Final Data Analysis

1. Theory and research question

In this analysis, I aim to investigate the impact of European residents' identity characteristics and attitudes towards the economic role of immigrants on their perceptions of the European Parliament, utilizing data from the latest round of the European Social Survey (round 10, 2020).

Given the current significance of immigration policy in public discourse, I anticipate that certain contextual variables may interact with attitudes toward immigration within the European Social Survey dataset. These variables could include country of residence, level of education, household number, or nativity.

One plausible consideration is that native-born citizens may experience distinct interactions with immigrant residents, thereby influencing their perceptions. Consequently, the research question guiding this analysis is: How does being native-born affect the relationship between EU residents' attitudes towards the economic role of immigrants and their evaluations of trust in the European Parliament?

2. Hypotheses

The alternative hypothesis (H1) and null hypothesis (H0) are formulated as follows:

H1: The effect of attitudes towards the economic role of immigrants on the rating of trust in the European Parliament differs between native residents and immigrant residents.

H0: The effect of attitudes towards the economic role of immigrants on the rating of trust in the European Parliament is the same for both native residents and immigrant residents.

3. Data and variables

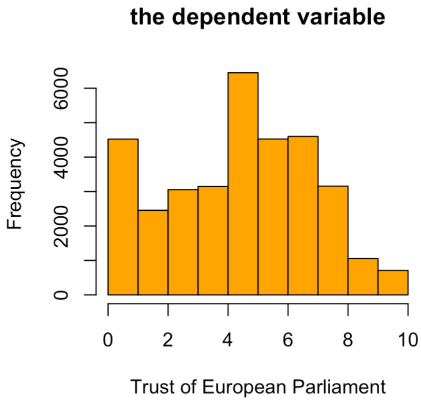
The dataset utilized in this analysis is derived from the latest iteration of the European Social Survey (round 10, 2020). We employed a simplified version of the dataset, encompassing country information, three categorical variables, and nine interval variables.

The dependent variable in our study is the European Parliament Trust rating, scaled from 0 to 10. The three categorical variables comprise gender, nativity, and whether respondents had experienced coronavirus.
The nine interval variables, along with their summary statistics, are detailed below:

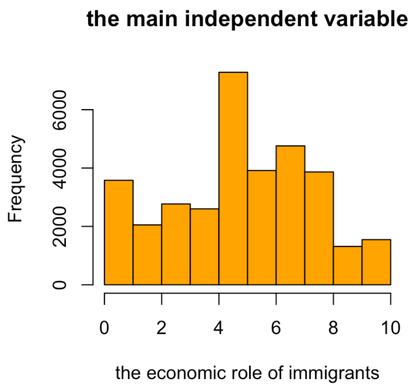
```
Summary Statistics
================================================================
Statistic    N     Mean Median St. Dev. Min Max Pctl(25) Pctl(75)
----------------------------------------------------------------
age        17,979 34.7    36     10.0    15  50     26        43
education  37,369 4.1     4      1.8      1   7     3         6
soctrust   37,490 5.0     5      2.5      0  10     3         7
news       15,040 21.2    30     13.6     0  50     10        30
```

parltrust	36,903	4.5	5	2.7	0	10	2	7
eptrust	35,366	4.8	5	2.6	0	10	3	7
lifesat	37,255	7.1	7	2.1	0	10	6	9
immig	36,553	5.2	5	2.6	0	10	3	7
household	37,467	2.5	2	1.3	1	13	2	3

Before proceeding with model specification, it is crucial to examine the distribution of these variables. The distributions of the dependent variable and the main independent variable are visualized below:

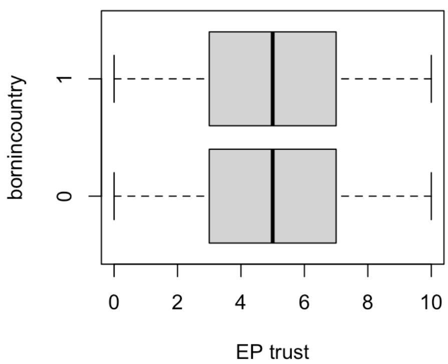


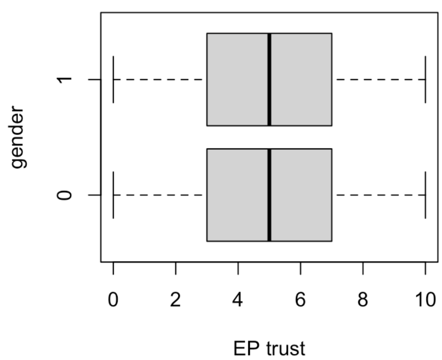

```
   0    1    2    3    4    5    6    7    8    9   10 
3056 1467 2456 3056 3150 6450 4525 4601 3157 1058  708 
```



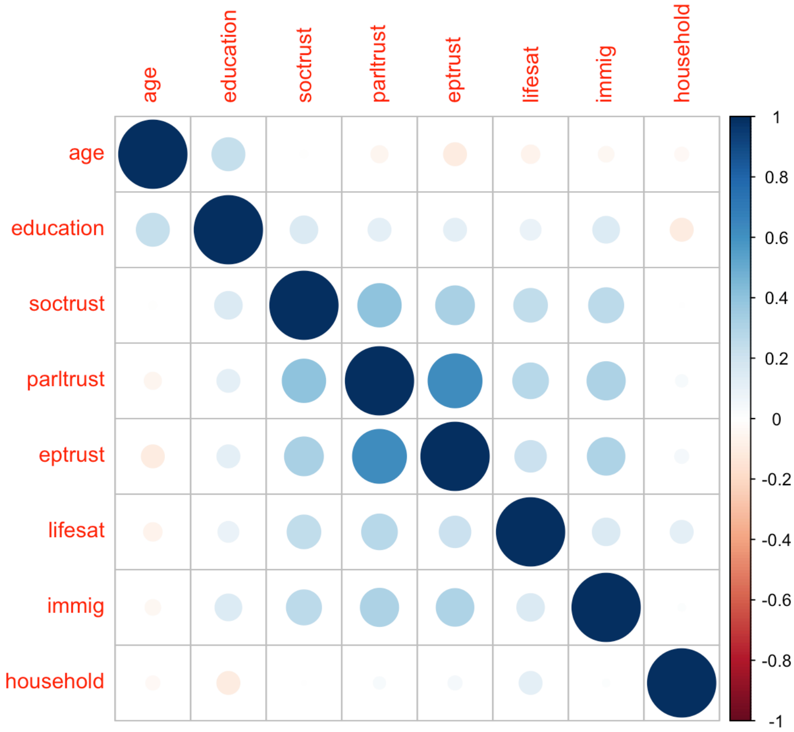

```
   0    1    2    3    4    5    6    7    8    9   10 
2422 1158 2051 2769 2602 7281 3915 4759 3866 1314 1547 
```

Boxplots are employed to illustrate the distribution of categorical variables such as "bornincountry" and "gender," as depicted below:





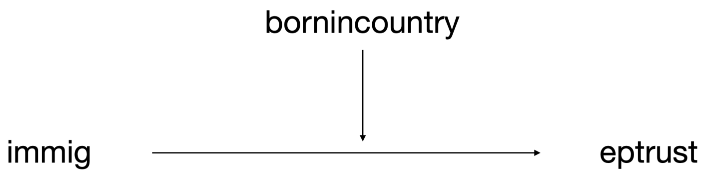
Furthermore, a comprehensive correlations plot is presented as a valuable reference for model specification:



This figure elucidates the correlation between independent and dependent variables, shedding light on potential multicollinearity issues.

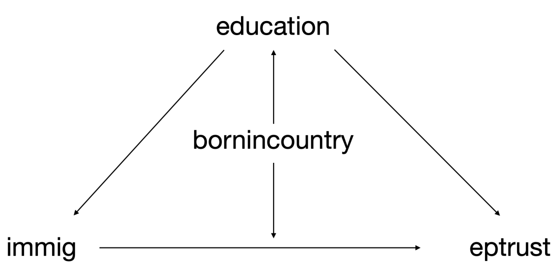
4. Model specification

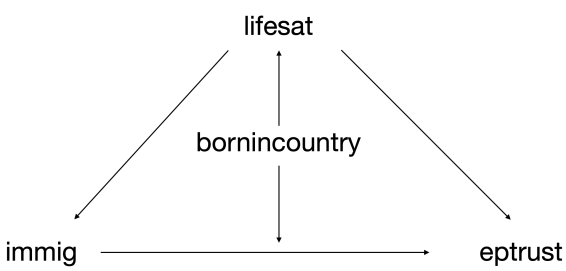
Our interest lies in exploring a conditional relationship represented by the equation:



$$\hat{y} = a + b_1 \cdot \text{bornincountry} + b_2 \cdot \text{immig} + b_3 \cdot \text{bornincountry} \cdot \text{immig} + e$$

Recognizing that education and life satisfaction are interconnected with variables such as “bornincountry”, “immig”, and “eptrust”, we refine the relationship to incorporate these factors:





$$\hat{y} = a + b_1 \cdot \text{bornincountry} + b_2 \cdot \text{immig} + b_3 \cdot \text{bornincountry} \cdot \text{immig} + b_4 \cdot \text{education} + b_5 \cdot \text{lifesat} + e$$

5. Descriptive Statistics

We undertake a stepwise model estimation approach to investigate the relationship between variables:

1. **Model 1:** This initial model begins without interaction and lacks additional control variables.
2. **Model 2:** The second model maintains the absence of interaction but incorporates relevant control variables.
3. **Model 3:** In the third model, interaction terms are introduced alongside control variables.

Summary Statistics for regressions

	Dependent variable: eptrust		
	(1)	(2)	(3)
bornincountry1	0.066	0.090*	-0.372***
	(0.051)	(0.050)	(0.138)
immig	0.275***	0.235***	0.167***
	(0.005)	(0.005)	(0.020)
education		0.079***	0.079***
		(0.008)	(0.008)
lifesat		0.198***	0.198***
		(0.006)	(0.006)
bornincountry1:immig			0.073***
			(0.020)
Constant	3.304***	1.754***	2.192***
	(0.059)	(0.075)	(0.144)
Observations	33,684	33,684	33,684
R2	0.077	0.107	0.107
Adjusted R2	0.077	0.106	0.107
Residual Std. Error	2.466 (df = 33681)	2.427 (df = 33679)	2.426 (df = 33678)
F Statistic	1,404.986*** (df = 2; 33681)	1,003.735*** (df = 4; 33679)	805.826*** (df = 5; 33678)
Note:			*p<0.1; **p<0.05; ***p<0.01

The rejection of the Null Hypothesis indicates a significant interaction between the "bornincountry" and "immig" variables.

Summary Statistics for regressions

	Dependent variable: eptrust			
	(1)	(2)	(3)	(4)
bornincountry1	-0.372***	-0.365***	-0.249*	0.108
	(0.138)	(0.138)	(0.136)	(0.119)
immig	0.167***	0.166***	0.149***	0.093***
	(0.020)	(0.020)	(0.019)	(0.017)
education	0.079***	0.078***	0.047***	0.031***
	(0.008)	(0.008)	(0.008)	(0.007)
lifesat	0.198***	0.193***	0.148***	0.035***
	(0.006)	(0.007)	(0.007)	(0.006)
household		0.064***		
		(0.010)		
soctrust			0.194***	
			(0.006)	
parltrust				0.505***
				(0.005)
bornincountry1:immig	0.073***	0.074***	0.050**	0.025
	(0.020)	(0.020)	(0.020)	(0.017)
Constant	2.192***	2.057***	1.793***	1.429***
	(0.144)	(0.145)	(0.142)	(0.124)
Observations	33,684	33,684	33,684	33,684
R2	0.107	0.108	0.138	0.342
Adjusted R2	0.107	0.108	0.138	0.342
Residual Std. Error	2.426 (df = 33678)	2.425 (df = 33677)	2.383 (df = 33677)	2.083 (df = 33677)
F Statistic	805.826*** (df = 5; 33678)	679.235*** (df = 6; 33677)	901.701*** (df = 6; 33677)	2,915.762*** (df = 6; 33677)
Note:				*p<0.1; **p<0.05; ***p<0.01

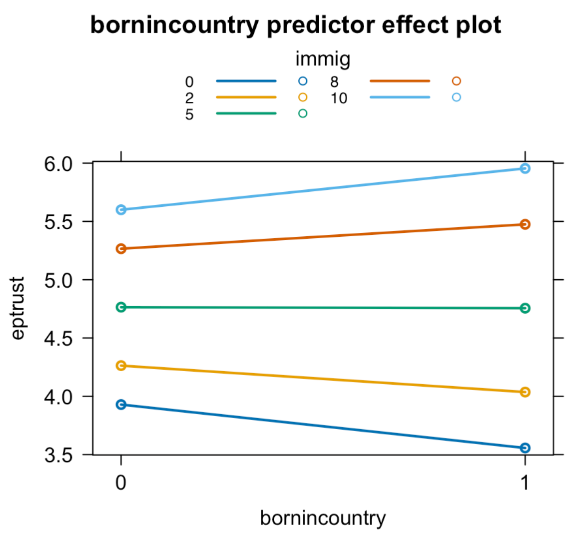
Subsequently, to further refine our understanding based on theoretical considerations, we conduct three additional regressions. These regressions introduce "household," "soctrust," and "parltrust" as potential controlling variables.

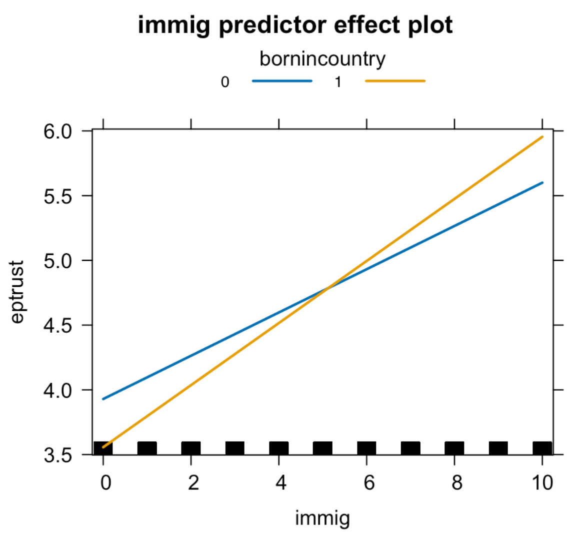
Among these additions, "household" emerges as a particularly significant variable in the model specification, demonstrating perfect statistical significance.

```
Call:
lm(formula = eptrust ~ bornincountry + immig + bornincountry *
    immig + education + lifesat + household, data = ess.clean)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8117 -1.6198  0.2298  1.6967  8.1023

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           2.056846   0.145088  14.177  < 2e-16 ***
bornincountry1       -0.365353   0.138394  -2.640 0.008295 **
immig                 0.165913   0.019633   8.451  < 2e-16 ***
education             0.077701   0.007662  10.141  < 2e-16 ***
lifesat               0.193475   0.006529  29.634  < 2e-16 ***
household             0.064252   0.009981   6.438 1.23e-10 ***
bornincountry1:immig  0.073882   0.020316   3.637 0.000277 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.425 on 33677 degrees of freedom
Multiple R-squared:  0.108,     Adjusted R-squared:  0.1078
F-statistic: 679.2 on 6 and 33677 DF,  p-value: < 2.2e-16
```

This approach allows us to systematically build upon our models, progressively incorporating variables and interactions to enhance the robustness of our analysis.

6. Model estimation



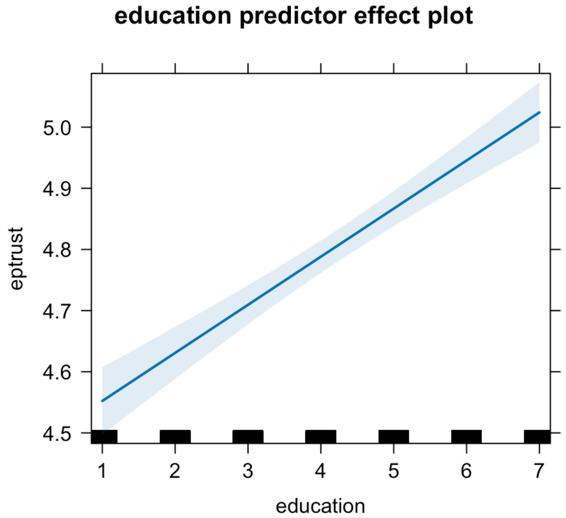


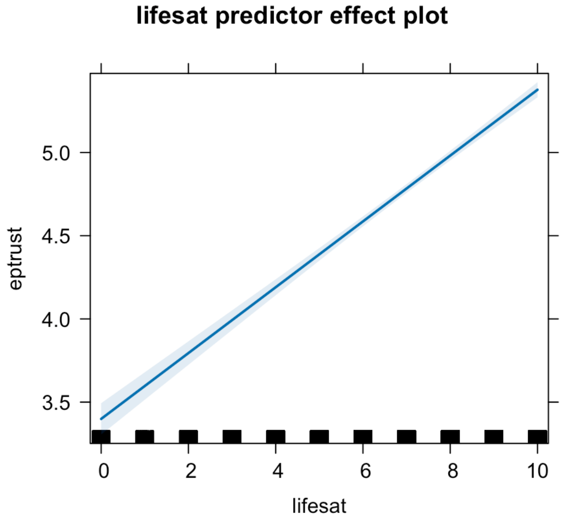
The graphical representation of our model estimation is presented in the following charts:

Left Chart: Interaction Effect for Native-Born Residents This chart illustrates a discernible interaction effect for individuals born in the resident country concerning their attitudes towards immigration.

Notably, there is a more pronounced impact on European Parliament (EP) trust scores, encompassing both elevated high scores and diminished low scores. This suggests a more nuanced and dynamic relationship between native-born residents and immigration attitudes.

Right Chart: Differential Coefficients for Immigrant and Native Residents In contrast, the right chart highlights divergent coefficients for immigrant residents and native residents. This divergence indicates that the influence of immigration attitudes on EP trust scores varies between these two groups. Understanding these distinct coefficients provides valuable insights into the nuanced nature of trust evaluations based on residency status and attitudes toward immigration.





Moreover, these two charts above demonstrate the impact of controlled variables, such as "education" and "lifesat." Both variables exhibit positive correlations with the dependent variable, further underscoring their role in shaping individuals' evaluations of trust in the European Parliament.

This integration offers a comprehensive overview of the main findings, including both interaction effects and the influence of controlled variables.

7. post-estimation diagnostics

In adherence to the Ordinary Least Squares (OLS) assumptions, a series of diagnostics were performed on the linear model.

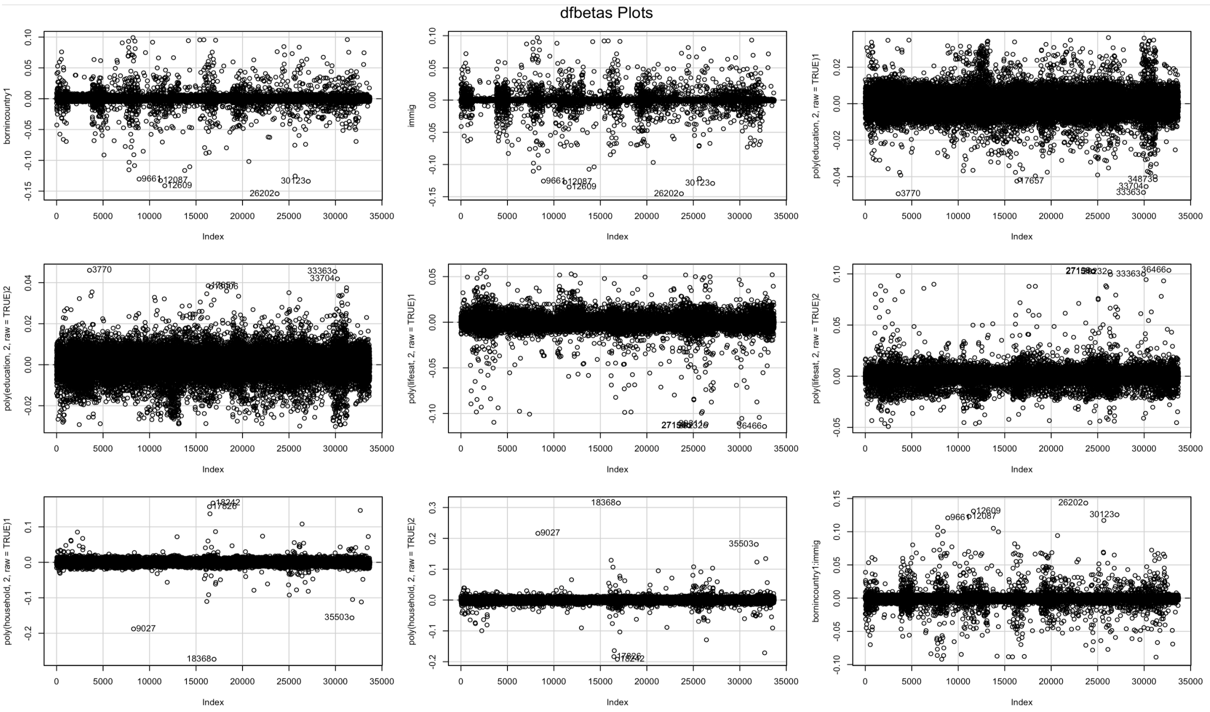
```
> residualPlots(reg_ess_4)
              Test stat Pr(>|Test stat|)
bornincountry
immig           -9.8346       < 2.2e-16 ***
education        3.4949       0.0004749 ***
lifesat        -10.1375       < 2.2e-16 ***
household       -2.9254       0.0034426 **
Tukey test      -4.8482       1.246e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Notably, variables marked with *** on the p-value prompted consideration of transformations to address fitting issues. Consequently, a quadratic transformation was introduced to enhance the model's fit:

```
> residualPlots(reg_ess_final)  # better fitting
                              Test stat Pr(>|Test stat|)
bornincountry
immig                           -8.3717          <2e-16 ***
poly(education, 2, raw = TRUE)
poly(lifesat, 2, raw = TRUE)
poly(household, 2, raw = TRUE)
Tukey test                       0.8855          0.3759
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

This transformation helps alleviate potential fitting problems associated with these variables.

Furthermore, to mitigate the influence of outliers, a careful examination was conducted, and the dfbetasPlots() function was employed to assess outliers across all variables simultaneously. Encouragingly, the graphs revealed that all |dfbetas| are less than 1, indicating the absence of outliers.

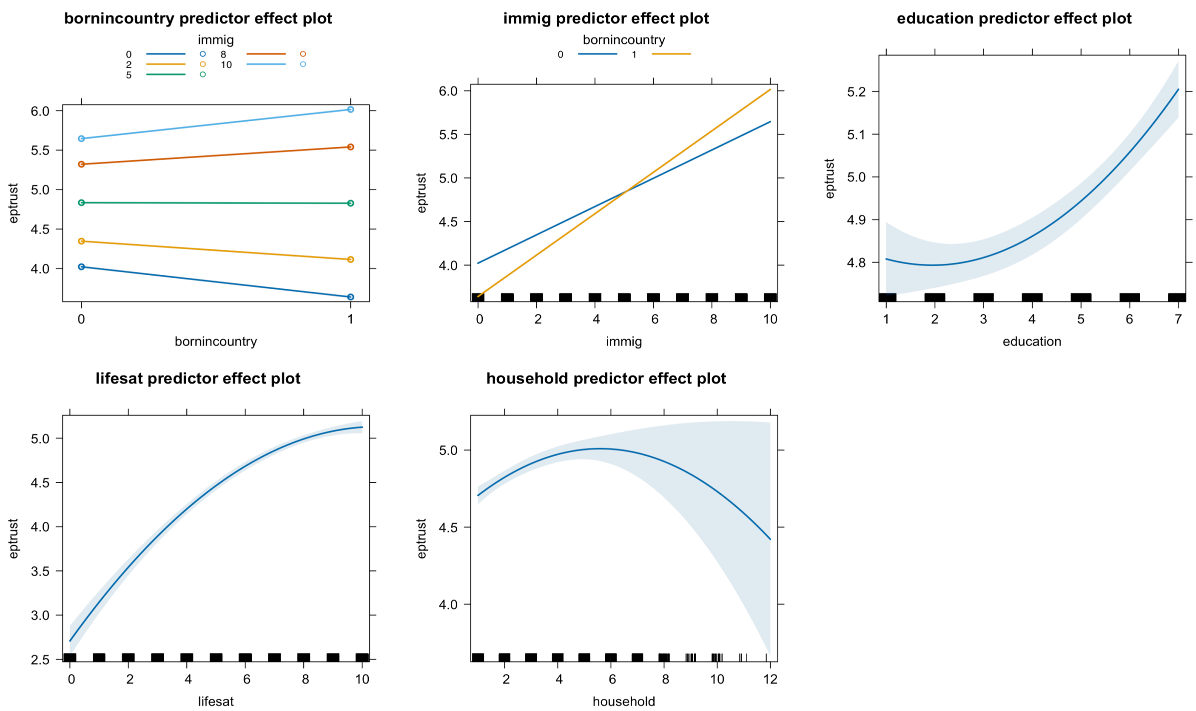


8. Reporting of results

Following adherence to OLS assumptions and comprehensive regression diagnostics, the summary statistics of the final linear regression are presented in the table below:

```
Call: lm(formula = eptrust ~ bornincountry + immig + bornincountry * immig + poly(education, 2, raw
         = TRUE) + poly(lifesat, 2, raw = TRUE) + poly(household, 2, raw = TRUE), data = ess.clean.out)

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     1.516605   0.173611   8.736  < 2e-16 ***
bornincountry1                 -0.382837   0.138204  -2.770 0.005607 **
immig                           0.162315   0.019615   8.275  < 2e-16 ***
poly(education, 2, raw = TRUE)1 -0.062956  0.035635  -1.767 0.077290 .
poly(education, 2, raw = TRUE)2  0.016147  0.004113   3.926 8.66e-05 ***
poly(lifesat, 2, raw = TRUE)1   0.460790   0.027071  17.021  < 2e-16 ***
poly(lifesat, 2, raw = TRUE)2  -0.021908   0.002143 -10.222  < 2e-16 ***
poly(household, 2, raw = TRUE)1  0.160551  0.033150   4.843 1.28e-06 ***
poly(household, 2, raw = TRUE)2 -0.014342  0.004952  -2.896 0.003781 **
bornincountry1:immig            0.075254   0.020287   3.710 0.000208 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard deviation: 2.42 on 33669 degrees of freedom
Multiple R-squared: 0.1112
F-statistic: 468.2 on 9 and 33669 DF,  p-value: < 2.2e-16
     AIC      BIC
155123.4 155216.1
```

Additionally, the plots derived from this regression are depicted below:



The results of the final linear regression demonstrate an overall robust fit, in line with the theoretical framework. Notably, the analysis reveals divergent attitudes between native-born individuals and immigrants concerning the economic contributions of immigrants, with a specific focus on their impact on European Parliament (EP) trust ratings. This conclusion holds TRUE even when controlling for education, household characteristics, and life satisfaction ratings.

These findings contribute valuable insights to our understanding of the nuanced dynamics surrounding trust evaluations within the European Parliament, particularly in the context of attitudes toward immigration.

Final Data Analysis

```
### Xingting Luo 238672 Final Data Analysis 2023

# Load packages
library(dplyr)
library(psych)
library(ggplot2)
library(corrplot)
library(GGally)
library(stargazer)
library(effects)
library(lattice)
library(car)  # vif test for MC
library(lmtest)  # BP test, coeftest


######## Section 0 - Initiate Data Set ########

## 1. Load data:
ESS <- read.csv("ESS10.csv")

## 2. Trim data set to only include the set of variables you can choose
from:
ESS <- ESS[ , c("cntry", "agea", "gndr", "eisced", "brncntr", "ppltrst",
"nwspol",
                "trstprl", "trstep", "stflife", "imbgeco", "hhmmb",
"respc19")]

## 3. Rename the variables to make them easier to work with:
ESS <- rename(ESS, country=cntry, age=agea, gender=gndr, education=eisced,
              bornincountry=brncntr, soctrust=ppltrst, news=nwspol,
              parltrust=trstprl, eptrust=trstep, lifesat=stflife,
              immig=imbgeco, household=hhmmb, covid=respc19)

## 4. Recode NA values.
# All NA values are by default coded as values in ESS.
# With the following lines of code, you tell R which values reflect NAs:
ESS[, c(2,4,6:12)][ESS[, c(2,4,6:12)] > 50] <- NA
ESS[, c(3,5,13)][ESS[, c(3,5,13)] > 5] <- NA

## 5. Recode dummy variables and define them as factors.
# The dummy variables are by default coded as 1,2 instead of 0,1 in ESS.
# Further, they are not automatically read as factors (i.e. as categorical
variables).
# The following lines of code will fix this:
```

```
# Gender:
ESS$gender <- ifelse(ESS$gender == 2, 1, 0) # value == 2 modify to 1, if
else, modify to 0
ESS$gender <- as.factor(ESS$gender)
# "Female" is now 1, "Male" is 0.

# Born in country:
ESS$bornincountry <- ifelse(ESS$bornincountry == 2, 0, 1)  # value == 2
modify to 0, if else, modify to 1
ESS$bornincountry <- as.factor(ESS$bornincountry)
# "Yes" is now 1, "No" is 0.

# Covid:
ESS$covid <- ifelse(ESS$covid==1, 1, 0)  # value == 1 modify to 1 (no
change), if else, modify to 0
ESS$covid <- as.factor(ESS$covid)
# "Yes" is now 1, "No" is 0.


######## Section 1 - Descriptive statistics ########

# 1.1 Data cleansing and preparation

# observe the NA distribution
summary(ESS)
# observe the data type of columns
str(ESS)

# age, news, covid variables have plenty Null values
# Three factor varibles: gender, bornincountry, covid
# country is a chr variable


# data cleansing exclude age, news, covid
ess.clean <- ESS[!is.na(ESS$education) & !is.na(ESS$bornincountry)
                 & !is.na(ESS$soctrust) & !is.na(ESS$parltrust)
                 & !is.na(ESS$eptrust) & !is.na(ESS$lifesat)
                 & !is.na(ESS$immig) & !is.na(ESS$household), ]
# 37611 -> 33684 observations
describe(ess.clean)


# data cleansing for age variable
ess.clean2 <- ess.clean[!is.na(ess.clean$age), ]
# 33684 -> 16231:  lost half of observations


# data cleansing for covid variable
```

```
ess.clean3 <- ess.clean2[!is.na(ess.clean2$covid), ]
# 16231 -> 13202
plot(ess.clean3$covid, ess.clean2$eptrust) # no obvious differences.


# data cleasing for news variable
ess.clean4 <- ess.clean[!is.na(ess.clean$news), ]
# 33684 -> 13202: lost 2/3 of observations


# 1.2 Descriptive statistics for IVs and DV

# observe the distribution of interval variables
stargazer(ESS, type = "text", out = "summary_stats.txt",
          summary.stat = c("n","mean", "median", "sd", "min", "max",
"p25", "p75"),
          title = "Summary Statistics",
          digits = 1)


######## Section 2 - Descriptive statistics: Visual analysis ########

# 2.1 Histograms for interval variables
# eptrust distribution
hist(ess.clean$eptrust,
     col = "orange",
     breaks = 10,
     xlab = "Trust of European Parliament",
     main = "the dependent variable")
table(ess.clean$eptrust)

# immig distribution
hist(ess.clean$immig,
     col = "orange",
     breaks = 10,
     xlab = "the economic role of immigrants",
     main = "the main independent variable")
table(ess.clean$immig)

# basically normal distributed but more 0 scores.

# 2.2 Boxplots for dummy variables

# In our regression model, we want to include a bornincountry interaction
term, so it
# makes sense to look at the descriptive measures within each category of
table(ess.clean$bornincountry)  # No: 2623; Yes: 31061
# bornincountry plot
plot(ess.clean$bornincountry, ess.clean$eptrust,
```

```
     horizontal = T, xlab = "EP trust", ylab = "bornincountry") # nearly
identical on a general level

# gender
plot(ess.clean$gender, ess.clean$eptrust,
     horizontal = T, xlab = "EP trust", ylab = "gender") # almost the same
in the overall level

# covid
plot(ess.clean2$covid, ess.clean2$eptrust, horizontal = T) # nearly
identical on a broad scale

# country
ess.clean$country <- factor(ess.clean$country)
plot(ess.clean$country, ess.clean$eptrust) # Nothing obvious

# 2.3 Scatter plots for coefficients

# age
scatterplot(eptrust ~ age, data=ess.clean2, pch=".")  # negative

# education
scatterplot(eptrust ~ education, data=ess.clean, pch=".")  # positive

# household
scatterplot(eptrust ~ household, data=ess.clean, pch=".")  # non-linear

# news
scatterplot(eptrust ~ news, data=ess.clean4, pch=".")  # positive, non-
linear

# age
scatterplot(eptrust ~ age, data=ess.clean3, pch=".")  # negative


# 2.4 Correlations
# check the correlation between variables
round(cor(ess.clean3[, c(2, 4, 6, 8:12)], use="complete.obs"),2)

corrplot(round(cor(ess.clean3[, c(2, 4, 6, 8:12)]),2))

ggpairs(ess.clean[, c("eptrust", "parltrust", "soctrust", "immig",
"lifesat")])
# the eptrust has moderate correlations with parltrust and soctrust,
# lower correlation between immig and lifesat.


######## Section 3 - Model estimation and interpretation ########
```

```r
# Theory: European residents' identity characteristics and
# attitudes towards the economic role of immigrants
# will affect their evaluation of the EP trust in this dataset.

# Research Question: How does whether or not being native-born will affect
# how EU residents' rating results for eptrust depend on immig?
# H1: The effect of the attitude towards the economic role of immigrants
on EP trust rating
# is different for native residents and immigrant residents.
# H0: The effect of the attitude towards the economic role of immigrants
on EP trust rating
# is the same for native residents and immigrant residents.


# 3.1 Models

# Cross Tabulation
xtabs(~ bornincountry + immig, data = ess.clean)

# compute cell means to compare between native and immigrated people.
Tapply(eptrust ~ bornincountry + immig, mean, data = ess.clean)
# Native people gave more extreme eptrust scores on either side.


# We estimate three models in step-wise fashion:
# Without interaction, without control (naive model):
reg_ess_1 <- lm(eptrust ~ bornincountry + immig, data = ess.clean)

# Without interaction, with control:
reg_ess_2 <- lm(eptrust ~ bornincountry + immig + education + lifesat,
data = ess.clean)
vif(reg_ess_2) # No Multicollinearity

# With interaction, with control:
reg_ess_3 <- lm(eptrust ~ bornincountry + immig + bornincountry*immig +
education + lifesat, data = ess.clean)

# Comparation
stargazer(reg_ess_1, reg_ess_2, reg_ess_3, type = "text",
          out = "summary_regressions.txt",
          title = "Summary Statistics for regressions")


# The effect plot for model reg_ess_3, which includes the bornincountry *
immig
plot( Effect( c("bornincountry", "immig"), reg_ess_3),
      main    = "Interaction Model (a)",
      confint = list(style = "bars"),
      lines   = list(multiline = TRUE),
```

```r
      ylim    = c (3, 6.5))

plot( Effect( c("bornincountry", "immig"), reg_ess_2),
      main    = "Main-Effects Model (b)",
      confint = list(style = "bars"),
      lines   = list(multiline = TRUE),
      ylim    = c (3, 6.5))

plot(predictorEffects(reg_ess_3), lines = list(multiline = TRUE))
# bornincountry * immig interaction shows the low immig scores affected by
bornincountry status

# With more control:
reg_ess_4 <- lm(eptrust ~ bornincountry + immig + bornincountry*immig +
education + lifesat + household, data = ess.clean)
reg_ess_5 <- lm(eptrust ~ bornincountry + immig + bornincountry*immig +
education + lifesat + soctrust, data = ess.clean)
reg_ess_6 <- lm(eptrust ~ bornincountry + immig + bornincountry*immig +
education + lifesat + parltrust, data = ess.clean)

stargazer(reg_ess_3, reg_ess_4, reg_ess_5, reg_ess_6, type = "text",
          out = "summary_regressions 2.txt",
          title = "Summary Statistics for regressions")

summary(reg_ess_4) # household can be added to the model.


######## Section 4 - OLS assumptions and regression diagnostics ########

## Specify 1 as the second argument to examine whether appear to have a
linear relationship:
plot(reg_ess_4, 1)

# We aim for a horizontal red line. This line looked a bit like a
parabola,
# it would be worth testing a quadratic relationship instead.


## Specify 2 as the second argument to examine whether the residuals are
normally
# distributed:
plot(reg_ess_4, 2)
# Alternatively, plot a histogram of the residuals:
hist(reg_ess_4$residuals)


## Specify 3 as the second argument to examine whether the residuals are
homoskedastic:
plot(reg_ess_4, 3)
```

```
# Again, we aim for a horizontal red line, implying no systematic pattern
in the
# residuals of the model.

bptest(reg_ess_4) # reject null hypothesis of homoscedasticity

reg_ess_4_robust <- coeftest(reg_ess_4, vcov = vcovHC(reg_ess_4, type =
"HC1"))
reg_ess_4_robust


# diagnostics for lm
residualPlots(reg_ess_4)
# the variables have *** marks on p-value, suggesting a transformation.

# Create residual plots with lifesat variable
residualPlots(reg_ess_4, ~ lifesat, fitted = FALSE, id = list(n = 3))
# Create residual plots with education variable
residualPlots(reg_ess_4, ~ education, fitted = FALSE, id = list(n = 3))


# use quadratic regression for education and lifesat variable
reg_ess_final <- lm(eptrust ~ bornincountry + immig + bornincountry*immig
                    + poly(education, 2, raw = TRUE)
                    + poly(lifesat, 2, raw = TRUE)
                    + poly(household, 2, raw = TRUE),
                    data = ess.clean)
summary(reg_ess_final)

# test again
residualPlots(reg_ess_final)  # better fitting

# marginalModelPlots for Marginal-Model Plots
marginalModelPlots(reg_ess_final) # no obvious problem.

Anova(reg_ess_final) # Type II tests: analysis-of-variance tables for
model **objects**


######## Section 5 - Outliers ########

# Studentized Residuals

# The generic qqPlot() function in the car package has a method for linear
models,
# plotting Studentized residuals against the corresponding quantiles of
t(n - k - 2).
qqPlot(reg_ess_final, id = list(n = 3))
```

```r
# A test based on the largest (absolute) Studentized residual,
# using the outlierTest () function in the car package,
outlierTest(reg_ess_final)
# No Studentized residuals with Bonferroni p < 0.05


# Leverage: Hat-Values

# function from car package
# includes index plots of Studentized residuals,
# the corresponding Bonferroni p-values for outlier testing,
# the hat-values, and Cook's distances.
influenceIndexPlot(reg_ess_final, id=list (n=3))
influencePlot(reg_ess_final, id=list (n=3))


# dfbetas plots for immig:
dfbetasPlots(reg_ess_final, ~immig,  id.n = 5, labels =
rownames(dfbetas(reg_ess_final)))
# dfbetas plots for all variables:
dfbetasPlots(reg_ess_final,  id.n = 5, labels =
rownames(dfbetas(reg_ess_final)))
# All |dfbetas| < 1 so no outliers!

# try remove 5 observations have biggest determination.
ess.clean.out <- ess.clean[-c(9661, 12609, 12087, 26202, 30123), ]

# Re-run regression, and report both results:
reg_ess_final_out <- lm(eptrust ~ bornincountry + immig +
bornincountry*immig
                    + poly(education, 2, raw = TRUE)
                    + poly(lifesat, 2, raw = TRUE)
                    + poly(household, 2, raw = TRUE),
                    data = ess.clean.out)
stargazer(reg_ess_final, reg_ess_final_out, type = "text")
# We notice that the coefficients have hardly changed (as expected),
# as the outliers don't have massive influence

S(reg_ess_final_out)
plot(predictorEffects(reg_ess_final_out), lines = list(multiline = TRUE))
```