

# Project Report

*Yichu Yan*

*11/20/2019*

## Factors affecting Euro exchange rates in different EU countries

### Background

Euro can be used in 19 EU countries: Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Portugal, Slovakia, Slovenia and Spain. But exchange rates of Euro to Dollar varies in these countries and are affected by Forex trades. In this way, I set project topic as investigating factors affecting Euro exchange rates in EU countries.

### Data collection

I collected 2000-2018 exchange rates, GDP, interest rates, and trade of goods data sheets of 10 EU countries (Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Spain) on the International Monetary Fund website and organized data in Excel to make it readable. In order to arrange data easily, I renamed variables with abbreviations while reading into RStudio.

Time: "Time"

Country: "Country"

RealER: "Real Effective Exchange Rate, based on Consumer Price Index"

GDP: "Gross Domestic Product, Nominal, Domestic Currency"

HCE: "Household Consumption Expenditure, incl. NPISHs, Nominal, Domestic Currency"

GCE: "Government Consumption Expenditure, Nominal, Domestic Currency"

GFCF: "Gross Fixed Capital Formation, Nominal, Domestic Currency"

CinI: "Change in Inventories, Nominal, Domestic Currency"

EGS: "Exports of Goods and Services, Nominal, Domestic Currency"

IGS: "Imports of Goods and Services, Nominal, Domestic Currency"

GDPV: "Gross Domestic Product, Volume"

GDPD: "Gross Domestic Product, Deflator"

GDPN: "Gross Domestic Product, Nominal, Seasonally Adjusted, Domestic Currency"

HCE2: "Household Consumption Expenditure, incl. NPISHs, Nominal, Seasonally Adjusted, Domestic Currency"

GFCE: "Govenment Final Consumption Expenditure, Nominal, Seasonally adjusted, Domestic Currency"

GFCF2: "Gross Fixed Capital Formation, Nominal, Seasonally Adjusted, Domestic Currency"

CinI2: "Change in Inventories, Nominal, Seasonally Adjusted, Domestic Currency"

EGS2: "Exports of Goods and Services, Nominal, Seasonally Adjusted, Domestic Currency"

IGS2: "Imports of Goods and Services, Nominal, Seasonally Adjusted, Domestic Currency"

GDPV2: "Gross Domestic Product, Volume, Seasonally Adjusted"

GDPD2: "Gross Domestic Product, Deflator, Seasonally Adjusted"

GvmBonds: "Government Bonds"

HEAR: "Harmonized Euro Area Rates, Outstanding Amounts, Deposits, Households, Agreed Maturity, Up to 2 Years"

HEAR2: "Harmonized Euro Area Rates, Outstanding Amounts, Deposits, Non-Financial Corporations, Agreed Maturity, Up to 2 Years"

HEAR3: "Harmonized Euro Area Rates, Loans, Households, Consumer Credit and Other, Up to 1 Year"

HEAR4: "Harmonized Euro Area Rates, New Business, Loans, Households, Consumption, Floating Rate and

up to 1 Year”

HEAR5: “Harmonized Euro Area Rates, Loans, Households, House Purchase, Over 5 Years”

GDPV8: “Harmonized Euro Area Rates, Loans, Non-Financial Corporations, Up to 1 Year”

ExportUS: “Goods, Value of Exports, US Dollars”

ExportNC: “Goods, Value of Exports, National Currency”

ImportUS: “Goods, Value of Imports, CIF, US Dollars”

ImportNC: “Goods, Value of Imports, CIF, National Currency”

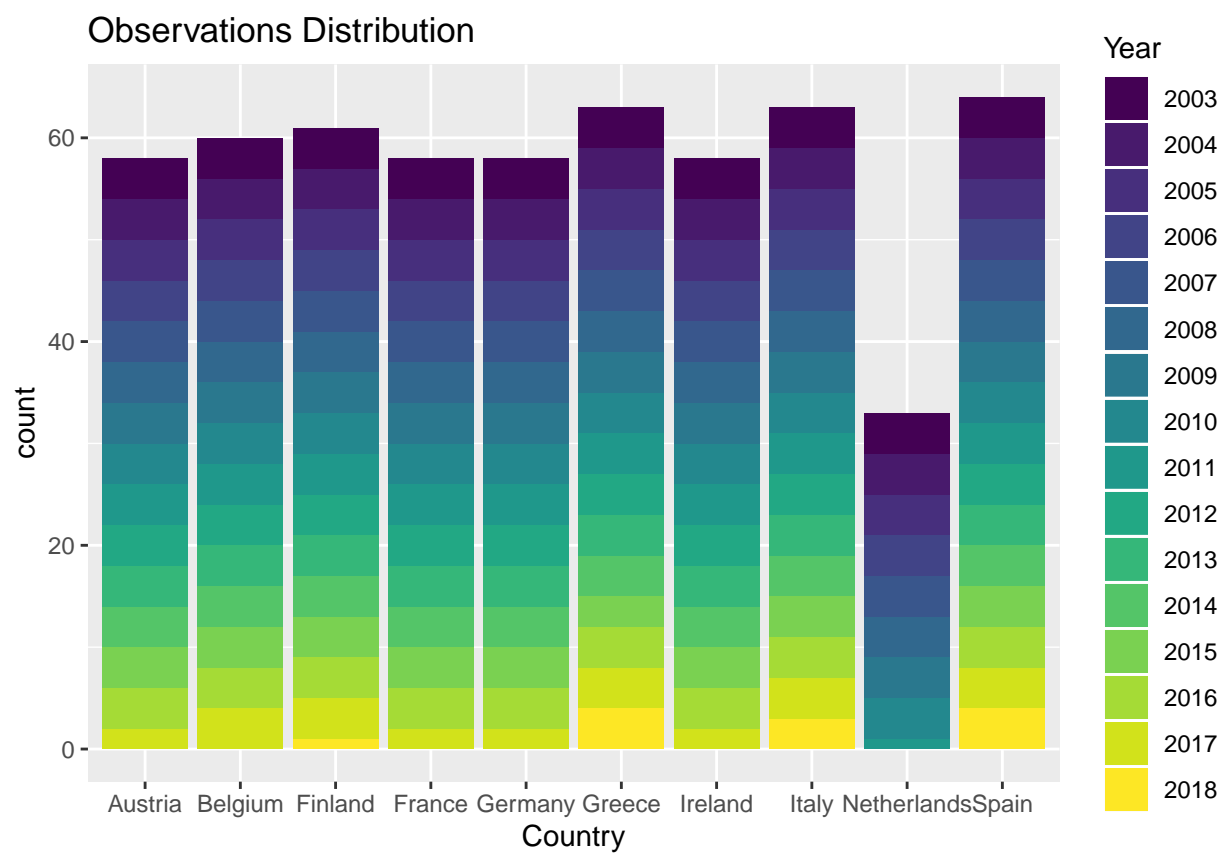
## Data cleaning

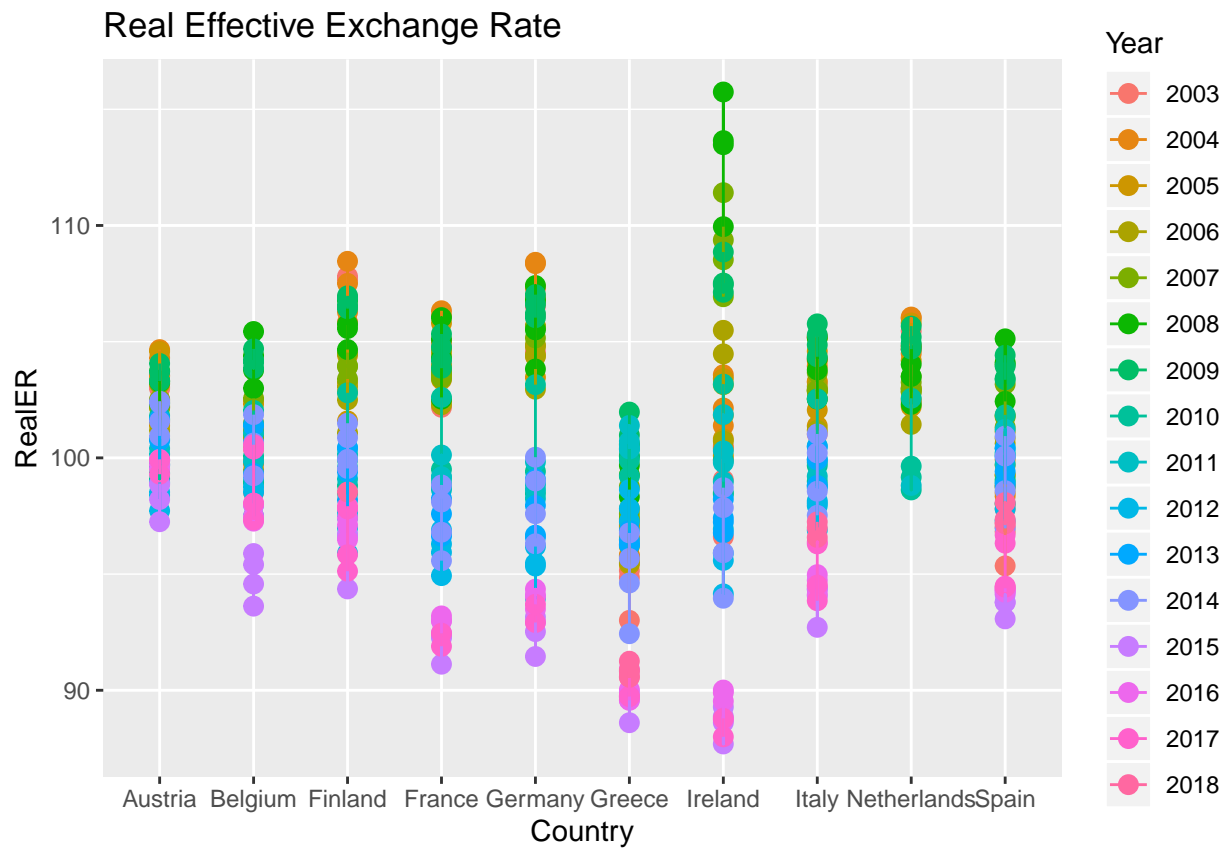
I summarized numbers of missing values in each variable and decided to remove such NAs from dataset.

```
## $vars_num_with_NA
##   variable q_na      p_na
## 1  GvmBonds   31 0.040789474
## 2    HEAR    120 0.157894737
## 3   HEAR2    120 0.157894737
## 4   HEAR3    120 0.157894737
## 5   HEAR4    152 0.200000000
## 6   HEAR5    120 0.157894737
## 7   GDPV8    120 0.157894737
## 8  ExportUS     4 0.005263158
## 9  ExportNC     4 0.005263158
## 10 ImportUS     4 0.005263158
## 11 ImportNC     4 0.005263158
##
## $vars_cat_with_NA
## [1] variable q_na      p_na
## <0 rows> (or 0-length row.names)
##
## $vars_cat_high_card
##   variable unique
## 1    Time       76
##
## $MAX_UNIQUE
## [1] 35
##
## $vars_one_value
## character(0)
##
## $vars_cat
## [1] "Time"      "Country"
##
## $vars_num
## [1] "RealER" "GDP"    "HCE"    "GCE"    "GFCF"   "CinI"
## [7] "EGS"    "IGS"    "GDPV"   "GDPD"   "GDPN"   "HCE2"
## [13] "GFCE"   "GFCF2"  "CinI2"  "EGS2"   "IGS2"   "GDPV2"
## [19] "GDPD2"  "GvmBonds" "HEAR"   "HEAR2"  "HEAR3"  "HEAR4"
## [25] "HEAR5"  "GDPV8"   "ExportUS" "ExportNC" "ImportUS" "ImportNC"
##
## $vars_char
## [1] "Time"      "Country"
##
## $vars_factor
## character(0)
##
## $vars_other
## character(0)
```

# EDA

## Data plots



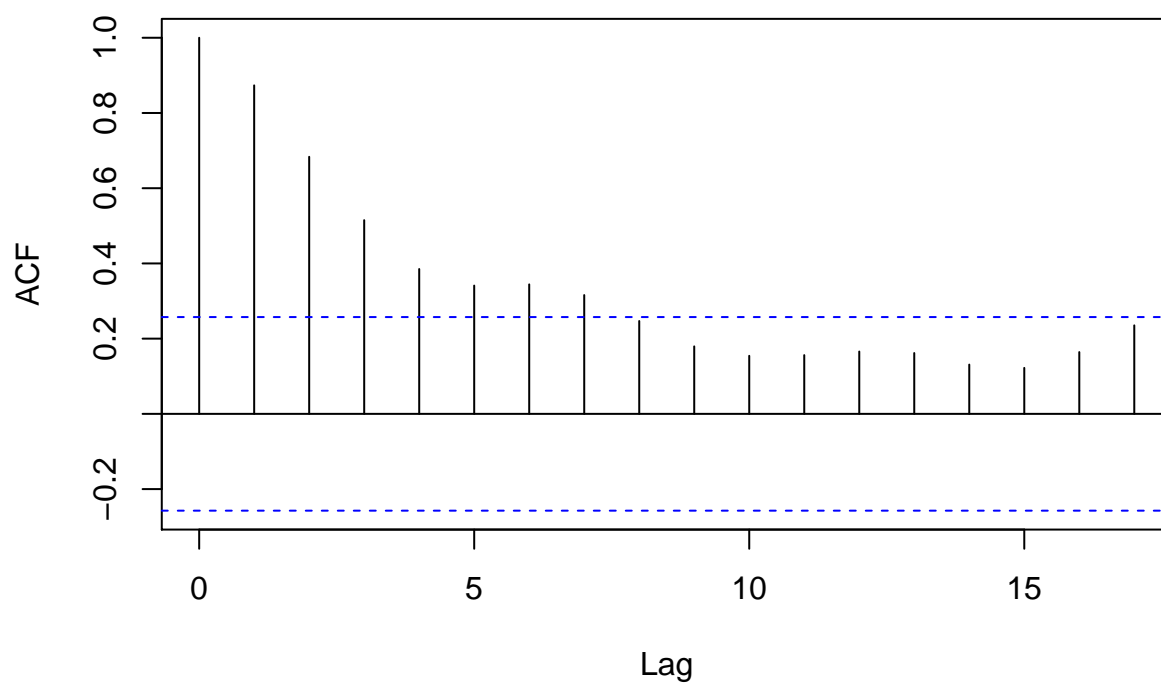


Time frame of cleaned data ranges from 2003 to 2018, although some factors related to exchange rates, GDP, interest rates, and trade of goods contains missing values in original data, after cleaning data, 10 countries still have plenty observations to fit models.

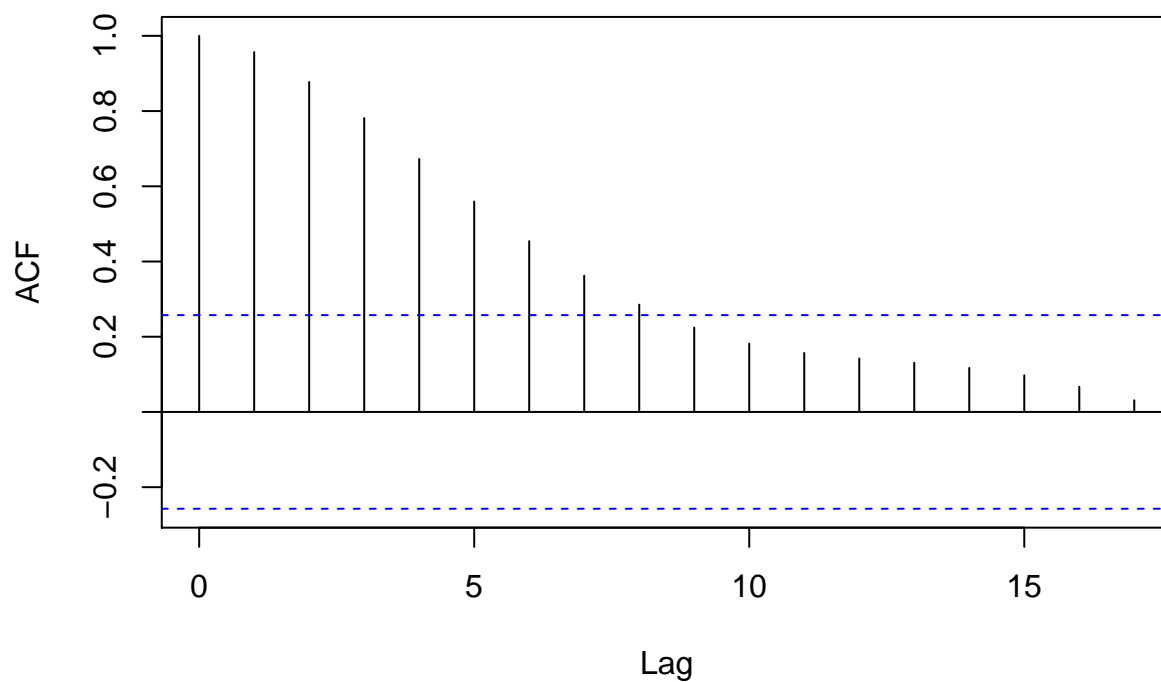
## Concerns

Check stationary or non-stationary

**Series data1\$RealER**



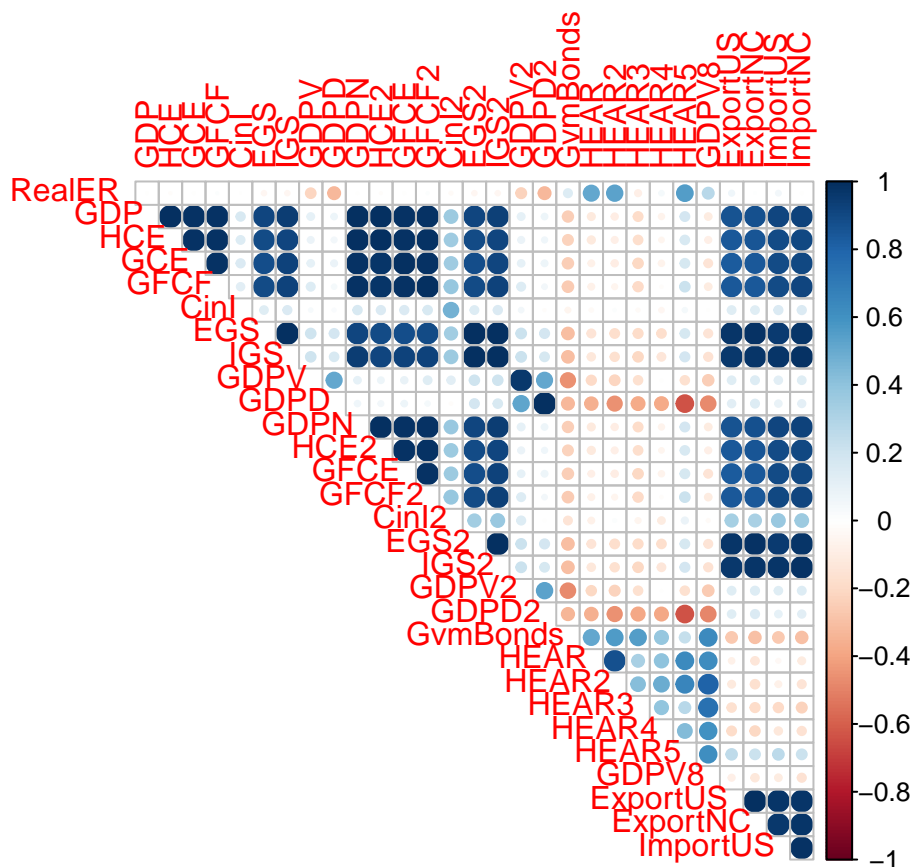
**Series data1\$HEAR2**



Non-stationary variables would give models large R square and makes estimates unreliable. Original data is time series data, even though I mixed 10 countries' data and deleted observations containing missing values, I still need to check ACF before and avoid using non-stationary variables to fit regressions, here are 2 examples.

## Correlation

##	Time	Country	RealER	GDP	HCE	GCE	GFCF	CinI
##	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
##	EGS	IGS	GDPV	GDPD	GDPN	HCE2	GFCE	GFCF2
##	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
##	CinI2	EGS2	IGS2	GDPV2	GDPD2	GvmBonds	HEAR	HEAR2
##	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
##	HEAR3	HEAR4	HEAR5	GDPV8	ExportUS	ExportNC	ImportUS	ImportNC
##	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

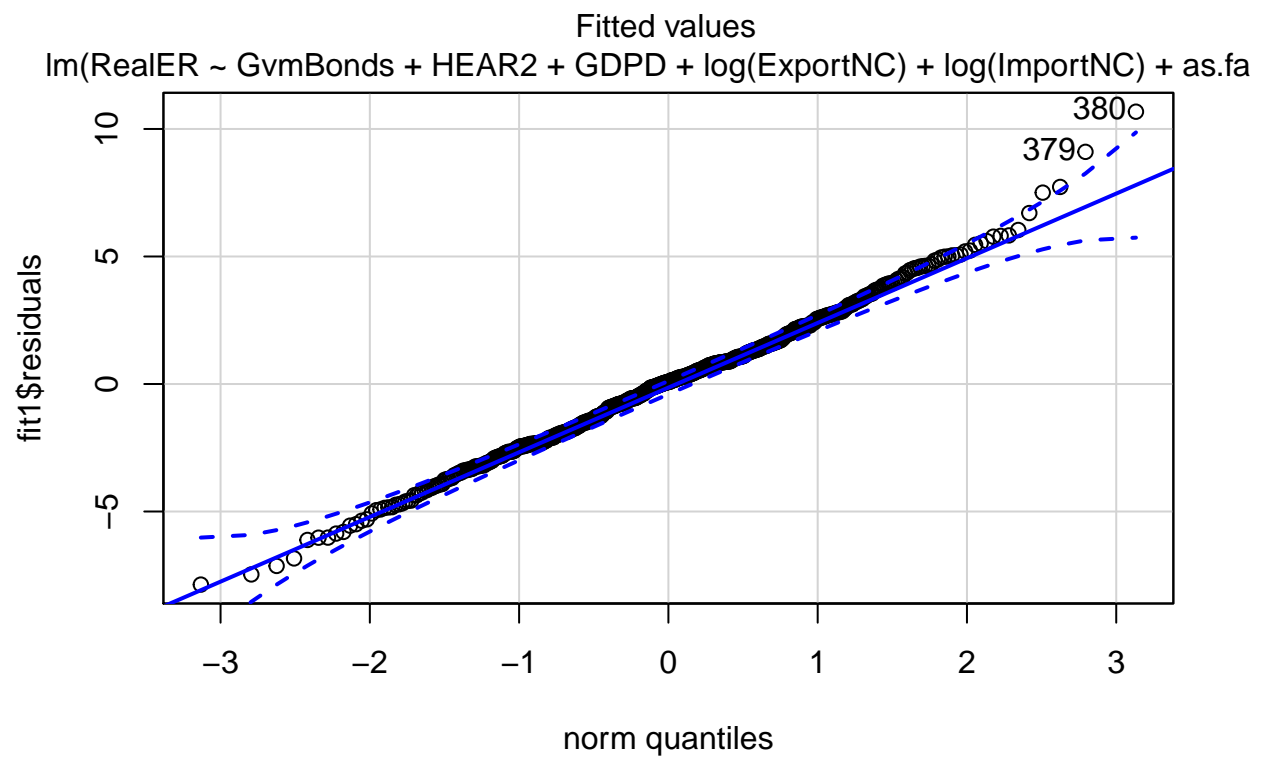
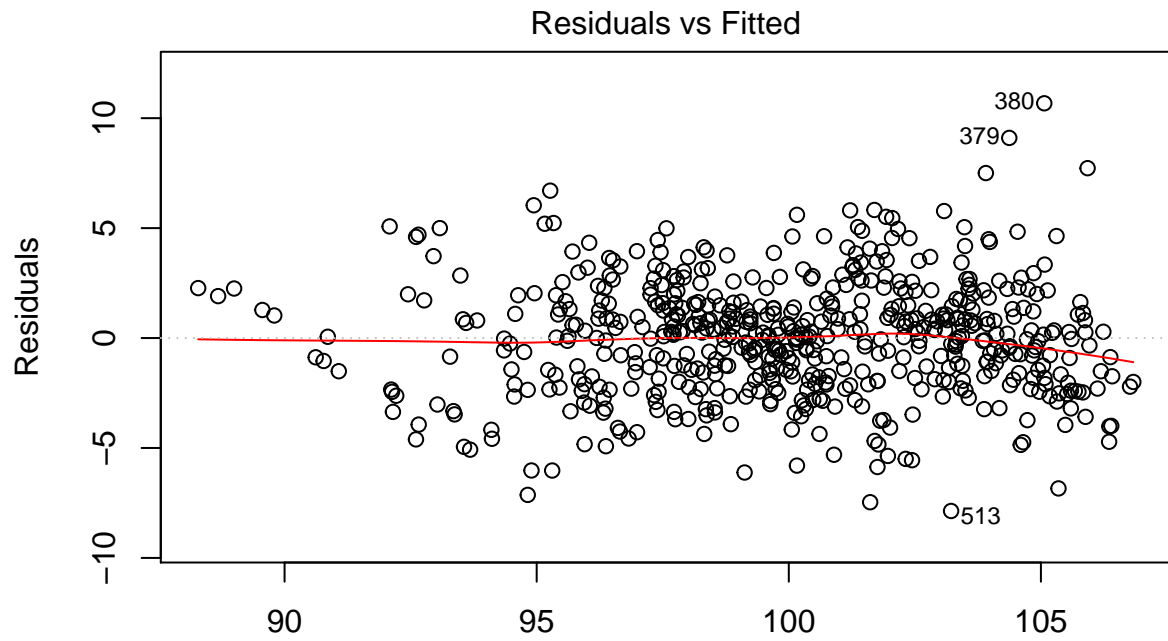


Some variables like ExportUS and ExportNC have high correlations because they are both used to describe values of export. In this way, we need to consider only use part of them in models.

## Models

```
##
## Call:
## lm(formula = RealER ~ GvmBonds + HEAR2 + GDPD + log(ExportNC) +
##     log(ImportNC) + as.factor(Country), data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.8652 -1.8523  0.0819  1.5686 10.6769
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    206.63215     9.54856   21.640 < 2e-16 ***
## GvmBonds         0.09868     0.07185    1.373  0.17020
## HEAR2           2.02171     0.13274   15.231 < 2e-16 ***
## GDPD            0.09900     0.03030    3.267  0.00115 **
## log(ExportNC)   -5.83821     1.08333   -5.389 1.04e-07 ***
## log(ImportNC)   -5.77247     1.40457   -4.110 4.55e-05 ***
## as.factor(Country)Belgium    10.64748     1.22079     8.722 < 2e-16 ***
## as.factor(Country)Finland    -8.42551     1.01820    -8.275 9.41e-16 ***
## as.factor(Country)France     13.32445     1.64349     8.107 3.26e-15 ***
## as.factor(Country)Germany    23.47793     2.28308    10.283 < 2e-16 ***
## as.factor(Country)Greece    -21.87098     1.45985   -14.982 < 2e-16 ***
## as.factor(Country)Ireland    -8.23730     1.08848    -7.568 1.57e-13 ***
## as.factor(Country)Italy      11.28357     1.38093     8.171 2.04e-15 ***
## as.factor(Country)Netherlands 11.70951     1.30216     8.992 < 2e-16 ***
## as.factor(Country)Spain       5.10566     1.06488     4.795 2.09e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.649 on 561 degrees of freedom
## Multiple R-squared:  0.6559, Adjusted R-squared:  0.6473
## F-statistic: 76.39 on 14 and 561 DF,  p-value: < 2.2e-16
```





```
## [1] 380 379
```

According to the residual plot, our model complies with the assumptions of normality and constant variance, so there is no issue about violating the model assumptions.

## Interpretation

The intercept is the predicted RealER in Austria if *GvmBonds*, *HEAR2*, *GDPD*,  $\log(\text{ExportNC})$  and  $\log(\text{ImportNC})$  are all equal to zero. Because these factors are never close to zero, the intercept has no direct interpretation.

The coefficient for *GvmBonds* is the predicted difference in RealER corresponding to a 1 unit difference in *GvmBonds*, if other variables equal to zero. Thus, the estimated predictive difference per unit of *GvmBonds* is 0.09868 for each country.

The coefficient for *HEAR2* is the predicted difference in RealER corresponding to a 1 unit difference in *HEAR2*, if other variables equal to zero. Thus, the estimated predictive difference per unit of *HEAR2* is 2.02171 for each country.

The coefficient for *GDPD* is the predicted difference in RealER corresponding to a 1 unit difference in *GDPD*, if other variables equal to zero. Thus, the estimated predictive difference per unit of *GDPD* is 0.09900 for each country.

The coefficient for  $\log(\text{ExportNC})$  is the predicted difference in RealER corresponding to a 1 unit difference in  $\log(\text{ExportNC})$ , if other variables equal to zero. Thus, the estimated predictive difference per unit of  $\log(\text{ExportNC})$  is 5.83821 decrease for each country.

The coefficient for  $\log(\text{ImportNC})$  is the predicted difference in RealER corresponding to a 1 unit difference  $\log(\text{ImportNC})$ , if other variables equal to zero. Thus, the estimated predictive difference per unit of  $\log(\text{ImportNC})$ s is 5.77247 decrease for each country.

Input variable that is used in these models is country, which is defined on a ten-point ordered scale:

Country = Austria: observations from Austria  
Country = Belgium: observations from Belgium  
Country = Finland: observations from Finland  
Country = France: observations from France  
Country = Germany: observations from Germany  
Country = Greece: observations from Greece  
Country = Ireland: observations from Ireland  
Country = Italy: observations from Italy  
Country = Netherlands: observations from Netherlands  
Country = Spain: observations from Spain

This parameterization of the model allows for different real exchange rate corresponding to each category of country. The “baseline” category (Country = Austria) corresponds to country Austria; the average real exchange rate for Austria is estimated by the intercept,  $206.63215 + 0.09868 \times GvmBonds + 2.02171 \times HEAR2 + 0.09900 \times GDPD - 5.83821 \times \log(ExportNC) - 5.77247 \times \log(ImportNC)$ .

The average real exchange rate in the other countries is found by adding the corresponding coefficient to this baseline average. This parameterization allows us to see that the real exchange rate in Germany achieve the highest average real exchange rate,  $206.63215 + 0.09868 \times GvmBonds + 2.02171 \times HEAR2 + 0.09900 \times GDPD - 5.83821 \times \log(ExportNC) - 5.77247 \times \log(ImportNC) + 10.64748$ .

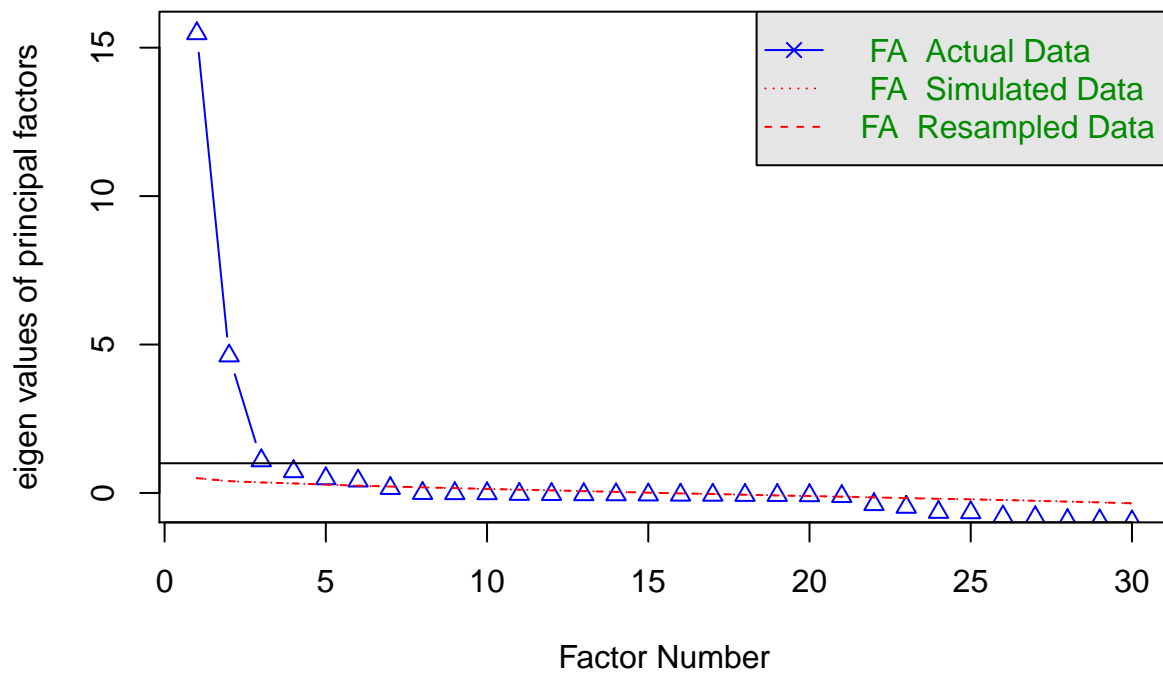
## Appendix

### Methods

#### Parallel Analysis

Dataset has 32 variables and I want to find out the number of factors that will be selected for later analysis.

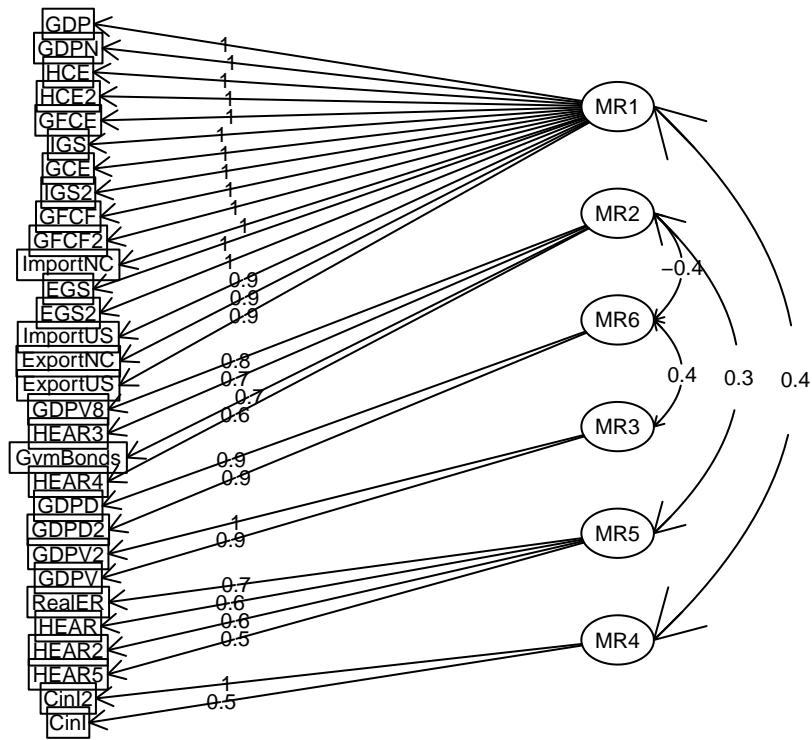
#### Parallel Analysis Scree Plots



## Parallel analysis suggests that the number of factors = 6 and the number of components = NA

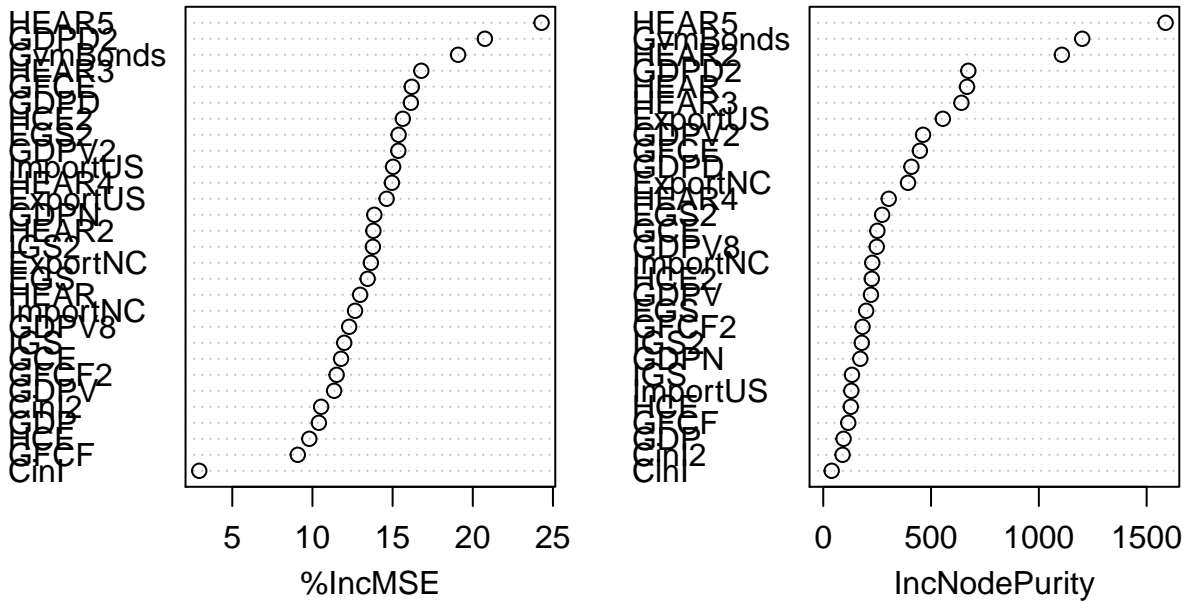
The blue line shows eigenvalues of actual data and the two red lines (placed on top of each other) show simulated and resampled data. Here we look at the large drops in the actual data and spot the point where it levels off to the right. Also we locate the point of inflection – the point where the gap between simulated data and actual data tends to be minimum.

## Factor Analysis

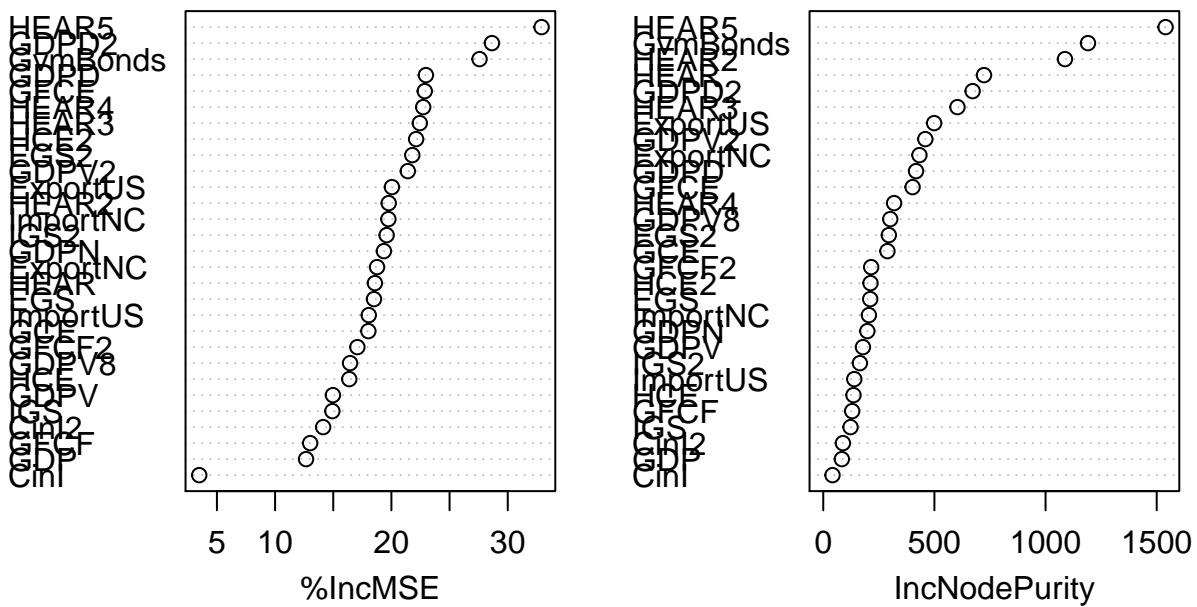


## Random Forest

model1



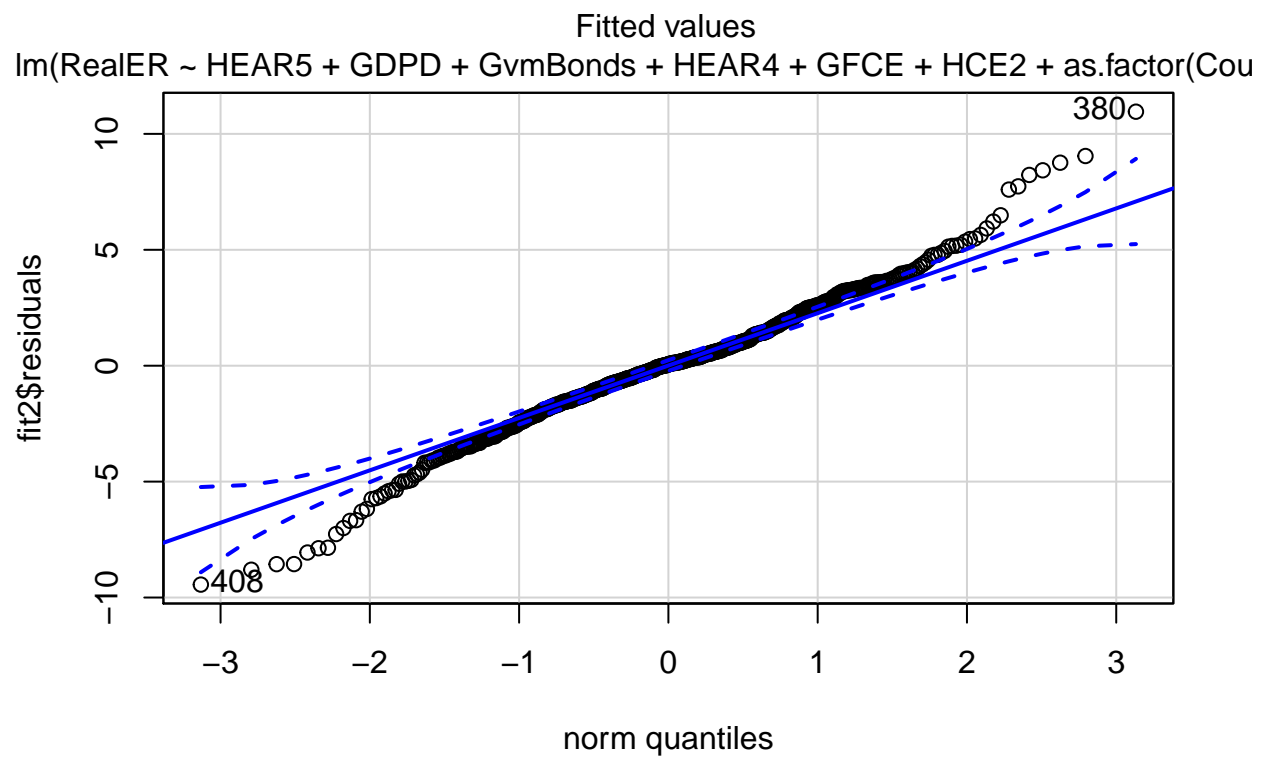
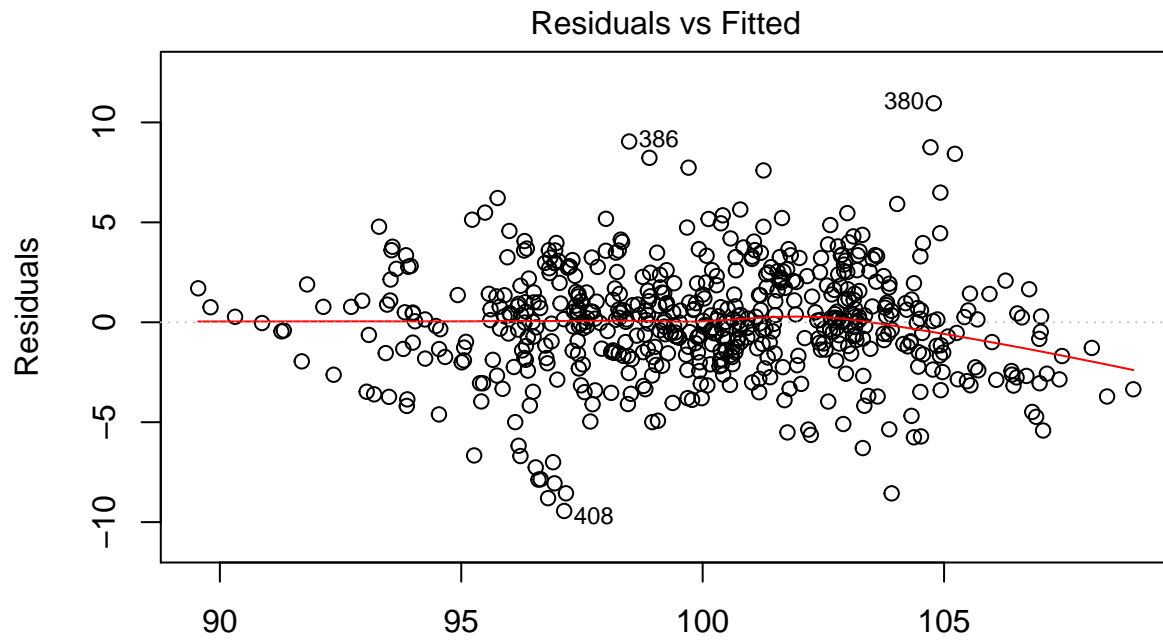
model2



Try to fit models using top factors

## Random Forest

```
##
## Call:
## lm(formula = RealER ~ HEAR5 + GDPD + GvmBonds + HEAR4 + GFCE +
##      HCE2 + as.factor(Country), data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4414 -1.5185  0.0602  1.5300 10.9569
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.016e+01  3.377e+00  23.734 < 2e-16 ***
## HEAR5          2.743e+00  1.913e-01  14.344 < 2e-16 ***
## GDPD           1.420e-01  3.124e-02   4.545 6.74e-06 ***
## GvmBonds       4.045e-01  5.691e-02   7.107 3.62e-12 ***
## HEAR4         -3.176e-01  9.558e-02  -3.323 0.00095 ***
## GFCE          -2.137e-05  5.752e-05  -0.372 0.71032
## HCE2          -5.847e-05  2.941e-05  -1.988 0.04726 *
## as.factor(Country)Belgium -1.390e+00  5.475e-01  -2.539 0.01138 *
## as.factor(Country)Finland  1.668e+00  5.962e-01   2.798 0.00532 **
## as.factor(Country)France   1.377e+01  2.543e+00   5.415 9.10e-08 ***
## as.factor(Country)Germany  1.814e+01  4.114e+00   4.411 1.24e-05 ***
## as.factor(Country)Greece  -6.541e+00  6.870e-01  -9.522 < 2e-16 ***
## as.factor(Country)Ireland -2.803e+00  6.163e-01  -4.548 6.63e-06 ***
## as.factor(Country)Italy    1.220e+01  2.931e+00   4.163 3.63e-05 ***
## as.factor(Country)Netherlands 1.963e+00  8.159e-01   2.406 0.01646 *
## as.factor(Country)Spain    7.416e+00  1.760e+00   4.212 2.94e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.775 on 560 degrees of freedom
## Multiple R-squared:  0.6231, Adjusted R-squared:  0.613
## F-statistic: 61.71 on 15 and 560 DF,  p-value: < 2.2e-16
```



```
## [1] 380 408
```

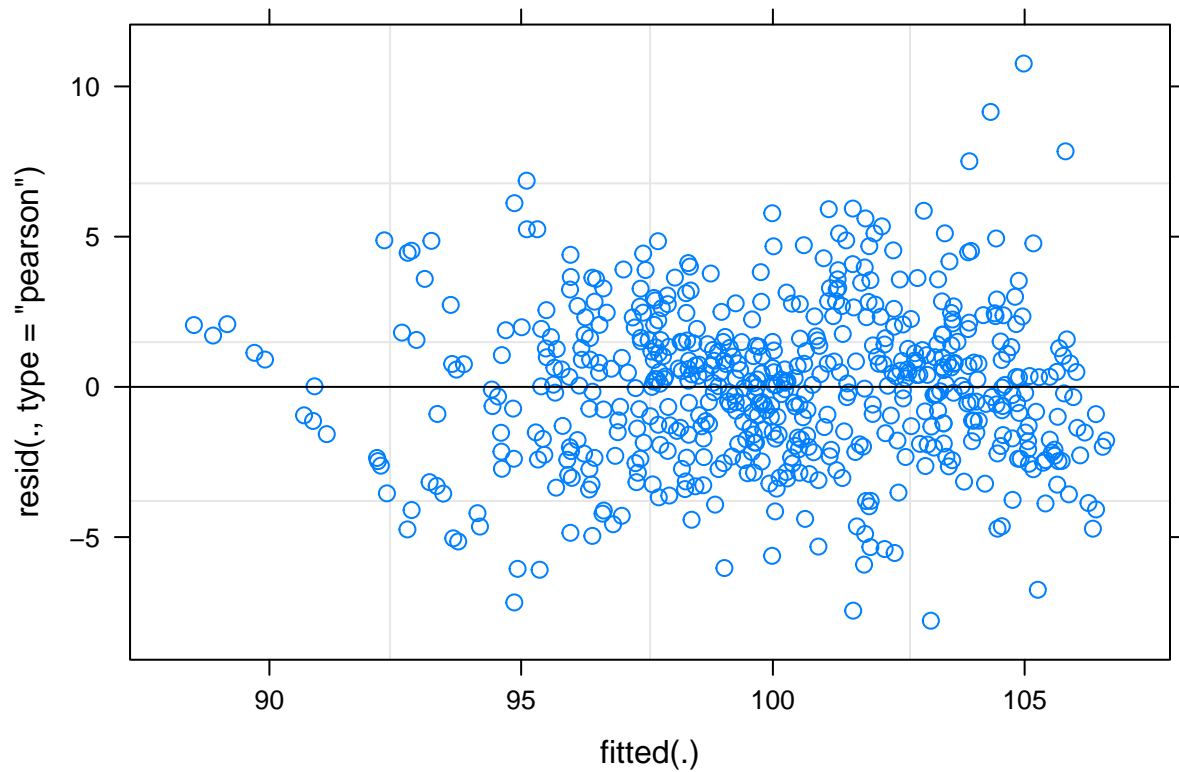
I try to fit a linear model using variables selected from random forest. Although results show most coefficients are significant, but R squared is not high. Residual plot and QQ plot also show this model is not good.

## Multilevel model varying across countreis

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
##
## Registered S3 methods overwritten by 'lme4':
##   method                                  from
##   cooks.distance.influence.merMod car
##   influence.merMod                    car
##   dfbeta.influence.merMod             car
##   dfbetas.influence.merMod            car
##
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ReaLER ~ GvmBonds + HEAR2 + GDPD + log(ExportNC) + log(ImportNC) +
##   (1 | Country)
##   Data: mydata
##
## REML criterion at convergence: 2827
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.9370 -0.6875  0.0257  0.5744  4.0608
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   Country (Intercept) 157.762   12.56
##   Residual              7.023    2.65
## Number of obs: 576, groups: Country, 10
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  202.72077   10.54146   19.231
## GvmBonds      0.10501    0.07171    1.464
## HEAR2         2.00401    0.13250   15.125
## GDPD          0.08080    0.02968    2.722
## log(ExportNC) -5.51714    1.07656   -5.125
## log(ImportNC) -5.21413    1.38543   -3.764
##
## Correlation of Fixed Effects:
##              (Intr) GvmBnd HEAR2  GDPD   1(ENC)
## GvmBonds     -0.245
## HEAR2         0.159 -0.608
## GDPD          0.548 -0.134  0.380
## lg(ExprtNC)  -0.146 -0.424  0.301 -0.196
## lg(ImprtNC)  -0.602  0.522 -0.429 -0.439 -0.630
##
## $Country
##              (Intercept) GvmBonds HEAR2 GDPD log(ExportNC)
## Austria      199.4545 0.1050081 2.00401 0.08079799 -5.517144
## Belgium      209.2189 0.1050081 2.00401 0.08079799 -5.517144
```



```
## Finland      191.7158 0.1050081 2.00401 0.08079799 -5.517144
## France       211.5679 0.1050081 2.00401 0.08079799 -5.517144
## Germany      221.1603 0.1050081 2.00401 0.08079799 -5.517144
## Greece       178.4808 0.1050081 2.00401 0.08079799 -5.517144
## Ireland      191.7775 0.1050081 2.00401 0.08079799 -5.517144
## Italy         209.7238 0.1050081 2.00401 0.08079799 -5.517144
## Netherlands  210.2352 0.1050081 2.00401 0.08079799 -5.517144
## Spain        203.8730 0.1050081 2.00401 0.08079799 -5.517144
##              log(ImportNC)
## Austria      -5.214128
## Belgium      -5.214128
## Finland      -5.214128
## France       -5.214128
## Germany      -5.214128
## Greece       -5.214128
## Ireland      -5.214128
## Italy         -5.214128
## Netherlands  -5.214128
## Spain        -5.214128
##
## attr(,"class")
## [1] "coef.mer"
```

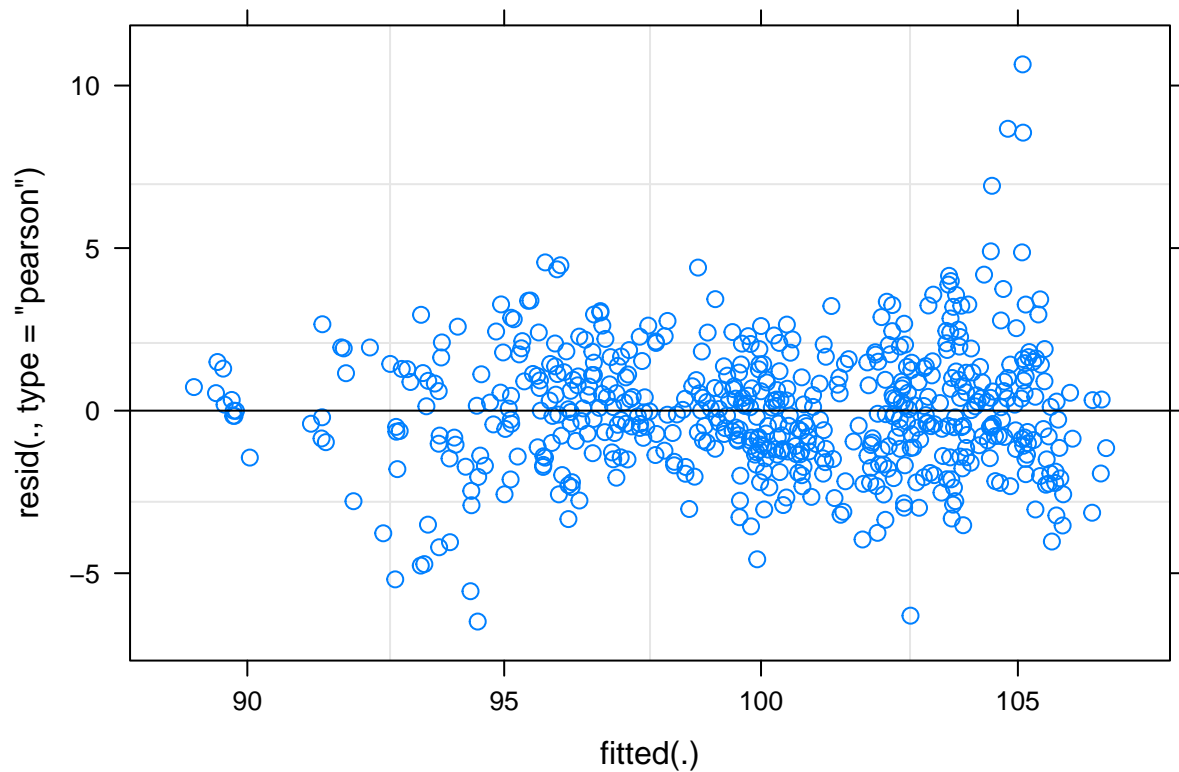


I set different country groups and try to fit a multilevel model. From the summary of this regression, I found one important predictor GvmBonds is not significant even though residual plot looks good.

## Multilevel model varying across countreis and years

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## ReaLER ~ GvmBonds + HEAR2 + GDPD + log(ExportUS) + (1 | Country) +
##      (1 | Year)
##      Data: mydata
##
## REML criterion at convergence: 2545.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2548 -0.5951 -0.0372  0.5693  5.3438
##
## Random effects:
##      Groups      Name      Variance Std.Dev.
##      Year      (Intercept) 16.753    4.093
##      Country  (Intercept)  2.486    1.577
##      Residual                3.972    1.993
## Number of obs: 576, groups:  Year, 16; Country, 10
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  59.64456    5.43891  10.966
## GvmBonds      0.32214    0.05887   5.472
## HEAR2         0.75540    0.22259   3.394
## GDPD          0.25088    0.02952   8.499
## log(ExportUS) 1.12898    0.41069   2.749
##
## Correlation of Fixed Effects:
##              (Intr) GvmBnd HEAR2  GDPD
## GvmBonds      0.001
## HEAR2        -0.044 -0.617
## GDPD         -0.517  0.106 -0.054
## lg(ExprrtUS) -0.811 -0.058  0.022 -0.031
##
## $Year
##      (Intercept) GvmBonds      HEAR2      GDPD log(ExportUS)
## 2003    64.34552  0.322137  0.7554043  0.2508818    1.128978
## 2004    65.46035  0.322137  0.7554043  0.2508818    1.128978
## 2005    63.93520  0.322137  0.7554043  0.2508818    1.128978
## 2006    62.40679  0.322137  0.7554043  0.2508818    1.128978
## 2007    62.14088  0.322137  0.7554043  0.2508818    1.128978
## 2008    62.50438  0.322137  0.7554043  0.2508818    1.128978
## 2009    64.53403  0.322137  0.7554043  0.2508818    1.128978
## 2010    60.03119  0.322137  0.7554043  0.2508818    1.128978
## 2011    58.58768  0.322137  0.7554043  0.2508818    1.128978
## 2012    55.40544  0.322137  0.7554043  0.2508818    1.128978
## 2013    57.71314  0.322137  0.7554043  0.2508818    1.128978
## 2014    58.24975  0.322137  0.7554043  0.2508818    1.128978
## 2015    53.08257  0.322137  0.7554043  0.2508818    1.128978
## 2016    54.38189  0.322137  0.7554043  0.2508818    1.128978
## 2017    54.74886  0.322137  0.7554043  0.2508818    1.128978
## 2018    56.78527  0.322137  0.7554043  0.2508818    1.128978
##
```

```
## $Country
##          (Intercept) GvmBonds    HEAR2    GDPD log(ExportUS)
## Austria      61.71834 0.322137 0.7554043 0.2508818    1.128978
## Belgium      59.79074 0.322137 0.7554043 0.2508818    1.128978
## Finland      62.62162 0.322137 0.7554043 0.2508818    1.128978
## France        59.01609 0.322137 0.7554043 0.2508818    1.128978
## Germany       59.08374 0.322137 0.7554043 0.2508818    1.128978
## Greece        57.60827 0.322137 0.7554043 0.2508818    1.128978
## Ireland       58.49511 0.322137 0.7554043 0.2508818    1.128978
## Italy          59.14478 0.322137 0.7554043 0.2508818    1.128978
## Netherlands   59.72650 0.322137 0.7554043 0.2508818    1.128978
## Spain         59.24040 0.322137 0.7554043 0.2508818    1.128978
##
## attr(,"class")
## [1] "coef.mer"
```



This model varying across countries and years and I got every variable significant. However, its AIC is higher than the linear model I selected.

## Discussion

According to the analysis using Linear Regression, we can conclude that different countries' Real Exchange Rates could be predicted by Government Bonds, Harmonized Euro Area Rates (Outstanding Amounts, Deposits, Non-Financial Corporations, Agreed Maturity, Up to 2 Years), Gross Domestic Product (Deflator), Goods Value of Exports in US Dollars and Goods Value of Import in US Dollars. And in Multilevel models, we choose to set up groups in different countries and different year and build linear regressions. Results of `coef()` show estimated model within each group. After comparing AIC of each model, we choose the linear regression with minimum AIC.