Pilot Research Project

Marko Prohaska and Matteo Bruno

January 2024

1 Introduction

In recent years, the dominant neoclassical economic system focused on growth and wealth creation has come under criticism. Some countries, like Bhutan, have instead decided to pursue maximizing happiness, with the official goal of the government to maximize Gross National Happiness.

The purpose of this paper is to look at which economic indicators are most relevant to the happiness of a country. Specifically, we are interested in investigating whether social or economic indicators are more relevant to a country's happiness. To do this, we will compile 10 development indicators, half of them social ones (such as literacy rate or life expectancy) and the other half economic (such as GDP growth rate and unemployment). We will then run a multivariate linear regression to see if our model explains a significant portion of the variance in happiness levels and analyze whether there is a significant difference between social and economic indicators in predicting happiness. To get more general results, we decided to consider data over 6 years, from 2016 to 2021. We didn't include data from 2022 since it was available for a limited number of countries only.

2 Data Set

2.1 Processing

We selected the following 5 social indicators:

- Life expectancy
- Literacy rate
- Infant mortality rate (per 1,000 live births)
- Prevalence of undernourishment
- Extreme poverty (percentage of the population)

and the following 5 economic indicators:

- GDP growth rate
- GDP per capita
- Unemployment rate
- Inflation
- Current account.

For measuring happiness, we used the data provided in the World Happiness Report. Next, we merged the data into one common data set by country name and year. We noticed how the data about the literacy rate and infant mortality rate was missing for the majority of countries. Therefore, to clean the dataset we would have had to fill the data with zeros or with the average of the values of these indicators. Since none of these option seemed particularly appealing, we decided to remove the indicators, considering a reduced set of eight covariates. Having done that, we still had a small number of missing entries, so we removed any countries which had incomplete data.

We also decided to normalize the indicator data due to the wide interval ranges of the individual indicators (for instance, GDP per capita was expressed in 1000s whilst life expectancy did not exceed 100).

Our final data set had dimensions 94x54.

2.2 Data Visualization

Next, we proceeded with data visualization. We opted for scatter plots to get an understanding of the correlation between each indicator and happiness (here we will show the data for 2018, for the rest of the years please see Appendix A). It is immediately clear that some development indicators are much more relevant to happiness than

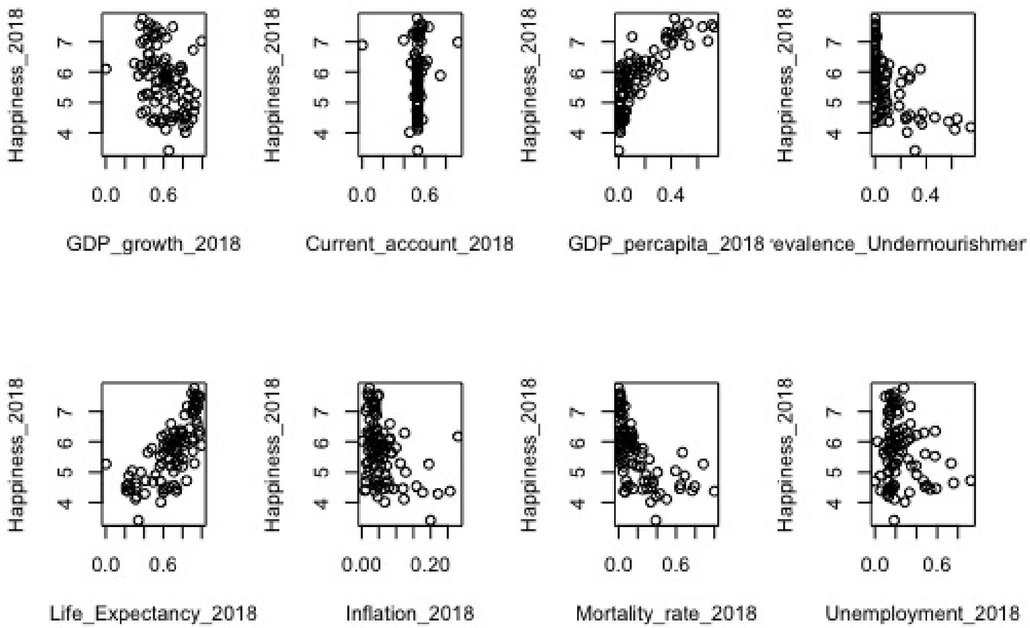


Figure 1: Scatter plots for year 2018

others. The plots indicate that current account has no correlation to happiness, as the values are distributed in a straight line. GDP growth and unemployment yield random scatters, implying practically no correlation. Whilst current account and GDP growth's lack of correlation with happiness are somewhat intuitive, since both do not speak much to the quality of life in a country (for instance, many developed nations run a current account deficit), the lack of correlation between unemployment and happiness is unexpected. Prevalence of undernourishment, inflation, and mortality rate have a vertical scatter for low values, indicating no correlation, but do reveal a negative correlation as values rise. Finally, GDP per capita and life expectancy are clearly positively correlated with happiness. Whilst life expectancy seems to be linearly correlated, GDP per capita, like some of the previous plots, seems to have a vertical scatter for low values before revealing a positive correlation. This pattern will be further discussed in the "Limitations" section. Moreover, it should be noted how these are only initial considerations based on a visual analysis only and we will further analyze the relevance of each covariate in the model selection section, sometimes coming to different conclusions.

3 Regression Diagnostic

We now turn to checking the normality of the residuals for our candidate model, for which we decided to run the Shapiro-Wilk test. We choose a significance level of 0.05 and, as it is possible to see from Appendix B, for years 2019 and 2021 the p-values are 0.03748 and 0.04952, while for every other year they are greater than 0.5. This indicates that, for two out of 6 years, we have sufficient significant statistical evidence to reject the null hypothesis that the residuals come from a Normal distribution. However, this is not surprising since the Shapiro-Wilk test can detect even mild deviations from normality, which might not be significant enough to discard the model. This is especially true when the sample size is large, as in our case. Furthermore, the QQ-plots of the residuals show how the residuals fit quite well to a straight line for every year. Therefore, even if, rigorously speaking, results from 2019 and 2021 reject normality, we can still safely work as if the residuals were normally distributed for every year.

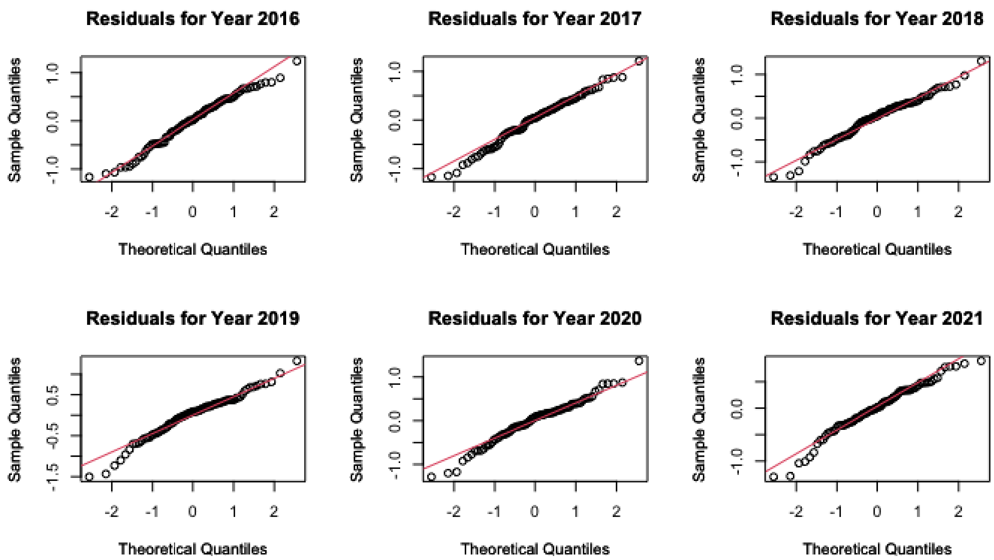
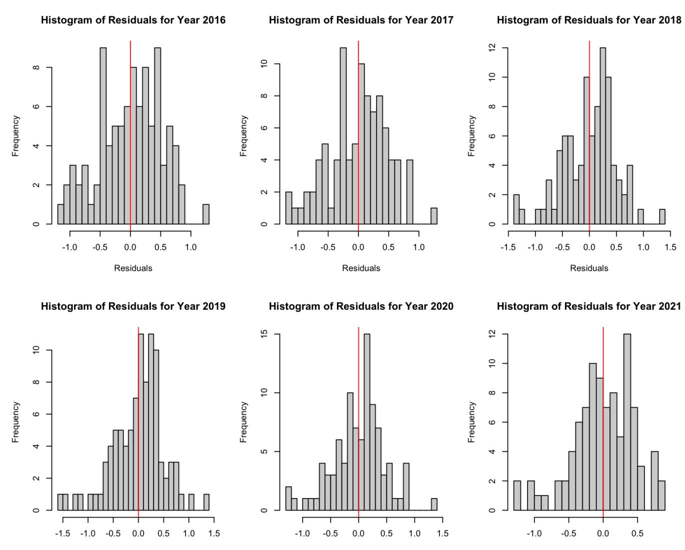
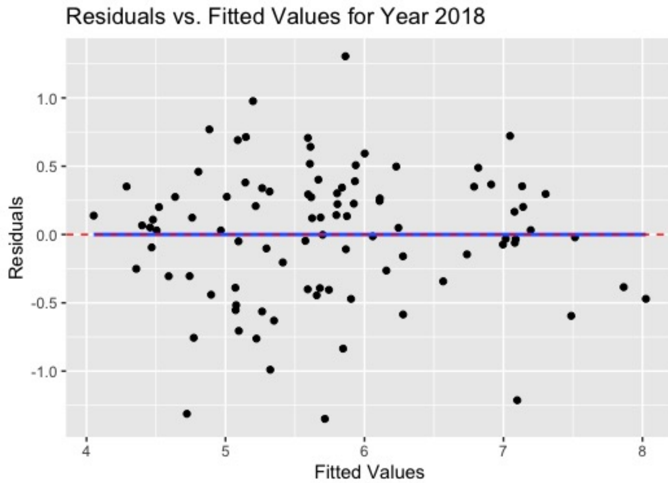


Figure 2: QQ-plots of residuals

The histograms of the residuals also seem to approximate a Gaussian curve, while the mean of the residuals is near zero (see R code for calculations), pointing to normality.



To check for homoscedasticity, we plotted the residuals in a scatter (Appendix C), from which we saw no evidence against constant variance.



4 Multivariate Linear Regression

Now that we showed that our model is reasonable we will proceed with the multivariate linear regression. Despite the fact that not all the development indicators are clearly correlated to happiness, we will nevertheless include them in the model, since they are relevant to our research question of which development indicators are most relevant to the happiness of a nation. Here, we will only display the results for 2018 (run R code to see results for other years). The regression yielded mixed results. The model's adjusted R-squared value ranged from 0.7327 to 0.7742, meaning

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                       6.02719    0.72864   8.272 1.60e-12 ***
GDP_growth_2018                  -1.17395    0.33574  -3.497 0.000752 ***
Current_account_2018             -0.46466    0.69672  -0.667 0.506627    
GDP_percapita_2018                2.65546    0.43448   6.112 2.88e-08 ***
Prevalence_Undernourishment_2018 -1.25718    0.46036  -2.731 0.007679 ** 
Life_Expectancy_2018              1.04201    0.70541   1.477 0.143327    
Inflation_2018                   -0.03802    1.16726  -0.033 0.974096    
Mortality_rate_2018               0.01902    0.59004   0.032 0.974366    
Unemployment_2018                -1.37053    0.35354  -3.877 0.000208 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5093 on 85 degrees of freedom
Multiple R-squared:  0.7747,	Adjusted R-squared:  0.7535 
F-statistic: 36.54 on 8 and 85 DF,  p-value: < 2.2e-16
```

Figure 3: Regression output - 2018

that it consistently explains more than 73% of the variance. Some of the covariates proved not to be statistically significant to the model ($p - value > 0.05$). Both current account and inflation were never statistically significant. Mortality also fared poorly as a covariate, only being statistically significant in the 2021 model. Undernourishment and life expectancy were each statistically significant for only two models (2018, 2019 and 2016, 2017, respectively). Surprisingly, GDP growth was statistically significant for all but two years, 2020 and 2021.

5 Model Selection

As discussed in the data visualization section and in the linear regression section, some covariates don't seem to be correlated to the Happiness level of a country. For this reason, we decided to check whether a simpler model could perform equally as well or better.

We decide to adopt three similar model selection techniques: step-up, step-down and step-both.

The step-up model starts form an empty model and add only relevant variables. The step-down model, on the other hand, starts from the full model, removing covariates that aren't relevant. The step-both method is a mixture of these two. We noticed how, for each year, the result form the three procedures were identical. However, the covariates included in the model vary from year to year. For this reason, we decided to keep count, for each covariate, of the number of years in which it was considered relevant by the model selection algorithm. We found out that "Inflation" and "Current account" weren't deemed relevant for any year, so we decided to consider a smaller model without the aforementioned independent variables.

Performing the linear regression on this model, we notice how the mean of the adjusted R-square for each year is higher than the one for the original model, going from 0.751999847059893 to 0.755442542171123, showing that we found a simpler yet more powerful model.

It's worth noting how GDP growth, which didn't seem to show high correlation with happiness from the scatter plots, appeared in the step-up model for 5 years out of 6.

6 Testing

Beside showing that the independent variable that we chose explain a relevant portion of the variance in happiness levels across the world, the main goal of our analysis is testing the hypothesis that social indicators are more relevant than economic indicators in predicting happiness.

We decided to use a two-sample t-test to check the alternative hypothesis that the percentage of variance in happiness explained by a model containing social variable only is greater than the one explained by a model containing only economic variables. To compute this we performed a linear regression between economic and then social covariates and happiness levels for each of the years we considered in our analysis. Then we computed the average of the adjusted R-square values and performed the t-test on the two values we obtained.

The p-value we obtained is 0.4368, way above any reasonable significance level and therefore we must retain the null hypothesis, that is, economic variables are at least as relevant as social variables. However, as we will further discuss in the "Limitations of the study" section, our sample of covariates was very small, containing eight indicators only, and the result we got could be entirely due to the covariates we selected.

Indeed, excluding Prevalence of Undernourishment from our pool of social indicators, we obtain a p-value of 2.895e-05 and we can therefore say that we have sufficient statistical evidence to reject the null hypothesis and accept the hypothesis that social indicators are more relevant. For the full results from the two t-test, please refer to Appendix F.

7 Limitations

One issue we faced with the dataset was that there were missing data for many of the years and countries. Since we opted to discard any countries with missing data, we ended up removing potentially valuable data that could have yielded clearer trends. This is particularly evident in the "Data Visualization" section, specifically by the aforementioned pattern of a 'bunching up' of data points for low values with trends emerging for greater values. This makes intuitive sense; take, for instance, undernourishment. If undernourishment is not prevalent in a country, it will not have a big effect on a population's happiness. But, the more common it is, the more it actively affects peoples' lives and therefore, also happiness scores. Furthermore, perhaps selecting a wider range of indicators could have also helped us in our analysis, since limiting the analysis to 5 social and 5 economic indicators, then reduced to 3 and 5 during data cleaning, means that we may not have captured the general trend for economic and social indicators.

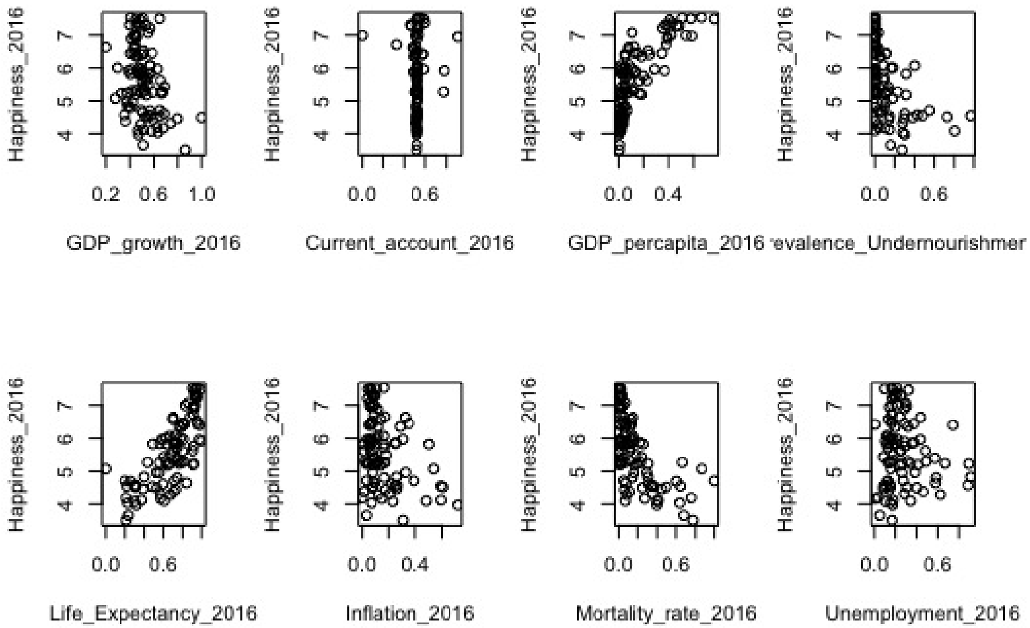
Another potential issue with the analysis is that, although we are treating the variables as independent, economic and social indicators are correlated to one another. To prevent this from affecting the results, perhaps it would help to generate a correlational matrix between variables and discard any above a threshold value.

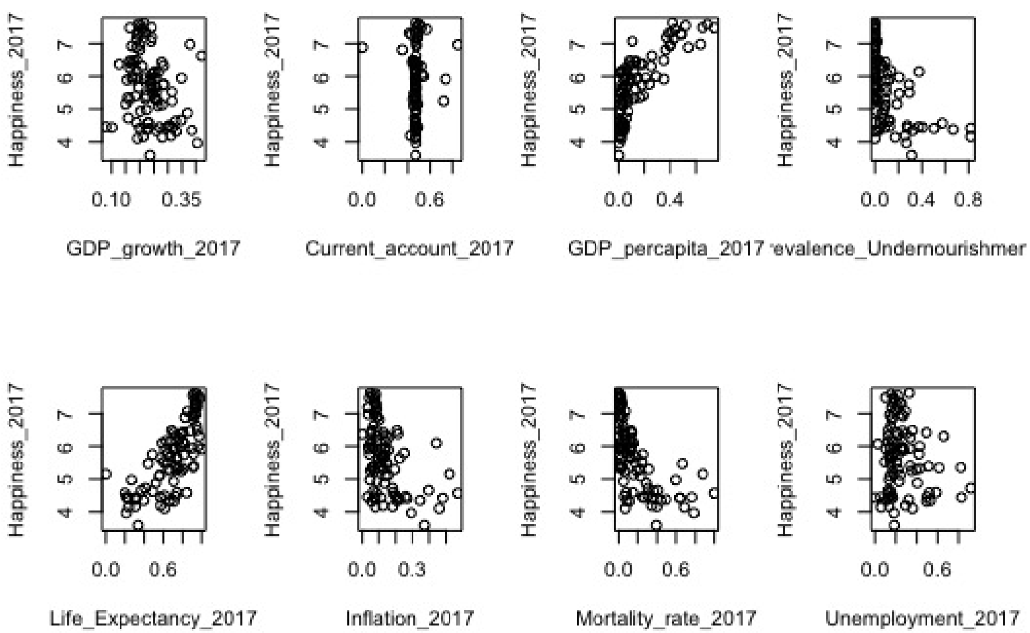
Moreover, as mentioned in the "Testing" section, it is not possible to draw any conclusion with strong statistical significance from our analysis of the relevance of social indicator against economic indicators since we only had 8 samples in total. This lack of data lead the test to be very sensible to the indicators that we chose, which might not be representative of economic and social variables as a whole. Indeed, as previously shown, changing slightly the selection of social covariates, excluding Unemployment, gives a completely different result, yielding a very low p-value and allowing us to draw a very different conclusion.

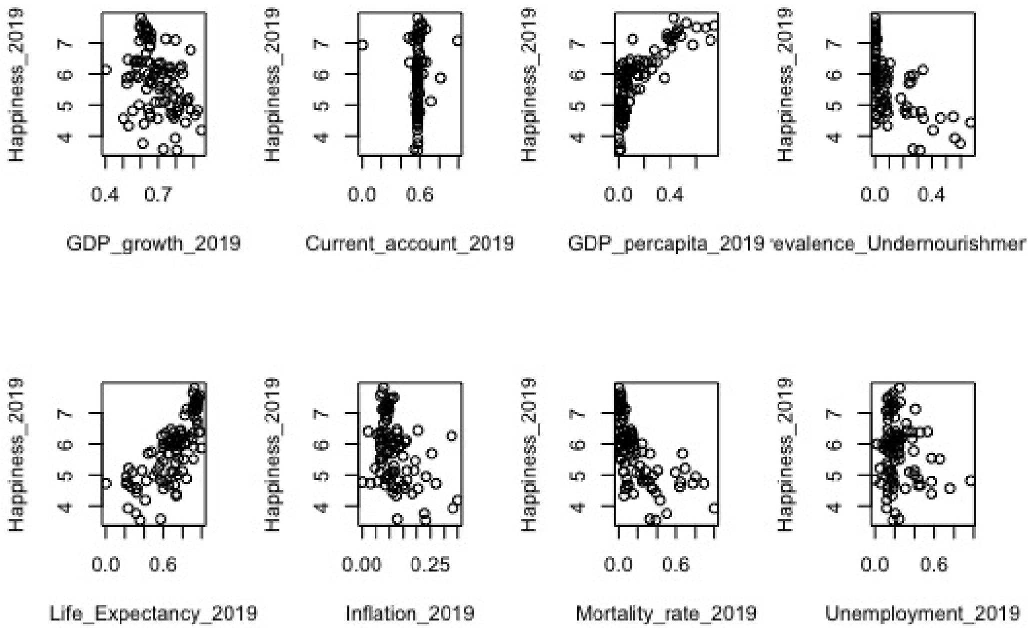
Finally, we would be remiss not to mention the fact that the data we were dealing with, happiness, is highly subjective and difficult to quantify. Therefore, all results with regards to predicting it should be taken with a grain of salt.

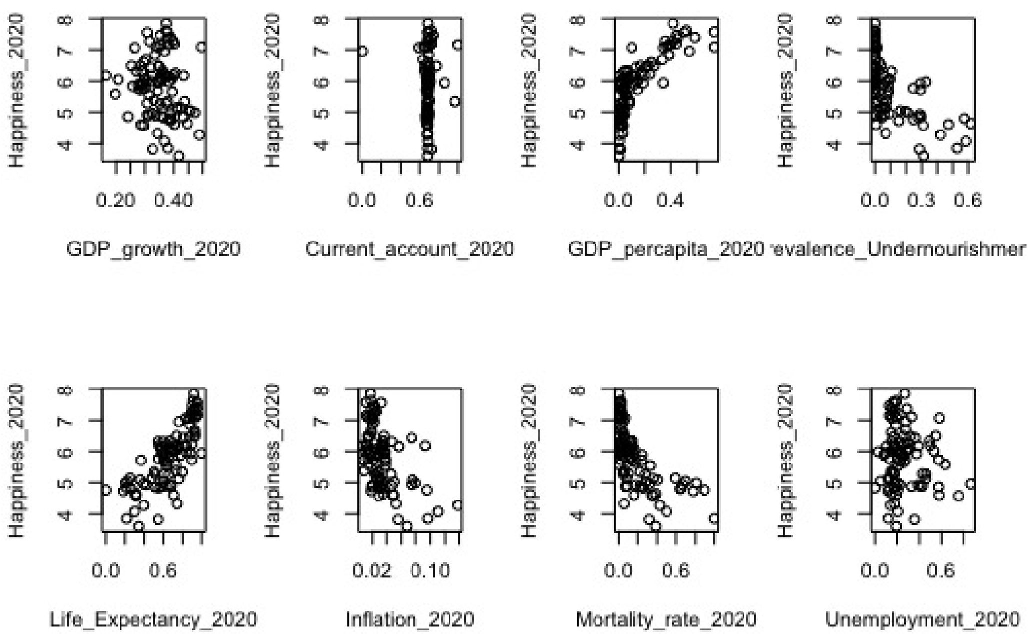
8 Appendix

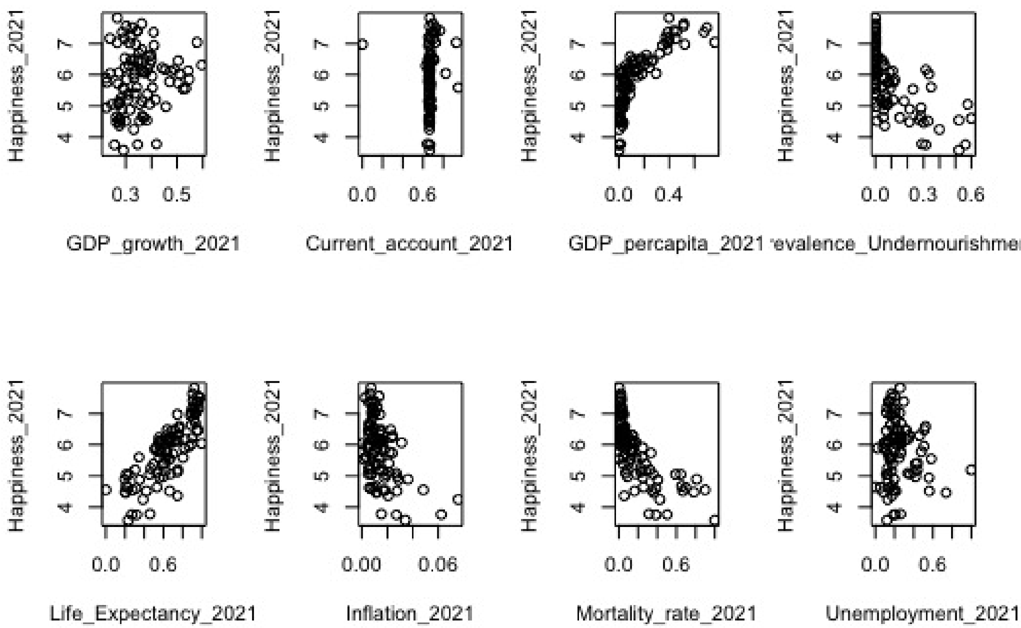
8.1 Appendix A: Scatter Plots






Happiness_2019
GDP_growth_2019
Current_account_2019
GDP_percapita_2019
'evalence_Undernourishmer
Life_Expectancy_2019
Inflation_2019
Mortality_rate_2019
Unemployment_2019


Happiness_2020
GDP_growth_2020
Current_account_2020
GDP_percapita_2020
'evalence_Undernourishmer
Life_Expectancy_2020
Inflation_2020
Mortality_rate_2020
Unemployment_2020

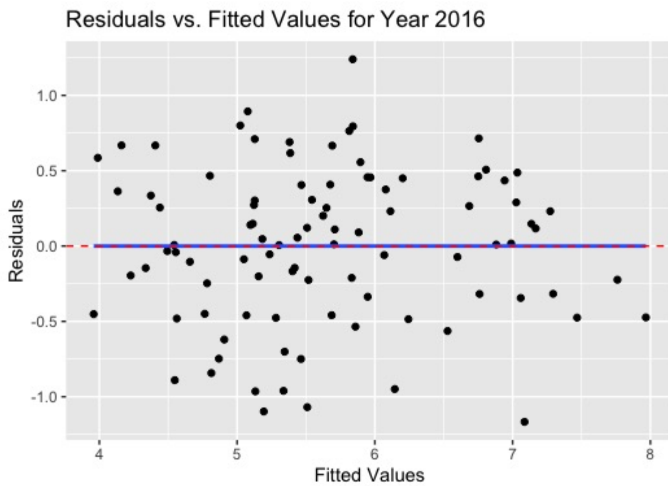


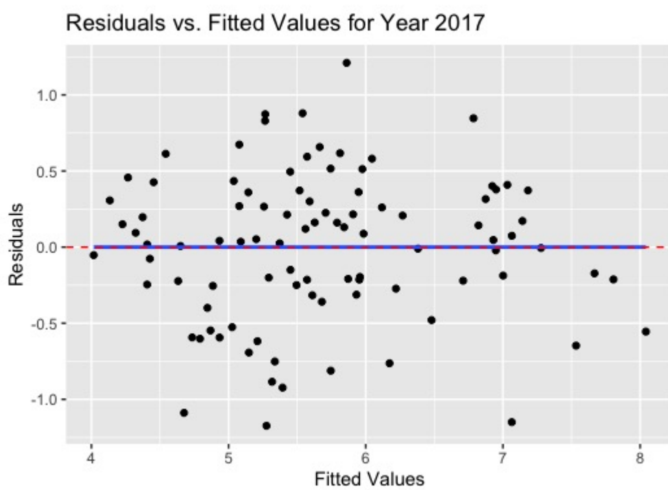
8.2 Appendix B: Shapiro-Wilk Test

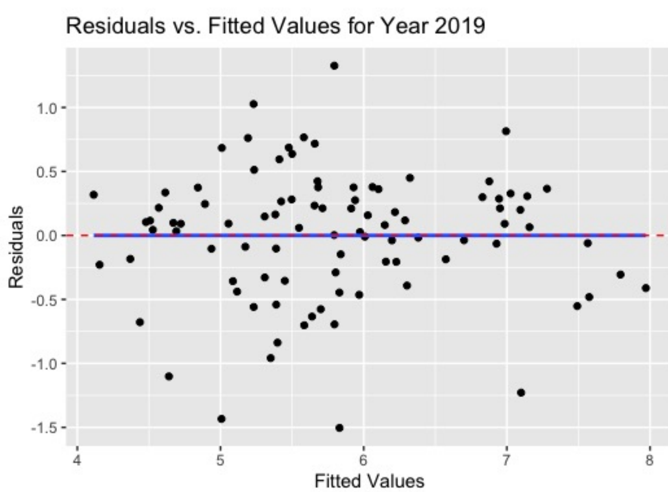
```
$`2016`

        Shapiro-Wilk normality test

data:  residuals
W = 0.98466, p-value = 0.3417


$`2017`

        Shapiro-Wilk normality test

data:  residuals
W = 0.98793, p-value = 0.5481


$`2018`

        Shapiro-Wilk normality test

data:  residuals
W = 0.97951, p-value = 0.1472
```

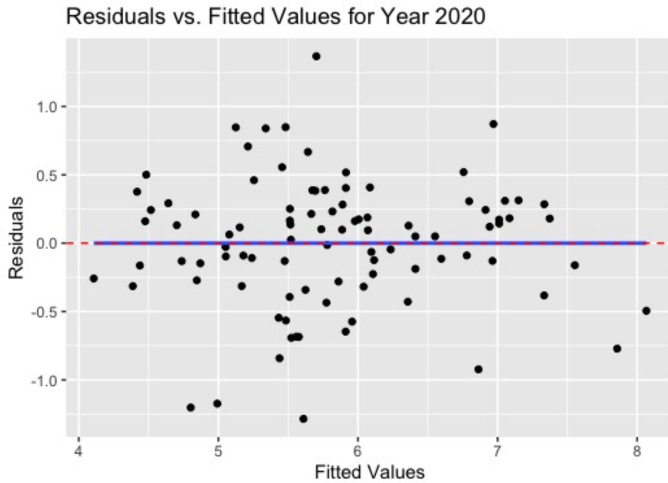
```
$`2019`

        Shapiro-Wilk normality test

data:  residuals
W = 0.97147, p-value = 0.03748


$`2020`

        Shapiro-Wilk normality test

data:  residuals
W = 0.97943, p-value = 0.1452


$`2021`

        Shapiro-Wilk normality test

data:  residuals
W = 0.97311, p-value = 0.04952
```

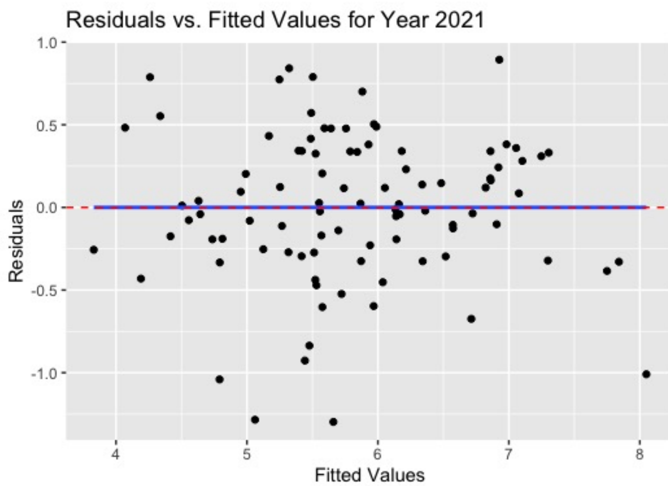
8.3 Appendix C: Residuals Scatter











8.4 Appendix D: t-tests results

```
data:  exp_var_nonecon and exp_var_econ
t = 0.16005, df = 46.938, p-value = 0.4368
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.1107806        Inf
sample estimates:
mean of x mean of y
0.2397118 0.2280311
```

Figure 4: t-test - full model

```
        Welch Two Sample t-test

data:  exp_var_nonecon_reduced and exp_var_econ
t = 4.6201, df = 32.511, p-value = 2.895e-05
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1726852       Inf
sample estimates:
mean of x mean of y
0.5006039 0.2280311
```

Figure 5: t-test - no Prevalence Undernourishment