Research Paper

What can we learn from Senegal's and South Korea's GDP per Capita evolution?

Maguette Paye |

January 15, 2021

Contents

1 Introduction

The Per Capita Gross Domestic Product (GDP) is commonly known as one of the best measures of development of a country and of the prosperity of its population. It is calculated by dividing the GDP by the population, so consequently it is a good indicator of the standard of living. Only accounting for the GDP of a country would limit ourselves to see how economically rich the country is, but it is fair to say that a wealthy country does not always rhyme with a wealthy population, that's when the GDP Per Capita comes into play. It is a good indicator of how well a population benefits from its country's economy. Please note that:

$$\text{GDP Per Capita} = [\text{Consumption} + \text{Investment} + \text{Government Spending} \\ + (\text{Exports - Imports})] \,/\, \text{Population}$$

This paper is about an interesting fact that I recently found out about. I never thought one day that I would be doing a comparative analysis on these two countries: South Korea and Senegal. At first glance, they seem to have close to nothing in common, but here's the thing; In 1960 the GDP per Capita of Senegal was twice higher than the GDP per Capita of South Korea, but surprisingly enough the trend has been dramatically inverted and today South Korea's is twenty-two times higher than Senegal's. This is a spectacular upswing for South Korea in 60 years. In this paper I will try to answer many questions:

- What can explain this change?
- Was SK already more developed than Senegal? Is GDP per Capita a great indicator of development as it is commonly said?

The goal of this research paper is not to do a classic socio-economic study, this paper is about applying machine learning and advanced statistics methods to gain the most insight possible from the data and answer the questions stated above. To start let's give a bit of context. In 1960, Senegal just gained independence from France after three centuries of colonisation and elected its first president Leopold Sedar Senghor. At the same time, South Korea was still in the recovery phase from the war that lasted from 1950 to 1953. In 1961, the country just underwent a coup from General Park Chun Hee...

2 Data Exploration

2.0.1 Evolution of each variable from 1960 to today

I created a dataset based on socio-economic indicators and features registered in the DataBank of the WorldBank website [11]. This dataset compiles information on South Korea and Senegal measured yearly from 1960 to 2019 (6.1 Data Dictionary).

The first step of the data description was to analyze the evolution of each variable throughout the years and compare the outputs for South Korea and Senegal. Our dataset comprises panel time series data on socio-economic indicators for each country (6.1 Data Dictionary)

The first step of our analysis will be to see how these predictors evolve overtime from the 60s to today. Firstly, when plotting the GDP Per Capita from 1960 to 2020 for both Countries, we can see that the GDP Per Capita of Senegal was twice South Korea's up until 1973 when the latter surpassed Senegal and skyrocketed to reach $32,761.98, which is a huge contrast with Senegal's current GDP of $1,446.83. Throughout this paper we will try to make sense of this finding (Appendix 1A).

Now that we observed the evolution of our variable of interest we will be looking for plausible reasons that explain this change. We decided to first take a look at the evolution of health through indicators such as the life expectancy and the infant mortality rate. The life expectancy of South Korea was already significantly higher than Senegal's in 1960. For both countries it increased; to 82 years old today in South Korea and to 68 years old in Senegal (Appendix 1B). Moreover the infant mortality rate per 1000 live births is a good indicator of the evolution of the healthcare system additionally to the life expectancy. We notice a significant decrease for both countries throughout the past 60 years (Appendix 1C).

The next exploration was in terms of evolution of the population. We notice that South Korea's population has always been bigger than Senegal's and the fertility rate of Senegal has always been bigger than South Korea's ever since the 1960s. Nevertheless there has been an immense drop in the number of children per woman in the latter, that went from 6 in 1960 to 1 today. For Senegal, the number dropped from 7 to 4.6 which is a smaller drop than South Korea's. It is natural to associate an increase of GDP per Capita to a potential decrease of the population (see equation), but our data shows us that this increasing GDP per Capita can be explained by the fact that the economic growth was more intense than the population growth in South Korea (Appendix 1D, 1E).

In terms of economy there are many indicators that we decided to take a look at, starting with agriculture and the industry value added of GDP that could give us some insight on any switch in their economy. Often we can notice that when getting developed, countries tend to focus less on the agricultural sector and more on the industry and services. According to the graphs we can see that there has been a dramatic decrease in the output of the agriculture sector as a percentage of the GDP going from more than 50%, to 2% today for South Korea. Comparatively for Senegal, it remained constant overtime with a sensible decrease to around 15%. This is important to note that before the “switch” in GDP per Capita in 1973, Korea’s Agricultural sector value added to GDP was almost double Senegal’s. Inversely, for the Industrial sector it has increased tremendously over time. We might suggest here that the sectors of focus played a role in this “switch”(Appendix 1F, 1G). Capital Investments is a good indicator of development (Data Dictionary) , we can see that there is an increasing trend for both countries. During the past 60 years, South Korea’s has been around twice higher than Senegal’s , and only recently the capital investment of Senegal has surpassed South Korea’s in 2018 (Appendix 1H).

A common experience both nations went through was being a development aid receiver. In the graph we can see looking at Aid per Capita that there has been more aid towards Senegal as early as 1964 and this skyrocketed up until 1990 when it was at its highest while South Korea’s received comparatively a smaller aid per capita which decreased up until becoming null in the 90s. (Appendix 1K) The augmentation of Development Aid for Senegal was simultaneous to a significant decrease in trade balance as a percentage of the GDP in the end of the 70s. Comparatively the trade balance of South Korea that was lower than Senegal’s up until 1973 significantly increased, and has been on an overall increasing trend until today (Appendix 1J). We might tend to think that the international aid would spark growth and when looking at the data we can actually see that it is at this moment that the gap between the two nations became wider. It is important to note that other factors such as political regime and policies put into place played in the development state of these two countries. However, a key aspect behind these numbers is the fact that around 60% of international aid given to the recipient country has to be used to buy and import products and equipment from the donating nations according to Emmanuel Okamba, a Congolese economist. This consequently plays on the development of the economy and of an internal market. Additionally, the Domestic credit to the private sector referring to the financial resources provided to the private sector by financial corporations, such as through loans, etc and that establish a claim for repayment, is very low in Senegal and only accounts for 30% of the GDP compared to South Korea where it exceeded 140% of the GDP. This indicator highlights us on the ease to develop a national economy and to create local businesses (Appendix 1I). This also sheds light on the issue with the local money. In fact, although a clear explanation has not been given, many research mention the very low Domestic Credit to the private sector of the countries in the CFA Franc zone, which is a money shared by most former french colonies of West Africa including Senegal.[14]

2.0.2 Principal Component Analysis

The descriptive analysis that we did was very useful to have an overview of the patterns surrounding the data but in order to capture the hidden functional relationship between variables, we conducted exploratory analyses. With many variables understanding the structure of the data is quite challenging, we therefore conducted a Principal component analysis (PCA). PCA is a tool that finds a low-dimensional representation of the data that captures variability as much as possible.It allows us to visualize highly complicated data in a simple manner. This tool extracts a set of Principal Components (PCs) that are small numbers of dimensions capturing variability of observations as much as possible. The contribution of each variable is represented by its coefficients called loading. The bigger the absolute value of the variable's loading, the most it contributes to the variance across the database. Because the variables in our data are different in scale, variables with larger scales will also have larger variances and overwhelm the PCA. Thus, we chose to standardize the variables at the beginning in order to have accurate results. The first two PCs account for more than 95% of the variations (rule of thumb is 85%), so 2 components are well representative of the data variability for both Countries (Appendix 2) There are two factors to consider when interpreting the plots, the length of the Eigenvectors and also the angle between them. Looking at the plot we can see that there is a clear distinction between South Korea and Senegal in terms of indicators. The first interesting aspect is the indicators that are highly correlated to the GDP per Capita, it is important to note that none of these variables are accounted for in the calculation of the GDP Per Capita except the population. Among these six variables the most surprising one is the credit to private sector indicator. It is very strongly correlated with the GDP per Capita and thus might be as we mentioned earlier the major difference that can explain the switch in the 70s. Additionally we notice a perfect negative correlation between the trade balance and the aid per capita. A positive and high trade balance means that the country exports more than it imports and therefore that there is a local market existing and a local industry, nevertheless the fact that it is negatively correlated with the amount of aid makes us think that the aid might be an obstacle to economic growth and development. The plotting of the PCA is highly insightful for our dataset because it portrays perfectly the switch in South Korea's economy seeing pink points from the left to the side (as from the 60s to today). The households' consumption is an indicator negatively correlated with the GDP per Capita which seems surprising because consumption is in the equation of the GDP as a positive term. From Appendix 2, we can infer that most of the variability found across the years for these two countries can be explained somewhat at an equal level by all the variables except the Aid per capita. This shows that the aid does not really foster development as we might tend to think.

3 Model Selection

3.0.1 Definition of Long Short Term Memory Model

Time series prediction has been improved thanks to Machine Learning and Deep Learning algorithms. When it comes to time series forecasts we often hear of statistical methods such as Auto regressive Moving Averages also called ARIMA, however in this paper I wanted to give a demonstration of how Recurrent Neural Networks (RNN) could be used for time series data forecasting. Forecasting time series using this Deep learning model especially on R was a complex task that I wanted to tackle to gain a deeper understanding of Neural Networks. RNN is a class of Artificial Neural Networks (ANN) designed to recognize, store in memory, and process patterns of sequential data. It builds relationships between segments of data it receives as input in order to understand its context and generates a subsequent output based on the patterns it observes. This is overall a great algorithm for forecasting because it is a Neural network with “memory, nevertheless it has its limitations. In this paper we will be focusing on a particular type of RNN called Long Short Term Models (LSTM) that overcomes vanilla RNN's limitations. The latter is an improved version of RNN because it is capable of learning long term dependencies and remembering information for longer periods of time is their default behavior.

LSTM are designed for applications when the input is in an ordered sequence where information from earlier in the sequence may be important. The LSTM units are composed of four main elements: the memory cell responsible for holding data and the three logistic gates define the flow of data inside the LSTM. The input gate is responsible for writing data into the memory cell, the read gate reads data from the information cell and sends this data back to the network, finally the keep gate maintains or deletes data from the information cell. These gates have the same structure as neurons in regular neural networks. They are Sigmoid activated nodes. By manipulating these gates an LSTM is able to remember what it needs and forget what is no longer useful. We will be using Keras and Tensorflow packages in R to build our model.

3.0.2 Data Preparation and Model Set Up

The model we will be building is an univariate LSTM. For this model, we will be needing only the GDP Per Capita of Senegal and also South Korea. Please note that we will repeat the same steps for both countries:

In order for the LSTM to be efficient with time series data we need to analyze its stationarity, and seasonality. Stationarity for a time series is very important, it stipulates that there is a constant mean, variance and covariance overtime, in other words these variables should not be dependent on time. Seasonality means that the data is showing a repetitive structure after every fixed time interval. In order to make sure the data is stationary and seasonal it is good to use a method called differencing. Differencing removes any trend or dependence of time of the times series, it consists of subtracting the previous observation (t-1) from the current observation (t).

After differencing our data goes from having a clear trend (see Appendix 1A) to looking stationary and somewhat seasonal as we can see in Appendix 3A and 3B.

Just before modelling, we verify that our model would be a good candidate for an LSTM checking for auto-correlation.[1] We therefore plotted an ACF plot which is a complete auto-correlation function portraying values of auto-correlation of any series with its lagged values. This plot describes how much present value is linked with previous value, it also accounts for trend seasonality and residual.It is recommended to use one of the high auto-correlation lags to develop an LSTM model.(Appendix 4A, 4B)

We therefore decided to focus on the first 45 (from 1960 to 2005) years to make our model the best possible The LSTM has a few exigences when it comes to the structure of the data the model is fed with:

- The LSTM expects its data to have the structure of a supervised learning model, so a dependent variable (y) and an independent variable (x). In order to achieve that the model required us to create a lag function that will allow us to have a column x gathering the t-1 lagged data and y gathering the t data. In our case if x is the GDP per Capita in 1980, y is the GDP Per Capita in 1981.

- The LSTM performs better when the data is normalized because it speeds up the learning process and also because the activation function is the Sigmoid function that performs better with values close to zero. Therefore, we normalized our differentiated and lagged dataset.

- The training and the test set must be evenly divisible in order to have the highest batch size possible

- The LSTM only accepts 3 Dimensional input variables, so since we have one input variable we change its shape to be a 3D array.

In our model we trained the data from 1960 to 1989 that we fed in our model and used the test data (from 1989 to 2004) in order to evaluate the accuracy of our model. The goal of this analysis was to see if we could have a good forecast of these two countries' GDP per Capita evolution. For South Korea, the GDP per Capita evolution surprised the international community, we will see if it was foreseeable with our forecast. For Senegal, the stagnation of their growth was very surprising as well so we will see also if our forecasting model could foresee it. We created the similar univariate LSTM models for South Korea and Senegal, the training and test set years chosen were the same because the output of the auto-correlation plot showed similar patterns for both in terms of auto-correlation before 2005. It is important to note that LSTM performs better when there is auto-correlation. Another aspect is that we wanted to have the highest number of batches possible without compromising the size of our training data. We decided to train 67% of our data and test on 33% of it. Another step we added is an additional layer of LSTM to improve the accuracy of our model.

3.0.3 Evaluation

Results for Senegal Our model was quite good at predicting the training data with an RMSE of 231 and a Mean absolute error percentage(MAPE) of 22%. Then, we can see that the green line (accurate GDP) shows a net and fast increase in the GDP per Capita of Senegal in the 70s and 80s. The model has a little trouble predicting the data when there are rapid increases or decreases. We then evaluated our model with the test data (1989 to 2004) and we can see that the model quite captures the pattern but at a smaller magnitude, with a RMSE of 234 and a MAPE of 37%. The close RMSE shows that there is a very negligible over-fitting of the data on the training set. (Appendix 5A, 6A)

Results for South Korea Our training set performs quite well with an RMSE of 392.1524 and a MAPE of 73%, and we can see below that the model has also a little trouble predicting the data when there are rapid increases or decreases, our model predicts the same overall trend as but with a smaller magnitude. This is quite good when we know that our data has a clear trend and is non stationary originally. Surely, since the model has been trained on this train set, it makes sense that it performs better than on the test set. Our test set performs surely a little more poorly than the training set naturally, with an RMSE of 7690.094 and a mape of 17% . In this graph we can see that the pattern is reproduced just at a different scale. (Appendix 5B, 6B)

4 Discussion

Forecasting is based on the idea that the past trends will continue into the future, nevertheless when it comes to Senegal and even more South Korea, there has been an increase so unexpected and exponential that the LSTM had trouble following the pattern. Identifying turning points is one of the most complex aspects of forecasting. The knowledge of past events helps somewhat the model to predict the GDP per Capita but sometimes unforeseen events can just not be predicted, especially in times of recession or rapid growth. When seeing the combined test and train for both countries (Appendix 7) we can see that overall our model does not perform too bad when we see it graphically. This makes us also question the reliability of the indicators MAPE and RMSE. It is important to note that there is no one size fits all indicator. When it comes to forecasting there are two things to take into account, the bias (overall direction of the error) and the precision (magnitude of the error). In our model we decided to optimize the mean squared error as it is often suggested in articles, but we decided to calculate the RMSE (square root of MSE) and the MAPE. The latter is a very popular accuracy measure but can be a poor accuracy indicator which can be explained by its formula:

$$MAPE = \frac{1}{n} \sum \frac{|e_t|}{d_t}$$

In fact, it divides each error by the GDP per Capita (d) , thus it is skewed and high errors occur when the GDP per Capita is low. This is therefore not reliable and one of the reasons we did not optimize with MAPE.

The MSE is a good indicator nevertheless it is not scaled to the original error since the error is squared:

$$MSE = \frac{1}{n} \sum e_t^2$$

Consequently , the best indicator depends on the data, and can also be determined with experience. The RMSE gives an idea of the average error but the higher value of the data the higher the potential RMSE, which can be problematic.

5 Conclusion

To conclude, our research paper has allowed us to determine the potential causes of the divergences between South Korea and Senegal's GDP per Capita. A key player is the Domestic Ratio credit to GDP that helped South Korea develop an inner market because financial institutions were capable of giving loans to business and entrepreneurs. Moreover, the development of the industry has been very beneficial to the development of the country. On the reverse, development Aid seems to be an important factor that slowed the development of Senegal because it increased tremendously the importations (thus, negative trade balance). Another assumption is that the local money had a role in the slow increase in GDP per Capita compared to South Korea because of the high correlation with Domestic Ratio credit to GDP and the GDP per Capita. Additionally, our forecasting models really show that the changes are hardly predictable because it does not follow a clear trend or pattern. This study also showed us that South Korea was already more developed than Senegal already in the 60s because of indicators such as life expectancy, mortality rate, industry value added of GDP, which can make one question how reliable the GDP per Capita can be as a measure of development. In my opinion it remains one of the best but is clearly only a quantitative indicator so it has its limitations.

6 Appendix

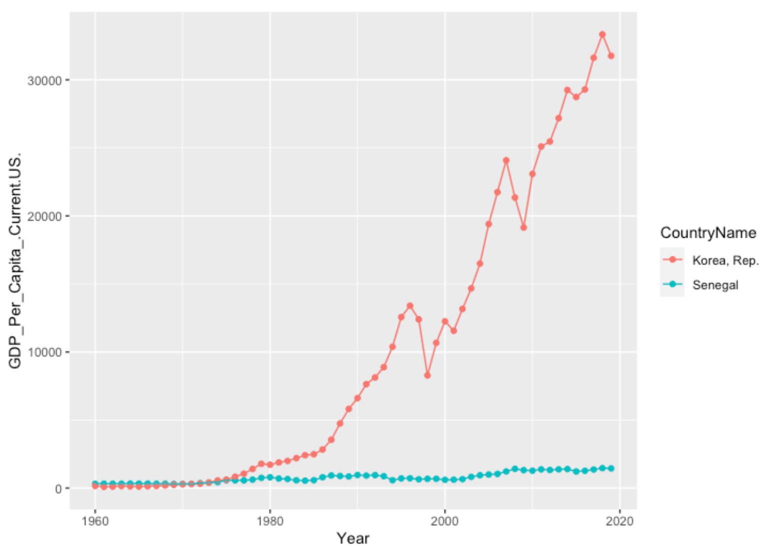
6.1 Data Dictionary

Feature	Definition
Year	Year
Population	All residents regardless of legal status or citizenship
GDP Per Capita (in current USD)	GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products.
Life expectancy	The number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life
Urban population (% of total population)	The people living in urban areas
Aid_Per_Capita	"Net official development assistance (ODA) consists of disbursements of loans and grants by official agencies of the members of the Development Assistance Committee (DAC), by multilateral institutions, and by non-DAC countries to promote economic development and welfare in countries and territories in the DAC list of ODA recipients." (WorldBank) "Net official aid refers to aid flows (net of repayments) from official donors to countries and territories" (WorldBank) Aid_Per_Capita = Net official development assistance and official aid received / Population
Agriculture, forestry, and fishing, value added (% of GDP)	Net output of the agriculture sector as a percentage of the GDP
Industry (including construction), value added (% of GDP)	Net output of the industry sector (mining, manufacturing, construction,electricity, water and gas)
Fertility rate	The number of children that would be born to a woman
Adolescent fertility rate	Number of births per 1,000 women ages 15-19
Household Consumption	"Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products (such as cars, washing machines, and home computers), purchased by households." (WorldBank)
Mortality Rate Infant	Mortality rate of infants per 1000 live births

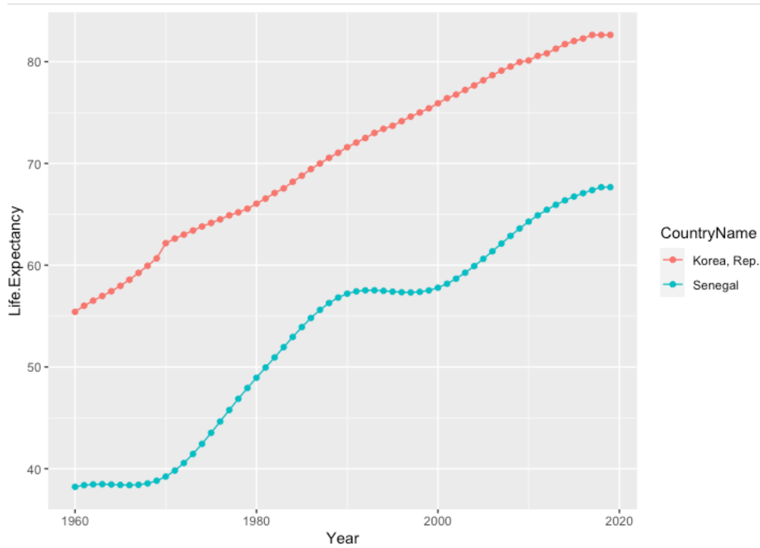
Table 1: Data Dictionary

6.2 Appendix 1: Evolution of Indicators from 1960 to 2019

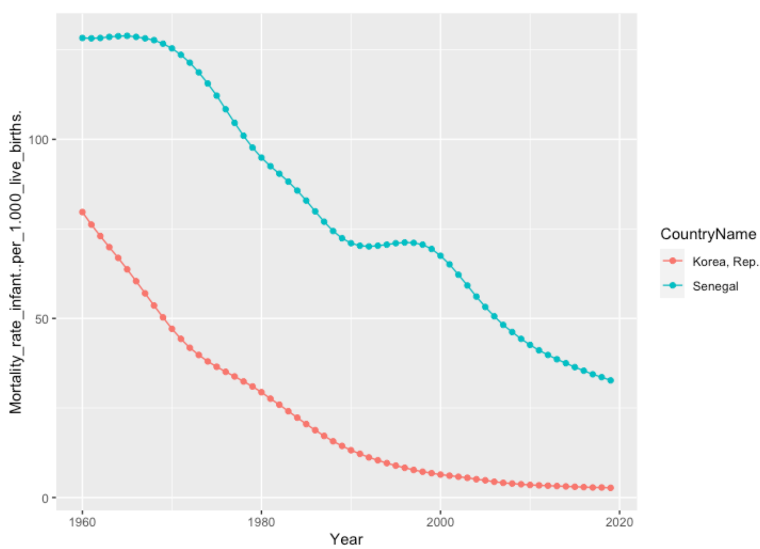
6.2.1 Appendix 1A: GDP Per Capita



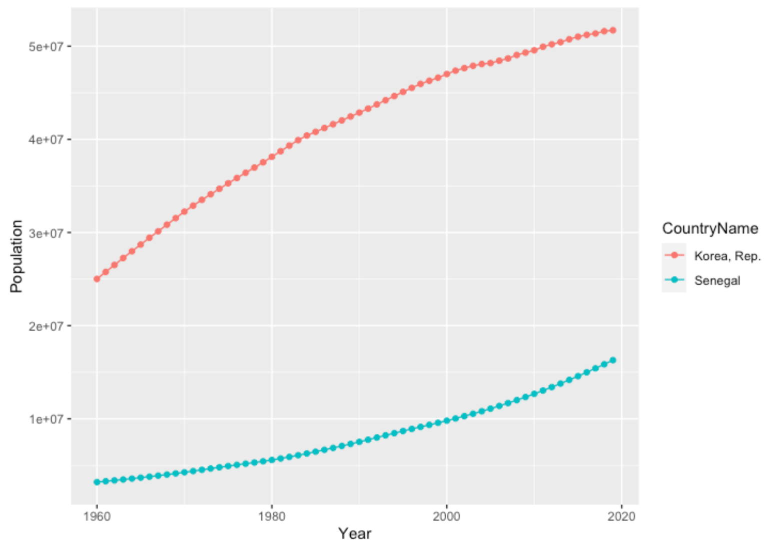
6.2.2 Appendix 1B: Life Expectancy



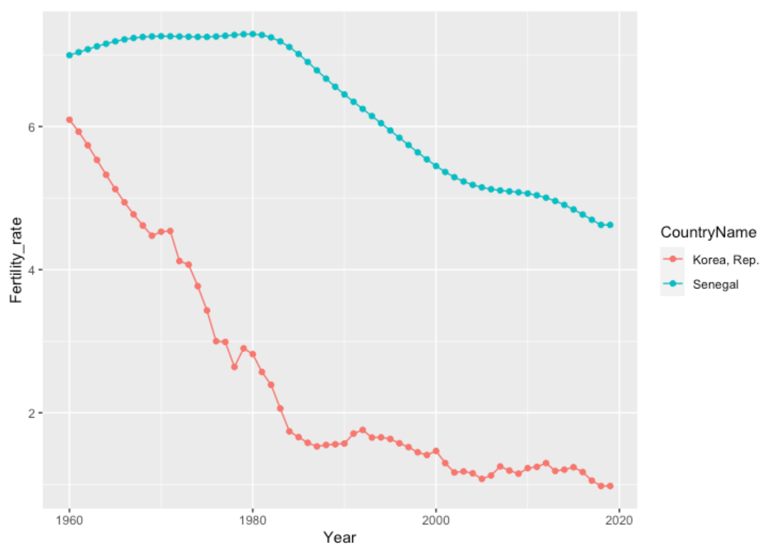
6.2.3 Appendix 1C: Infant Mortality Rate



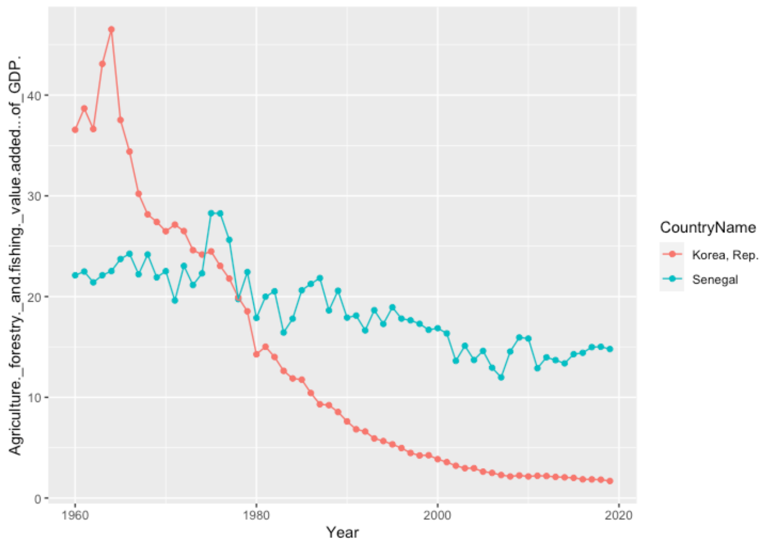
6.2.4 Appendix 1D: Population



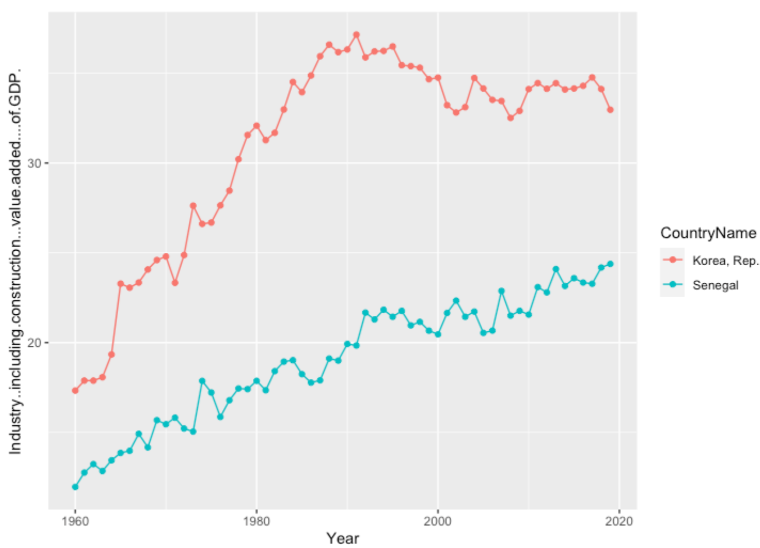
6.2.5 Appendix 1E: Fertility Rate



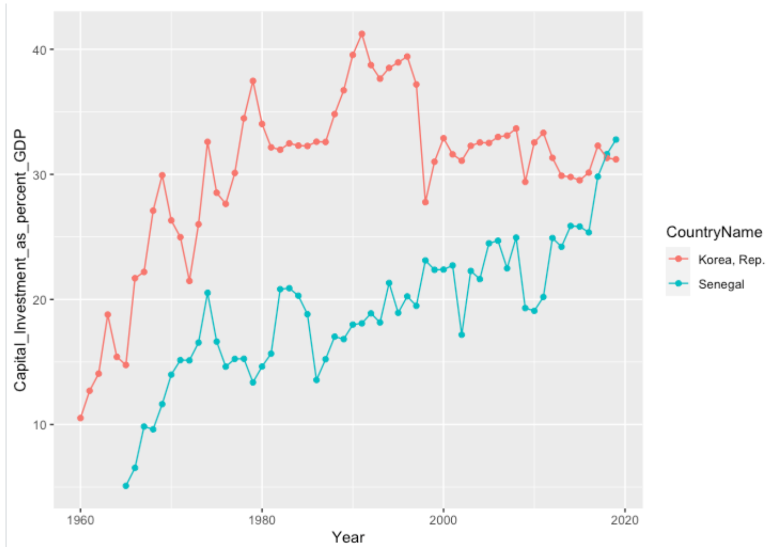
6.2.6 Appendix 1F: Agriculture



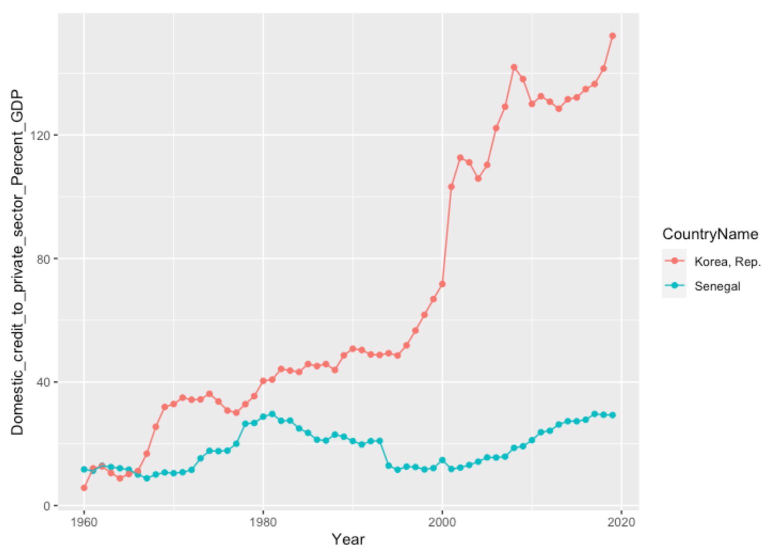
6.2.7 Appendix 1G: Industry



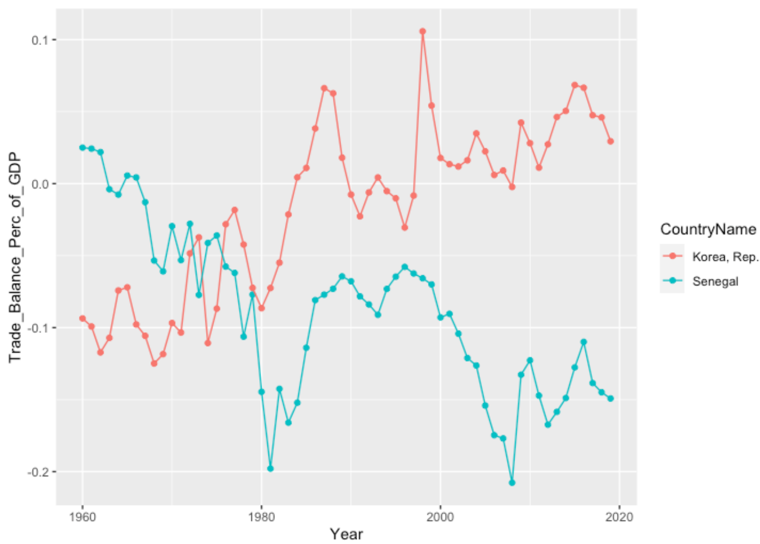
6.2.8 Appendix 1H: Capital Investment



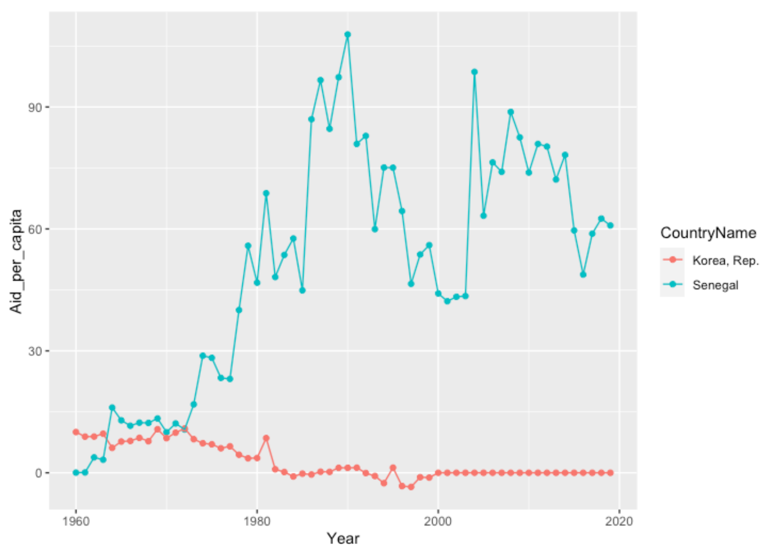
6.2.9 Appendix 1I: Domestic Credit to Private Sector



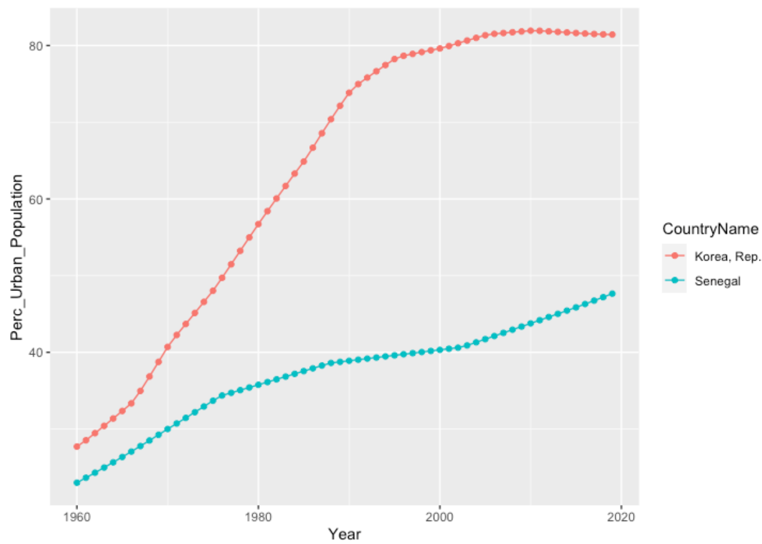
6.2.10 Appendix 1J: Trade Balance



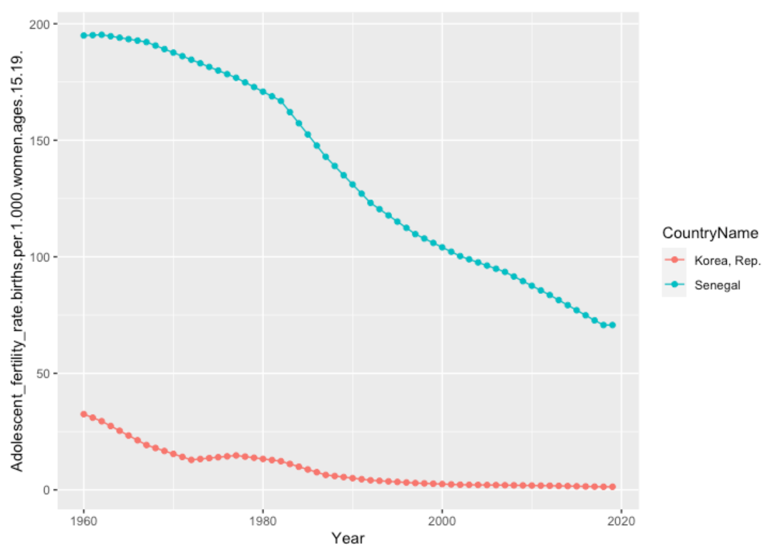
6.2.11 Appendix 1K: Aid Per Capita



6.2.12 Appendix 1L: Percentage of Urban Population

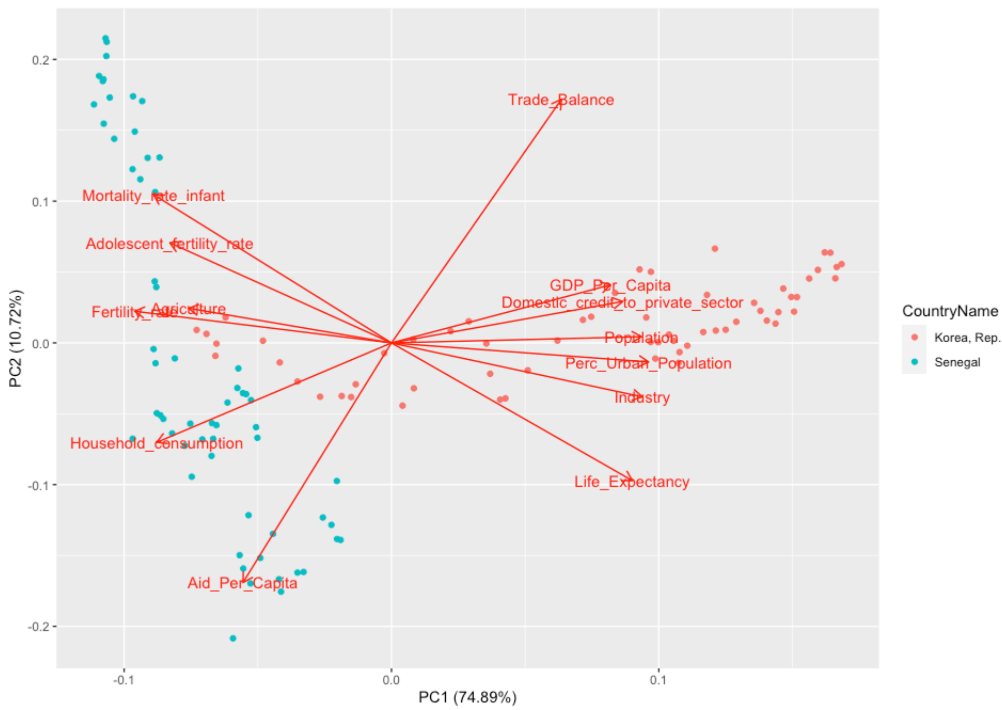


6.2.13 Appendix 1M: Adolescent Fertility



6.3 Appendix 2: Principal Components Analysis

6.3.1 Appendix 2A: PCA Plots

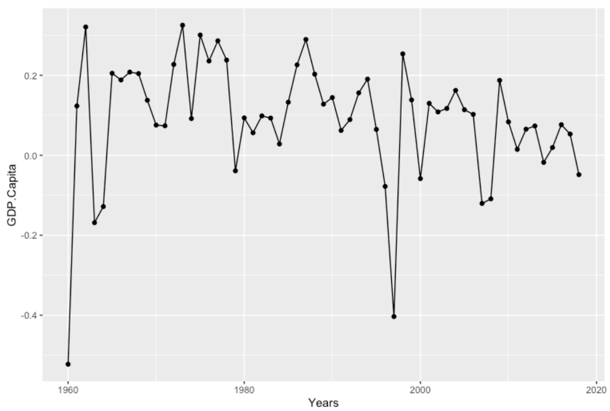


6.3.2 Appendix 2B: PCA Loadings

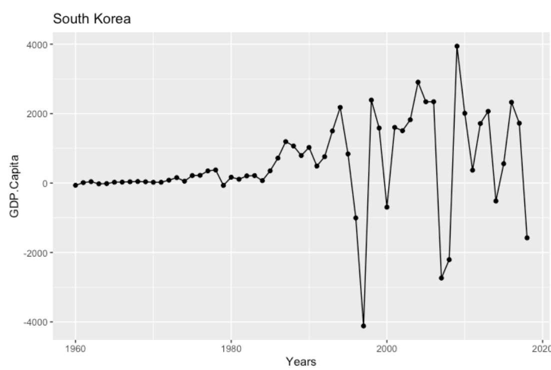
```
Rotation (n x k) = (13 x 13):
                                         PC1         PC2
Population                         0.3055365  0.01431061
Life_Expectancy                    0.2943807 -0.31743321
GDP_Per_Capita                     0.2677006  0.13379088
Perc_Urban_Population              0.3142074 -0.04389690
Aid_Per_Capita                    -0.1814490 -0.55236692
Agriculture                       -0.2478026  0.07982593
Industry                           0.3064864 -0.12436081
Fertility_rate                    -0.3136783  0.07302787
Adolescent_fertility_rate         -0.2708291  0.23078491
Mortality_rate_infant             -0.2905061  0.34149760
Trade_Balance                      0.2077591  0.56164474
Household_consumption             -0.2867124 -0.22910742
Domestic_credit_to_private_sector  0.2825413  0.09543081
```

6.4 Appendix 3: Pre-Model Exploration

6.4.1 Appendix 3A: Distribution of Senegal GDP per Capita differenced

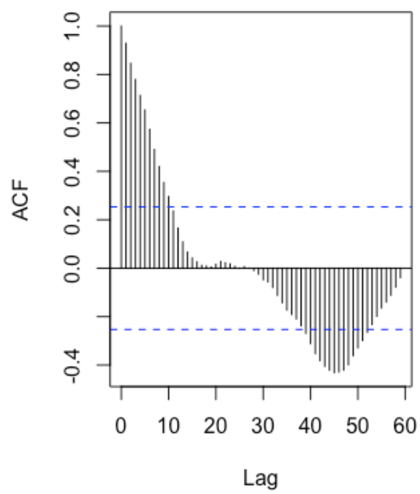


6.4.2 Appendix 3B: Distribution of South Korea GDP per Capita differenced

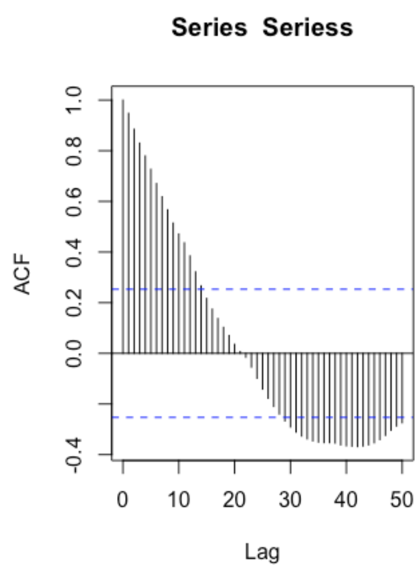


6.5 Appendix 4: ACF Plots

6.5.1 Appendix 4A: Senegal

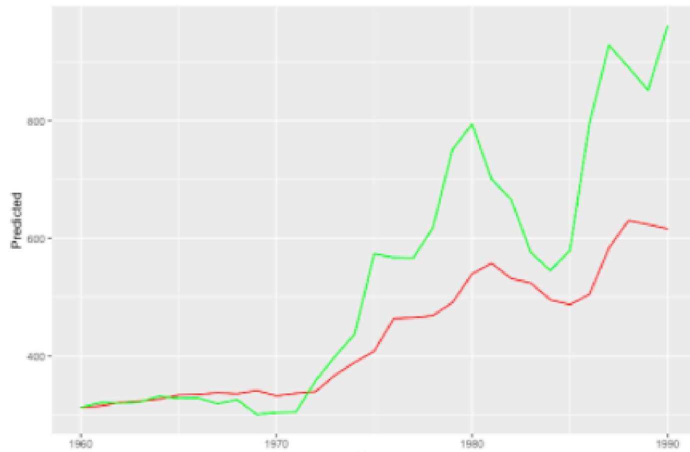


6.5.2 Appendix 4B: South Korea



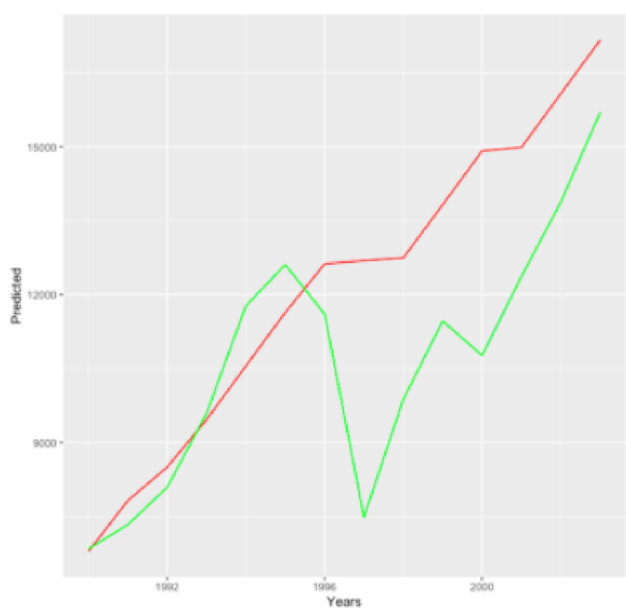
6.6 Appendix 5: Training Data Predictions

6.6.1 Appendix 5A: Senegal



Green: Actual Values - Red: Values predicted by the model

6.6.2 Appendix 5B: South Korea



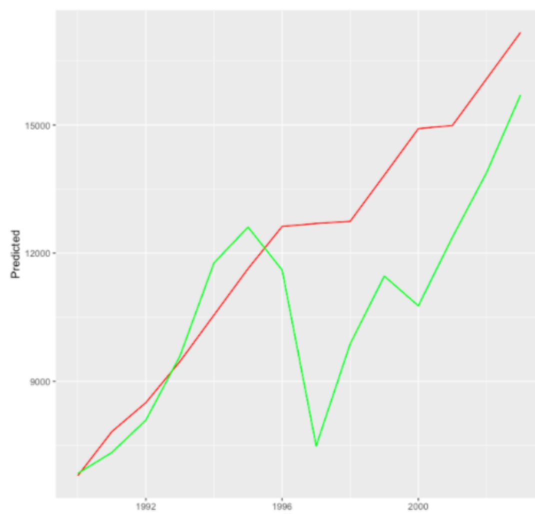
Green: Actual Values - Red: Values predicted by the model

6.7 Appendix 6: Test Data Predictions

6.7.1 Appendix 6A: Senegal

r Green: Actual Values - Red: Values predicted by the model

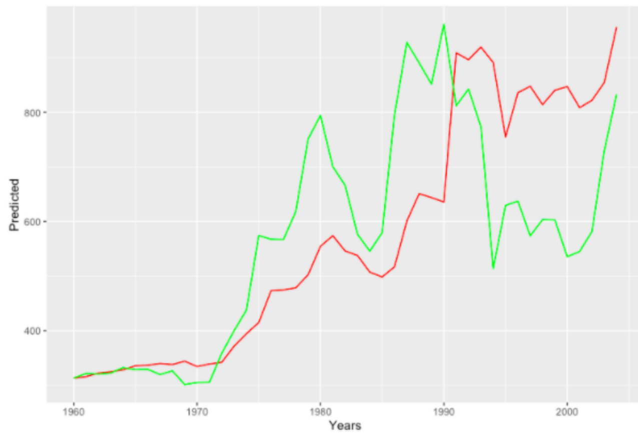
6.7.2 Appendix 6B: South Korea



Green: Actual Values - Red: Values predicted by the model

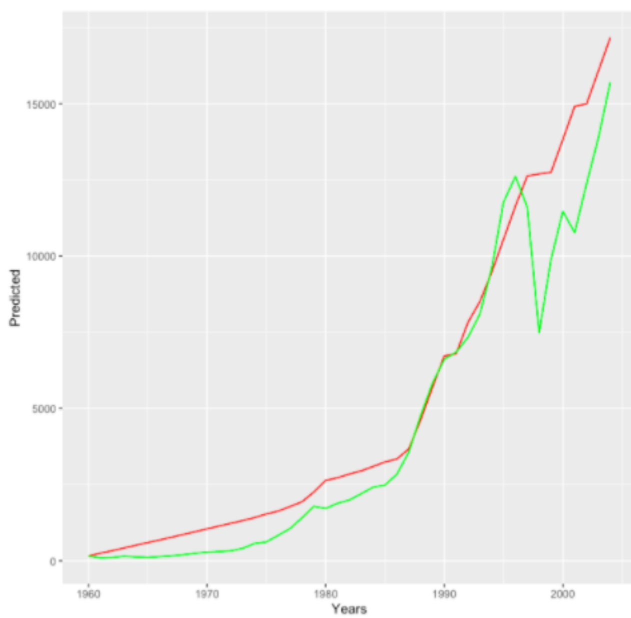
6.8 Appendix 7: Combined Train and Test

6.8.1 Appendix 7A: Senegal



Green: Actual Values - Red: Values predicted by the model

6.8.2 Appendix 7B: South Korea



Green: Actual Values - Red: Values predicted by the model

7 References

[1] Jayesh Salvi. *Significance of ACF and PACF Plots In Time Series Analysis* (March 26, 3019) - Retrieved from https://towardsdatascience.com/significance-of-acf-and-pacf-plots-in-time-series-analysis-2fa11a5d10a8

[2] Andrey Markin.*LTSM time series forecasting with Keras*. Retrieved from https://rpubs.com/andreasme/keras-lstm-notebook

[3] Jason Brownlee. *How to Develop LSTM Models for Time Series Forecasting* (November 14, 2018). Retrieved from https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

[4] Les Picker.*The Economic Decline in Africa* (January 1, 2014) Retrieved from https://www.nber.org/digest/jan04/economic-decline-africa

[5] Africa 24.*POLITITIA - Afrique: Faut-il rompre avec l'aide publique au développement?* (October 18, 2017). Retrieved from https://www.youtube.com/watch?v=HQoNqshyRuwt=902s&ab_channel=Africa24

[6] François Chollet and J.J. Allaire*Time Series Forecasting with Recurrent Neural Networks* (December 19, 2017). Retrieved from https://blogs.rstudio.com/ai/posts /2017-12-20-time-series-forecasting-with-recurrent-neural-networks/

[7] Jason Brownlee *How to Remove Trends and Seasonality with a Difference Transform in Python* (July 10, 2017). Retrieved from https://machinelearningmastery.com/remove-trends-seasonality-difference-transform-python/

[8] Nicolas Vandeput *Forecasting KPIs: RMSE, MAE, MAPE Bias.*(July 5, 2019). Retrieved from https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias

[9] Anais Dotis.*Autocorrelation in Time Series* (May , 2019). Retrieved from https://medium.com/@dganais/ autocorrelation-in-time-series-c870e87e8a65

[10] Yves Croissant and Giovanni Millo *Panel data econometrics in R:*. Retrieved from https://cran.r-project.org/web/packages/plm/vignettes/plmPackage.html

[11] Worldbank Databank. Retrieved from https://data.worldbank.org/country

[12] The Global Economy. Retrieved from https://www.theglobaleconomy.com/Senegal/

[13] Kenan Malik *As a system, foreign aid is a fraud and does nothing for inequality.*The Guardian. (September 02, 2018). Retrieved from https://www.theguardian.com/

commentisfree/2018/sep/02 /as-a-system-foreign-aid-is-a-fraud-and- does-nothing-for-inequality

[14] *Le franc CFA freine le développement de l'Afrique.*(July 7, 2020) Retrieved from https://www.lemonde.fr/afrique/article/2015/07/08/le-franc-cfa-freine-le-developpement-de-l-afrique$_4$675137$_3$212.$html$