

**STATS II: Research Paper Version I**

*How does happiness level affect people's immigration sentiment and their level of trust for politicians?*

Candidate:



Date: 6<sup>th</sup> December 2019

Body Word Count: 1650

Full Word Count: 2937

The recent growth in support of far-right and anti-establishment ideology has prompted a great debate in the political research world regarding the exact causes that motivate the public to entertain such ideas.

Thus, this paper aims to investigate two research questions: a) how does people's happiness level affect their immigration sentiment? and b) how does people's happiness level affect their trust of politicians?

I will use the European Social Survey Wave 8 (2016) to test my two hypotheses:

H1: Happier people are more likely to be sympathetic towards immigrants and have a positive immigration sentiment.

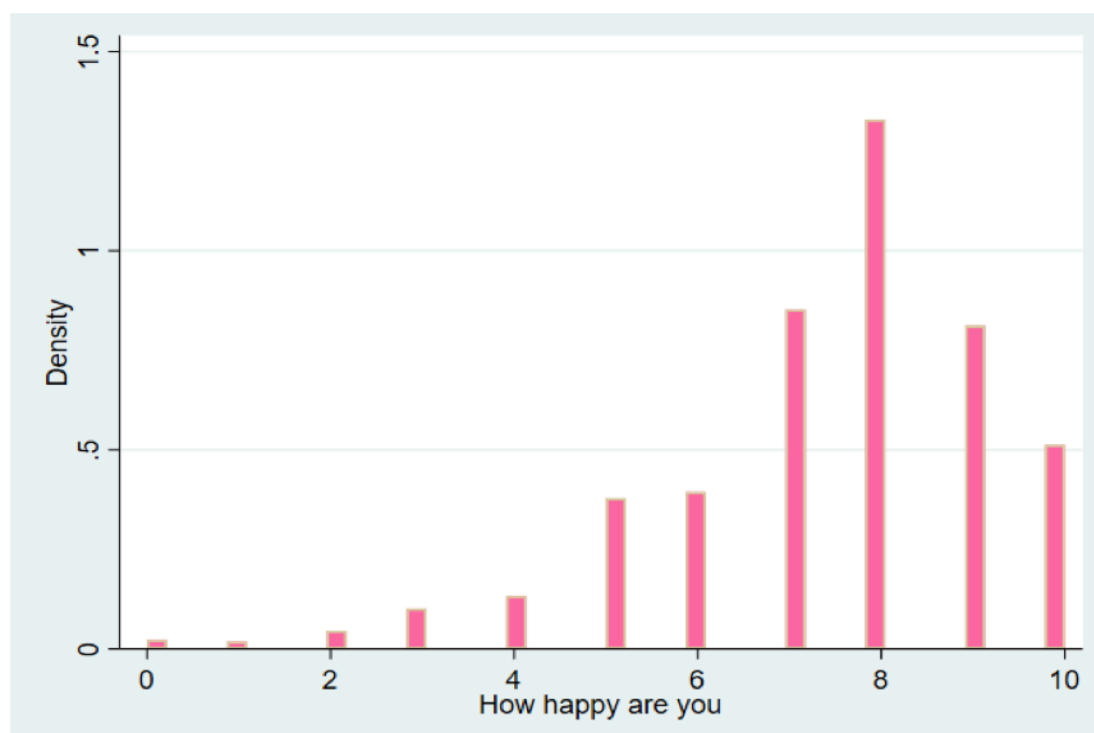
H2: Happier people are more likely to have a high level of trust of politicians.

### ***Variable Description:***

The variables are taken from the dataset produced by the ESS Wave 8 (2016), which conducted a cross-national survey on 23 countries asking citizens about a range of social questions and issues. The dataset is of such a magnitude that it allows for a sample size of over 10,000 data points, which is beneficial as the parameters of  $\beta$  cannot be considered significant unless the sample size is of that dimension.

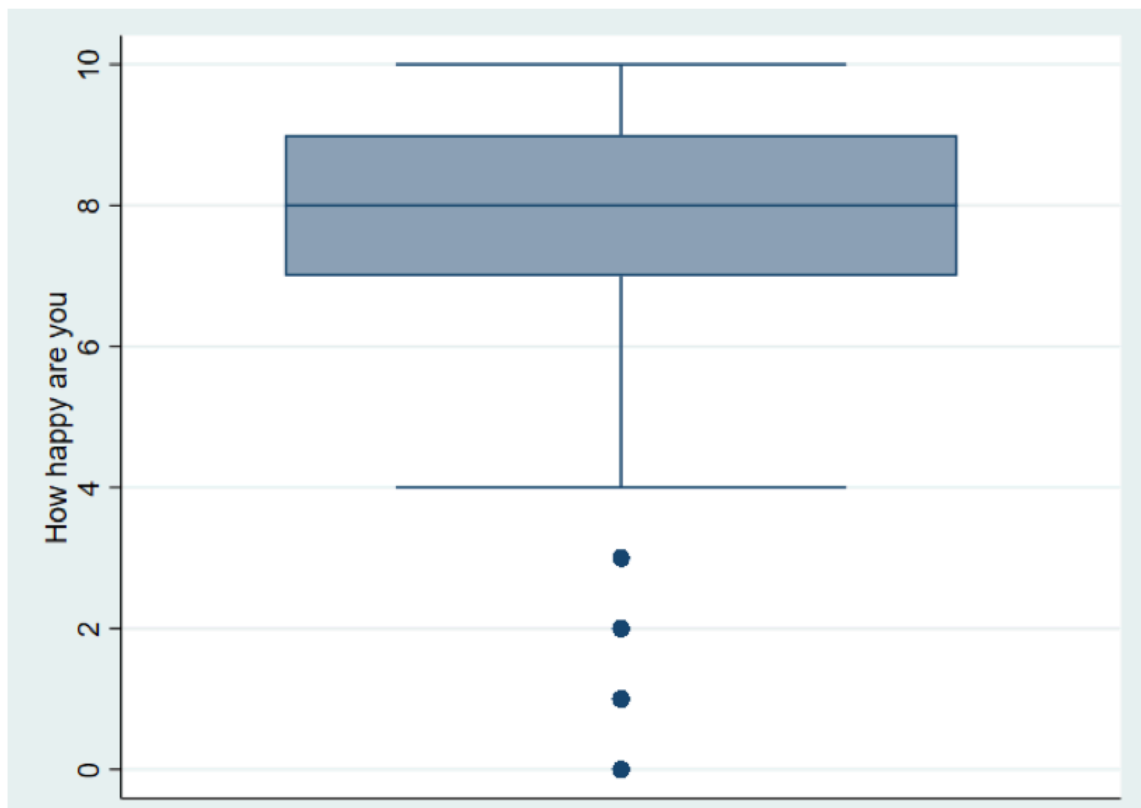
The independent variable in this study is happiness, which is measured in the dataset through the question: "taking all things together, how happy would you say you are?" The scale ranges from 0 (extremely unhappy) to 10 (extremely happy) (ESS 2016), visualized in Figure 1.

*Figure 1: Histogram of Level of Happiness*



Using the command `sum`, I conclude that the mean level of happiness is 7.44 with a median of 8. The fact that  $\text{mean} < \text{median}$  indicates a skew to the left, indicating that the majority of people lean toward identifying as happier than not. A box plot of the variable (figure 2) indicates outliers with the values that fall outside of the lower whisker. I will take this into consideration and control for outliers in regression, to eliminate potential bias in the model.

*Figure 1.2: Boxplot of Level of Happiness*

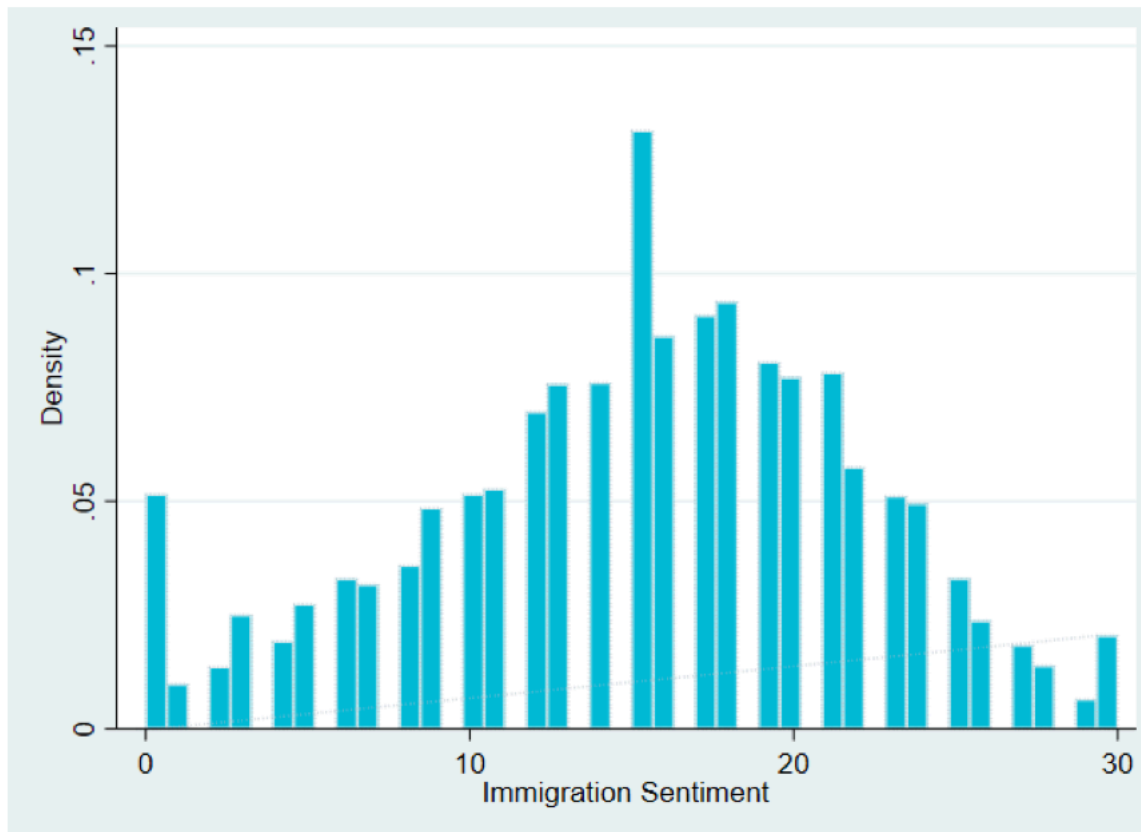


The first dependent continuous variable measured is immigration sentiment. In efforts to create one general variable that denotes the general immigration sentiment, I combined the following variables:

- Imbegco: immigration is “bad/good for the country’s economy”?
- Imueclt: immigration undermines/enriches a “country’s cultural life”?
- Imwbcnt: immigrants make the country better/worse place to live?

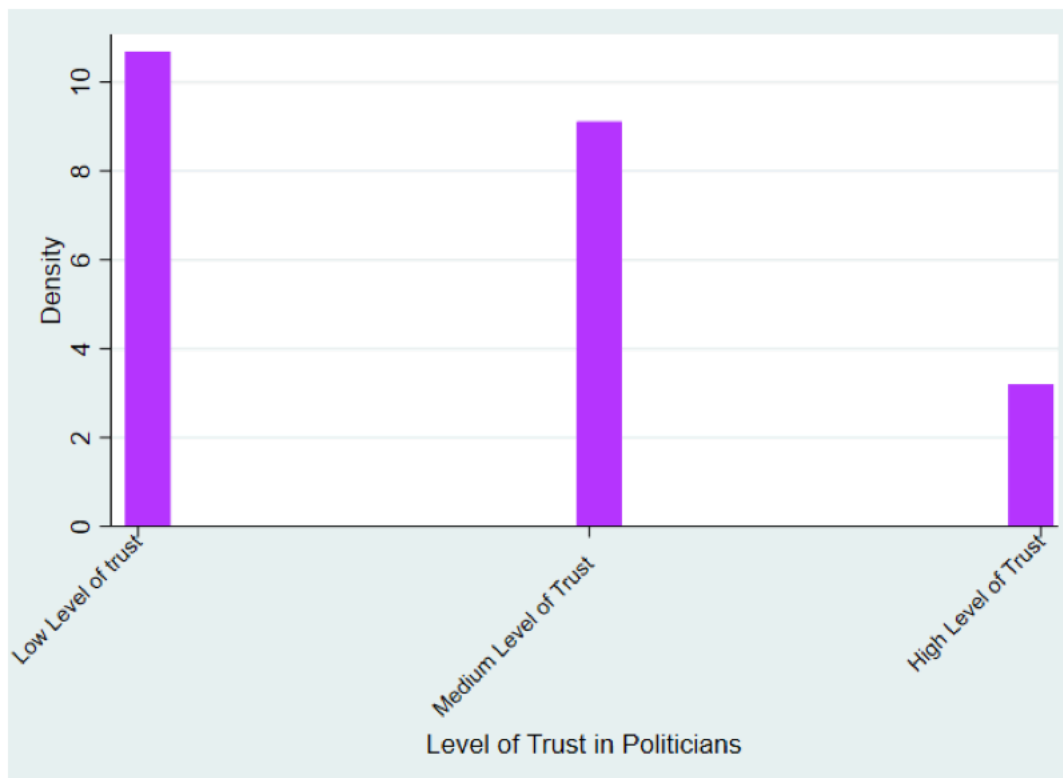
These three variables are highly correlated, thus can be combined into one variable “Immigration Sentiment” which ranges responses from 0=extremely anti-immigrant to 30= extremely pro-immigrant. The distribution of the variable can be observed in Figure 2, and upon summarizing in STATA, the mean (average immigration sentiment) is at 15.3 with a median of 16. This indicates an almost normal distribution, with most data concentrated around a medium level of immigration sentiment.

Figure 2: Histogram of Immigration Sentiment



The second dependent variable measured is the level of trust in politicians. I use the variable *trstplt* from the data set, which ranks how much the respondent personally trusts politicians from 0 to 10 (0=No trust, 10=Complete Trust). I created a new variable named “Level of Trust in Politicians” which organizes the values of *trstplt* into three categories: 1. Low level of trust (0-3), 2. Medium level of trust (4-6) and 3. High level of trust in politicians (7-10). Figure 3 shows the variable in categorical form; the majority of values are grouped in the categories of low and medium trust level, and a description of the variable in STATA reveals the mean is 1.67, indicating that the average of values lies between the first and second categories (low and medium trust level).

Figure 3: Histogram of Level of Trust in Politicians



I am also controlling for multiple variables to ensure that any effect observed on the dependent variables is as a result of a change in the independent variable, rather than an external unmeasured variable. I am controlling for age, education level, ethnicity, and employment within the last 5 years. All of these variables could influence both immigration sentiment and trust in politicians, so by controlling for them I hope to isolate the effect to the independent variable being studied. Other control variables that could strengthen the model include rural/urban upbringing, etc., however the dataset does not include these variables in its survey.

### Happiness and Immigration: OLS Regression

To test the relationship between happiness and immigration sentiment, I will use the Ordinary Least Square method of regression, due to the dependent variable being continuous. The OLS method generates an estimation of the parameters  $\beta_0$  and  $\beta_1$ , which represent the y-intercept and the slope of regression respectively. The OLS method also seeks to minimize the sum of squared errors in the data.

The regression equation for this variable will be as followed: Immigration Sentiment (Y) =  $\beta_0$  +  $\beta_1$  (Level of Happiness) +  $\beta_2$  (Employment for past 5 years) +  $\beta_3$  (Age) +  $\beta_4$  (Education Level) +  $\beta_5$  (Ethnicity) + error.

I have created several models to fully grasp the effect of happiness on immigration sentiment. To begin with, Models 1-3 are simple OLS regressions; Model 1 has no control variables, Model 2 is controlled for age and education level and Model 3 is controlled for all four variables. As can be observed from Table 1, as

the regression includes more control<sup>1</sup> variables, the  $r^2$  value increases from an insignificant 0.06 to 0.25, indicating that while model 1 accounts for 6% of the data, model 3 accounts for 25%, making it more accurate.

*Table 1: OLS Regression*

Degree of Immigration Sentiment	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>	<i>Model 4</i>
Level of happiness	0.899*** (0.017)	0.804 *** (0.017)	0.565 *** (0.03)	0.553 *** (0.033)
Age		-0.012 *** (0.001)	0.0062 (0.004)	0.007 * (0.004)
Education Level		Controlled	Controlled	Controlled
Ethnicity			Controlled	Controlled
Employment within last 5 years			0.199 (0.126)	0.263 ** (0.124)
Robust Standard Error	NO	NO	NO	NO
Outliers Exist	YES	YES	YES	NO
Fixed effects	NO	NO	NO	NO
Constant	8.62	8.85	10.35	9.90
N	41,207	40,993	11,370	10,980
R <sup>2</sup>	0.06	0.11	0.25	0.26

Entries are parameters estimated with OLS regression

Standard errors are in parentheses

\*\*\* p<0.01, \*\* p< 0.05, \*p<0.1

In model 4, I used the Cook's D method to isolate and eliminate outliers in the independent variable that could skew the effect away from the observed pattern. I saw a change in both  $\beta_0$  (decrease to 9.90),  $\beta_1$  (decrease to 0.553) and the  $r^2$

---

<sup>1</sup> The coefficients of control variables ethnicity and education level are not included in the tables due to too many categories.

increase to 0.26. Although Model 4 is more reliable than the previous ones due to having no outliers and having a higher  $r^2$ , this regression does not account for two other important assumptions in data analysis: multi-level data and heteroscedasticity.

The data is gathered on a cross-national scale, thus there is second level of analysis added to the first level of individual cases. Otherwise we assume people from one country will be the same as from another, which can lead to wrong conclusions. Thus, we need to account for these levels by creating a fixed effect regression, through controlling for the country of the respondents (see Appendix).

Another important test is for homoscedasticity of data, which occurs when variance of residuals is identical for each value of  $y$  (Bol 2019). This can be observed in the graphing of residuals (figure 4), where there is a distinct pattern in the plot, indicating that there is heteroscedasticity and that the standard error of the models is incorrect causing an erroneous prediction of  $y$ . In order to make the standard error unbiased, I apply a robust command to Models 5,6&7.

**Figure 4: Graph of Residuals against Fitted Values**

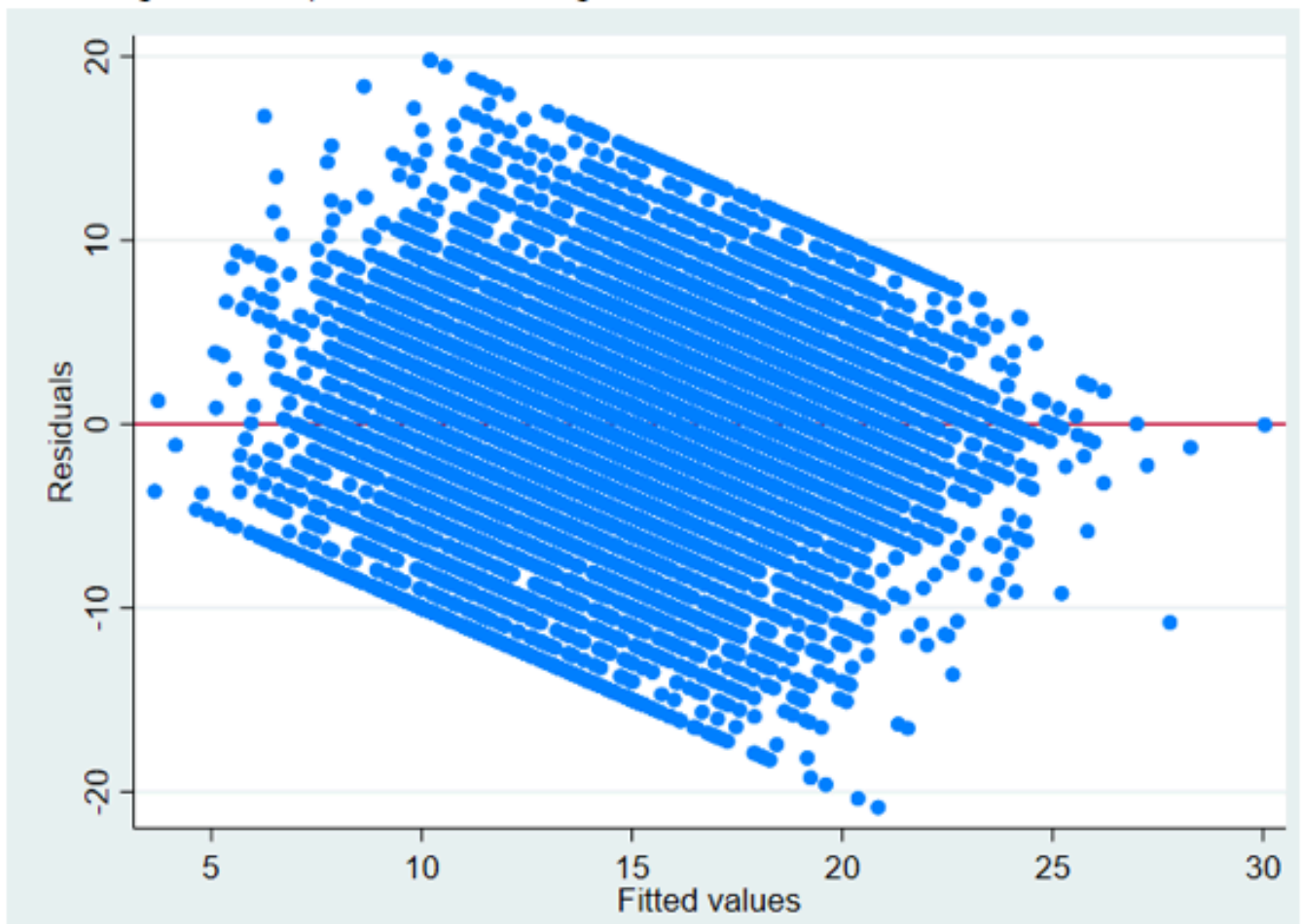


Table 2: OLS Regression, with Fixed Effect and Homoscedasticity

Degree of Immigration Sentiment	Model 5	Model 6	Model 7
Level of happiness	0.638 *** (0.03)	0.539 *** (0.03)	0.533 *** (0.03)
Age		0.0056 (0.003)	0.0075 * (0.004)
Education Level		Controlled	Controlled
Ethnicity			Controlled
Employment within last 5 years			0.266 ** (0.125)
Robust Standard Error	YES	YES	YES
Outliers Exist	NO	NO	NO
Fixed effects	YES	YES	YES
Constant	7.823	6.818	8.05
N	10,980	10,980	10,980
R <sup>2</sup>	0.17	0.24	0.27

Entries are parameters estimated with OLS regression

Standard errors are in parentheses

\*\*\* p<0.01, \*\* p< 0.05, \*p<0.1

Thus, Model 7 contains all the control variables, fixed effect, robust standard error and no outliers, making it the most representative model of the data. This is illustrated by the fact that it accounts for 27% of data, indicated by the  $r^2$  of 0.27, which is higher of all the other models. It concludes that a unit increase in happiness level correlates with a 0.533 increase on the immigration sentiment scale towards pro-immigration beliefs, confirming H1 with an effect that is statistically significant to  $p<0.01$ . The constant, or  $\beta_0$ , shows that when happiness level is at 0 (extremely unhappy), immigration sentiment is at 8.05, which out of 30 is quite low and anti-immigrant.



Before concluding that Model 7 is the best regression to capture the effect between happiness and immigration, I needed to test whether the model is best represented in linear form. I tested whether the model would be better suited in Curvilinear effect, by using a regression where the independent variable was squared, and I also tested a Logarithmic (ceiling) effect by transforming x in a logarithmic function. I concluded that the linear effect is the best fit for describing the effect, as the linear regression has an  $r^2$  of 0.2747 (accounts for 24.47% of data), while the other two models account for 24% and 27.23% respectively.

### Happiness and Trust: Multinomial Logistic Regression

In order to analyze the effect of happiness on trust in politicians, I use a multinomial logistic regression due to the variable's categorical nature. In order to compare probability of high level of trust in politicians, I have decided to set low level of trust as the reference category.

The parameters ( $\beta$ ) of this regression cannot be interpreted further than just denoting whether the effect is positive or negative, without calculating the relative risk ratios, which give the probability of y if x changes by one unit (shown Table 3).

*Table 3: Multinomial Logistic Regression*

<b>Trust in Politicians</b>	<b>Model 8</b> (Measuring high level of trust against low level as reference category)	<b>Model 9</b> (Measuring high level of trust against low level as reference category)
<b>Level of happiness</b>	1.26 *** (0.012)	1.25 *** (0.023)
<b>Age</b>		1.00 ** (0.002)
<b>Education level</b>		Controlled
<b>Employment within last 5 years</b>		1.044 (0.073)
<b>Fixed effect</b>	YES	YES
<b>N</b>	44,172	12,274
<b>Pseudo R<sup>2</sup></b>	0.06	0.07
<b><math>\beta</math> Coefficient (not rrr)</b>	Positive	Positive

Entries are relative risk ratios (rrr) estimated with logistic regression  
Standard Errors are in parentheses  
\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

For both models, the  $\beta > 0$ , indicating a positive effect. For Model 9, which contains all the control variables, the rrr is 1.25. This is interpreted as the probability of having a high level of trust in politicians increases by 0.25 or 25% (1.25-1) with each unit increase in happiness (x), compared to a low level of trust in politicians (the reference category), proving H2. The  $r^2$  value of 0.07 indicates that 7% of the data is represented by the model, which is not as significant as it was for the OLS, but given the statistical significance of the results the effect is still worth considering.

Multinomial and logistic regressions do not need to be tested for heteroscedasticity or outliers.

## **Conclusion**

Both hypotheses were supported by the data gathered, indicating that an increase in happiness correlates with increase in immigration support and trust for politicians. Both regressions are statistically significant to  $p < 0.01$ , had a low standard error and an  $R^2$  value of 0.27 and 0.07 respectively, which represent great models for an experiment in social science. However, despite having good models, one cannot infer causality from the data gathered and state that a) happiness causes pro-immigration or b) happiness causes people to trust politicians, as causality is much more complex and requires further in-depth investigation.

## Appendix

### Bibliography:

Bol, Damien (2019). Statistics II Lectures.

### Dataset Access/ Download:

<https://www.europeansocialsurvey.org/data/download.html?r=8>

### STATA Do-file:

```
//PART 1: VARIABLES
*tab and clean variables*
//DEPENDENT VARIABLE NR 1 - Level of immigration sentiment
(continuous)
tab imbgeco //this variable means immigrants are bad --->good for
the economy//
tab imbgeco, nolab //no missing variables//
tab imueclt //cultural life is undermined ---> enriched by
immigrants//
tab imueclt, nolab //no missing variables
tab imwbcnt //immigrants make the country worse ---> better place
to live//
tab imwbcnt, nolab //no missing variables
//need to transform this into one variable on anti-immigration
sentiment
//1 is extremely anti-immigrant, 10 is extremely pro-immigrant
corr imbgeco imueclt imwbcnt
gen antiimmigration=(imbgeco+imueclt+imwbcnt)
label variable antiimmigration "Immigration Sentiment"
label define IMMIGRATION2 0 "Extremely Anti-Immigrant" 30
"Extremely Pro-Immigrant"
label val antiimmigration IMMIGRATION2
tab antiimmigration
sum antiimmigration, det
//mean= 15.3, standard deviation= 6.76, Skewness= -0.29
hist antiimmigration //shows the distribution of the continuous
dependent variable,
//anti-immigration sentiment, as a close to normal distribution//

//DEPENDENT VARIABLE NR 2 - Level of trust in politicians
(categorical)
tab trstplt //how much do you personally trust politicians (0
means not at all,
//10 means completely)
hist trstplt
sum trstplt, det
//mean = 3.649, stdv = 2.418, skew= 0.08
```

```

//mainly concentrated around 0 and 5, results appear skewed to
the right
//need to categorize into 3 categories = low level of trust,
medium and high
gen trustcategory=.
replace trustcategory=1 if trstplt<4
replace trustcategory=2 if trstplt>=4 & trstplt<=6
replace trustcategory=3 if trstplt>6
label variable trustcategory "Level of Trust in Politicians"
label define TRUST1 1 "Low Level of trust" 2 "Medium Level of
Trust" 3 "High Level of Trust"
label val trustcategory TRUST1
tab trustcategory
hist trustcategory
sum trustcategory, det
//mean=1.67, stdv= 0.7

//INDEPENDENT VARIABLE: LEVEL OF HAPPINESS
//"taking all things together, how happy would you say you are?"
//0= extremely unhappy, 10=extremely happy
tab happy
hist happy
sum happy, det
graph box happy //there are outliers, so we need to be careful
when doing the
// regression

//CONTROL VARIABLES - any other variables that could influence
the effect on
//immigration sentiment
//highest level of education
tab edulvlb, nolab //(categorical)
replace edulvlb=. if edulvlb==5555
tab edulvlb

//age
tab agea //(continuous)

//ancestry & ethnic groups
tab anctry1, nolab
replace anctry1=. if anctry1==444444
tab anctry1

//employed within the last 5 years (categorical)
tab uemp5yr
//ideally, would include whether the person lives in urban or
rural landscape as
//a control variable, but it is not in the data set

//PART II: OLS REGRESSION (ANTI-IMMIGRATION)
reg antiimmigration happy
//antiimmigration(y) = B0+B1(happy)
//statistically significant because the p value is really small
***

```

```
//every time happiness increases by one value, pro-immigration
sentiment increases
// by 0.899.
//B0 (represents level of immigration sentiment when happiness is
0)
//when happiness is 0 is 8.62 so quite low and anti-immigrants
(on a scale of 1-30)
//error of 0.0175
//r squared is low = 0.0602 (model accounts for 6% of data)
//N=41,207
```

```
reg antiimmigration happy agea i.edulvlb
//b1= 0.804- increase 1 in happiness = increase in immigrant
tolerance by 0.804
//B0= 8.85
//statistically significant *** with standard error of 0.0172
//r squared is slightly higher, 0.11 (model accounts for 11% of
the data)
//N=40,993
```

```
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
//with all the control variables, B1= 0.565 (change in y for
every unit change in x)
//B0= 10.35
//still statistically significant *** with standard error of
0.031
//r squared is higher and 0.25 (accounts for 25% of the data)
//can be fixed by more control variables that aren't in the study
-- ie rural/urban
//N= 11,370
```

```
//However, the result could be skewed as a result of outliers:
//OUTLIERS
```

```
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
predict cook, cooksd
gen outlier=0
replace outlier=1 if cook>(0.00035) //the cook d number is
(4/11,370) = 0.00035
```

```
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1 if
outlier==0
//r squared is now 0.26, with model accounting for 26% of the
data
//standard error of 0.03 and statistically significant
```

```
//FIXED EFFECT: To account for not having independent variable at
level 2
//(cross national survey)
tab cntry //country variable, need to recode with letters
gen ccntry=.
replace ccntry=1 if cntry=="AT"
replace ccntry=2 if cntry=="BE"
replace ccntry=3 if cntry=="CH"
replace ccntry=4 if cntry=="CZ"
replace ccntry=5 if cntry=="DE"
```

```

replace ccntry=6 if cntry=="EE"
replace ccntry=7 if cntry=="ES"
replace ccntry=8 if cntry=="FI"
replace ccntry=9 if cntry=="FR"
replace ccntry=10 if cntry=="GB"
replace ccntry=11 if cntry=="HU"
replace ccntry=12 if cntry=="IE"
replace ccntry=13 if cntry=="IL"
replace ccntry=14 if cntry=="IS"
replace ccntry=15 if cntry=="IT"
replace ccntry=16 if cntry=="LT"
replace ccntry=17 if cntry=="NL"
replace ccntry=18 if cntry=="NO"
replace ccntry=19 if cntry=="PL"
replace ccntry=20 if cntry=="PT"
replace ccntry=21 if cntry=="RU"
replace ccntry=22 if cntry=="SE"
replace ccntry=23 if cntry=="SI"
tab ccntry

reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
i.ccntry
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
i.ccntry if outlier==0

//DETECTING HETEROSKEDASTICITY
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
i.ccntry if outlier==0, robust
rvfplot, yline(0)

//MODEL WITH 4 CONTROL VARIABLES (AGE, LEVEL OF EDUCATION,
ETHNICITY, EMPLOYMENT
//IN LAST 5 YEARS)
//WITH FIXED EFFECT AND NO OUTLIERS
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
i.ccntry if outlier==0, robust
//B0= 8.05: when level of happiness is at 0 (extremely unhappy),
immigration
//sentiment is at 8.05 out of 30 (0= extremely anti-immigrant)
//B1= 0.533: with every unit increase in happiness, the
immigration sentiment will
//increase by 0.533 (becoming more sympathetic towards
immigrants)
//statistically significant ***
//standard error= 0.03
//R squared= 0.27 = model account for 27% of the data
//N= 10,980

//MODEL WITH 2 CONTROL VARIABLES (AGE & EDUCATION LEVEL)
//WITH FIXED EFFECT AND NO OUTLIERS
reg antiimmigration happy agea i.edulvlb i.ccntry if outlier==0,
robust
//B0= 6.818
//B1= 0.539

```

```

//statistically significant ***
//standard error= 0.03
//R squared= 0.24
//N=10,980

//MODEL WITH NO CONTROL VARIABLES
//WITH FIXED EFFECT AND NO OUTLIERS
reg antiimmigration happy i.ccntry if outlier==0, robust
//B0= 7.823
//B1= 0.638
//statistically significant ***
//standard error= 0.03
//R squared: 0.17
//N= 10,980

//DETERMINING WHETHER INDEPENDENT VARIABLE IS LINEAR:
//LOGARITHMIC EFFECT:
gen happy_log=log(happy)
reg antiimmigration happy_log i.uemp5yr agea i.edulvlb i.anctry1
i.ccntry if outlier==0
//R-squared is 0.2723, Adjusted R-Squared is 0.2585

//CURVILINEAR EFFECT:
gen happy_sq=happy*happy
reg antiimmigration happy_sq i.uemp5yr agea i.edulvlb anctry1
i.ccntry if outlier==0
//R-Squared is 0.2470, Adjusted R-Squared is 0.2434

//LINEAR EFFECT:
reg antiimmigration happy i.uemp5yr agea i.edulvlb i.anctry1
i.ccntry if outlier==0
//R-squared is 0.2747, Adjusted= 0.2610
//The linear model has a higher r-squared value, indicating that
a linear trend account
//for a larger percentage of the data (therefore the effect is
linear)

//PART III: MULTINOMIAL LOGISTIC REGRESSION (TRUST IN POLITICIANS)
mlogit trustcategory c.happy, baseoutcome (1)
//coefficient is positive, therefore the probability of y
increases w every increase in x
mlogit trustcategory c.happy i. ccntry, baseoutcome (1) rrr
//testing: the happier people are, the more likely they are to
have a high level of trust
//in politicians, compared to a low level of trust
//coefficient= 1.26
//1.26-1= 0.26
//every time happiness increases by one unit, the probability of
people having a high level
//of trust for politicians increases by 26%
//standard error: 0.012

//N=44,172
//stat significant ***
//standar error 0.0089 high level

```

```
//Pseudo R squared= 0.0139
```

```
mlogit trustcategory c.happy c.agea i.edulvlb i.ccny,  
baseoutcome (1) rrr  
//coefficient = 1.121  
//1.121-1= 0.121  
//every time happiness increases by one unit, the probability of  
people having a high level  
//of trust for politicians decreases by 12%  
//standard error= 0.0069  
//N= 43,835  
//Pseudo R squared= 0.0712
```

```
mlogit trustcategory c.happy i.uemp5yr c.agea i.edulvlb i.ccny,  
baseoutcome (1) rrr  
//coefficient= 1.112  
//1.112-1= 0.112  
//every time happiness increases by one unit, the probability of  
people having a high level  
//of trust for politicians decreases by 11.2%  
//standard error= 0.012  
//N=12,274  
//Pseudo R squared= 0.0712
```