US Income Inequality art project

Mirko Febbo

edx personal project

PS please excuse this document I had some unfortunate last-minute probleme with the knit function. I hope this try hard PDF will be adequate nonetheless

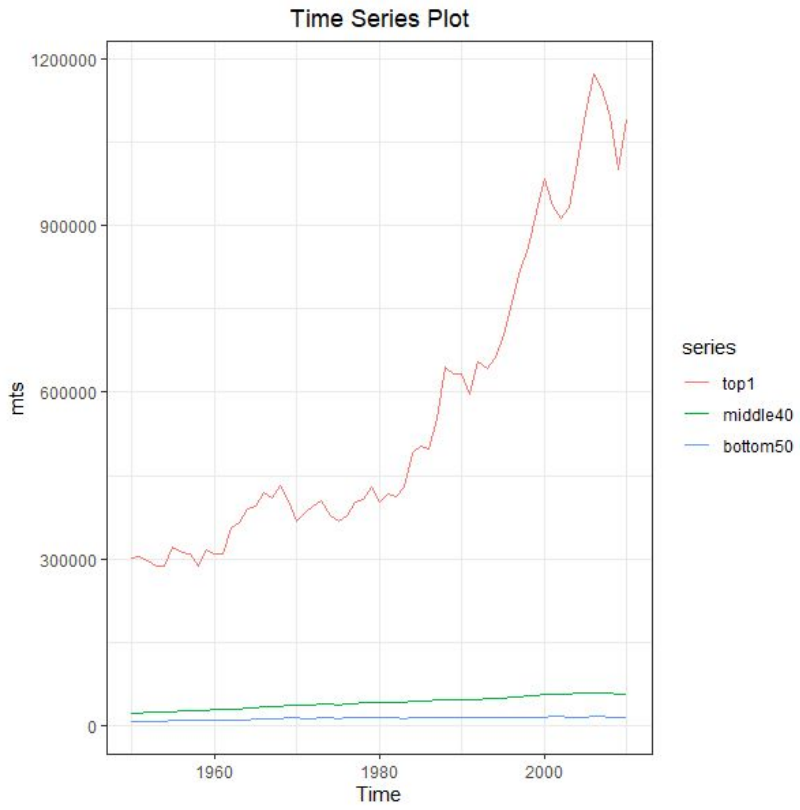
Introduction

This project and ambition to take the certificate came from the idea to create art pieces with the tools explored in class data and machine learning. This would help my art practice to conceptualise sensitive idea with data. In this particular project I want to look at the income inequality that grew since the 1970 and it's intimate link with neoliberalism. With the prediction, the data and the process I am planing make creative visual that might be 3D printable.

For this report, we will only be looking at the data science and forecasts parts of the project.

Data wrangling

For the data I combined two sources of data "World Inequality Database" (WID) and "United Nation: World Population Prospect" (UN). The WID website let me select the country, their respective class share of income by year and the average individual GDP. Mixed with the yearly population database from the UN I could make the average yearly income of the United States classes. Which consist of 1%, 10%, 40%, 50% of the population, I have chosen not to consider the 10% as I mainly want to compare the lower classes with the top one.



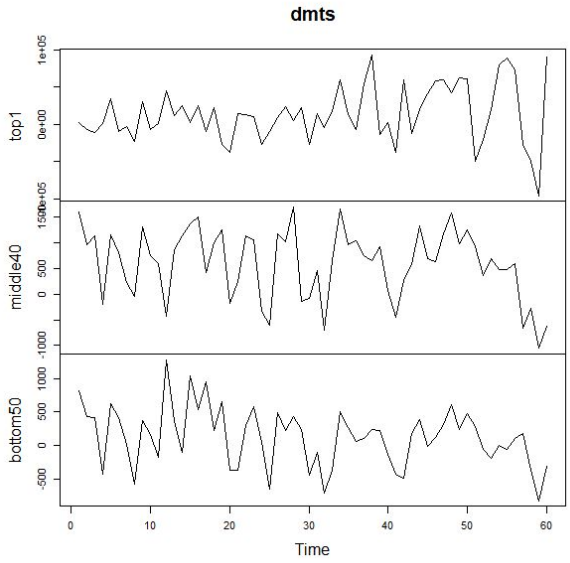
Since we are working with prediction in the future, I have chosen to work with a time serie data to better make my predictions.

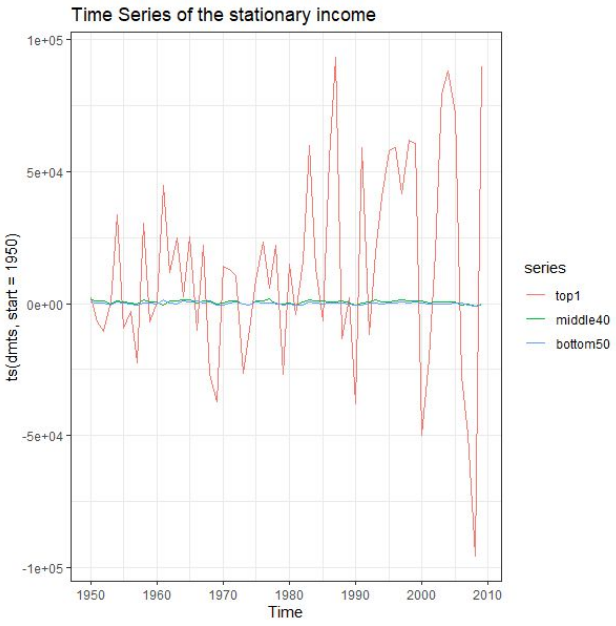
Here we have a first time series plot that show the extravagant income inequality in the US.

Causality Forecasts

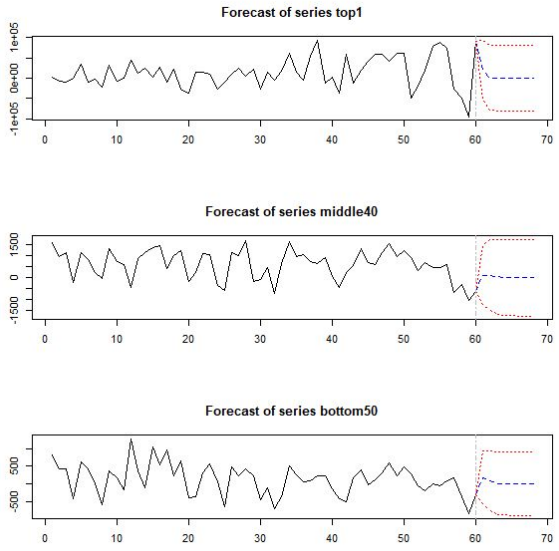
To start off we needed to figure out if the data is stationary. Thus we run an augmented Dicley-Fuller test and a stationary test on each income. Looking at the augmented Dickey=Fuller test we see that the p-value are very different from one income to another. The stationary test gives us more details and que can observe the p-value and ADF correlated with their lags, and compared with different type of settings.

To make the data stationary we will be working with the diffrence of our income. we do the ADF test to see that the p-value is closer to 0 for all the income bracket.



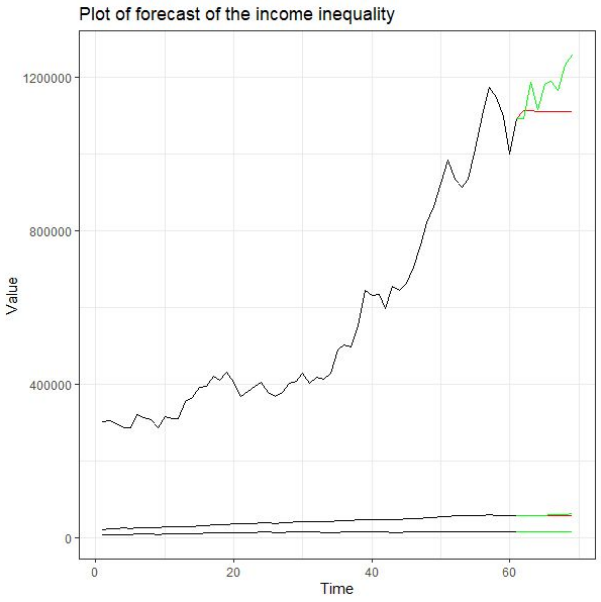


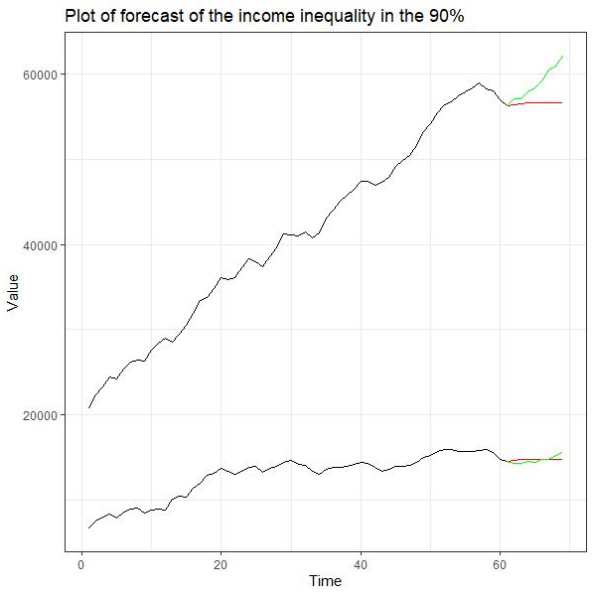
We then identify the lag order, and make the dataset a variable so we can do a serial test. The serial test will make a residual diagnostics. Thus we will be able to select the variable with a granger test for causality. And for this to give reliable result we need all the variables of the multivariate time serie to be stationary.



We can then apply a forecast on the stationary data. Now let's see how it compete with the actual data.

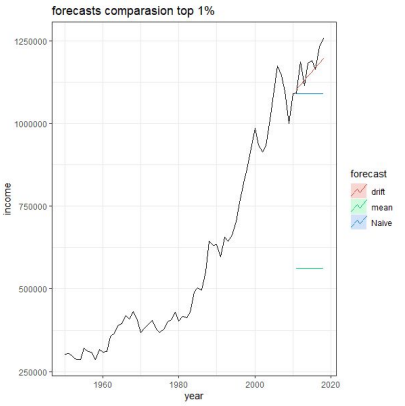
Looking at the graph below, we understand that a forecasting model based on causality would be doing poorly.

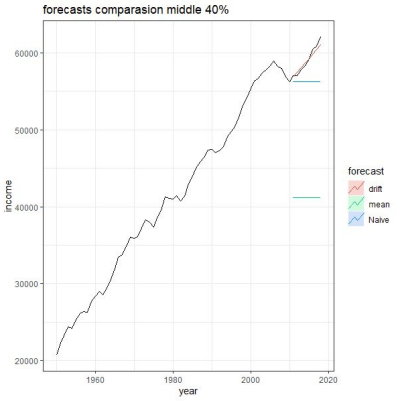


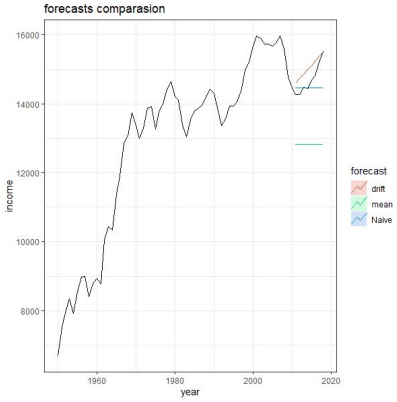


Mean, Naive and Random Walk Forecasting

Here we have multiple simple forecast comparison and we can see that the random walk (ARIMA) with drift is the model with the most success on all the income class.

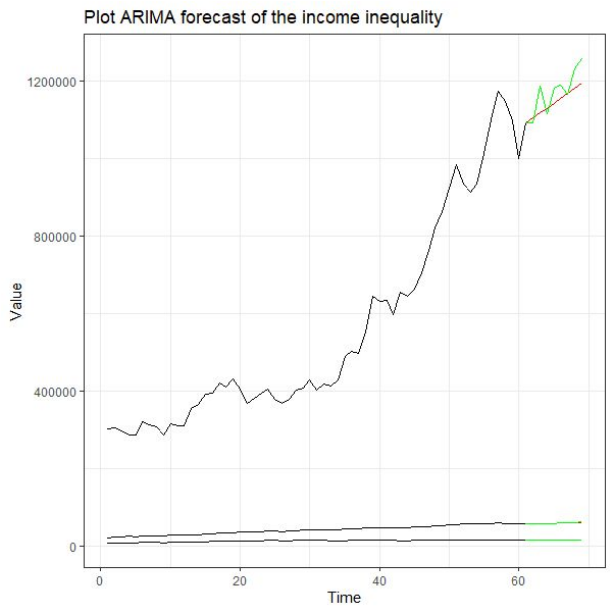


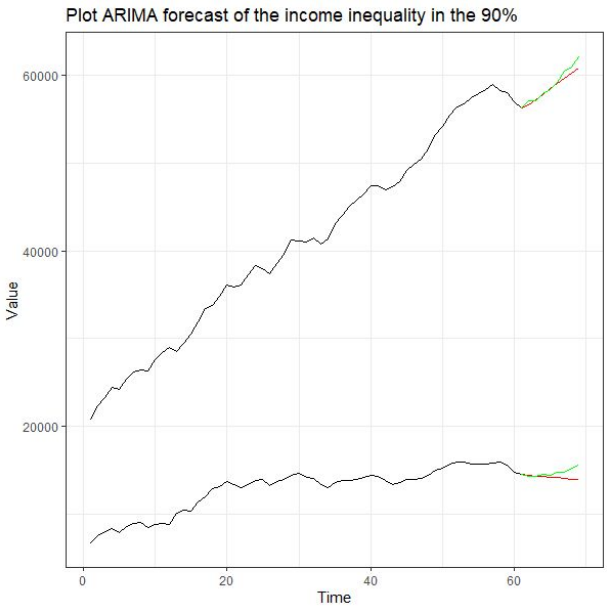




Auto Select ARIMA

Seeing the ARIMA model performance we improved upon it. To build an autonomous AIRMA model where the best value of the ARIMA where selected depending on the data. I started by running a couple of ACF and PACF on the data to better understand the relationship of the lags.





Conclusion

By the end of all of our iterations and improvements, we were able to achieve fairly good. We looked at if a correlation forecast could work, to then proceed to test multiple simple forecast to understand that the ARIMA would be our best bet. Thus all that was needed to do is find the best value of the ARIMA that would better fit the model. With this model in the pocket it could very well compliment my work considering that we could predict indefinitely. By making the model predict on himself wich will give us a nice caricatured prediction of a possible reality.

https://wid.world/data/
https://population.un.org/wpp/Download/Standard/CSV/
https://www.mirkofebbo.com/