

**Report Title:**

Complementing EIU’s predictive techniques, to predict future stability of a country’s exchange rate

**BC2406 Analytics I: Visual & Predictive Techniques**

**Seminar Group**: 6

**Group**: 3

**Team Members**:

Faith Lee Kai Ling (U2022307C)

Jensen Lim Chang Sheng (U2021364D)

Sim Miao Xuan (U2010155G)

Teresa Zhang Han Yu (U2022886C)

**Professor**:

Vivek Choudhary

**Table of Contents**

[**Executive Summary**](#_w8847cheek6p) **1**

**1. Introduction**

[1.1. Introduction](#_hneip4lfn6mo) 2

[1.2. Why Machine Learning: Competitor Analysis](#_gze2xq1ii9gj) 2

**2. Data**

[2.1. Data Selection](#_nh2eha8nlhyi) 3

[2.2. Data Preparation](#_n89cgo80uz6s) 4

[2.3. Data Exploration and Visualisation](#_7efcskqaozas) 4

**3. Linear Regression**

[3.1. Creating and Selecting a Linear Regression Model](#_yk8y1zan2r4s) 5

[3.2. Evaluating Linear Regression Models using Diagnostic Plots](#_p6q3awxas6uz) 7

[**4. Creating and Selecting a CART model**](#_rr5pjm4ownpz) **8**

**5. Model Evaluation**

[5.1. Selecting between Linear Regression and CART](#_6wow03obxl0k) 11

[5.2. Explaining the Variables Selected by our Chosen Model – CART](#_5ndn1k5aisq1) 11

[**6. Obtaining a Scale of Stability by Calculating Year-by-Year Fluctuations**](#_tkuznuopb1vt) **14**

[6.1. Formulating our scale](#_pe0sy640n8bs) 14

[6.2. Obtaining the Degree of Fluctuation in our Predicted Data](#_e572jbsgt8c2) 15

[**7. Pilot Studies in EIU and Link to Country Report**](#_l8dyo367714j) **15**

[**8. Challenges regarding Assumptions Made**](#_qbp66cl437v4) **16**

[**9. Conclusion**](#_mxsmzq6h0b9i) **16**

[**References**](#_e2jkhwzb7kun) **17**

[**Appendix**](#_lemnor8d44r) **18**

#### 

#### Executive Summary

Our team has identified the potential for the Economist Intelligence Unit (EIU) to increase value in their country reports, so as to differentiate the organisation from their competitors such as Capital Economics and Oxford Economics. Hence, our team’s proposal is to use machine learning to not only enhance EIU’s forecasting accuracy of their country’s report exchange rate, but also to include a scale which helps determine the degree of fluctuation in the future exchange rates.

In order to develop our models to predict different country’s future exchange, we have taken EIU’s past data to build both a linear regression model and Classification and Regression Tree (CART) model that EIU can use. However, after much trial and consideration, our team believes that the best model in predicting future exchange rates would be the continuous CART model. To further add on, our team also devised a simple method of calculating percentage changes of other countries' exchange rate to determine a scale that predicted exchange rates can be judged upon.

We propose that the continuous CART model and degree of fluctuation scale to be added to the country reports in a new section titled “Exchange Rate Stability”. With this new section it will help bring more value to the EIU country reports; potentially attract more clients creating new sources of revenue.

#### 1.1. Introduction

A stable currency is often indicative of sound monetary policy, which is likely due to a strong government and dynamic economy. Stable currencies also help to attract capital from foreign investors, meanwhile currency risk deters them (Segal, 2021). Accurately predicting exchange rates is hence especially important for countries that are highly dependent on trading partners, since unforeseen fluctuations in their trading partner’s currency could cost heavy economic losses. In view of how different economies are becoming more connected globally, exchange rate is certainly a pertinent economic consideration.

While the EIU does well in using expert opinion and domain knowledge to make their predictions, our team’s goal was to more quantitatively predict exchange rates with machine learning. We tapped into the extensive amount of economic data collected by the EIU to provide a concrete and unbiased perspective into future predicted exchange rates using machine learning, to complement their existing methodologies.

In this report, we also further process past exchange rate data to determine a scale of quantifying the degree of exchange rate fluctuation. Then, we will predict exchange rates and calculate the degree of fluctuation, to determine the future stability of a country’s currency.

Not only is this information useful for companies who wish to evaluate future opportunities and risk regarding their planned or current international business ventures, governments can use this data to determine future economic health and make decisions accordingly.

We chose to conduct linear regression and CART since our chosen dependent variable is continuous.

#### 1.2. Why Machine Learning: Competitor Analysis

Machine learning has proven to minimise forecasting error, making forecasting more efficient (Addepto, 2021). For EIU to further expand their reach to more customers, our team believes that EIU should utilise machine learning to make faster and more accurate predictions to their data. Contrasting with EIU’s current system of country specialists that provide country-specific insight and analysis, machine learning is able to make use of the huge amount of data that EIU already has, to identify potential trends and predict future trends. As machine learning requires a good amount of time for the algorithm to learn the patterns that are present in past data bases and improve their predictions, it is essential that EIU adopts machine learning as early as possible to boost their country reports.

According to Deloitte’s machine learning report, it states that machine learning will help businesses gain a competitive advantage, especially so for companies with data analytics capabilities such as EIU. A study also showed that companies using data analytics were 5% more productive and 6% more profitable than competitors (Deloitte, 2017). Our team believes that with machine learning, EIU will be able to create more value for their flagship product: the country reports.

One example of a company successfully using machine learning highlighted in the article above would be Pluribus Labs(McAfee & Brynjolfsson, 2012). Pluribus Labs uses machine learning techniques to deliver predictive investment information. Hence it is obvious to see that even with just a small team of 8 data scientists, Pluribus Labs was able to achieve so much in terms of revenue and growth. Therefore machine learning should be applied into EIU reports allowing EIU to grow and expand faster than their competitors.

#### 2.1. Data Selection

To develop a predictive model to present a machine learning Proof of Concept (POC) for EIU senior managements’ consideration, we first researched what response variable that we should predict. As explained, we chose exchange rates since they are indicative of overall economic health.

We did some background research and found out that there are a few relevant x-variables that affect exchange rates, namely, inflation, interest rate, public debts, political stability and economic performance, terms of trade, and current account deficit (CompareRemit, 2020).

Seeing that most of these variables are economic data, we tried to source the dataset from EIU. The dataset we selected from EIU consists of 26 variables including exchange rate. The data span from 1993 to 2023 and were recorded on a quarterly basis. Thus, it provides us with sufficient data to work with (see Data Sources in Resources).

Instead of just developing a predictive model on one country, we decided to try it on multiple countries. There are a few reasons for that. Firstly, we are able to compare the results and analysis between multiple countries. Secondly, we wanted to see if our model will be able to work on different countries. By showing that we can do it on multiple countries, it is more likely that it will be able to work on most, if not all countries, thus our model will appeal more to EIU.

The countries we decided on are Canada, Netherlands, Sweden, Russia, and United Kingdom. We chose these countries as they are large economies with a substantial hand to play in the global economy. Based on metrics such as Gross Domestic Product (GDP) or the sum value of goods and services in a country, Canada and United Kingdom are in the top 10 ranks globally, with Russia just out of the rank (Tures, 2020). Sweden and Netherlands ranked right below Russia for the top 25 economies in the world (Silver, 2020). Furthermore, we were able to find the most complete datasets on these countries, making our predictive analysis more robust as the data we obtained could paint a more complete picture in predicting exchange rates.

Lastly, we read in the data into RStudio in excel format as that prevented loss of precision in our analysis. This is because .csv files only retain data up to one decimal place.

#### 2.2. Data Preparation

Firstly, we check for data that could possibly be missing values, for example "-". We replaced them with NA so that R can recognise it as missing values. We assume that the data was simply not reported and hence not present in our dataset. Then, we removed dates from 2021Q3 onwards as those were forecasted data and we solely wished to work on actual data for more accurate predictive analysis. Lastly, we converted the `Date` column to be of date type.

We created a new data table to store the variables without the `Date` to simplify the rest of our analysis.

Next, we eliminated variables by logically evaluating them. We removed duplicate columns that measure the same variables, the only difference being their units or how they were measured. As such, we removed variables measured at “end-period” when their “average” for the period was present, variables measured in domestic currency (“LCU”) when their “USD” counterpart was present, and percentage change when the actual values were present (for example, we removed percentage change in imports since imports was present). This is because these variables are perfectly multicollinear with each other and we would be unable to perform regression analysis with all of them present.

We also checked if negative values were present where they should not be, for example in interest rate, unemployment rate, nominal GDP, and exchange rate.

Finally, we dealt with the NA values. We have decided on two ways to deal with them, by either dropping the entire column or replacing the NA values with the respective mean of each column. We decided against dropping an entire row, unless most of the data in the row were NA, as a lot of other data from other columns will be dropped, reducing the amount of data we were able to work with.

To determine whether the entire column should be dropped or the NA values should be replaced with the mean, we select the respective columns along with exchange rate, omit the NA values, and check the correlation. We decided that if the value of correlation is more than 0.1 and the percentage of the NA values out of the rest of the data in the columns is less than 30%, we keep the columns and replace the NA values with mean, else, we drop the columns.

After dealing with the NA values, we look at the correlation of all the x variables with exchange rate and if the value is less than 0.1, we remove those variables, as they do not significantly impact exchange rate and we assume that they will not be of help during machine learning.

However, we only need to deal with NA values when we are making a linear regression model since CART models are able to be constructed even when the dataset has NA values due to its surrogate function. Thus, in order to improve accuracy, we did not remove or replace NA values for CART analysis. Lastly, for linear regression, we removed variables if they have relatively low correlation values as they would prove to be unhelpful in predicting exchange rates.

#### 2.3. Data Exploration and Visualisation

With that, we plotted exchange rate against date to visualise how exchange rate changes over the years for our selected countries (Figures 1-5 in Appendix). From the graphs, we could see an apparent appreciation in exchange rate for Russia, while the other countries’ exchange rates fluctuated about a relatively constant mean.

We then plotted all the independent variables against exchange rate to get a gist of the relationships between them (Figures 6-11 in Appendix). There were weak linear relations between the variables and exchange rate, which follows from the relatively weak correlation values we found between them. Immediately, we can tell that a linear regression model may not be very representative of the actual relationship. However, the data looked to be well-segregated, making CART an obvious choice.

For creating the linear regression model, we wished to visualise the correlations better, hence we also plotted a correlation heatmap to be able to tell what were the variables we should look out for in the formation of our models (Figures 12-16 in Appendix). Lastly, we also plotted a correlation plot of all the other variables against exchange rate, from the variable with the most negative correlation, to the variable with the most positive correlation (Figure 17-21 in Appendix). From the plots, we can see that variables like deposit interest rates and public debt have the highest absolute correlation values and those should be the variables we should include in our linear regression model, and vice versa for the variables with a low absolute correlation value.

#### 3.1. Creating and Selecting a Linear Regression Model

We will look at UK to explain how we performed linear regression. In creating the linear model, first we included all variables of the data that was specifically cleaned for linear regression.

Then, we removed variables based on variance inflation factor (VIF) values. High VIF values reveal multicollinearity between the different variables. Therefore, we removed the variable with the highest value, and did so until VIF was less than 10 (see Figure 22 in Appendix for UK’s linear model summary). While the RMSE for both train and test set increased, removing variables based on high VIF is still important to do in order to remove instances of multicollinearity which is essential for a good linear regression model.

Then, we used the step function to further remove variables based on Akaike Information Criterion (AIC), meaning that variables that did not provide a significant contribution to the prediction were removed (see Figure 23 in Appendix for UK’s linear model summary).

For UK, after AIC, the train and test set root mean square error (RMSE) remained similar. However, the p-value of all variables were under 5%, making all of them significant. In fact, after AIC, all variables in each country’s model had a p-value of less than 5% (except for `Gross fixed investment (% real change pa)` for Sweden’s model). Hence, our decision to remove based on VIF, then AIC was sound. Please see Figure 24 in Appendix for the final linear regression models created for each country.

To evaluate the models, we conducted a train-test split of a 7:3 ratio. Below displays our findings.

| **Country** | | **Model** | **Trainset Error** | **Testset Error** |
| --- | --- | --- | --- | --- |
| **Canada** | **1** | Complete | 0.034374 | 0.037026 |
| **2** | VIF | 0.074768 | 0.078237 |
| **3** | AIC | 0.074768 | 0.078237 |
| **Netherlands** | **4** | Complete | 0.035298 | 0.039499 |
| **5** | VIF | 0.047603 | 0.072332 |
| **6** | AIC | 0.047603 | 0.072332 |
| **Russia** | **7** | Complete | 7.779643 | 12.124396 |
| **8** | VIF | 9.443954 | 10.172301 |
| **9** | AIC | 9.595250 | 9.324460 |
| **Sweden** | **10** | Complete | 0.322075 | 0.407612 |
| **11** | VIF | 0.500180 | 0.512381 |
| **12** | AIC | 0.508453 | 0.439762 |
| **UK** | **13** | Complete | 0.024581 | 0.029124 |
| **14** | VIF | 0.038313 | 0.031272 |
| **15** | AIC | 0.038705 | 0.030144 |

*Highlighted in green are the train/test errors of our finalised linear regression models*

Generally, train and test error were not hugely different, showing that there was no significant under or overfitting in the models. Furthermore, the test error was in general always higher, which is to be expected. Exceptions include Russia at row 7 and Netherlands’ final regression model at row 6, in which the testset error was notably higher than the trainset error. However, after VIF and AIC, this issue at row 7 was remedied at row 9 for Russia’s final model. For Netherlands, the difference suggested slight underfitting since the trainset and testset error increased from row 4 to 6.

Another exception was test error being lower than train error (at rows 12, 14, and 15). This could just be due to chance occurrences that the model was able to predict accurately on the test data.

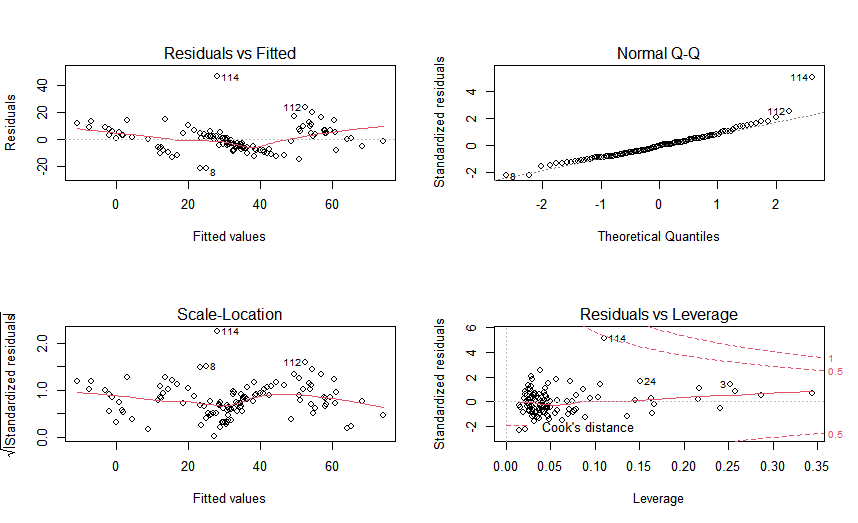
The reason for Russia’s seemingly larger error is because its exchange rate is higher than all the other countries (measured in USD) hence the absolute values of the residuals are larger.

If we were to build a linear regression model in general for all countries, we would probably select the variables with an occurrence greater than or equal to 3 (see Figure 25 in Appendix). Some of the more notable variables include unemployment rate, and exports. This proves the legitimacy of the machine learning conducted. Currency appreciation makes imports cheaper and exports less competitive, causing local companies to reduce costs by cutting jobs, causing unemployment to rise. As for exports, a weaker currency stimulates exports and conversely, a stronger currency puts a damper on exports.

However, since our sample size is small, we decided to build individualised linear models for each country. If we were to build individualised models for a larger number of countries, we would be able to more accurately select the variables with high occurrences to craft a linear model in general for predicting exchange rates. Nevertheless, there are merits to building individualised models as different countries may experience different factors affecting their exchange rate.

#### 3.2. Evaluating Linear Regression Models using Diagnostic Plots

Linear regression makes several assumptions about the data, such as linearity, normality of residuals, constant variance of residuals, no multicollinearity, and no influencing outlier in the dataset. We eliminated multicollinearity when creating our models by removing based on VIF. However, we need to examine the diagnostic plots to check if the models made work well for the data we used.

**

*Linear Regression Diagnostic Plot for Russia*

We will examine Russia’s diagnostic plot to explain the various plots. The assumption of linearity can be checked by examining the **Residuals vs Fitted** graph on the top left. A good plot will show no fixed pattern; that is, the red line should be approximately horizontal at zero. In Russia’s case, the line appears to curve at both ends, which may indicate a problem with the linear model made. Perhaps a linear model is not so suitable to be used in this case since the variables may not have a linear relationship.

As for the **Normal Q-Q** plot in the top right, we can visually verify the normality assumption. The plot should approximately follow a straight line. For Russia, the points all fall nicely on a straight line, except the point labelled 114. Thus, we can assume that the residuals follow a normal distribution.

Looking at the **Scale-Location** graph in the bottom left, we can check to what extent the residuals have constant variance. An ideal plot would have a relatively horizontal red line and points spread equally across the range. We can see in Russia’s plot that the points are not very well distributed throughout the range, but the red line is relatively horizontal. Thus, we can conclude that the residuals have a moderately but not completely constant variance.

Lastly, inspecting the **Residuals vs Leverage** graph in the bottom right, we can pick out influential outliers. Not all outliers are influential in linear regression analysis, hence the Cook’s distance is used to determine the influence of a value. Ideally, all points should be well inside the Cook’s distance lines. However, for Russia, point 114 lies very close to the Cook’s distance boundary, indicating that it could be an influencing outlier.

We could consider removing 114 as it appears to stand out in all our diagnostic plots, this could improve the accuracy of our linear regression model for Russia. Finally, we can repeat this diagnostic for all the other countries to arrive at a more accurate model, please see the other diagnostic plots at Figure 26-29 in Appendix.

#### 4. Creating and Selecting a CART model

We created a regression tree model using the function rpart which recursively partitions the data to grow the regression tree. Firstly, we split the data (specifically cleaned for CART) into train and test data 7:3, using the same seed that was used for train-test split in linear regression, this ensures that train and test data sets for both models have a similar distribution and hence can be compared relatively more fairly.

We will explain how we created the optimal tree using Canada as an example. We first created the CART model using Canada’s train set data; we have to apply the “anova” method since our response variable is continuous. From there, we were able to create the maximum tree model for Canada such that no variables were left out.

Then we printed the complexity parameter (cp) table and obtained the cross validated (CV) error at each prune trigger. We found the minimum CV error and its 1 standard deviation error (1SE) to get our error cap. We are also able to plot the CV error on a graph against the complexity parameters and observe which is the optimal tree, otherwise known as the first tree with a CV error that falls below the error cap.

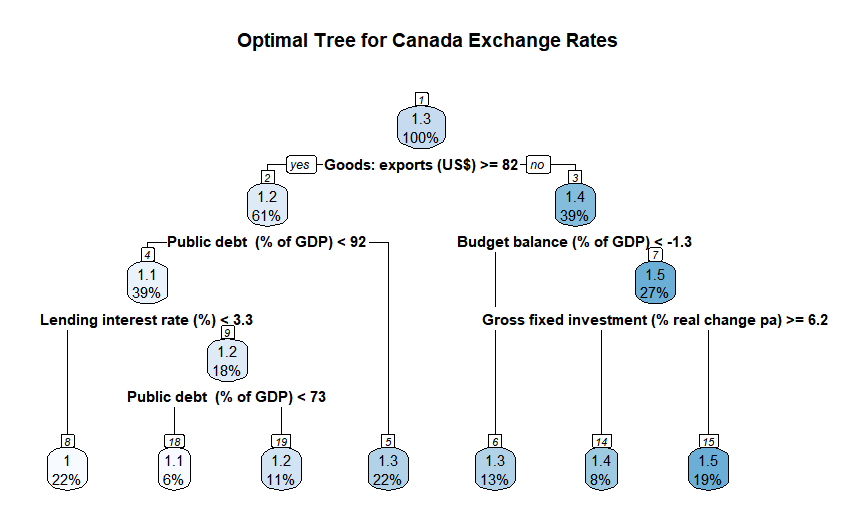
To obtain our optimal cp value, we calculated the geometric mean of the cp of the tree that first falls right below the CV error cap line and that of the tree before (see Figure 30 in Appendix). By using that we then prune our maximum tree with the optimal cp value resulting in our optimal tree shown in the figure below:

| n= 79   node), split, n, deviance, yval  \* denotes terminal node   1) root 79 2.29781700 1.273703   2) Goods: exports (US$)>=81.788 48 0.77321710 1.171463   4) M1 Money supply (LCU)< 755.805 29 0.16220800 1.082101   8) M1 Money supply (US$)>=383.1716 20 0.03182998 1.039565 \*  9) M1 Money supply (US$)< 383.1716 9 0.01377722 1.176626 \*  5) M1 Money supply (LCU)>=755.805 19 0.02596961 1.307856 \*  3) Goods: exports (US$)< 81.788 31 0.24594900 1.432011   6) Budget balance (% of GDP)< -1.3135 10 0.02212507 1.345856 \*  7) Budget balance (% of GDP)>=-1.3135 21 0.11425310 1.473036   14) Gross fixed investment (% real change pa)>=6.2265 6 0.00767030 1.381378 \*  15) Gross fixed investment (% real change pa)< 6.2265 15 0.03601178 1.509700 \* |
| --- |

Now, we are able to print details of the optimal regression tree, such as the variables used and the decision at each split. For Canada, `Budget balance (% of GDP)`, `Goods: exports (US$)`, `Gross fixed investment (% real change pa)` and `M1 Money supply (US$)` are the variables used to construct the optimal tree model. From the display, we can also see the relative error at the terminal nodes as well as the root node error which is 0.0291.

| Regression tree: rpart(formula = `Exchange rate LCU:US$ (av)` ~ ., data = trainsetc,   method = "anova", control = rpart.control(minsplit = 2, cp = 0))  Variables actually used in tree construction: [1] Budget balance (% of GDP) Goods: exports (US$) Gross fixed investment (% real change pa) [4] M1 Money supply (LCU) M1 Money supply (US$)   Root node error: 2.3/79 = 0.0291  n= 79    CP nsplit rel error xerror xstd 1 0.5565 0 1.0000 1.025 0.1143 2 0.2546 1 0.4435 0.537 0.0617 3 0.0507 2 0.1889 0.303 0.0608 4 0.0477 3 0.1382 0.281 0.0576 5 0.0307 4 0.0905 0.237 0.0522 6 0.0169 5 0.0598 0.158 0.0447 |
| --- |

Next, we can print out the optimal tree for Canada for visual understanding of what the regression tree looks like. In order to read the tree, we recall that predictions are only made at terminal nodes, every internal node is for making a decision, starting from the root node.



After obtaining the CART model using the trainset data, we can then use the testset data to test the accuracy of the model we have constructed. Firstly, we will use the CART model to predict values of exchange rates with the testset data’s given variables.

Lastly, we are then able to find the RMSE value and from there determine the accuracy of our model. We then applied the same approach for creating the CART model across the other four countries. Their resulting RMSE values are shown in the table below. The optimal decision tree of each country can be found in the Appendix (Figures 31-34 in Appendix).

| **Country** | **Testset Error** |
| --- | --- |
| Canada | 0.08927858 |
| Netherlands | 0.09855255 |
| Russia | 11.97997 |
| Sweden | 0.8715446 |
| United Kingdom | 0.03243715 |

*Table of testset RMSE for each country’s CART model*

#### 5.1. Selecting between Linear Regression and CART

In our data visualisation, we observed that, in general, the relationship between the variables chosen and exchange rate tend to not be linear. However, we could see that the data were segregated in a way that CART would likely be able to handle. Hence, we suggest to EIU to use CART as the machine learning methodology of choice for predicting exchange rates.

We acknowledge that the test error was mostly higher for CART than for linear regression. However, this could probably be explained by how we handled missing values for data preparation for linear regression. Since we replaced NA values with the mean of the respective columns, this might be why our linear models were able to produce a lower testing error. This not only improved the test error, but may have made the model formed using the train data inaccurate. This is because the mean of the column may have been very different from the actual data.

We also note the advantages of using CART for EIU. CART is able to handle NA values because of its surrogate functionality, this eliminates the need for EIU to clean NA data. We recognise the complexity of collecting such a large amount of economic data; the data may be unrecorded, impossible to derive, or not available to the public. Hence, the surrogate function is a great advantage of CART over linear regression.

Furthermore, our team's idea to increase the value of EIU’s flagship country reports is by predicting future exchange rates and testing its degree of stability. To achieve this we will need EIU to accurately predict the independent variables to include into our model to predict our exchange rates. We understand that not all independent variables that are required in our linear regression model may be easy to predict, hence the surrogate function in CART gives EIU the freedom to choose which variables they can predict more accurately, making the future exchange rate values more accurate. Overall, CART is a better choice for predicting future exchange rates whereas linear regression may struggle to do so.

The last advantage is that CART selects variables from the maximum tree more automatically to create the optimal model. In linear regression, we had to manually select and remove the variables, which when done on a large scale, can lead to much human error. Automation is an important consideration when choosing a model to create for many more countries.

#### 5.2. Explaining the Variables Selected by our Chosen Model – CART

CART’s algorithm selects variables that minimise the errors in predictions of exchange rates. Initially, a maximum regression tree would be created using all the available variables in the dataset. However, the maximum tree would only be accurate with the trainset data that was used to create it since it would overfit the data, hence it would not be accurate with other new data sets. Therefore, it is necessary to prune the tree to a model which has a right balance between fitting for the train and test datasets, achieving the minimum error for both. During the pruning process, we find the optimal cp and rpart will automatically select the variables it deems most important in predicting exchange rates to create the most simple yet accurate tree.

We will analyse Canada’s optimal tree to explain why CART chose the variables that it did. As seen in its optimal tree, it relied on `Goods: exports (US$)` to determine the first decision. This can be seen in the real world as exchange rates have an effect on the trade surplus or deficit, and a weaker domestic currency stimulates exports but makes imports more expensive (Kramer, 2020). If a country exports more than it imports, there is a high demand for its goods, and thus, for its currency. When demand is high, prices rise and the currency appreciates in value (Lioudis, 2020), making exports an important indicator of where exchange rates would lean towards.

Following that, at node 2 and 3, the tree is split by `Public Debt (% of GDP)` and `Budget balance (% of GDP)` respectively. A large debt encourages inflation, and if inflation is high, the debt will be serviced and ultimately paid off with cheaper real dollars in the future. Foreigners will be less willing to own securities denominated in that currency if the risk of default is great (Twin, 2021), inevitably affecting exchange rate values. On the other hand, a budget deficit can easily result in an inflow of foreign financial capital, a stronger exchange rate and thus, a trade deficit, since a strong exchange rate makes it more difficult for exporters to sell their goods abroad while making imports cheaper (Openstax, 2016).

Node 4 and 7 made use of `Lending Interest rate (%)` and `Gross Fixed Investment (% real change pa)` to further split into terminal notes. Higher interest rates offer lenders in an economy a higher return relative to other countries, attracting foreign capital and causing exchange rates to rise (Twin, 2021), making lending interest rate an important factor. Gross fixed investments are also closely related to exchange rates as they are acquisitions of produced assets, including the production of such assets by producers for their own use, minus disposals. The relevant assets are related to assets that are intended for use in the production of other goods and services for a period of more than a year, in which over this period of time exchange rates would play a major role in determining the value of the said assets.

Among the other optimal trees produced, these are common variables that are repeated throughout, such as `M1 Money Supply (US$)`, `Lending Interest rate (%)`, `Public Debt (% of GDP)` and `Goods: exports (US$)`. This proves that the variables are consistently important in predicting exchange rates throughout other countries as well. We could make a CART model in general using such variables, similar to how we planned for linear regression.

Additionally, there are variables such as `Unemployment rate (%)` and `Deposit interest rate (%)` which appear frequently in the other optimal trees (can be viewed at Figures 31-34 in Appendix). In the long term, a country with high unemployment rates may boost productivity and competitiveness, creating jobs and increasing demand for exports, reducing unemployment and thus strengthening the exchange rate (Pettinger, 2011). High unemployment may be caused by a lack of competitiveness, which thus reduces the value of the exchange rates over time. The deposit rate is the interest rate paid by commercial banks or financial institutions on cash deposits of account holders. If a country can achieve a successful balance of increased interest rates without an accompanying increase in inflation, its currency's value and exchange rate are more likely to rise (Lioudis, 2021). These two variables are also weighed as highly important and very relevant in the current economy.

However, if we move on to analyse the variables for Canada using the variable importance function, we observe that the variables that we have determined to be relevant using the optimal tree differs from that in CART.

| M1 Money supply (US$)  1.69434873  Long-term bond yield (%)  1.57397202  Stock of money M2 (US$)  1.52110151  Goods: exports (US$)  1.35535030  Goods: imports (US$)  1.28381379  Nominal GDP (US$)  1.22552619  Public debt (% of GDP)  0.69543073  Deposit interest rate (%)  0.35788558  Gross fixed investment (% real change pa)  0.32734256  Budget balance (% of GDP)  0.24912526  Lending interest rate (%)  0.14803019  Unemployment rate (%)  0.13470401  Short term interest rate (%; average)  0.11866110  Private consumption (real % change pa)  0.06465458 |
| --- |

Upon obtaining the variables for Canada, we can see that the most important variable with the highest value is `M1 Money Supply (US$)`. Increases in money supply could lead to a depreciation in the exchange rate due to inflation and lower interest rates (Pettinger, 2017). Any increase in the money supply is likely to cause inflation as with more currency chasing the same quantity of goods, firms will respond by putting up prices. Furthermore, higher money supply puts downward pressure on interest rates, leading to depreciations in exchange rates. This makes it evident that `M1 Money Supply (US$)` would be the most important variable as it can be said to be directly linked to exchange rates.

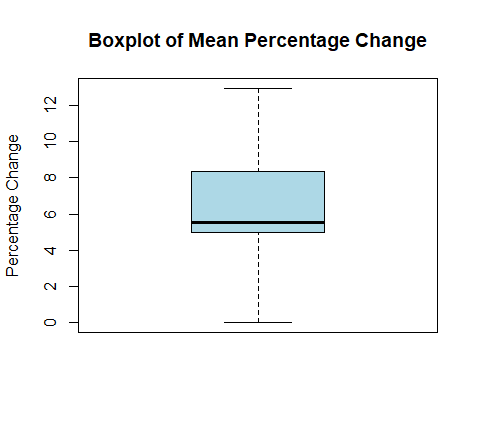
We can also observe that the variables used in the tree did not consist of the three least important variables, meaning that all the variables used are relatively relevant in predicting exchange rates. This further proves that our CART model has been accurate in sifting out the variables after pruning since those chosen have been proven to be important and realistically applicable to the current economic situation. Therefore, we can safely rely on the optimal tree in selecting variables automatically. The other countries’ variable importance can also be viewed at Figures 35-38 in Appendix.

#### 6. Obtaining a Scale of Stability by Calculating Year-by-Year Fluctuations

This next section details the creation of a new section i.e. “Degree of Stability”, which we believe will be useful to be included in EIU’s country reports in future reports.

#### 6.1. Formulating our scale

To determine the degree of fluctuations in our predicted model, we first decided that we would make a scale that we could base our decision on. To formulate this scale, we took data from the World Bank which provided us with exchange rates of 203 different countries from the year 1960 to 2020 (see Data Sources in Resources). For data cleaning, we first decided to remove the data from the years 1960 to 1999 as we wanted to use more recent data, making our scale more relevant for our current day use. We then removed all the constant exchange rates as we assume that this is erroneous data, since it is almost impossible that an exchange rate is constant for 20 years (Gishen, 2021). We further removed countries that had 0 values in their exchange rate as the World Bank indicated that they were unable to retrieve those country’s exchange rate during those years. After our data cleaning, we calculated the year on year (excluding 2000) percentage change of the exchange rate by taking the difference of the current year’s exchange rate and the previous year’s exchange rate divided by the previous year’s exchange rate multiplied by 100. We also then take the absolute value of this figure as we are only interested in the magnitude of the percentage change. Having attained the percentage changes for the years 2001 to 2020, we then calculated the mean percentage change of each country's exchange rate. We then plotted a boxplot using this data (shown in the figure below).



This boxplot then helps us to create a scale for the degree of fluctuation of exchange rates for countries where any mean percentage change lower than 5.00%, the 1st quartile, can be deemed as the lowest degree of fluctuation of exchange rate while any value above 8.36%, the 3rd quartile, can be considered as the highest degree of fluctuation.

#### 6.2. Obtaining the Degree of Fluctuation in our Predicted Data

Having decided that our CART model is more suitable, we used the CART model to predict the exchange rates from 2000 to 2020 of the five countries we have chosen. We then calculate the percentage change of each year using the same method stated above that we have used for our scale. The mean absolute percentage change for the predicted exchange rates for the five countries are shown in the table below.

| **Country** | **Mean Percentage Change of Exchange Rate** |
| --- | --- |
| Canada | 2.543745 |
| Netherlands | 1.766449 |
| Russia | 11.52225 |
| Sweden | 2.804483 |
| United Kingdom | 0.8351119 |

*Degree of fluctuation in exchange rate from 2000 to 2020*

From the scale we can then deduce that Russia has a high degree of fluctuation from the year 2000 to 2020, which is true if you were to observe the exchange rate in Figure 3 in Appendix of the time-series graph of Russia’s exchange rate. Russia experienced huge changes in their exchange rates especially in the 3rd quarter of 2014, in which their exchange rate increased from 36.20 to 47.3433 in the very next quarter. Overall, Russia experienced the greatest percentage change from 2000 to 2020. Thus, this result not only shows the usefulness of our scale in being able to help us deduce the degree of fluctuation of the exchange rate in a certain country, we are also able to prove the accuracy of the CART model in helping us predict the exchange rate such that we were able to calculate the mean percentage change year on year exchange rate.

Furthermore this scale can be made flexible for the data EIU is trying to study. For instance if EIU were to determine the degree of fluctuation of exchange rate of a country from a different year range other than 2000 to 2020. New data can be fed into the scale, making the scale more applicable to the period that EIU wants to examine.

#### 7. Pilot Studies in EIU and Link to Country Report

With that, we have shown CART to be a useful model in predicting future exchange rates, and the degree of stability scale to be useful in determining future stability using the predicted exchange rate. Thus, EIU should strongly consider using CART to supplement their predictions. First, we suggest that they use machine learning to predict exchange rates. This allows for them to arrive at a concrete, quantified value, after which the forecasted values can be further refined by experts' inputs and surveys. Then, determine the degree of stability using the scale created to arrive at a final exchange rate stability index value.

These findings can constitute a new section in EIU’s flagship country reports, under a section titled “Exchange Rate Stability” in their forecasting reports. This section would particularly be of value to multinational companies and companies looking to expand overseas or source from overseas. After reading this section of the report, they can make decisions accordingly to ensure that international ventures do not suffer losses in times of predicted high currency risk.

EIU can test the accuracy of the CART model by comparing past predicted values against actual values when those values are released. Alternatively, EIU could train the CART model with data from previous years, for example up until 2015, then use the created model to predict exchange rates for 2016 and beyond. Then they can calculate the difference between the actual and predicted to evaluate the effectiveness of the machine learning. EIU can do something similar to verify the accuracy of the stability scale created as well. Overall, EIU should try machine learning for predicting exchange rates, and get feedback from customers as to how this new section to be created has been useful to them.

#### 8. Challenges regarding Assumptions Made

In our report, we have provided a Proof of Concept (POC) of the functionality of using the CART model to forecast exchange rates for different countries, and using those predictions to determine the degree of fluctuation in future exchange rates. We concede that good predictions from our machine learning CART model for future exchange rates will be dependent on the accuracy of EIU predictions of the selected independent variables. Hence, our team recommends EIU to further explore multiple machine learning models to predict the other data in their country reports, so that the predicted independent variables that they feed into CART will be as accurate as possible for the prediction of exchange rates. However, we understand that some of the independent variables in our decision tree may be difficult to predict. This is also an additional reason why we chose the CART model over the linear regression model. CART can use surrogates when the data at a decision split is unavailable or even majority rule when all data is absent, allowing EIU to be flexible in predicting independent variables to be put in the CART model to predict future exchange rates.

Finally, we acknowledge the downfalls of our model. The metrics chosen were all “internal” factors of an economy. We were unable to find sufficient data for the time period we wanted to examine for other factors, such as Covid-19 contraction rate, global financial crisis data, and so on. We recognise that exchange rate could be very well predicted by such external factors. For example, with the current Covid-19 pandemic, EIU has data regarding the rating of how well governments are reacting to the pandemic, and it has been proven that this data has a direct impact on the economy itself. Hence, we also suggest EIU to consider quantifying and collecting such external economic measures that are likely to have impacts on exchange rate. This would greatly improve the accuracy of the model since we know that exchange rate is highly susceptible to external factors. We encourage EIU to continuously include data from their special reports that they might find have a correlation with exchange rate to reinforce the CART model, making it more accurate.

#### 9. Conclusion

Having recognised that there may be some shortfalls to our CART model, we still believe that the CART machine learning model is still a good consideration for EIU in predicting their exchange rate data. Our team believes instead of using machine learning as a replacement to their current multiple forecasting methodologies, we see machine learning as a valuable addition to the current methodologies used, helping to further improve the accuracy of their overall country report. We also believe that the method used to get a scale is a very valuable addition to the EIU reports as it helps their client to be able to foresee degrees of fluctuation on a scale, providing greater value to the clients and hence creating a new source of revenue for EIU. With machine learning, we believe that EIU can keep that step ahead of their competitors such as Capital Economics and remain the main company that provides forecasting and advisory services to assist entrepreneurs, financiers, and government officials.

#### References

Addepto. (2021, September 14). *Machine learning in Economics - How is it Used?*

<https://addepto.com/machine-learning-in-economics-how-is-it-used/>

Deloitte. (2017). *Business impacts of machine learning*.

<https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/process-and-operations/TG_Google%20Machine%20Learning%20report_Digital%20Final.pdf>

McAfee, A., & Brynjolfsson, E. (2014, October 8). *Big Data: The Management Revolution*.

<https://hbr.org/2012/10/big-data-the-management-revolution>

CompareRemit (2020, April 9). 8 Key Factors that Affect Foreign Exchange Rates.

<https://www.compareremit.com/money-transfer-guide/key-factors-affecting-currency-exchange-rates/>

Segal, T. (2021, January 4). Currency Fluctuations: How they Affect the Economy.

<https://www.investopedia.com/articles/forex/080613/effects-currency-fluctuations-economy.asp>

Silver, C. (2020). The Top 25 Economies in the World. Retrieved 2 November 2021, from

<https://www.investopedia.com/insights/worlds-top-economies/>

Tures, J. (2020). Is the US Economy ‘Just OK’ Compared to Other Countries?. Retrieved 2 November 2021, from

<https://observer.com/2020/01/us-economy-vs-other-countries-how-america-stacks-up/>

Kramer, L. (2020, March 20). How Importing and Exporting Impacts the Economy..

<https://www.investopedia.com/articles/investing/100813/interesting-facts-about-imports-and-exports.asp>

Lioudis, N. (2020, October 19). How the Balance of Trade Affects Currency Exchange Rates.

<https://www.investopedia.com/ask/answers/041515/how-does-balance-trade-impact-currency-exchange-rates.asp>

Gishen, S. M. (2021, January 7). How often do exchange rates fluctuate?

<https://www.investopedia.com/ask/answers/08/how-often-to-exchange-rates-fluctuate.asp>

Openstax. (2016, May 18). Principles of Economics.

<https://opentextbc.ca/principlesofeconomics/chapter/31-4-fiscal-policy-and-the-trade-balance/>

Twin, A., 2021. 6 Factors That Influence Exchange Rates. Investopedia. Available at:

<https://www.investopedia.com/trading/factors-influence-exchange-rates/>

Pettinger, T., 2017. Money supply and the exchange rate - Economics Help. [online] Economics Help.

<https://www.economicshelp.org/blog/11550/currency/money-supply-and-the-exchange-rate/>

Pettinger, T., 2011. Question: what effect would higher unemployment have on a country's currency? - Economics Help. [online] Economics Help. Available at:

<https://www.economicshelp.org/blog/2899/economics/unemployment-and-exchange-rate/>

Lioudis, N., 2021. How Do National Interest Rates Affect Currency Values and Exchange Rates?. Investopedia. Available at:

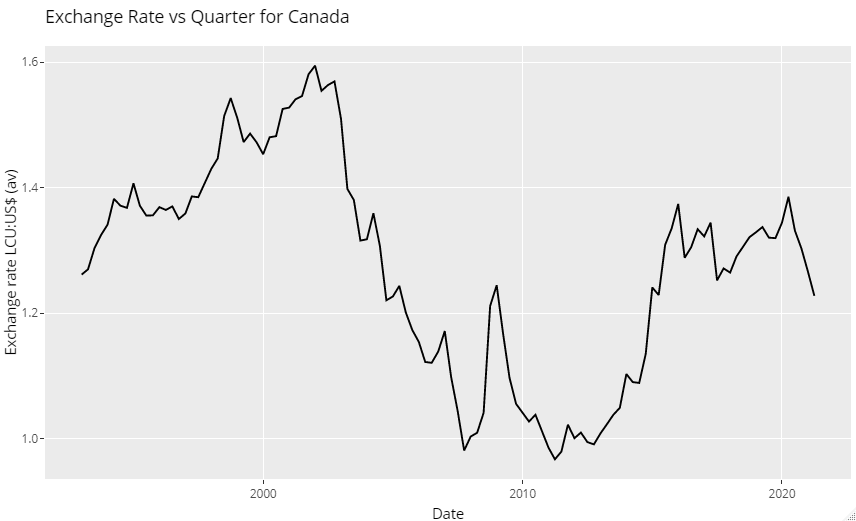
<https://www.investopedia.com/ask/answers/040315/how-do-changes-national-interest-rates-affect-currencys-value-and-exchange-rate.asp>

**Data Sources**:

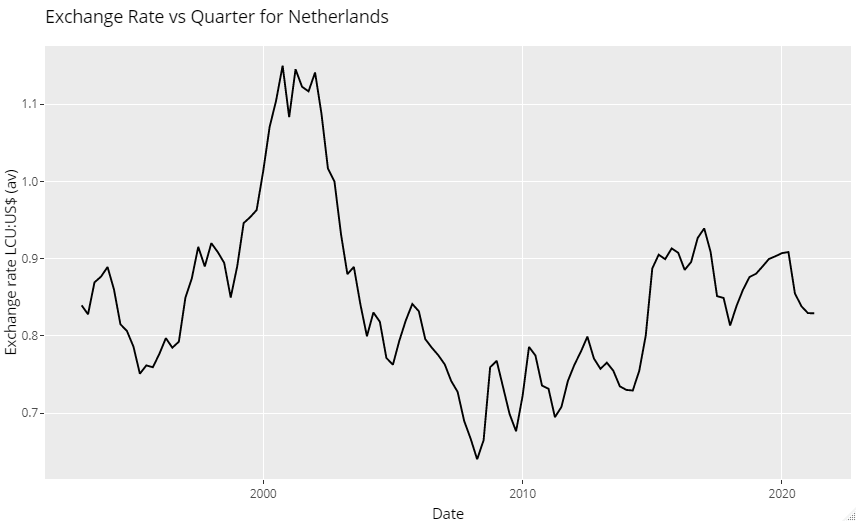
<https://data.eiu.com/>

[https://data.worldbank.org/indicator/PA.NUS.FCR](https://data.worldbank.org/indicator/PA.NUS.FCRF)

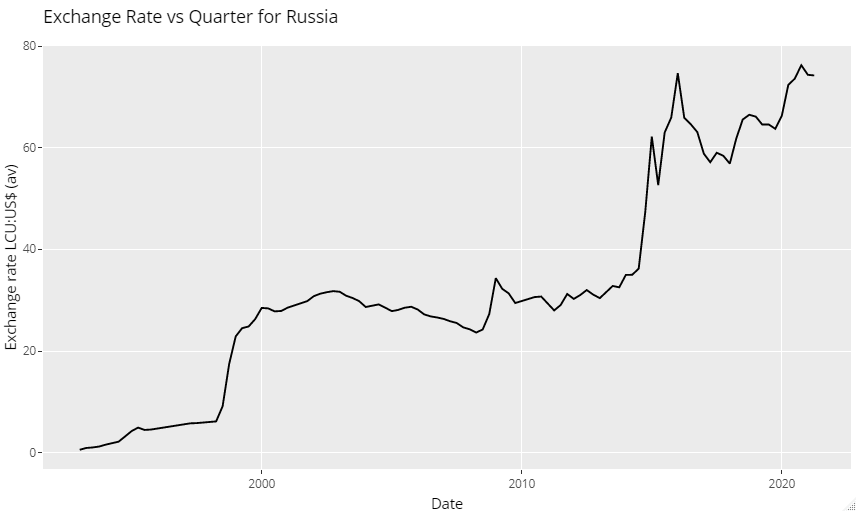
#### Appendix



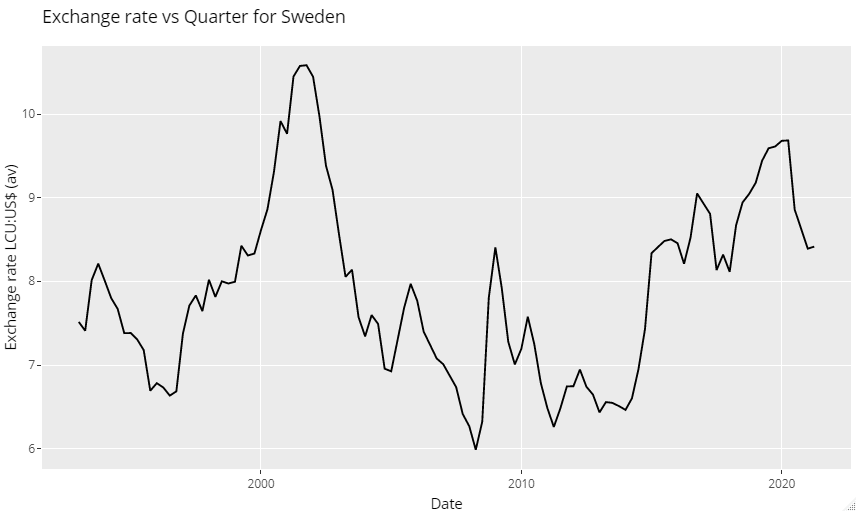
*Figure 1: Time-series graph of exchange rate over the years for Canada*



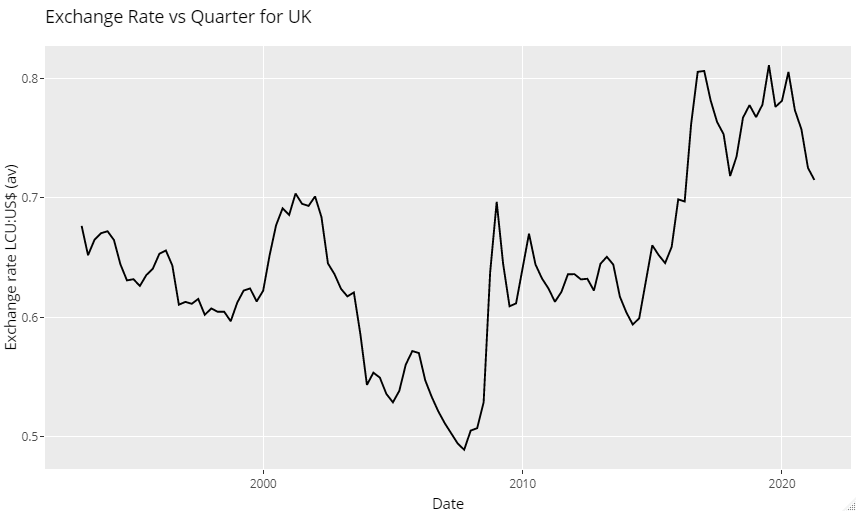
*Figure 2: Time-series graph of exchange rate over the years for Netherlands*



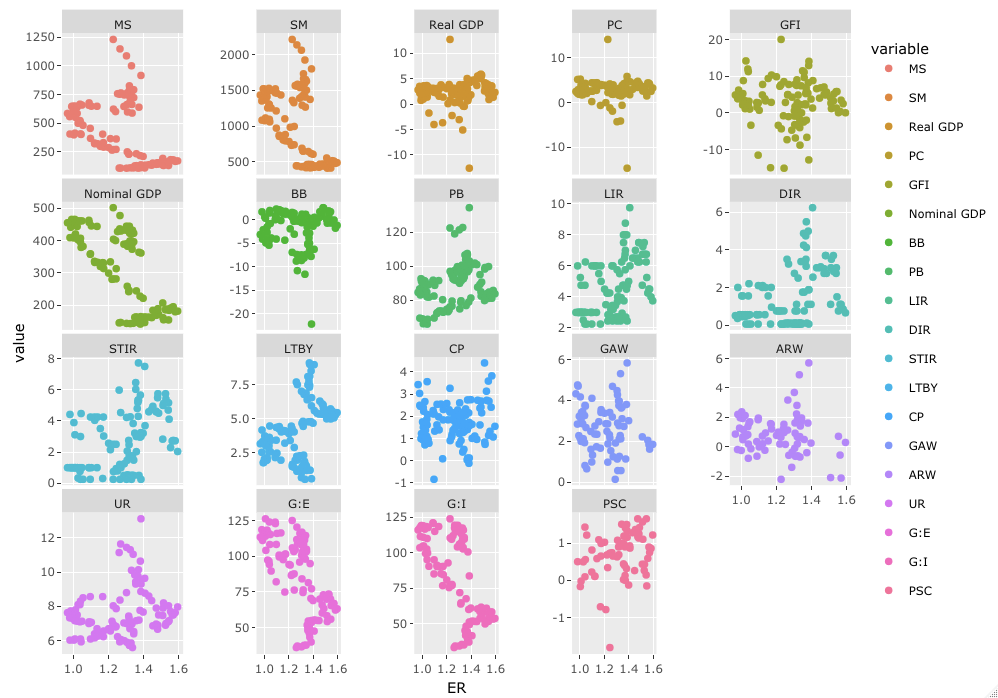
*Figure 3: Time-series graph of exchange rate over the years for Russia*



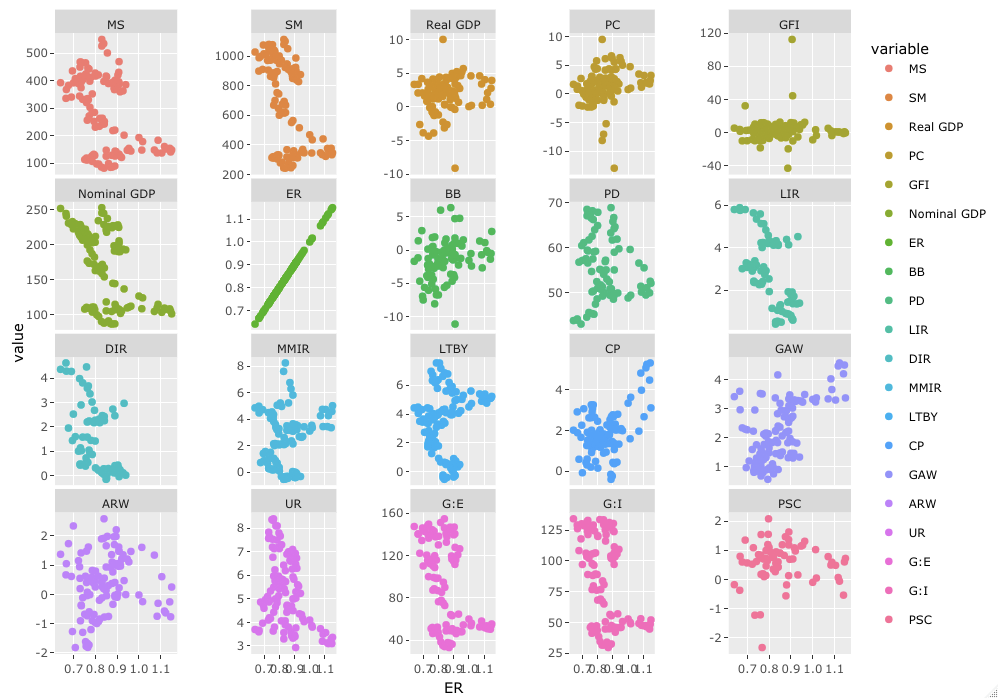
*Figure 4: Time-series graph of exchange rate over the years for Sweden*

****

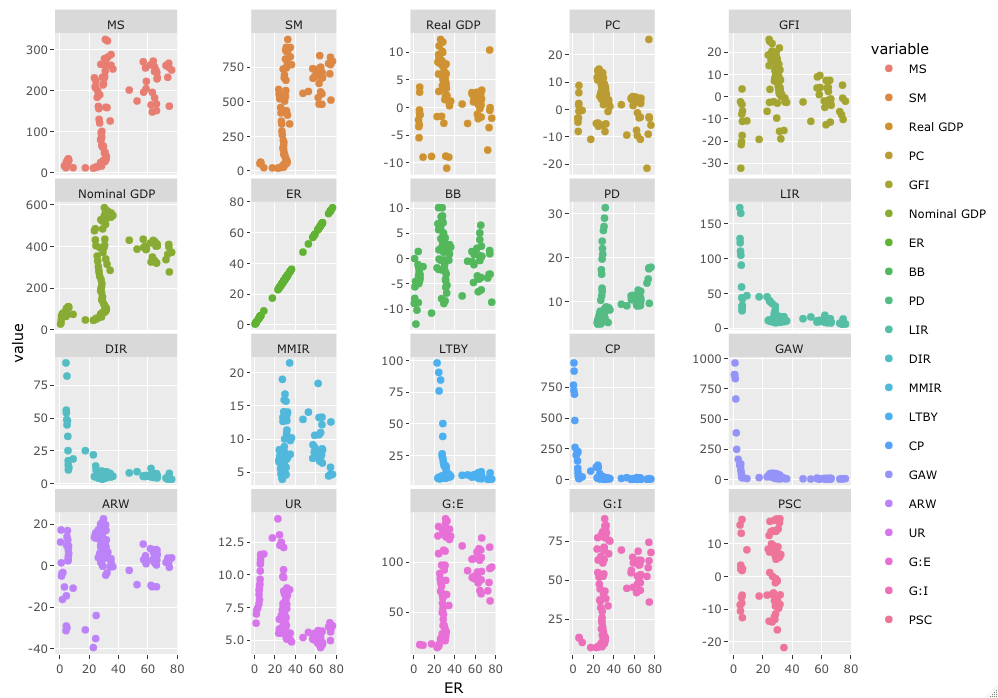
*Figure 5: Time-series graph of exchange rate over the years for United Kingdom*

****

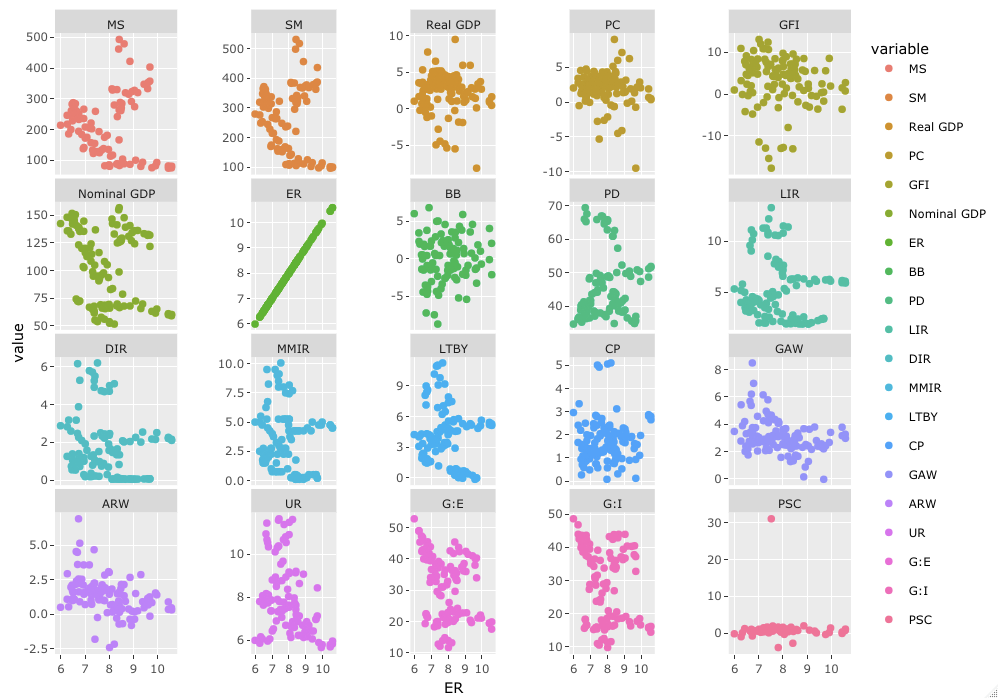
*Figure 6: Scatter plot with facet wrap of all independent variables against exchange rates for Canada*

****

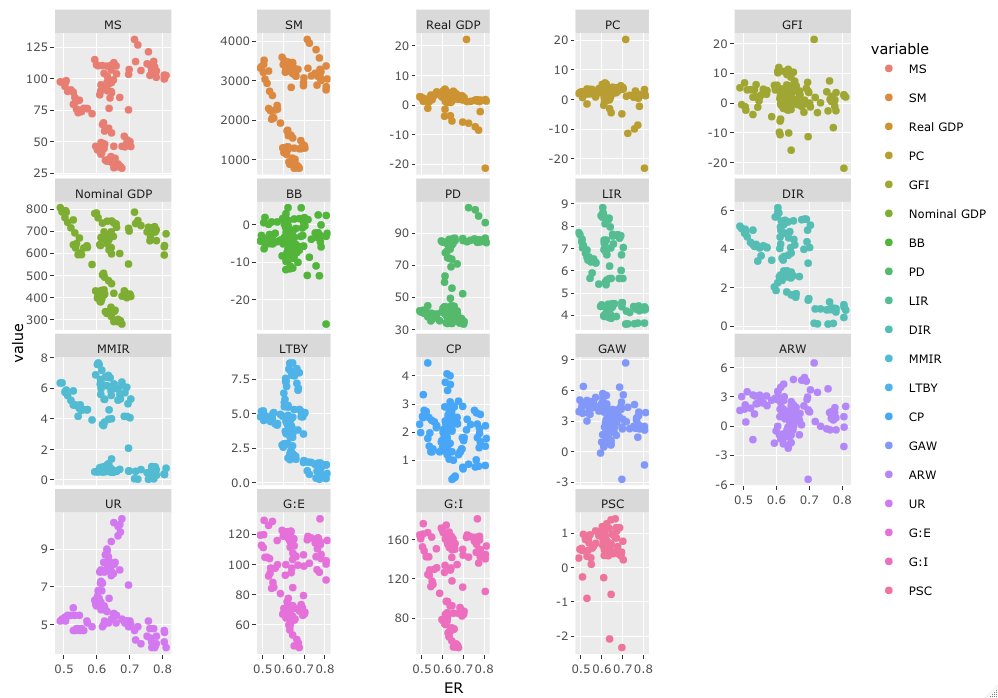
*Figure 7: Scatter plot with facet wrap of all independent variables vs exchange rates for Netherlands*

****

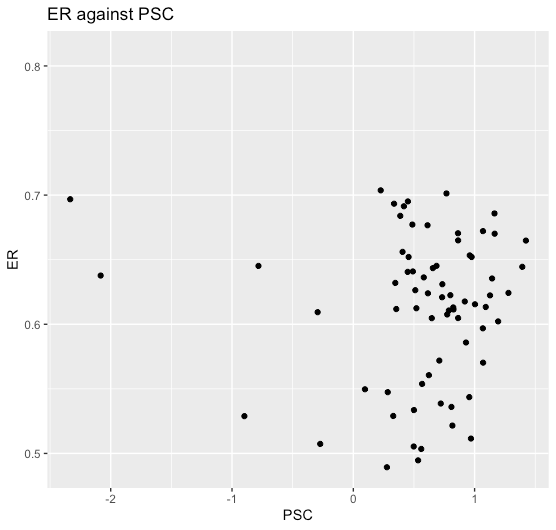
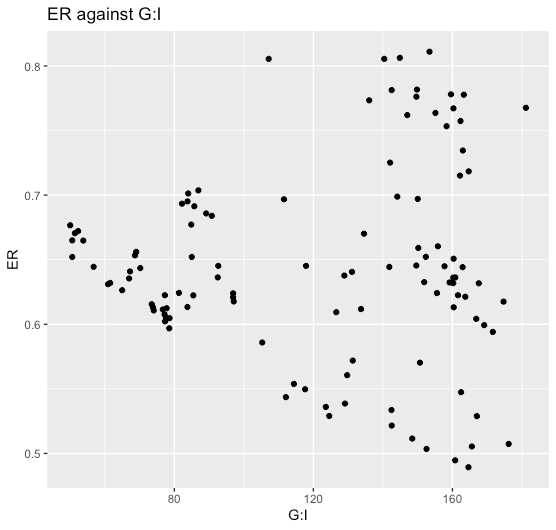
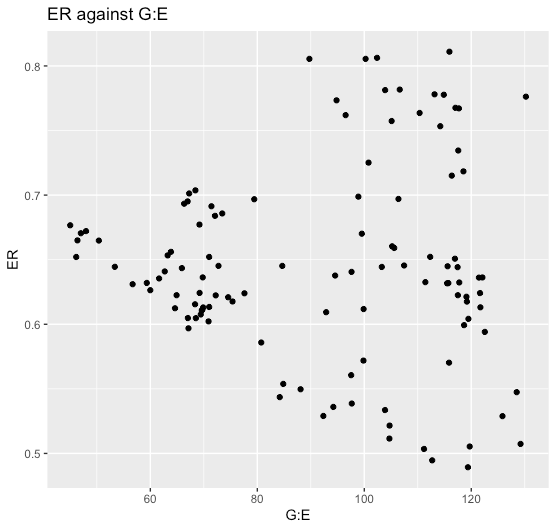
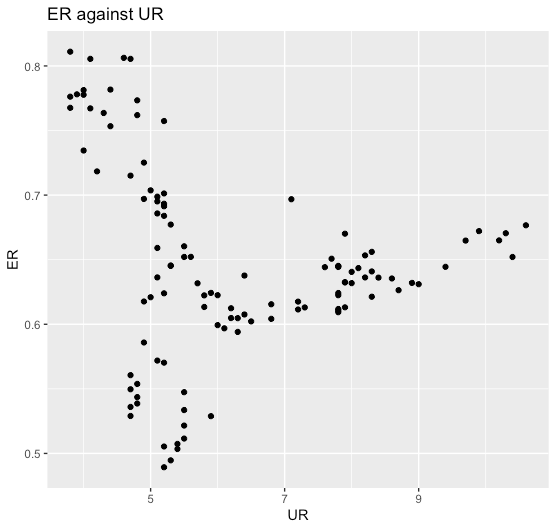
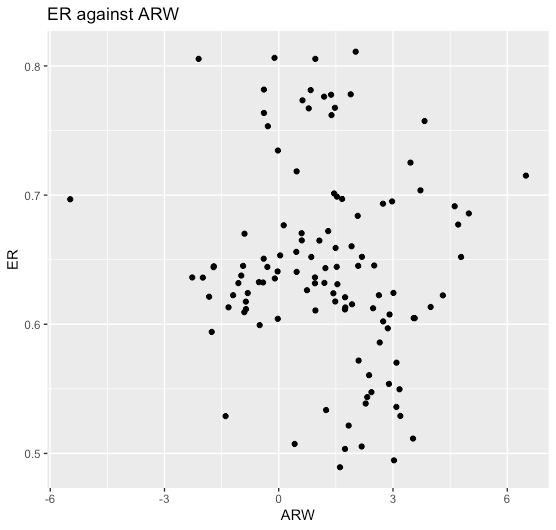
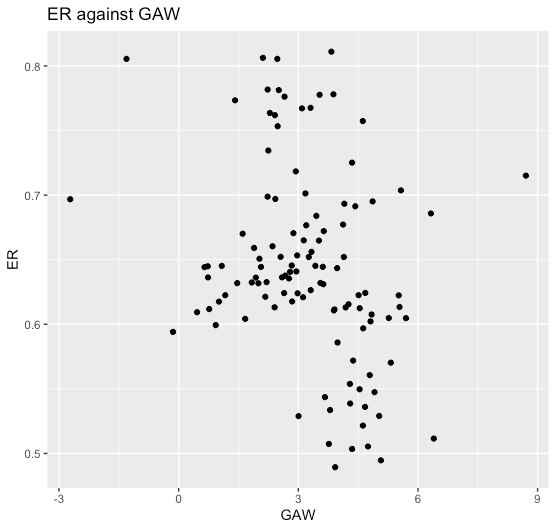
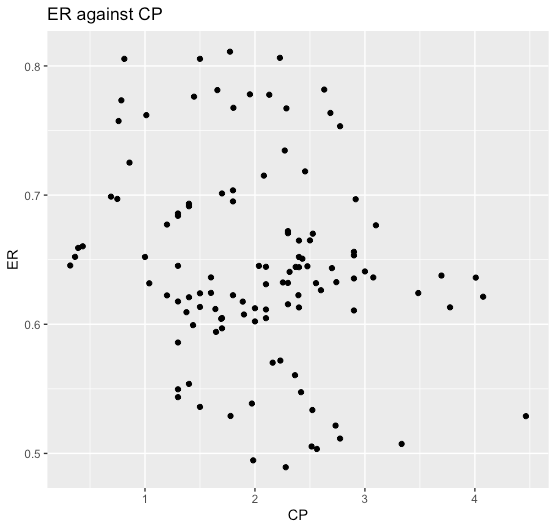
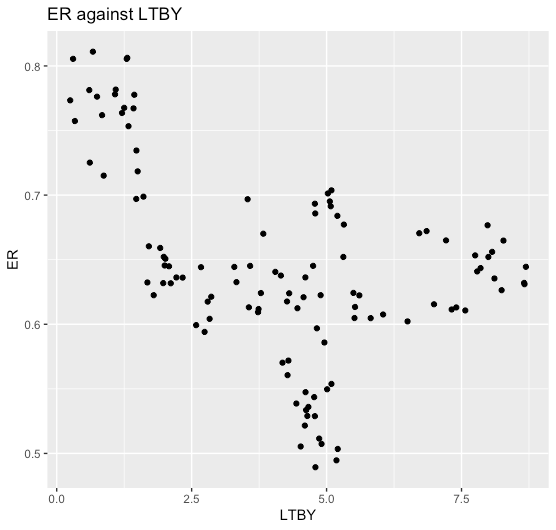
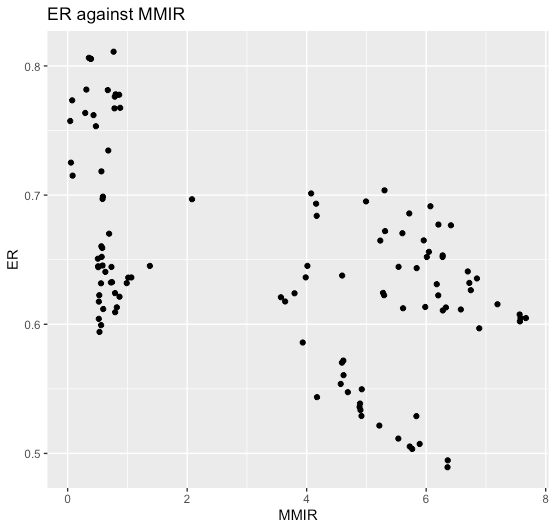
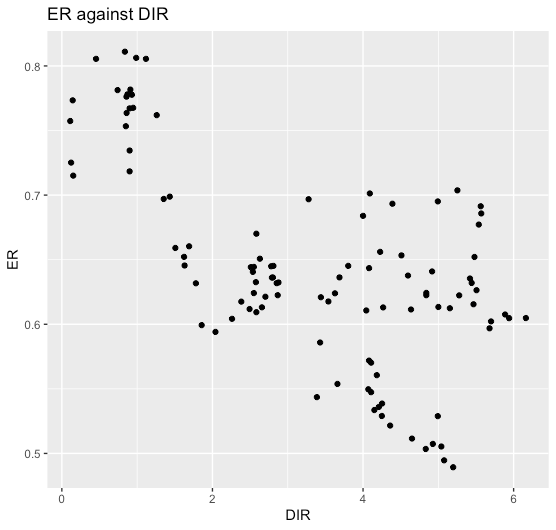
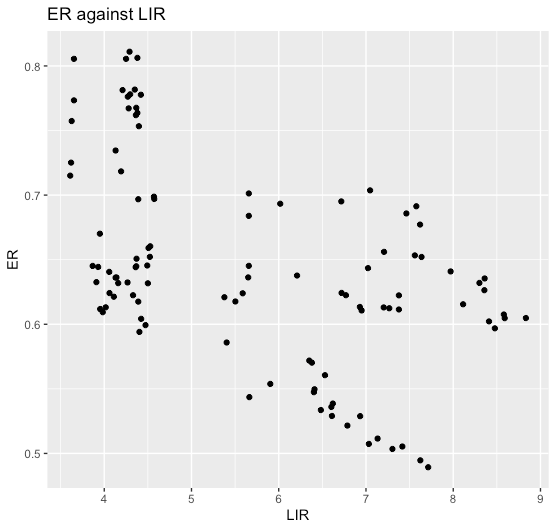
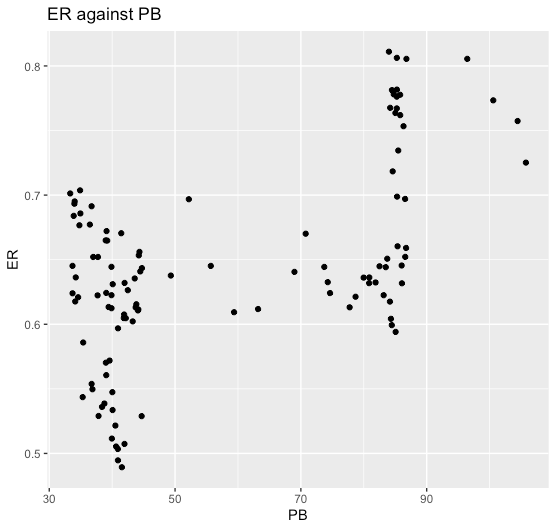
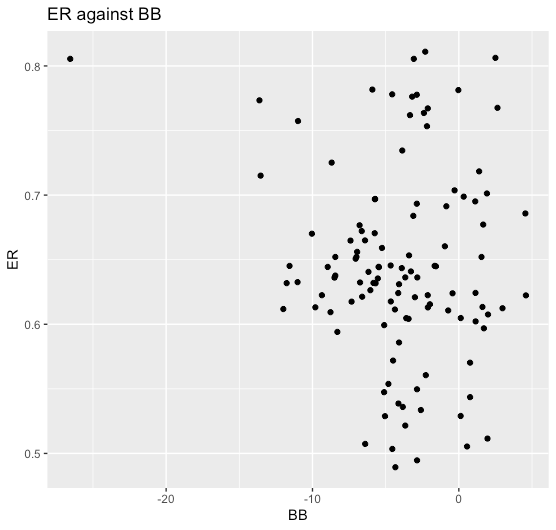
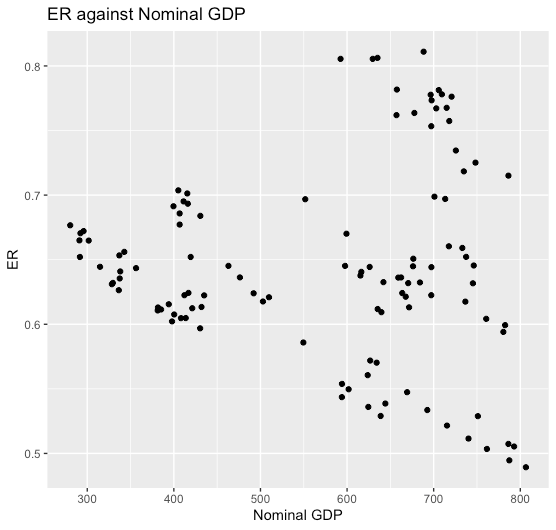
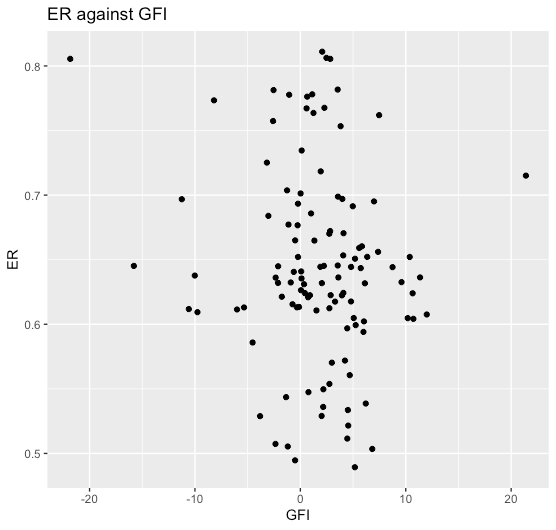
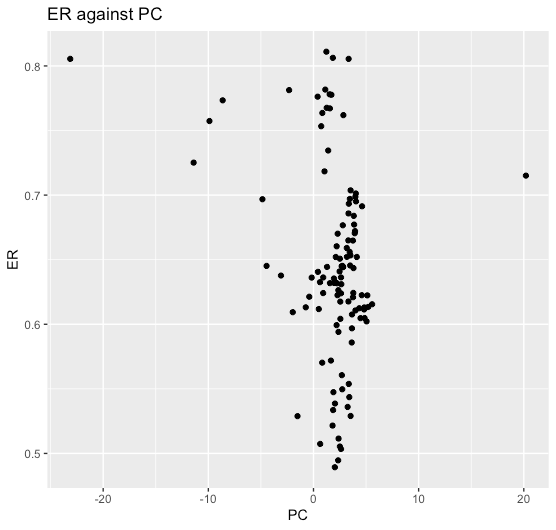
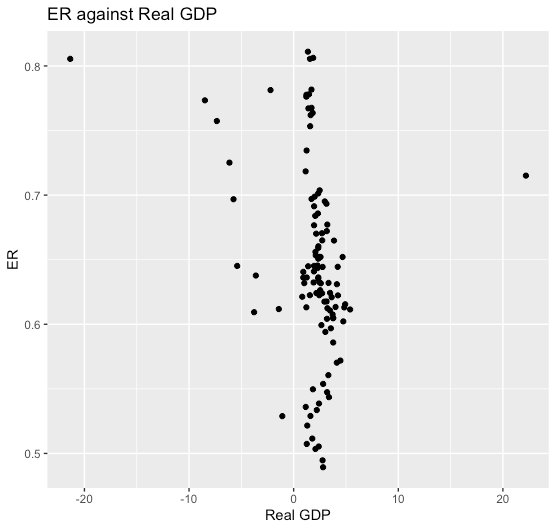
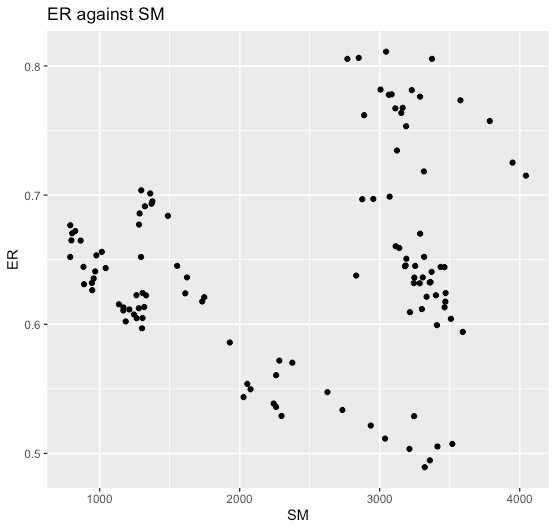
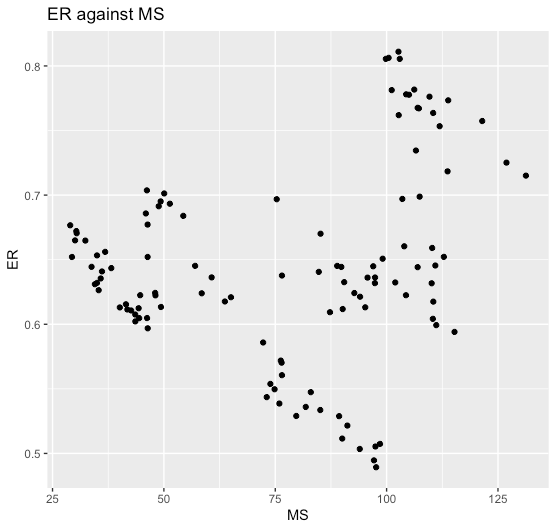
*Figure 8: Scatter plot with facet wrap of all independent variables against exchange rates for Russia*

****

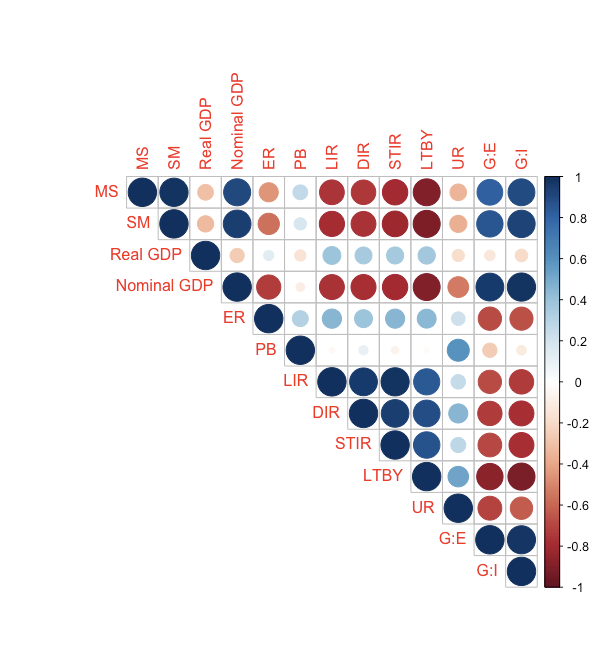
*Figure 9: Scatter plot with facet wrap of all independent variables against exchange rates for Sweden*

****

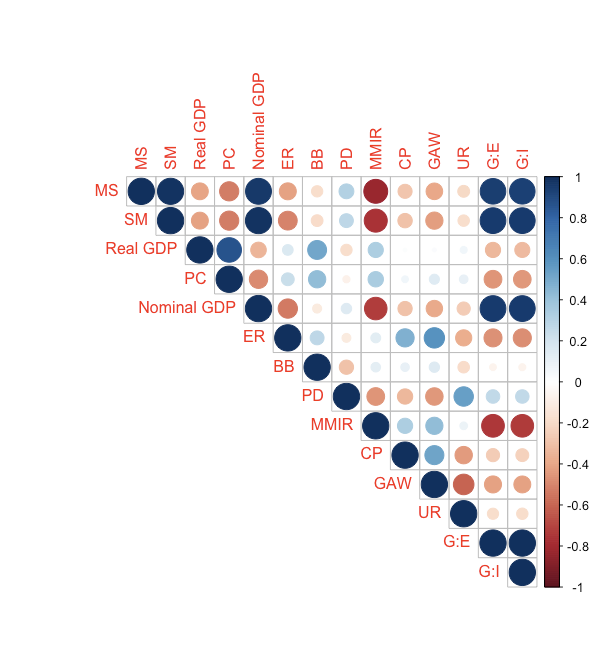
*Figure 10: Scatter plot with facet wrap of all independent variables against exchange rates for UK*



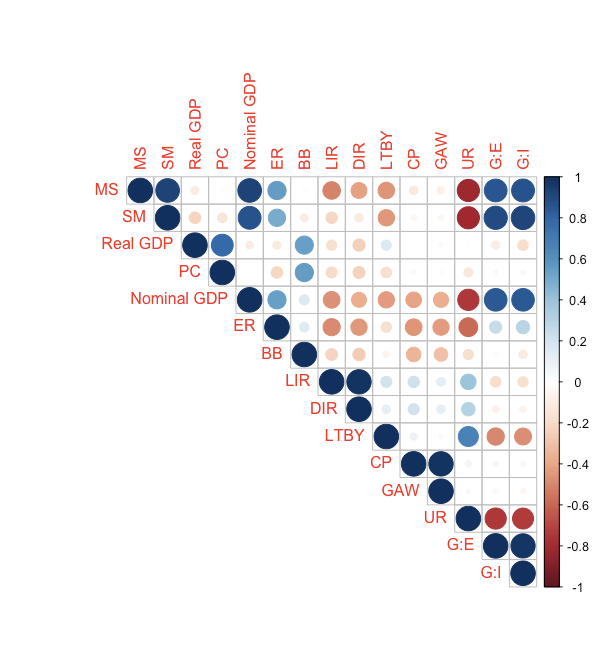
*Figure 11: Individual scatter plot of exchange rates against all independent variables for UK*



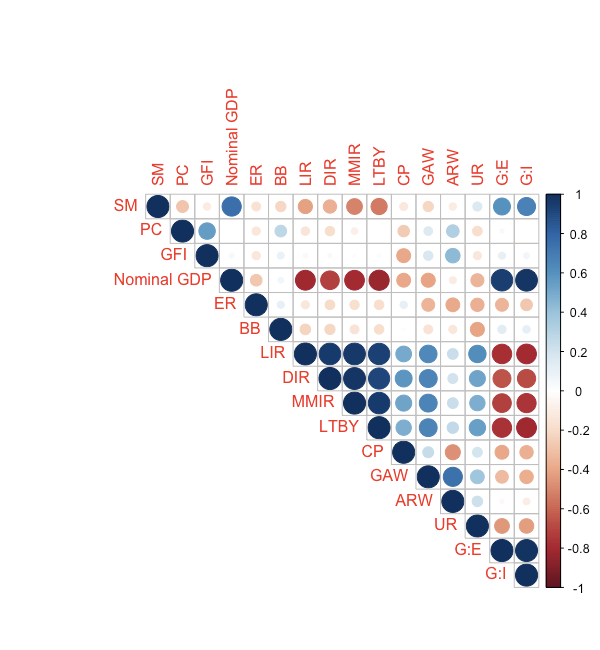
*Figure 12: Correlation heatmap for Canada, focus on the line “ER”*



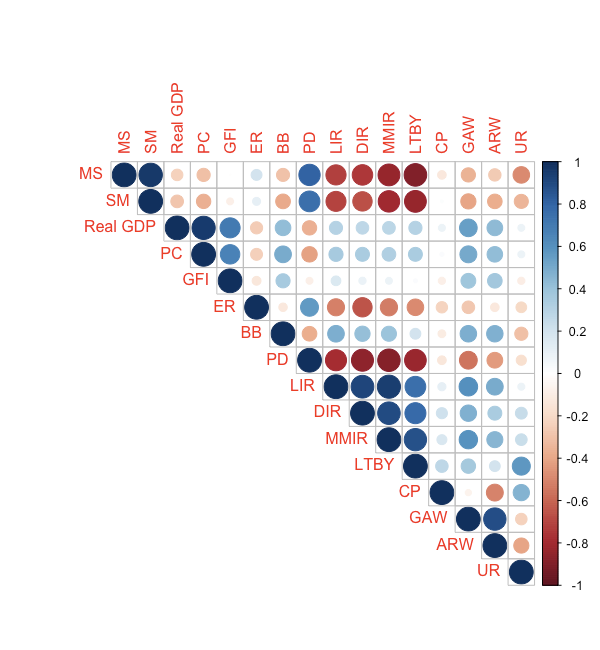
*Figure 13: Correlation heatmap for Netherlands, focus on the line “ER”*



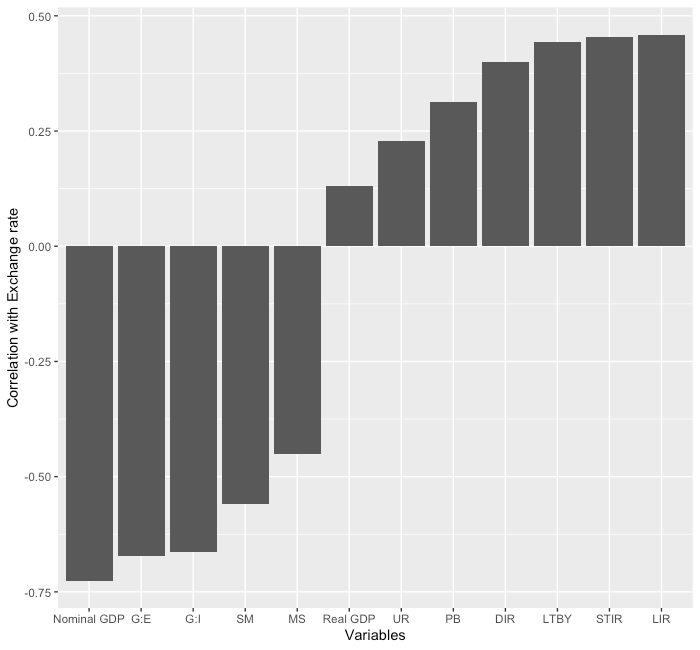
*Figure 14: Correlation heatmap for Russia, focus on the line “ER”*



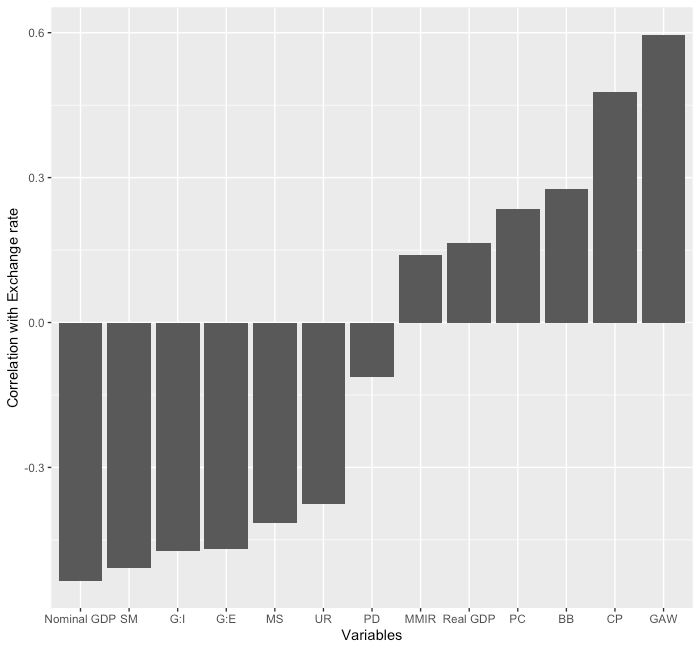
*Figure 15: Correlation heatmap for Sweden, focus on the line “ER”*



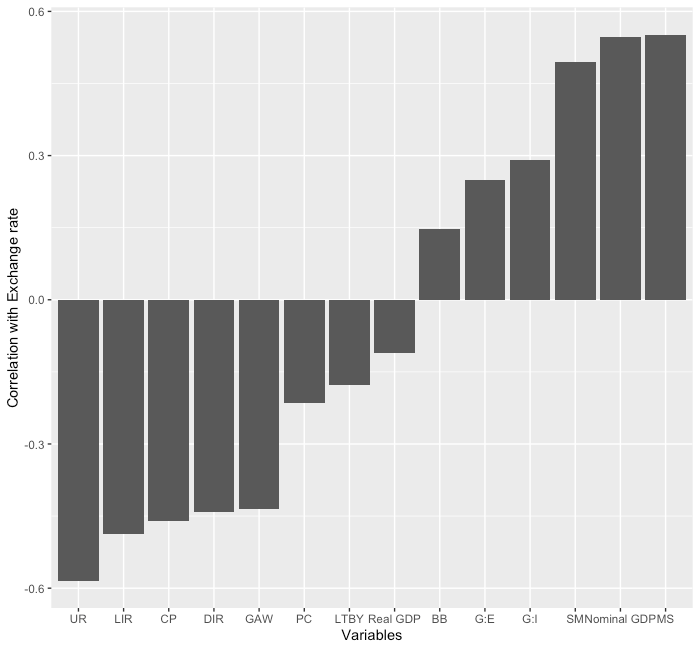
*Figure 16: Correlation heatmap for United Kingdom, focus on the line “ER”*



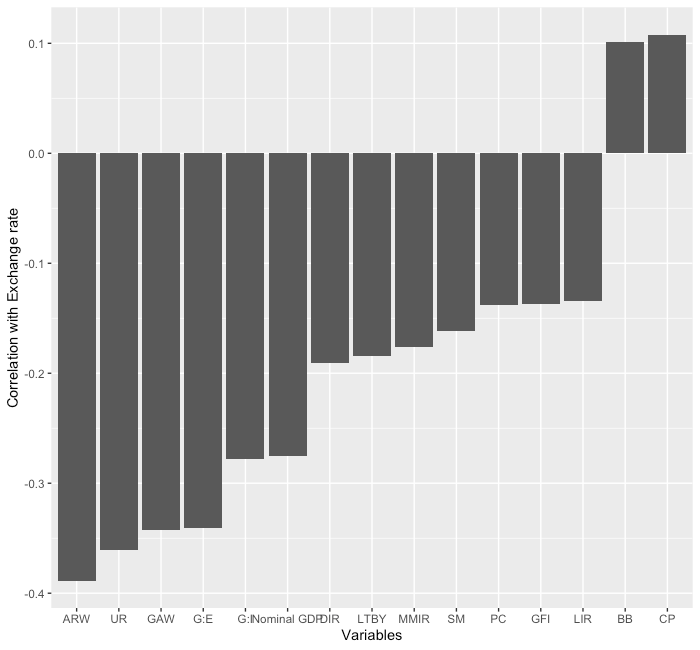
*Figure 17: Correlation graph for Canada*



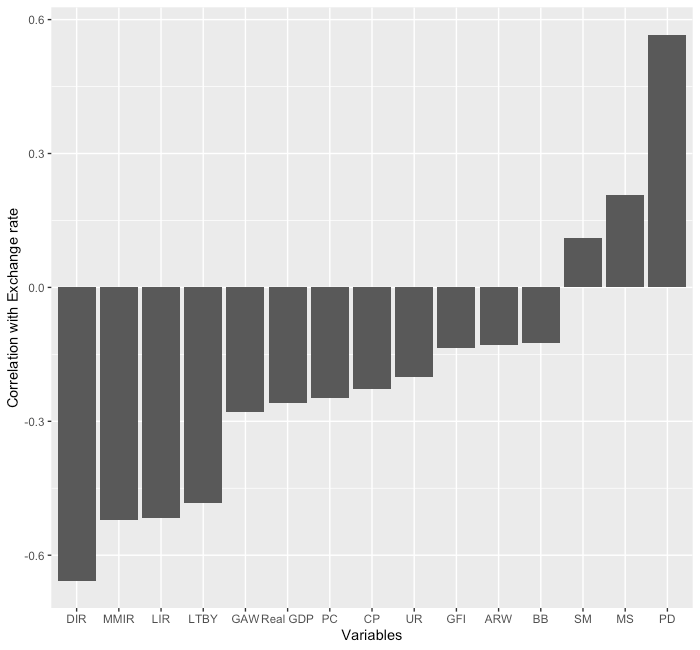
*Figure 18: Correlation graph for Netherlands*



*Figure 19: Correlation graph for Russia*

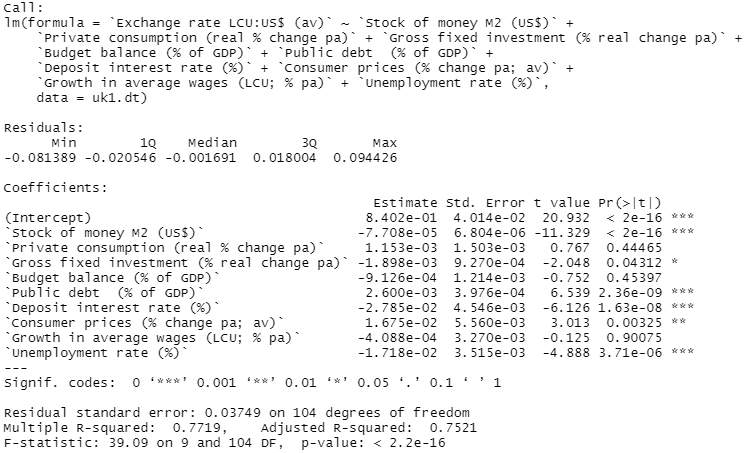


*Figure 20: Correlation graph for Sweden*

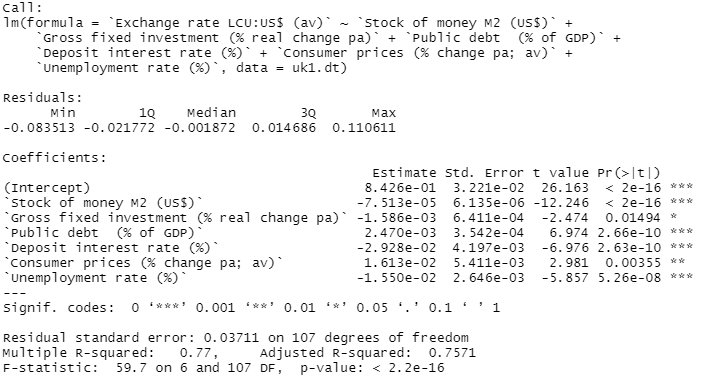


*Figure 21: Correlation graph for United Kingdom*

| Legend for Figure 6-21: | |
| --- | --- |
| MS: "M1 Money supply (US$)"  SM: "Stock of money M2 (US$)"  Real GDP: "Real GDP (% change pa)"  PC: "Private consumption (real % change pa)"  GFI: "Gross fixed investment (% real change pa)"  Nominal GDP: "Nominal GDP (US$)"  ER: "Exchange rate LCU:US$ (av)"  BB: "Budget balance (% of GDP)"  PD: "Public debt (% of GDP)"  LIR: "Lending interest rate (%)"  DIR: "Deposit interest rate (%)" | STIR: "Short term interest rate (%; average)"  MMIR: "Money market interest rate (%; average)"  LTBY: "Long-term bond yield (%)"  CP: "Consumer prices (% change pa; av)"  GAW: "Growth in average wages (LCU; % pa)"  ARW: "Average real wages (% change pa)"  UR: "Unemployment rate (%)"  G:E: "Goods: exports (US$)"  G:I: "Goods: imports (US$)"  PSC: "Private sector credit/GDP" |



*Figure 22: Summary of linear regression model after removing based on VIF for UK*



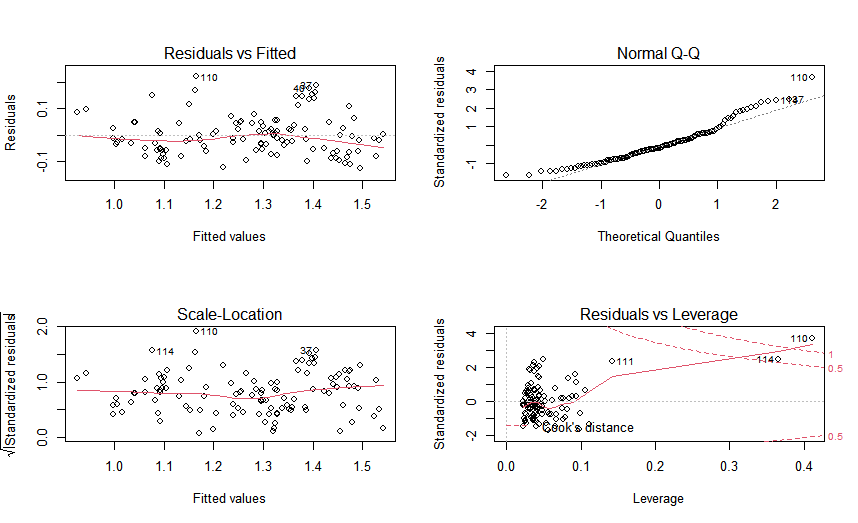
*Figure 23: Summary of linear regression model after removing based on VIF then AIC for UK*

| Canada: lm(formula = `Exchange rate LCU:US$ (av)` ~ `Real GDP (% change pa)` +  `Public debt (% of GDP)` + `Short term interest rate (%; average)` +  `Unemployment rate (%)` + `Goods: exports (US$)`, data = canada1.dt)  Netherlands: lm(formula = `Exchange rate LCU:US$ (av)` ~ `Real GDP (% change pa)` +  `Private consumption (real % change pa)` + `Budget balance (% of GDP)` +  `Public debt (% of GDP)` + `Money market interest rate (%; average)` +  `Consumer prices (% change pa; av)` + `Growth in average wages (LCU; % pa)` +  `Unemployment rate (%)` + `Goods: imports (US$)`, data = netherlands1.dt)  Russia: lm(formula = `Exchange rate LCU:US$ (av)` ~ `Private consumption (real % change pa)` +  `Deposit interest rate (%)` + `Long-term bond yield (%)` +  `Consumer prices (% change pa; av)` + `Unemployment rate (%)` +  `Goods: exports (US$)`, data = russia1.dt)  Sweden: lm(formula = `Exchange rate LCU:US$ (av)` ~ `Stock of money M2 (US$)` +  `Private consumption (real % change pa)` + `Gross fixed investment (% real change pa)` +  `Deposit interest rate (%)` + `Consumer prices (% change pa; av)` +  `Unemployment rate (%)` + `Goods: exports (US$)`, data = sweden1.dt)  UK: lm(formula = `Exchange rate LCU:US$ (av)` ~ `Stock of money M2 (US$)` +  `Gross fixed investment (% real change pa)` + `Public debt (% of GDP)` +  `Deposit interest rate (%)` + `Consumer prices (% change pa; av)` +  `Unemployment rate (%)`, data = uk1.dt) |
| --- |

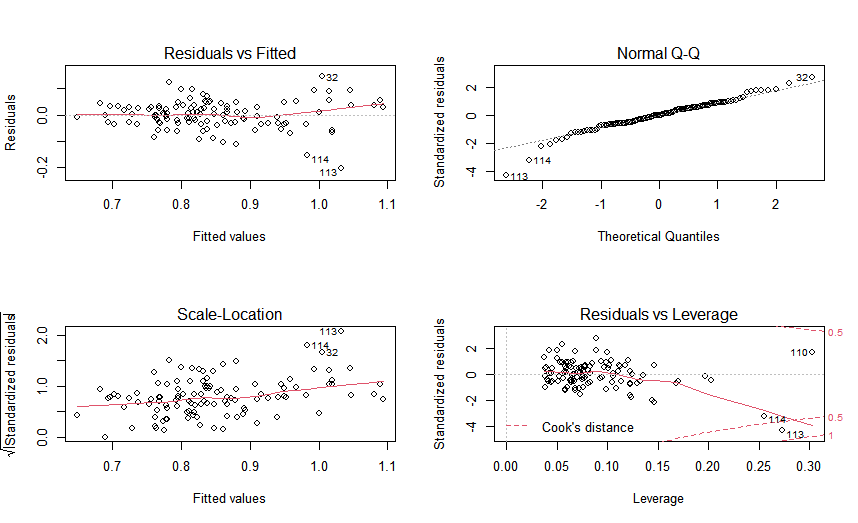
*Figure 24: Resulting linear regression models for each country*

| X Variable | Private Consumption | Budget Balance | Public Debt | Short Term Interest Rate |
| --- | --- | --- | --- | --- |
| Occurrences | 3 | 1 | 2 | 1 |
| X Variable | Consumer Prices | Exports | Real GDP | Long-term Bond Yield |
| Occurrences | 4 | 4 | 2 | 1 |
| X Variable | Money Market Interest Rate | Growth in Average Wages | Unemployment Rate | Imports |
| Occurrences | 1 | 1 | 5 | 1 |
| X Variable | Stock of Money | Gross Fixed Investment | Deposit Interest Rate |  |
| Occurrences | 2 | 2 | 3 |  |

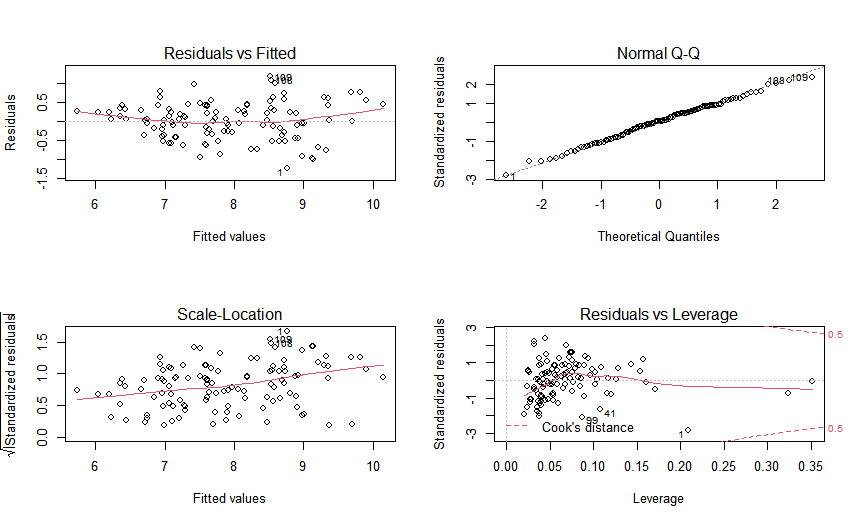
*Figure 25: Variable occurrences in each of the 5 models created*

**

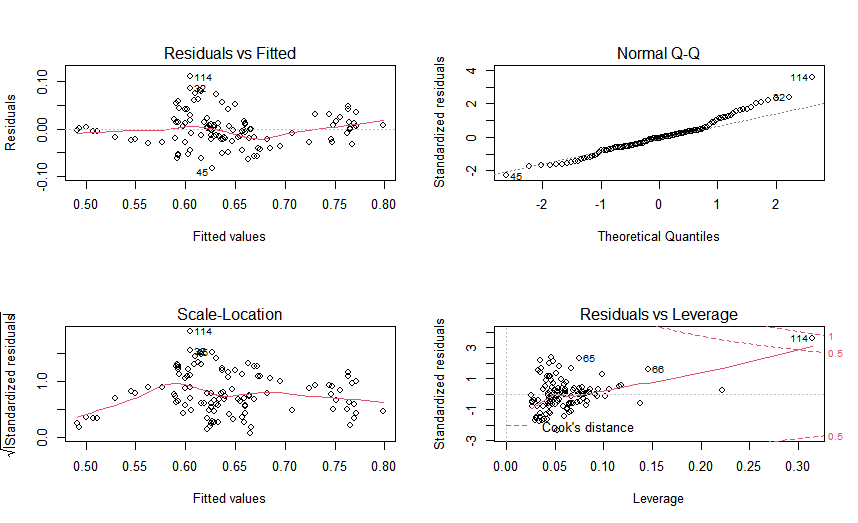
*Figure 26: Linear Regression Diagnostic Plot for Canada*

**

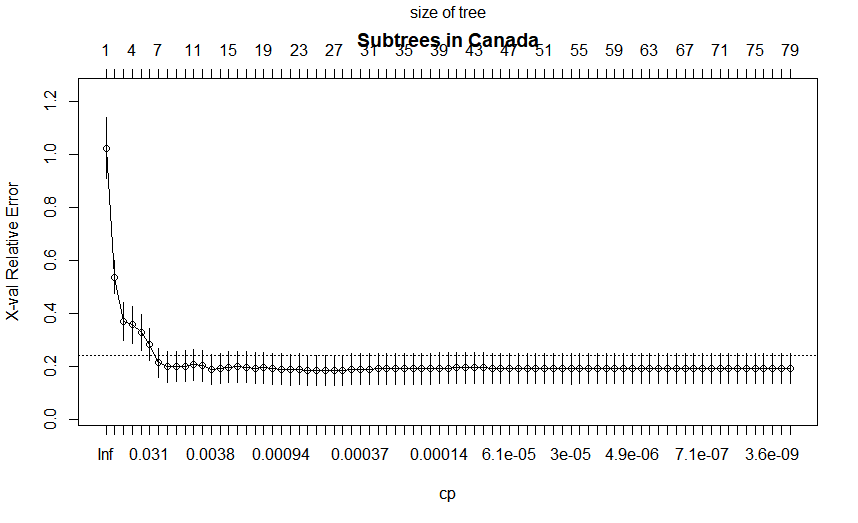
*Figure 27: Linear Regression Diagnostic Plot for Netherlands*

**

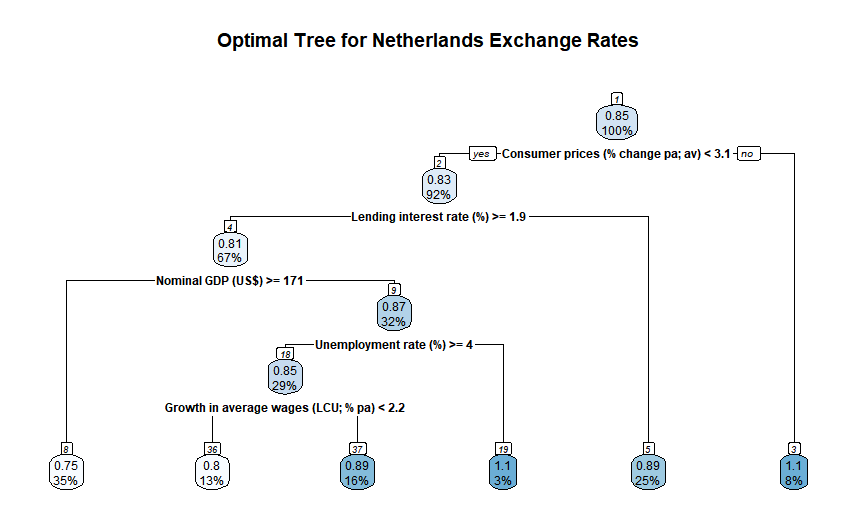
*Figure 28: Linear Regression Diagnostic Plot for Sweden*

**

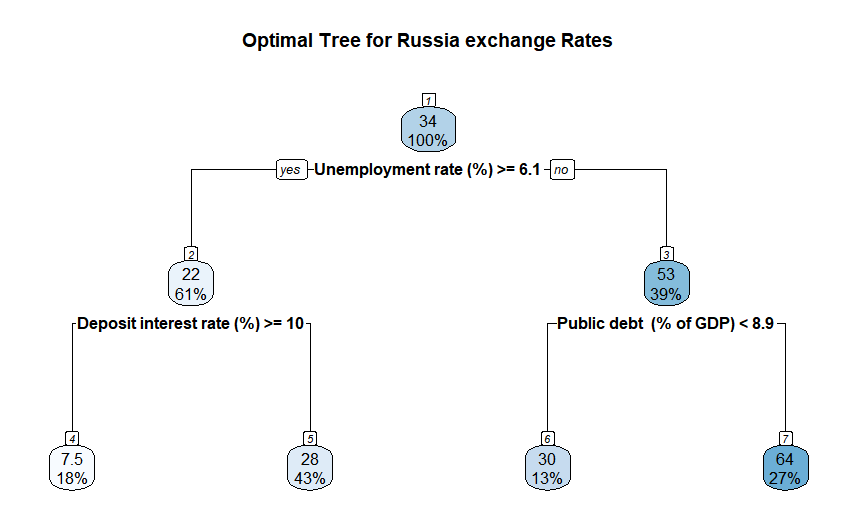
*Figure 29: Linear Regression Diagnostic Plot for UK*

**

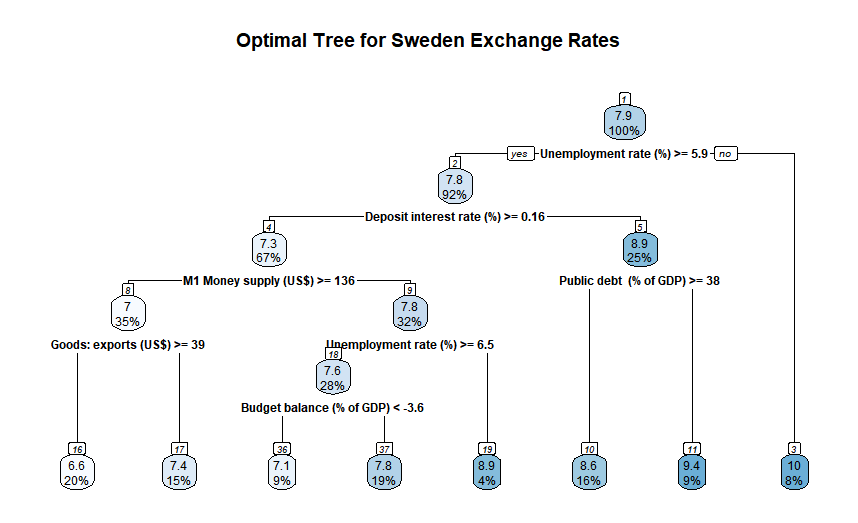
*Figure 30: CV error against cp value for maximum regression tree of Canada*

**

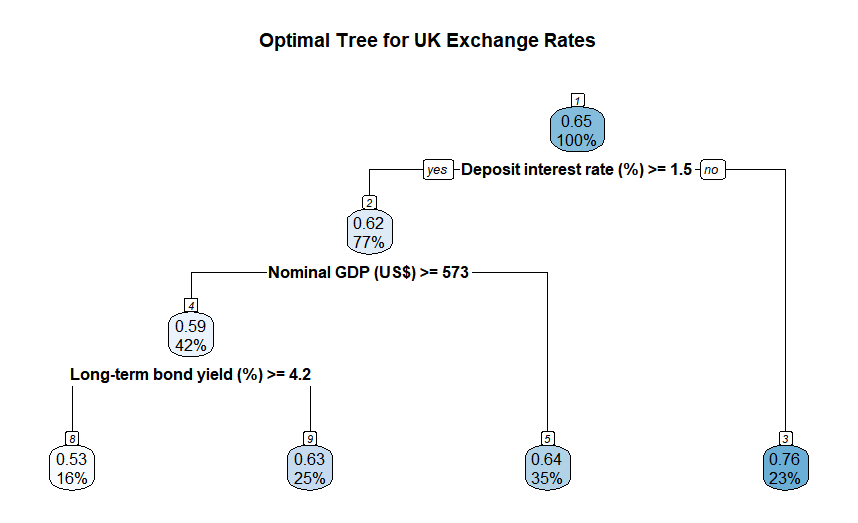
*Figure 31: Netherlands CART Optimal Tree*

**

*Figure 32: Russia CART Optimal Tree*

**

*Figure 33: Sweden CART Optimal Tree*

**

*Figure 34: UK CART Optimal Tree*

| Consumer prices (% change pa; av)  0.47960510  Growth in average wages (LCU; % pa)  0.37005717  Long-term bond yield (%)  0.35634484  Unemployment rate (%)  0.35514462  Nominal GDP (US$)  0.19985587  Goods: exports (US$)  0.19186163  Goods: imports (US$)  0.19186163  M1 Money supply (US$)  0.19186163  Stock of money M2 (US$)  0.19186163  Lending interest rate (%)  0.17929417  Money market interest rate (%; average)  0.17929417  Deposit interest rate (%)  0.15240004  Private consumption (real % change pa)  0.09687788  Average real wages (% change pa)  0.04025635  Real GDP (% change pa)  0.03019226 |
| --- |

*Figure 35: Netherlands CART Variable Importance*

| Unemployment rate (%)  20528.480  Nominal GDP (US$)  18849.826  Growth in average wages (LCU; % pa)  14543.091  M1 Money supply (US$)  11818.931  Consumer prices (% change pa; av)  11227.985  Stock of money M2 (US$)  10046.091  Public debt (% of GDP)  7812.106  Goods: exports (US$)  6249.685  Goods: imports (US$)  5468.474  Private consumption (real % change pa)  4687.263  Deposit interest rate (%)  4102.682  Lending interest rate (%)  3471.500  Budget balance (% of GDP)  1893.545  Gross fixed investment (% real change pa)  1893.545  Real GDP (% change pa)  1893.545 |
| --- |

*Figure 36: Russia CART Variable Importance*

| Long-term bond yield (%)  39.201832  Unemployment rate (%)  37.243352  Deposit interest rate (%)  36.764835  Money market interest rate (%; average)  36.764835  Lending interest rate (%)  33.966101  Nominal GDP (US$)  25.836386  M1 Money supply (US$)  25.181126  Public debt (% of GDP)  14.350192  Goods: exports (US$)  14.342600  Goods: imports (US$)  14.321849  Stock of money M2 (US$)  14.301098  Average real wages (% change pa)  11.820244  Private consumption (real % change pa)  7.655821  Budget balance (% of GDP)  2.992709  Real GDP (% change pa)  2.974292  Gross fixed investment (% real change pa)  1.827748 |
| --- |

*Figure 37: Sweden CART Variable Importance*

| Long-term bond yield (%)  0.41296363  Deposit interest rate (%)  0.37924321  Public debt (% of GDP)  0.28231526  M1 Money supply (US$)  0.20893662  Money market interest rate (%; average)  0.20786781  Unemployment rate (%)  0.20565048  Growth in average wages (LCU; % pa)  0.07666478  Lending interest rate (%)  0.07076749  Goods: exports (US$)  0.03895238  Goods: imports (US$)  0.03895238  Nominal GDP (US$)  0.03895238  Stock of money M2 (US$)  0.03756122 |
| --- |

*Figure 38: UK CART Variable Importance*