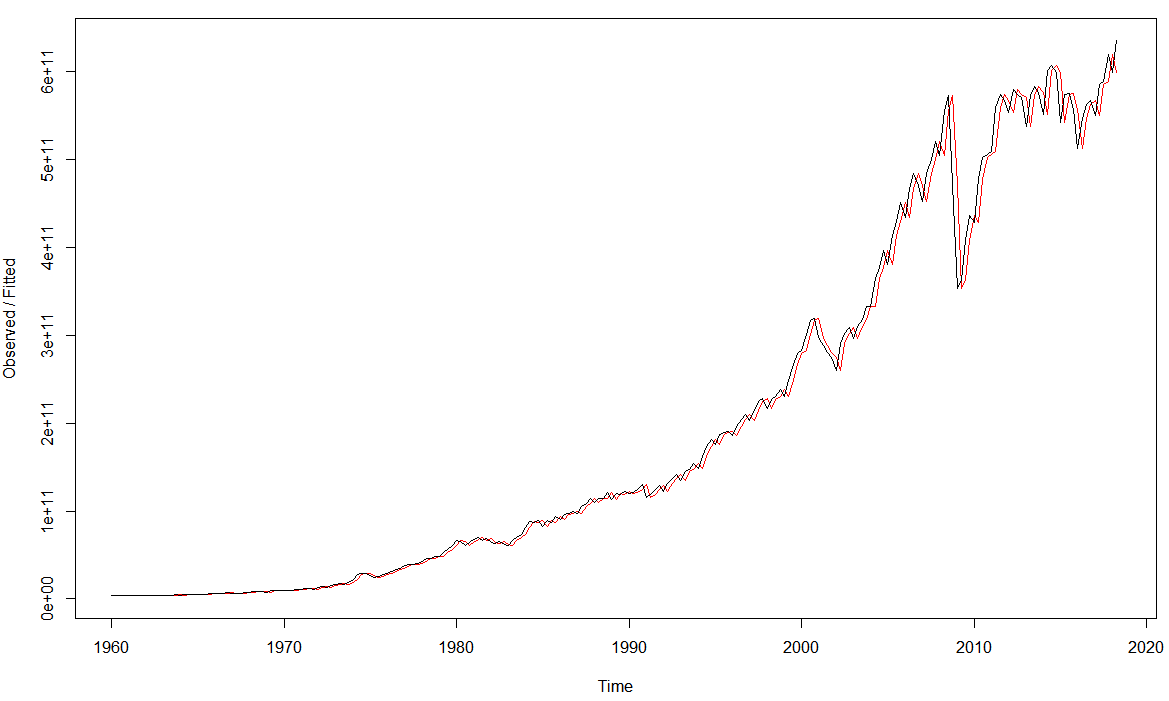
U.S. quarterly net balance 1960-2018 – Time Series Decomposition

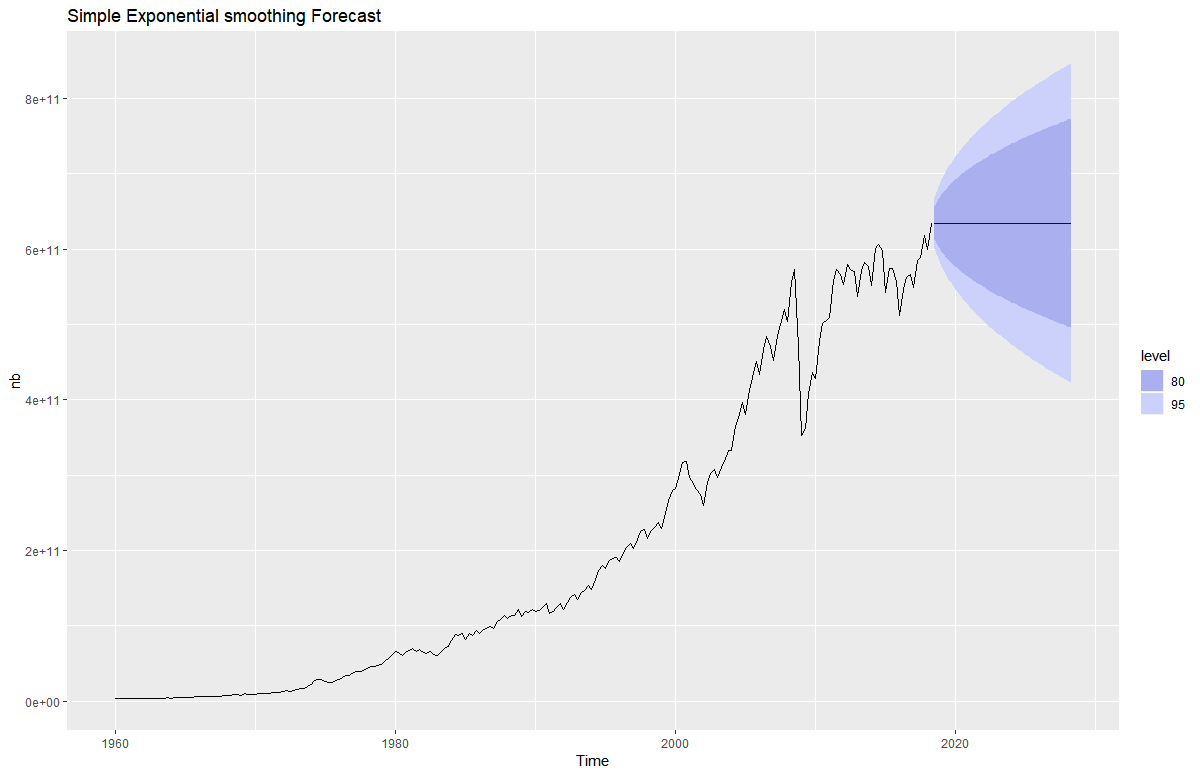


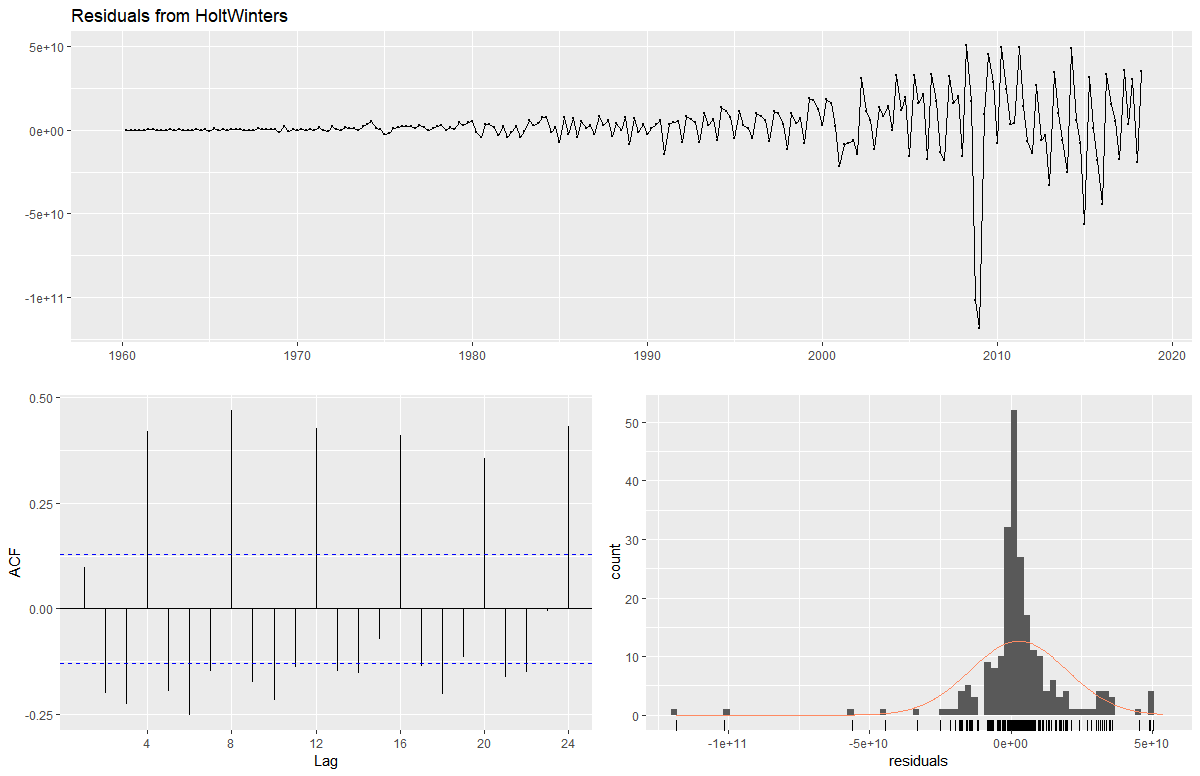
# Appendix 8

U.S. quarterly net balance 1960-2018 – Fitting the best model

## Simple Exponential Smoothing

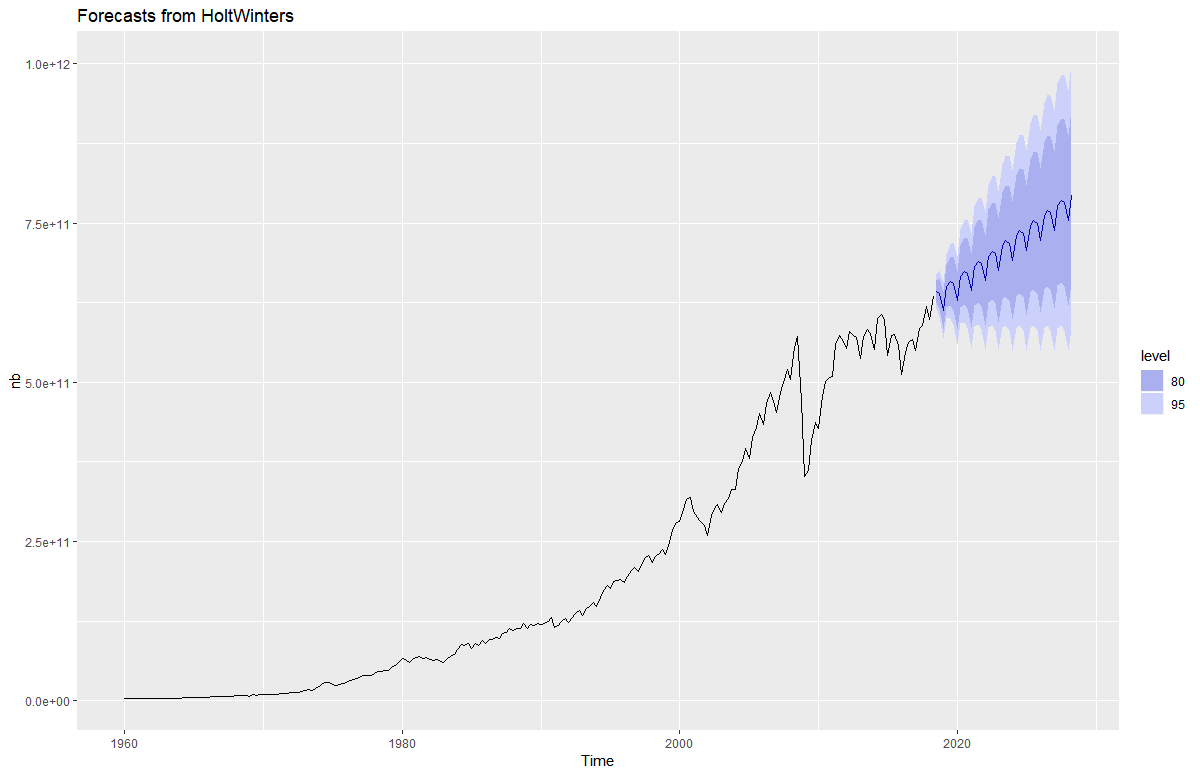


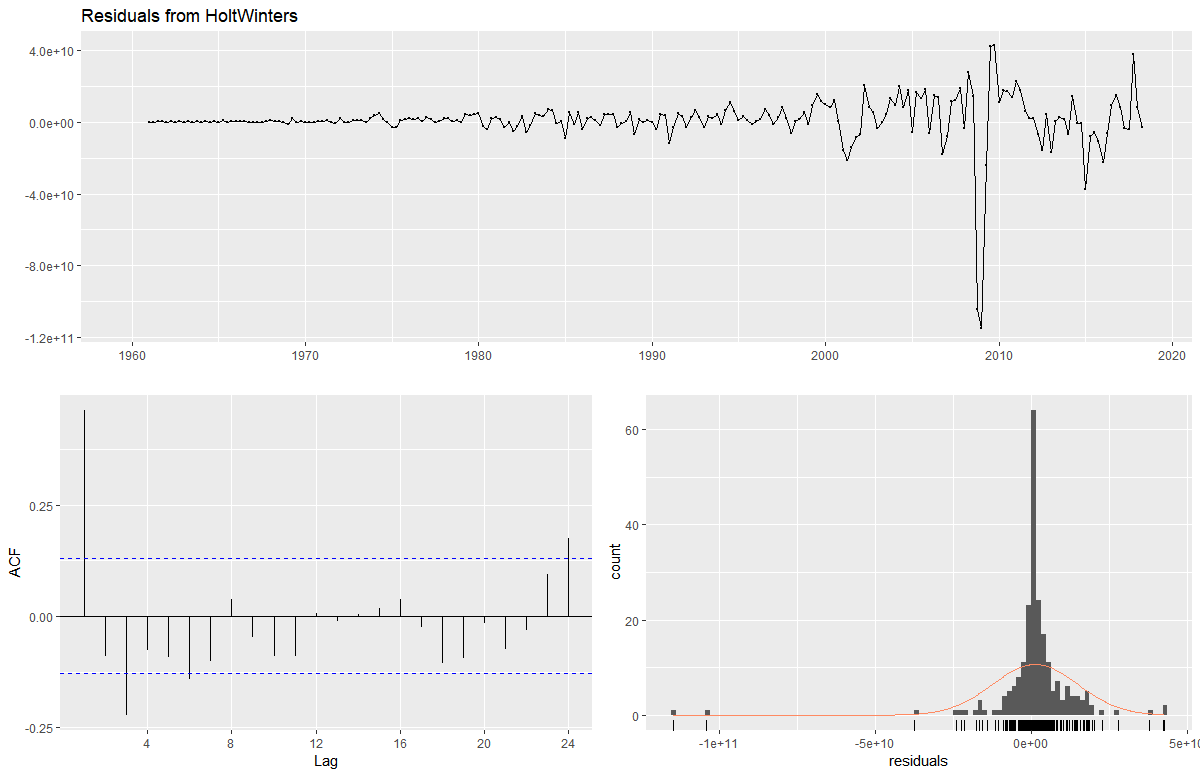






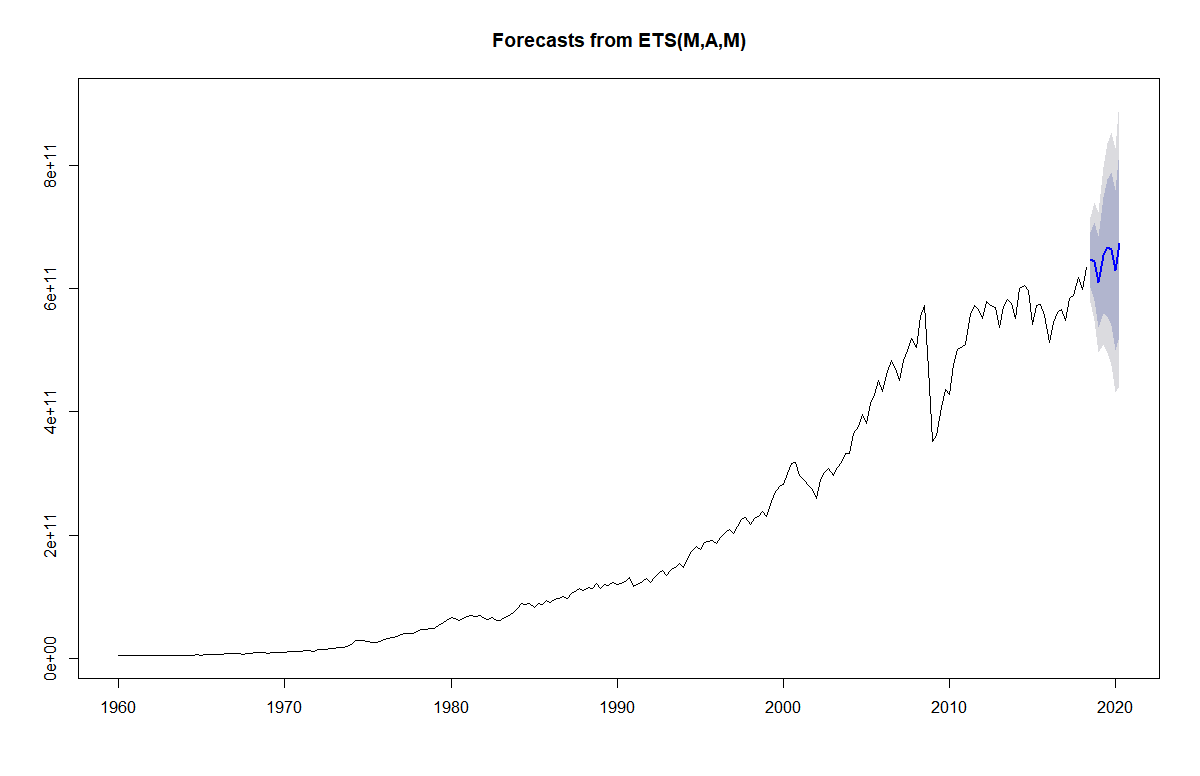
## Holt Winters

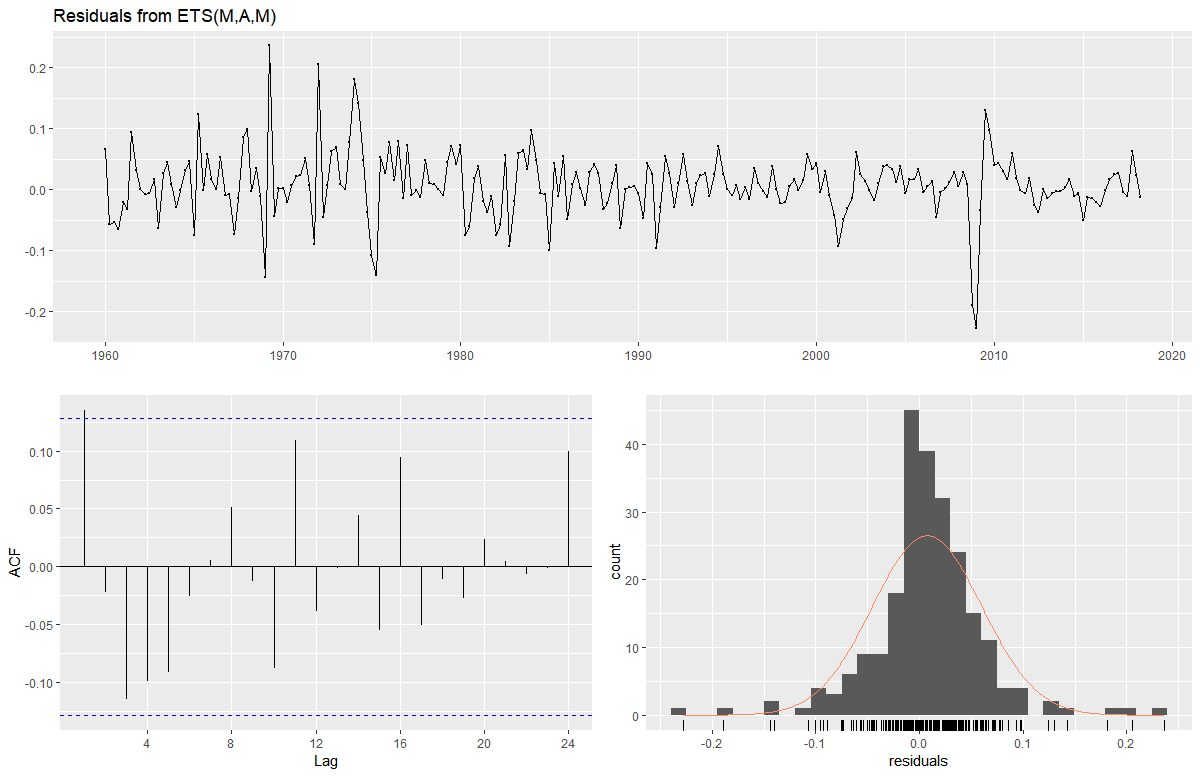




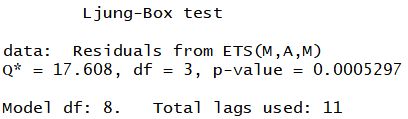


## ETS

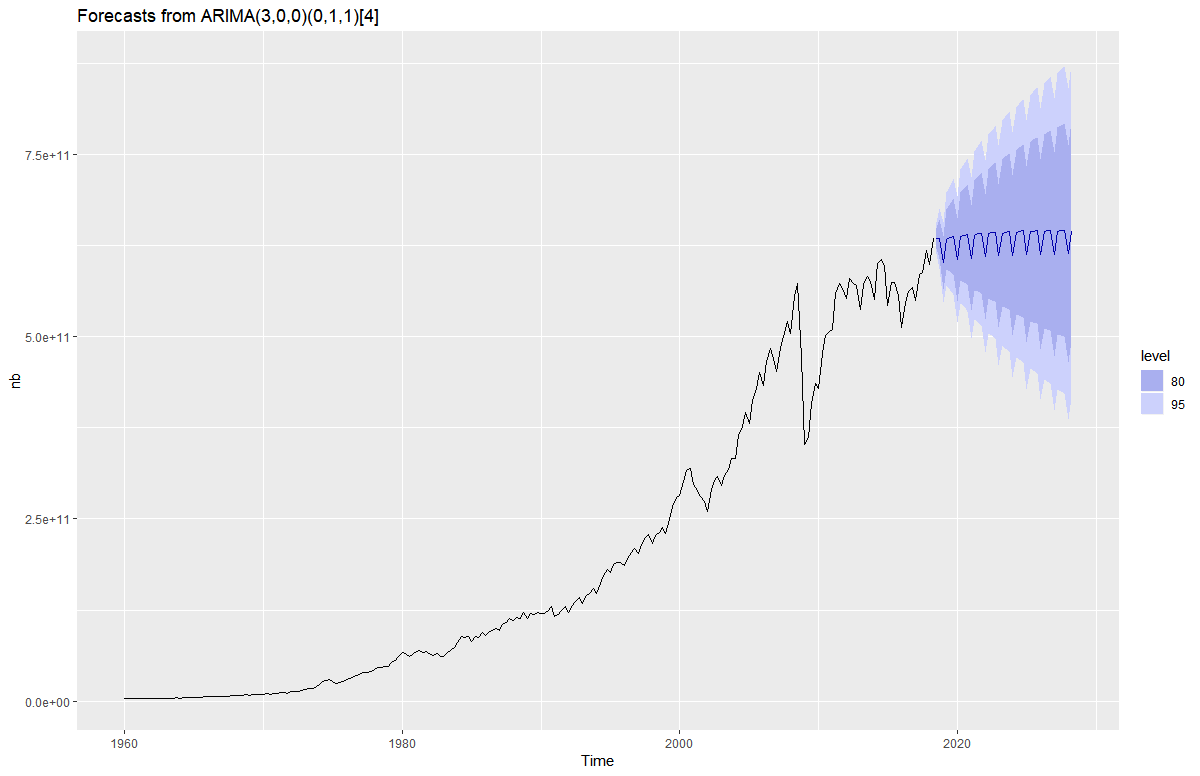


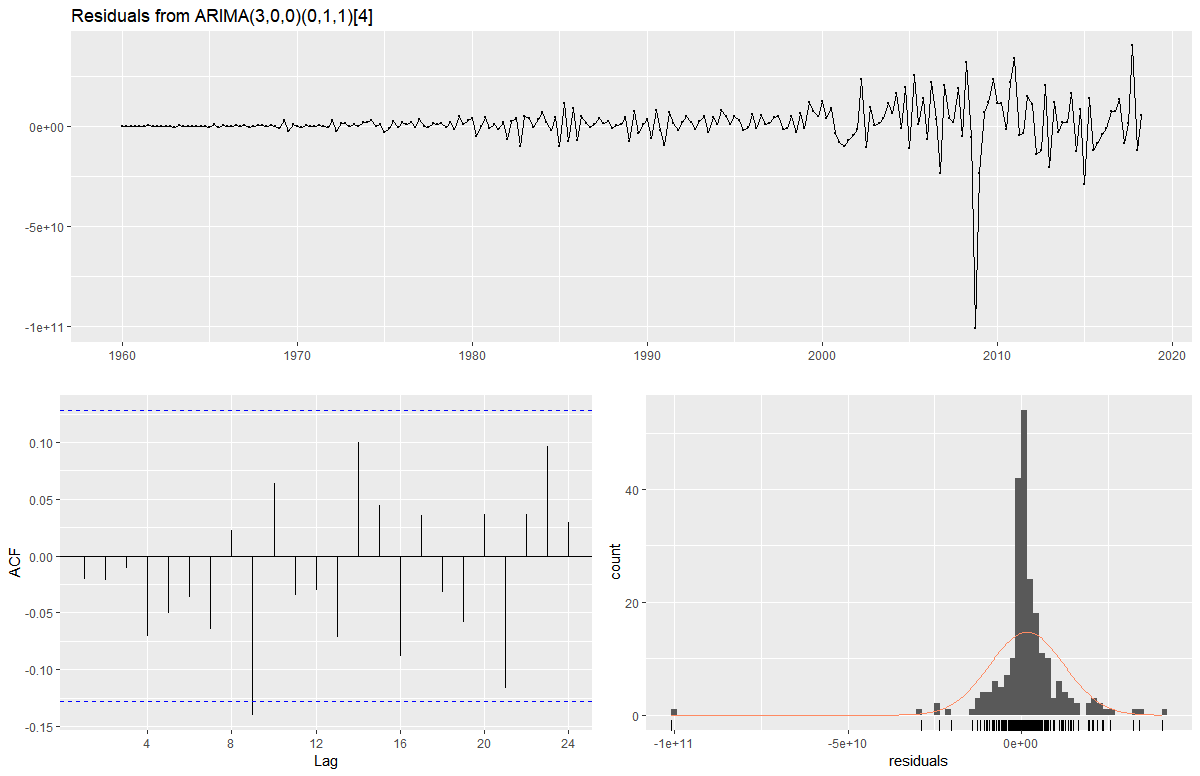


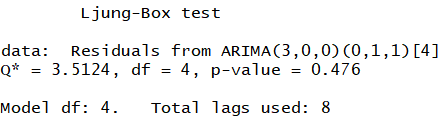




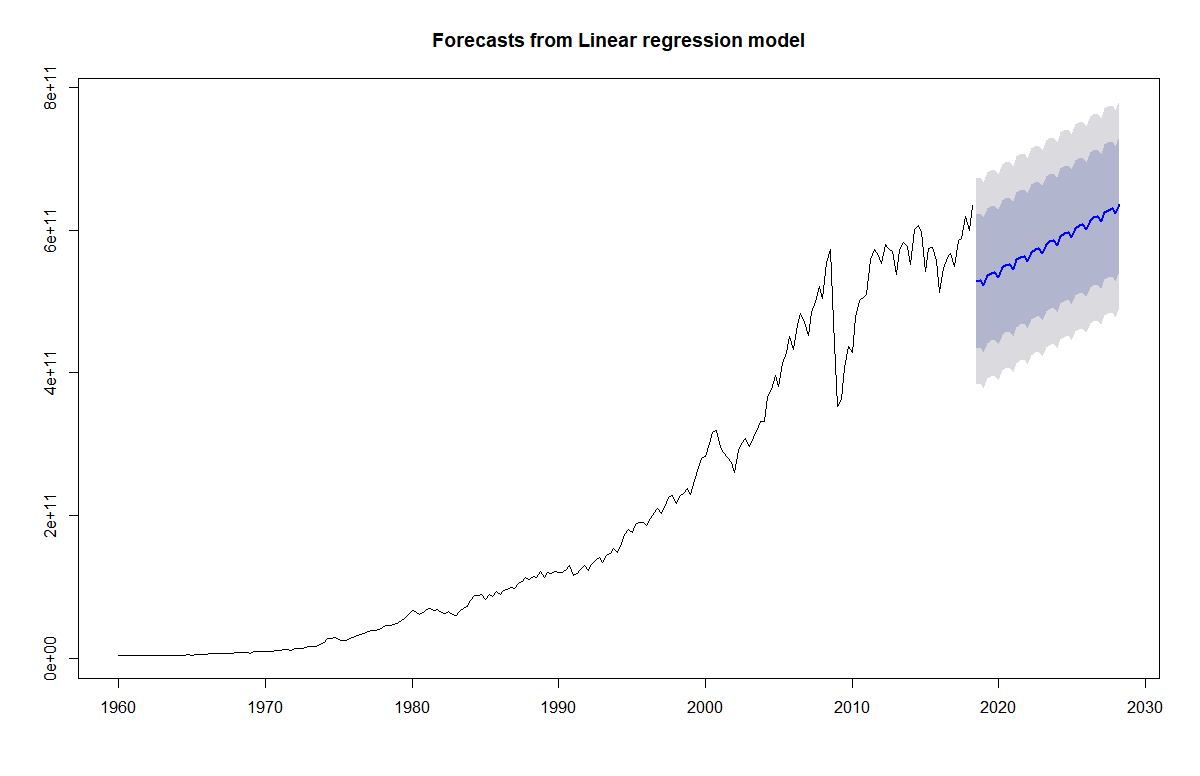
## ARIMA

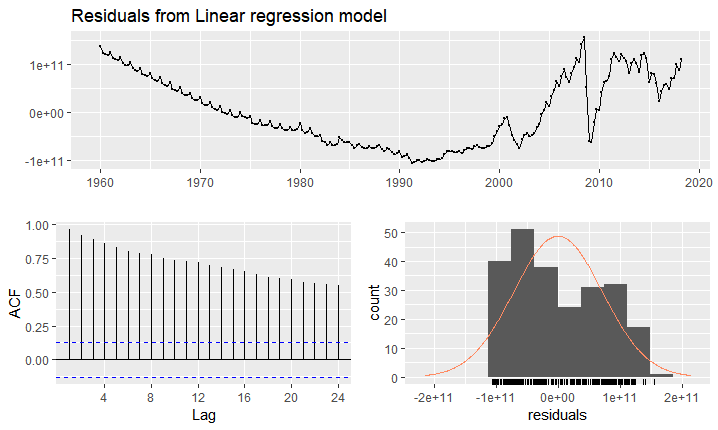




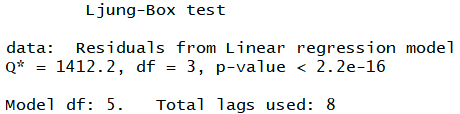


## Linear Regression



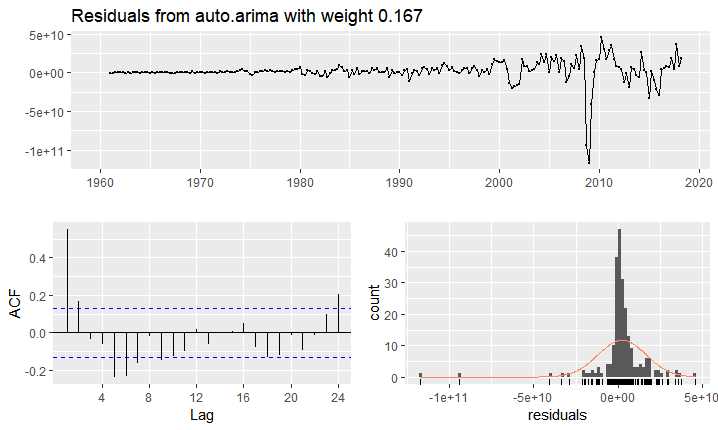
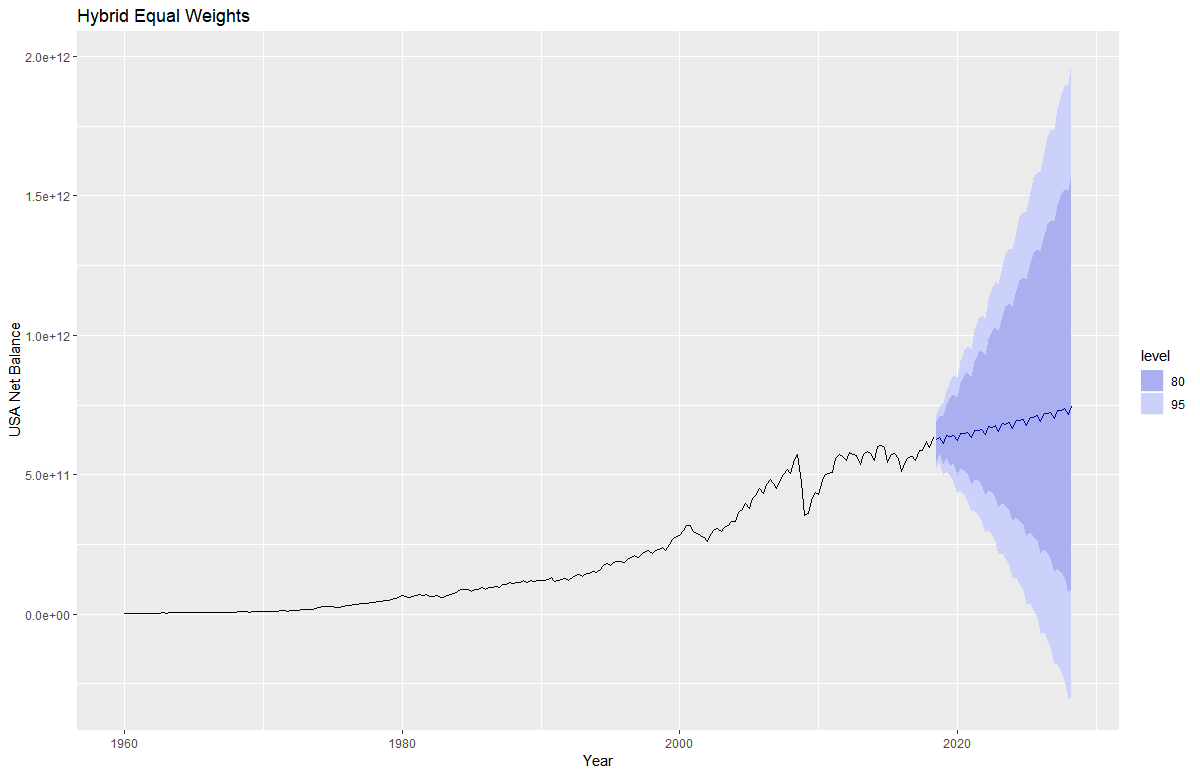




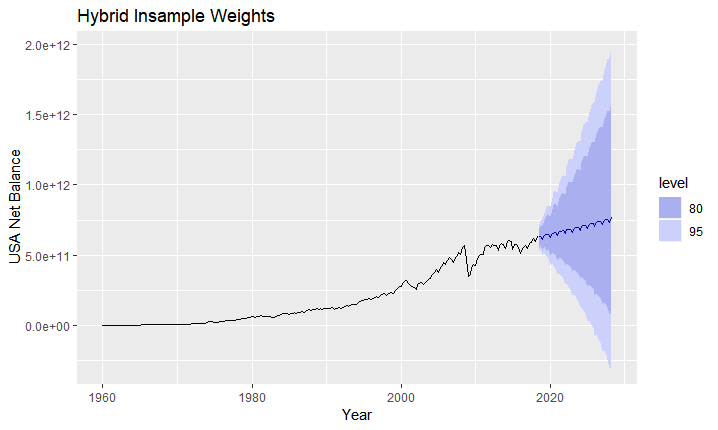


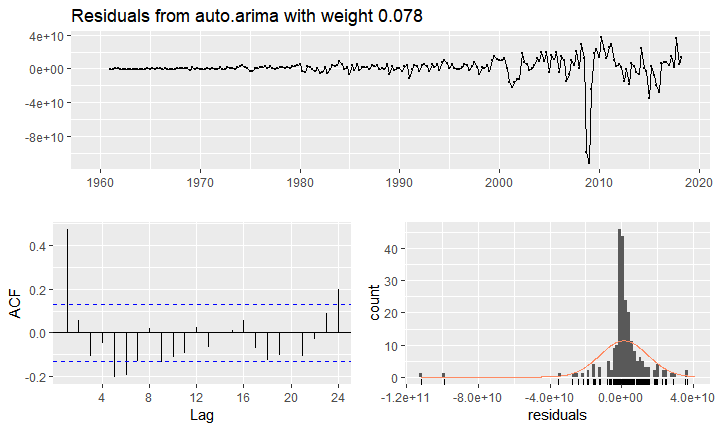
## Combinations

### Forecast Hybrid with Equal Weights



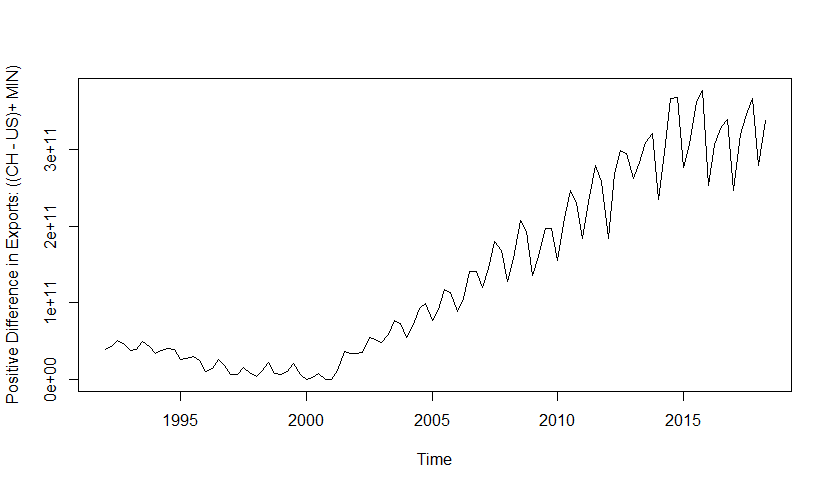
### Forecast Hybrid with In-Sample Weights





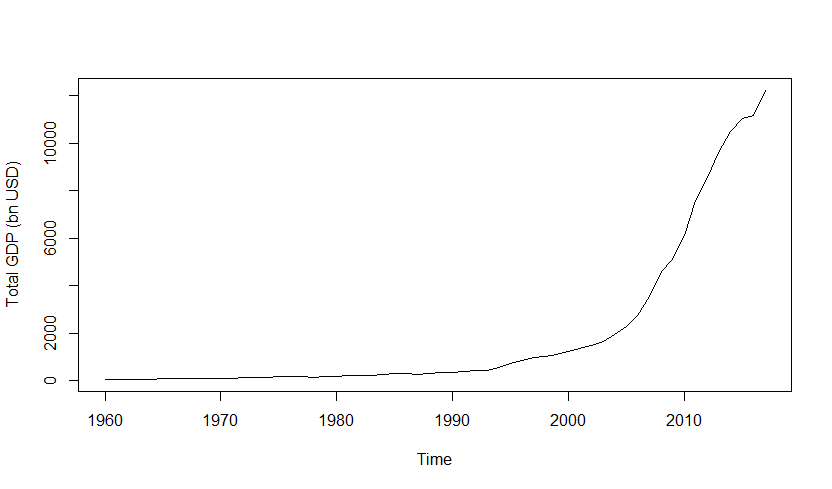
# Appendix 9

Positive Difference in Value of Export Data for Regression Fit

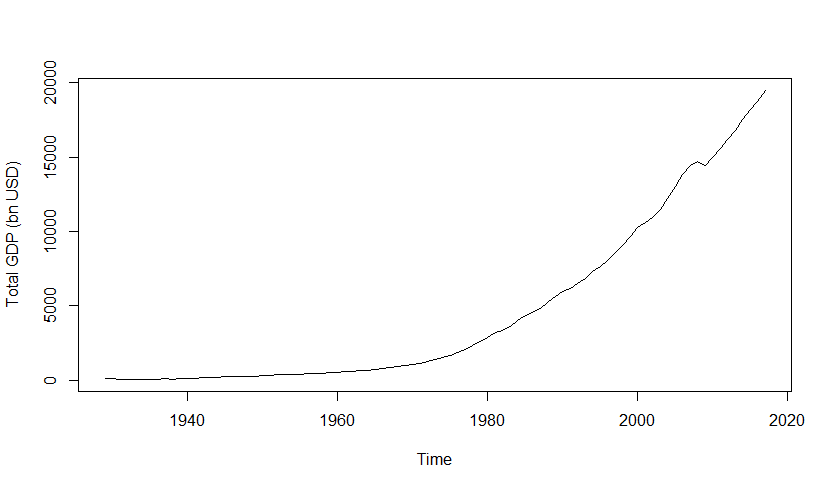


# Appendix 10

Yearly GDP for China

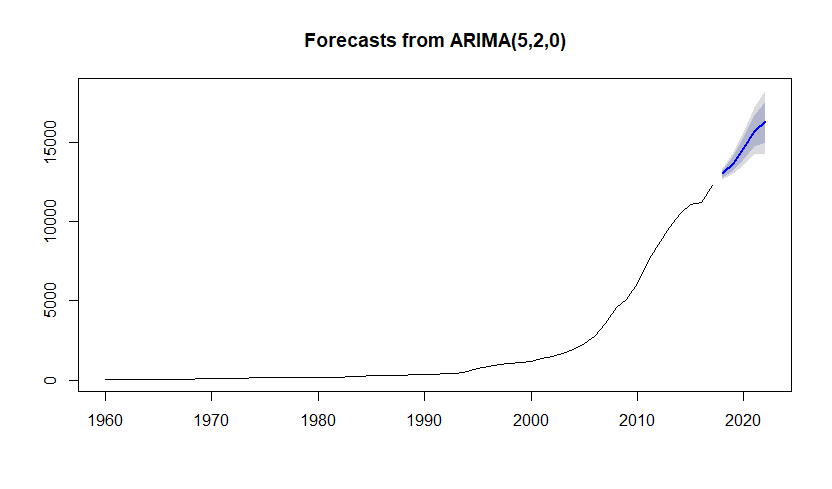


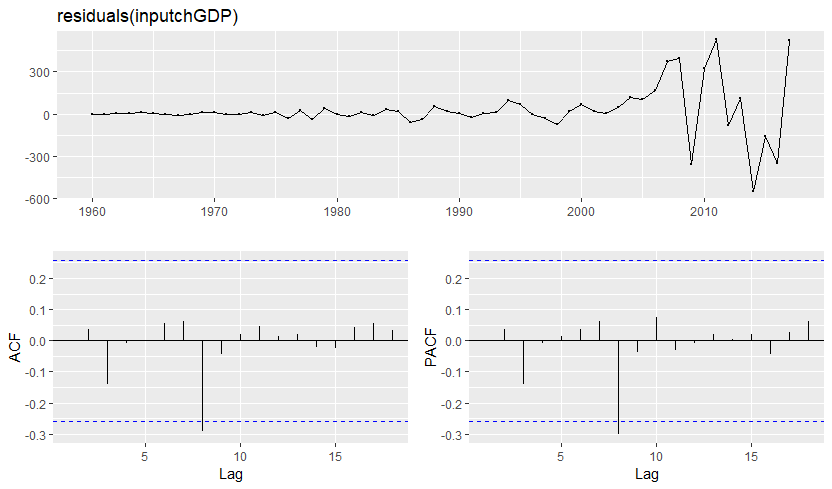
Yearly GDP US



# Appendix 11

5 year forecast for Chinese GDP:





Series: chGDP

ARIMA(5,2,0)

Coefficients:

ar1 ar2 ar3 ar4 ar5

-0.3351 -0.2025 0.0774 0.1051 -0.3958

s.e. 0.1494 0.1599 0.1610 0.1709 0.1641

sigma^2 estimated as 32404: log likelihood=-368.49

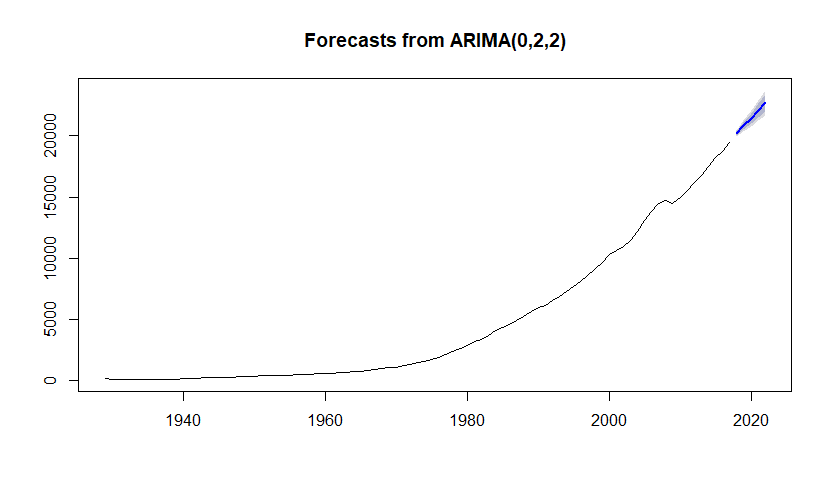
AIC=748.99 AICc=750.7 BIC=761.14

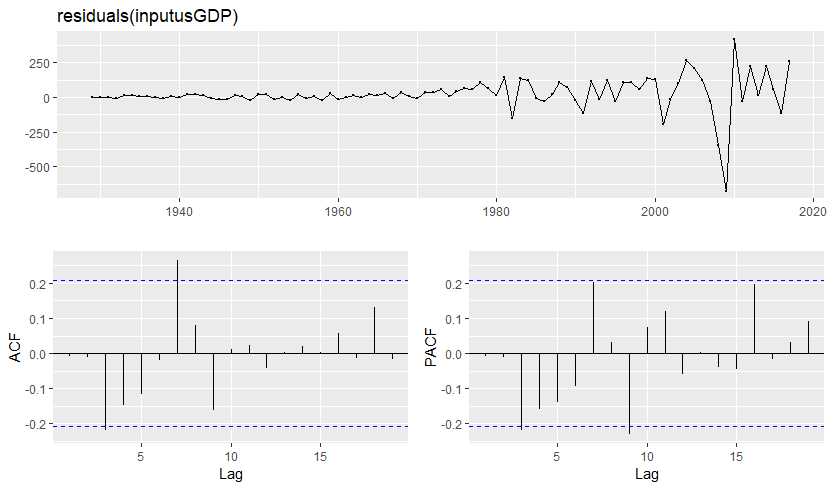
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 24.73592 168.8003 88.0751 2.05054 7.29542 0.406252 -0.001744568

5 year forecast for US GDP:





Auto.arima for US GDP:

Series: usGDP

ARIMA(0,2,2)

Coefficients:

ma1 ma2

-0.4118 -0.3083

s.e. 0.0965 0.0869

sigma^2 estimated as 16065: log likelihood=-544.03

AIC=1094.05 AICc=1094.34 BIC=1101.45

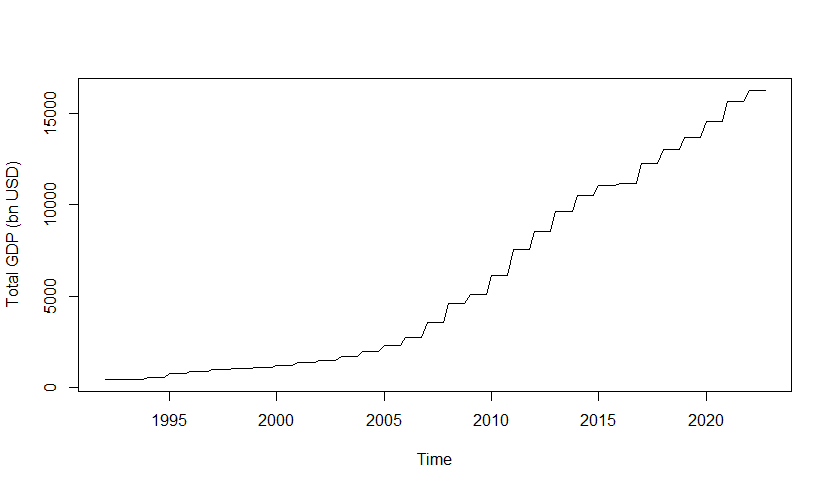
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

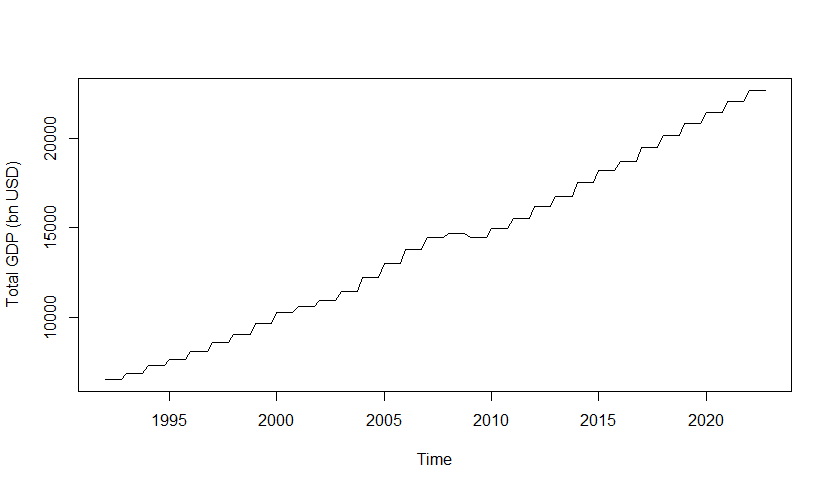
Training set 25.57823 123.8666 68.40446 1.88507 3.548316 0.3006835 -0.008145399

# Appendix 12

Chinese GDP as predictor variable (forecast horizon included)

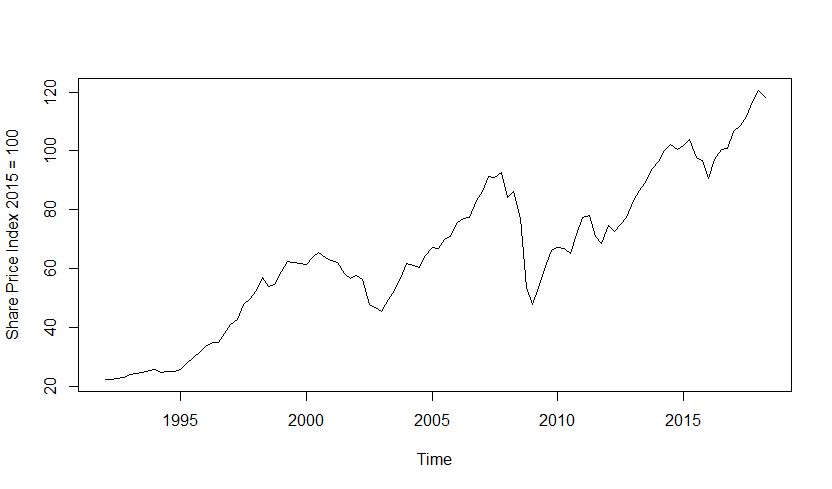


US GDP as predictor variable (forecast horizon included)



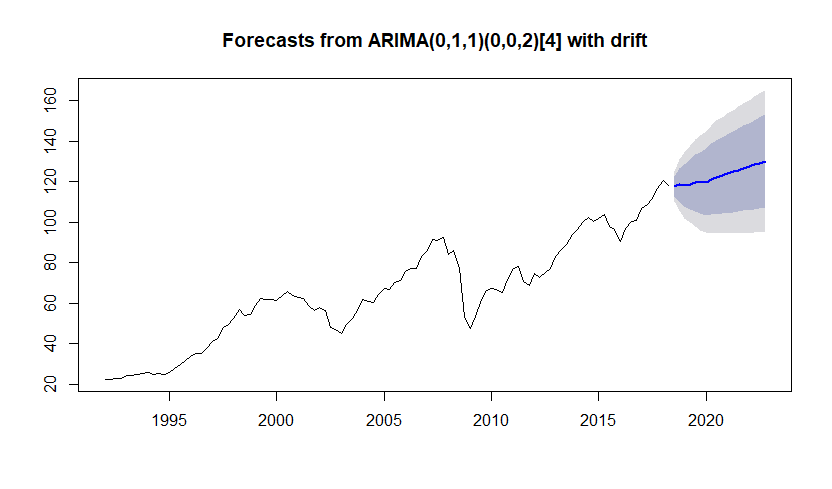
# Appendix 13

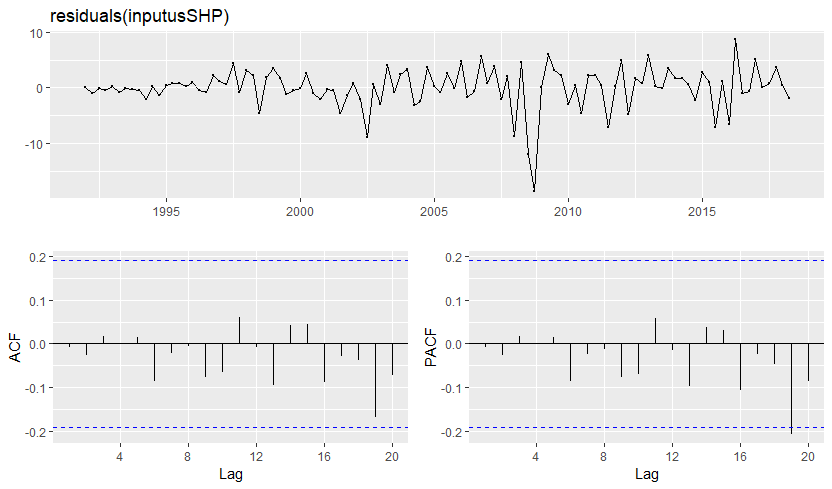
Total Prices for all Shares for the US Quarterly



# Appendix 14

5 year forecast of Total Share Prices for US





Series: usSHP

ARIMA(0,1,1)(0,0,2)[4] with drift

Coefficients:

ma1 sma1 sma2 drift

0.3687 -0.0710 -0.2307 0.8786

s.e. 0.0945 0.0945 0.0918 0.3548

sigma^2 estimated as 14.24: log likelihood=-286.69

AIC=583.39 AICc=583.99 BIC=596.66

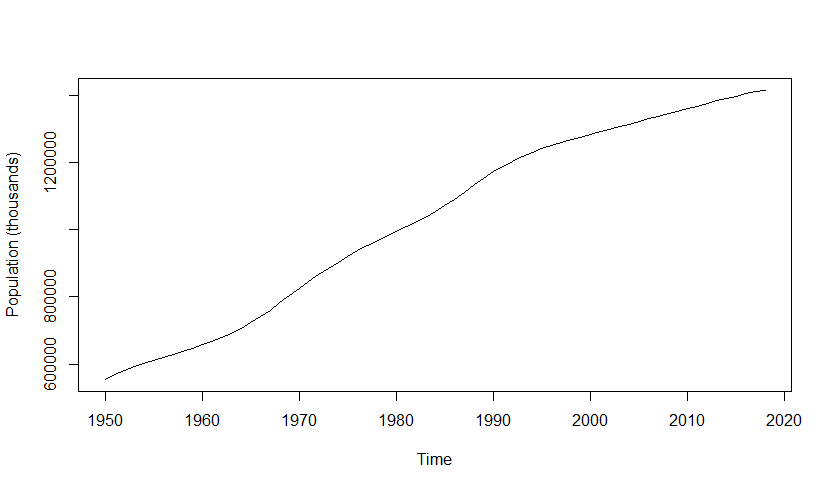
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

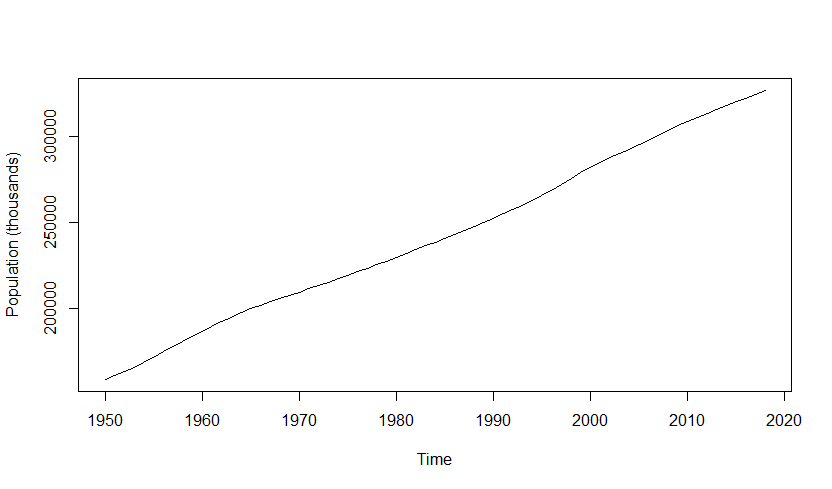
Training set -0.00373768 3.683213 2.42978 -0.2352762 3.824772 0.2828589 -0.009082655

# Appendix 15

Chinese Population data

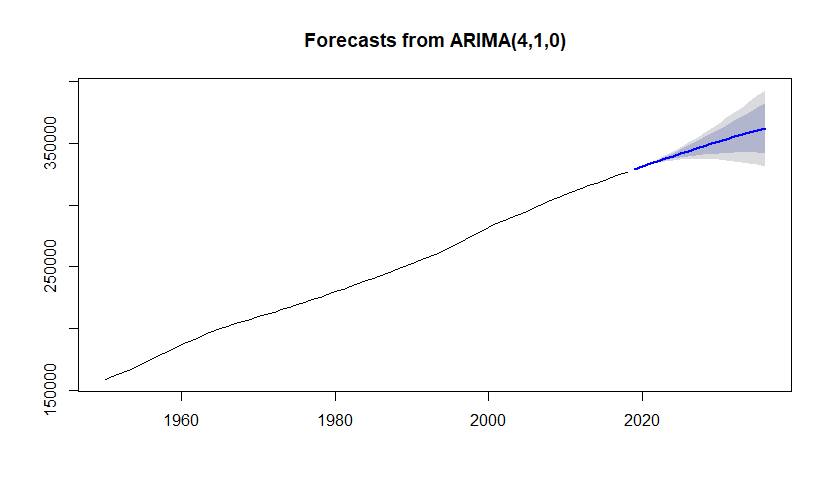


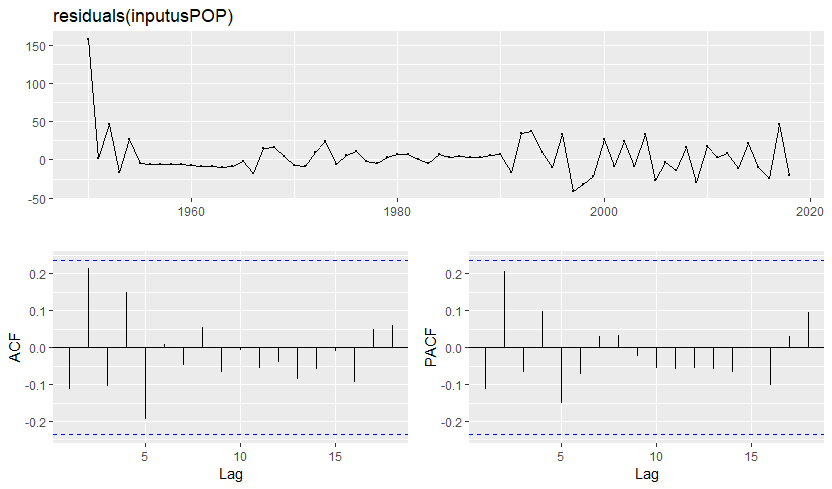
US Population data



# Appendix 16

US Population forecast





Series: usPOP

ARIMA(4,1,0)

Coefficients:

ar1 ar2 ar3 ar4

3.2876 -4.2561 2.5950 -0.6273

s.e. 0.0990 0.2731 0.2757 0.1017

sigma^2 estimated as 732.1: log likelihood=-296.65

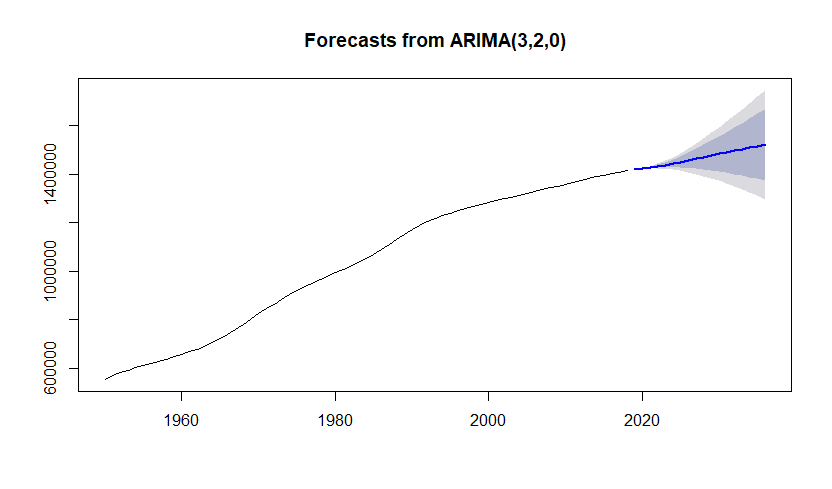
AIC=603.29 AICc=604.26 BIC=614.39

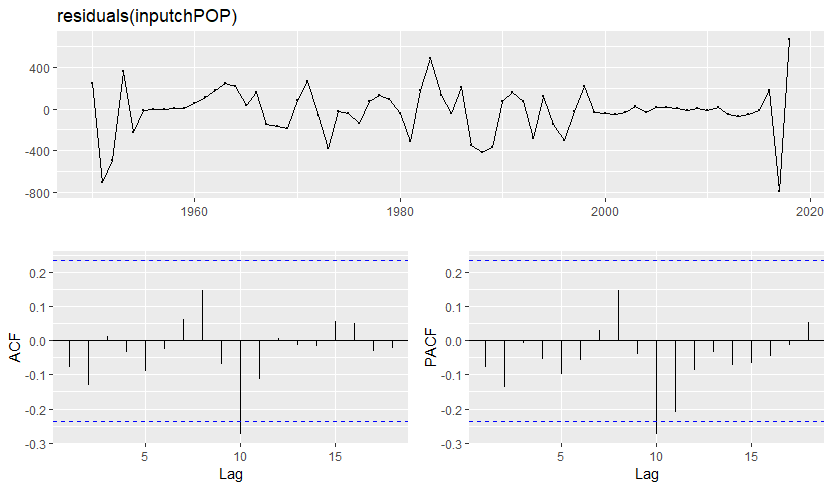
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 3.825057 26.05809 15.98585 0.002133681 0.007033149 0.006473787 -0.1134771

Chinese Population forecast





Series: chPOP

ARIMA(3,2,0)

Coefficients:

ar1 ar2 ar3

1.9919 -1.3960 0.3218

s.e. 0.1222 0.2248 0.1283

sigma^2 estimated as 59703: log likelihood=-459

AIC=926.01 AICc=926.66 BIC=934.83

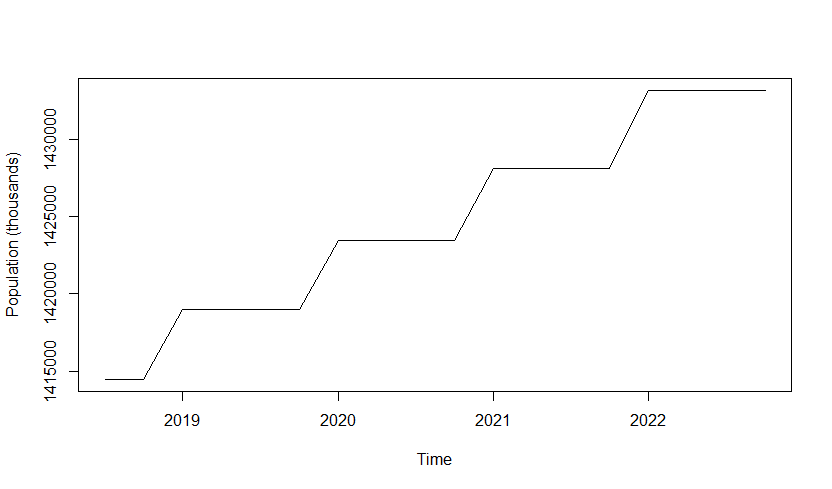
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

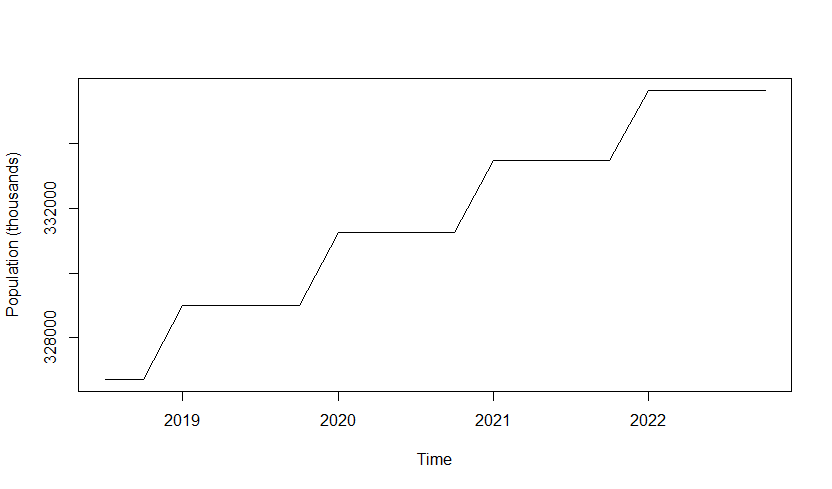
Training set -16.75988 235.3231 158.831 -0.001844313 0.01759377 0.01255801 -0.07689315

# Appendix 17

Chinese Population predictor series for forecast horizon



US Population predictor series for forecast horizon



# Appendix 18

Best subset combinations for dynamic regression model selection

Predictors:

1) GDP US

2) GDP China

3) Population US

4) Population China

5) US Total Share Prices

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Combination # | | | | | | | | | | | | | |
| Predictor | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 1 | Y | Y | Y | Y | y |  |  |  |  | Y | Y | Y |  |
| 2 | Y | Y | Y | y |  |  |  |  | y |  | Y | Y | Y |
| 3 | Y | Y | Y |  |  |  |  | Y | Y | Y |  | Y |  |
| 4 | Y | y |  |  |  |  | y | Y | Y | Y | Y |  |  |
| 5 | Y |  |  |  |  | y | Y | Y | Y | Y | Y | Y |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Combination # | | | | | |
| Predictor | 14 | 15 | 16 | 17 | 18 | 19 |
| 1 |  | Y |  |  |  |  |
| 2 |  | Y | Y |  | Y |  |
| 3 |  |  | y | Y | Y | y |
| 4 | Y |  |  | y | Y |  |
| 5 |  |  |  |  |  |  |

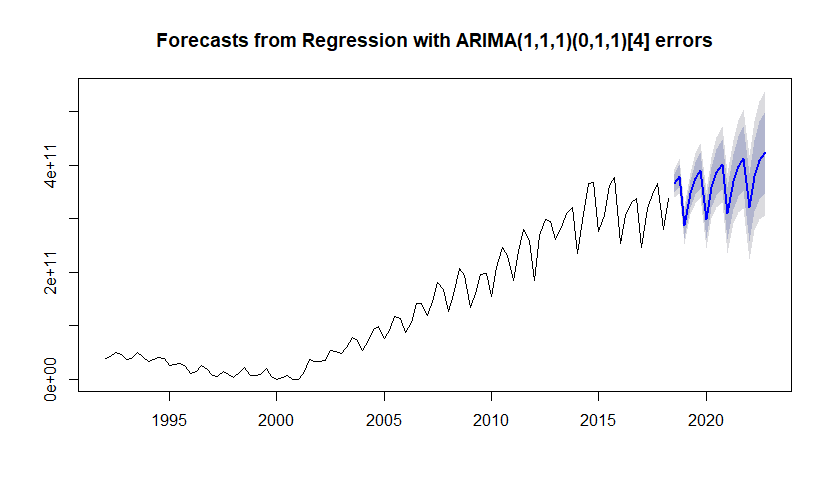
# Appendix 19

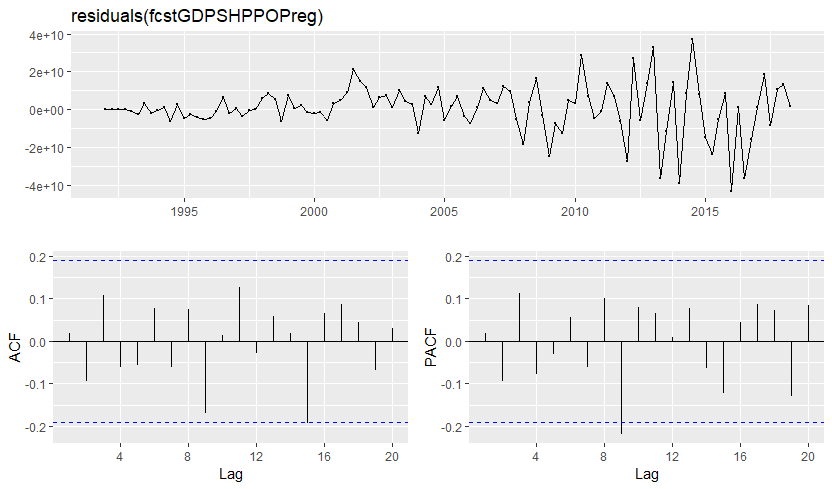
Best subsets forecast metric comparison table

|  |  |  |
| --- | --- | --- |
| Combination # | RMSE | AICc |
| 1 | 15462970701 | 5038.85 |
| 2 | 15757400822 | 5094.04 |
| 3 | 15817441263 | 5092.61 |
| 4 | 15817833387 | 5092.62 |
| 5 | 15326174204 | 5033.1 |
| 6 | 21577791254 | 5151.67 |
| 7 | 15448527723 | 5036.71 |
| 8 | 15462970701 | 5038.85 |
| 9 | 15983857976 | 5096.96 |
| 10 | 17357612031 | 5115.45 |
| 11 | 17864580610 | 5119.43 |
| 12 | 15763018843 | 5094.11 |
| 13 | 16090527137 | 5091.81 |
| 14 | 1.1984e+10 | 4998.61 |
| 15 | 15817833387 | 5092.62 |
| 16 | 16060337802 | 5093.55 |
| 17 | 15495904797 | 5037.04 |
| 18 | 15991126121 | 5094.84 |
| 19 | 15476506192 | 5035.07 |

# Appendix 20

Best performing model: subset combination 14 of dynamic regression





Series: nonnegdiffEXP

Regression with ARIMA(1,1,1)(0,1,1)[4] errors

Coefficients:

ar1 ma1 sma1 Population (thousands)

0.5364 -0.8264 -0.4295 -315479.9

s.e. 0.1618 0.1136 0.0915 2575475.8

sigma^2 estimated as 1.842e+20: log likelihood=-2498.49

AIC=4997.98 AICc=4998.61 BIC=5018.05

Error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 506445987 1.1984e+10 8744722384 -Inf Inf 0.4792741 0.01649968

References

CIA. (2018). COUNTRY COMPARISON :: IMPORTS. (n.d.). Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2087rank.html>

CIA. (2018). COUNTRY COMPARISON :: EXPORTS. (n.d.) Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2078rank.html>

Coccari, R L. (1978). Alternative Models for Forecasting U. S. Exports. Palgrave Macmillan Journals. Retrieved from <https://www.jstor.org/stable/pdf/154634.pdf>

Genberg, H., Martinez, A., Salemi, M. (2014). The IMF/WEO Forecast Process. Washington, D.C., USA: Independent Evaluation Office of the International Monetary Fund. Retrieved from <http://www-ieo.imf.org/ieo/files/completedevaluations/BP%2014%2003%20-%20The%20IMF%20WEO%20Forecast%20Process%20-%20Genberg%20Martinez%20Salemi.pdf>

Keck, A., Raubold, A., Truppia, A., (2009). Forecasting International Trade: A Time Series Approach. World Trade Organization. Retrieved from <https://www.econ-jobs.com/research/36056-Forecasting-international-trade-A-time-series-approach.pdf>

Levitt, A. (2018, November 05). U.S. dairy export volume rises 10% in 2018 Q3. Retrieved from <https://www.feedstuffs.com/markets/us-dairy-export-volume-rises-10-2018-q3>

Mahmoud, E., Motwani, J., Rice, G. (1989). Forecasting US Exports: An Illustration Using Time Series and Econometric Models. Great Britain: Pergamon Press. Retrieved from <https://ac.els-cdn.com/0305048390900277/1-s2.0-0305048390900277-main.pdf?_tid=e66b9a4e-649b-4184-85b5-15506b05c2af&acdnat=1542824665_cf0867cd00ec4de06fcf15338a48f286>

Martin, W. (2018, November 21). Crops are rotting in fields as Trump's trade war bites US farmers. Retrieved November 22, 2018, from <https://www.businessinsider.com/us-china-trade-war-soybeans-left-rotting-thanks-to-lack-of-storage-2018-11>

Martin, W. (2018, July 20). Trump threatens tariffs on all $505 billion of Chinese goods coming to America. Retrieved from <http://uk.businessinsider.com/trump-trade-war-tariffs-500-billion-chinese-imports-2018-7>

Reuters. (2018, November 08). US arms exports up 13% over 2017 as Trump champions deals. Retrieved from <https://www.channelnewsasia.com/news/world/us-arms-exports-up-13-over-2017-as-trump-champions-deals-10911548>

Shane, D. (2018, September 10). China exports growth slows as trade war heats up. Retrieved from <https://money.cnn.com/2018/09/10/news/economy/china-exports-trade-war/index.html>

Weinraub, M. (2018, November 21). Harvesting in a trade war - U.S. crops rot as storage costs soar. Retrieved from <https://uk.reuters.com/article/uk-usa-trade-china-grains/harvesting-in-a-trade-war-u-s-crops-rot-as-storage-costs-soar-idUKKCN1NQ0GR>

World Bank Group. (2018). Exports of goods and services (% of GDP). (n.d.). Retrieved from <https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?locations=US-CN&name_desc=false>

# Personal Ethics Statement Concerning Telfer School Assignments

By signing this Statement, I am attesting to the fact that I have reviewed not only my own work, but the work of my colleagues, in its entirety.

I attest to the fact that my own work in this project meets all of the rules of quotation and referencing in use at the Telfer School of Management at the University of Ottawa, as well as adheres to the fraud policies as outlined in the Academic Regulations in the University’s Undergraduate Studies Calendar. Academic Fraud Webpage

To the best of my knowledge, I also believe that each of my group colleagues has also met the rules of quotation and referencing aforementioned in this Statement.

I understand that if my group assignment is submitted without a signed copy of this Personal Ethics Statement from each group member, it will be interpreted by the Telfer School that the missing student(s) signature is confirmation of non-participation of the aforementioned student(s) in the required work.

./Desktop/Screen%20Shot%202018-10-09%20at%208.\_\_\_\_\_\_\_\_\_ November 26, 2018\_\_\_\_\_\_\_\_\_

Signature                                                Date

SAMAHA, TINA\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  8610185\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

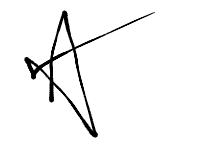
Last Name (print), First name (print)    Student Number

https://lh6.googleusercontent.com/uz_aOmBWWuWVY7iKQZ420Bup_U7_2ADkYHKTuTNRh1QFEFo_NT6YaKmJR2AQH6RXeeq-wptZf9VnQV2ZHKw3U7U1UijzHA4R0L5WdEA49KY8uY9oBr_IIiEIP4bqZNDxoNHhTMuem14qxPW23Q\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ November 26, 2018\_\_\_\_\_\_\_\_\_

Signature                                                Date

\_GUAY, GENEVIEVE\_\_\_\_\_\_\_\_\_\_\_\_ 8821581\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name (print), First name (print)    Student Number

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ November 26, 2018\_\_\_\_\_\_\_\_\_

Signature                                                Date

\_Sheik-Ahmed, Ayoub\_\_\_\_\_\_\_\_\_\_\_\_ 7370221\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name (print), First name (print)    Student Number

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ November 26, 2018\_\_\_\_\_\_\_\_\_\_\_

Signature                                                Date

Detlor, Thomas\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 8319090\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name (print), First name (print)    Student Number

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature                                                Date

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name (print), First name (print)    Student Number