**OLS Regression Analysis of Factors Affecting Child Poverty Rates in OECD Countries**

Rameasa Arna

2023-03-26

# Introduction

This report provides an OLS regression analysis conducted on a subset of the OECD dataset to explore the potential predictors associated with child poverty rates. Child poverty rates are an essential indicator of the economic well-being of individuals and the economy in OECD countries.

The outcome variable selected for the OLS model is Children Poverty Rate, which represents the percentage of child living in poverty in OECD countries. This outcome variable is significant as it reflects the economic well-being of people and the state of the economy in these countries.

The analysis considers three predictors, namely total fertility rate, maternity leave duration, and youth unemployment rate. Each predictor variable’s significance and potential impact on the percentage of people living in poverty are explained below:

1. Total Fertility Rate: The total fertility rate measures the number of children born to women of childbearing age. This predictor is crucial as a high fertility rate can lead to a larger population, which may put pressure on resources and increase the likelihood of poverty.
2. Maternity leave: Maternity leave duration indicates the length of time a mother is entitled to take off work after giving birth. This predictor is important because a longer maternity leave duration can provide women with the necessary time and support to take care of their children, reducing the likelihood of poverty.
3. Unemployment Rate of Youth: The unemployment rate represents the percentage of the labor force that is unemployed. This predictor is essential as a high unemployment rate may indicate a weak economy, increasing the likelihood of poverty.

# 

# Model

**Table 1 : Summary of Multiple OLS Regression Analysis for Predicting Child Poverty using Total Fertility, Maternity Leave and Unemployment Rate of Youth Data**

|  |
| --- |
| Independent Variable Coefficient prob. |

tot.fert -8.71\*\*\*(1.92) 0.001

mleave -0.04\*\*(0.012) 0.003

unempl.y 0.41\*\*(0.100) 0.001

R-Square = 0.81

Adjusted R-square = 0.77

\*p< 0.05. \*\*p<0.01. \*\*\*p<0.001

Our analysis shows that the percentage of children living in poverty in OECD countries is influenced by several factors. Firstly, we found a negative relationship between the total fertility rate and the percentage of children living in poverty. This means that as the total fertility rate increases, the percentage of children living in poverty decreases. Specifically, for every additional child per woman, the percentage of children living in poverty decreases by 8.71%.

Secondly, we found a negative relationship between maternity leave duration and the percentage of children living in poverty. This means that as maternity leave duration increases, the percentage of people living in poverty decreases. Specifically, for every additional week of maternity leave, the percentage of people living in poverty decreases by 0.04%.

Finally, we found a positive relationship between youth unemployment rate and the percentage of people living in poverty. This means that as the youth unemployment rate increases, the percentage of children living in poverty also increases. Specifically, for every one percentage increase in the youth unemployment rate, the percentage of children living in poverty increases by 0.41%.

Overall, our model shows that these three factors explain a substantial amount of the variability in the percentage of children living in poverty in OECD countries. We are confident in our findings, as our model’s fit was good with a high adjusted R-squared value of 0.7751 and a low residual standard error of 1.974. The model as a whole was statistically significant at the 0.001 level.

# Regression Diagnostics

In order to assess the accuracy and reliability of our regression model, we will conduct various diagnostic tests.

## 1. Normally-Distributed, Mean Zero Errors:

We will use the Zero-Mean Assumption, Q-Q plot and Shapiro-Wilk Normality Test to verify the distribution.

### Zero-Mean Assumption:

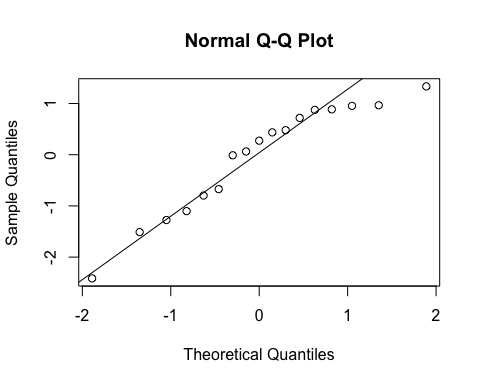
If the mean of the residuals is close to zero, then the zero-mean assumption of the regression model is satisfied.

**Table 2: Descriptive Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min | 1st Quartile | Median | Mean | 3rd Quartile | Max. |
| -2.41 | -0.79 | 0.27 | -0.04 | 0.87 | 1.33 |

The summary of the residuals shows that the mean value is close to zero (-0.04), which is a good indication that the zero-mean assumption is met. Additionally, the minimum and maximum values of the residuals are within acceptable ranges (-2.41 to 1.33). Therefore, we can conclude that the zero-mean assumption is valid.

### Q-Q plot :

The q-q plot is a graph that plots the quantiles of the residuals against the theoretical normal distribution. If the residuals are normally distributed, the points on the q-q plot should form a straight line. 

The q-q plot shows that the residuals follow a roughly normal distribution.

### The Shapiro- Wilk Normality Test :

The Shapiro-Wilk normality test is a statistical test used to determine whether a dataset follows a normal distribution. In this case, the null hypothesis is that the residuals follow a normal distribution, and the alternative hypothesis is that they do not.

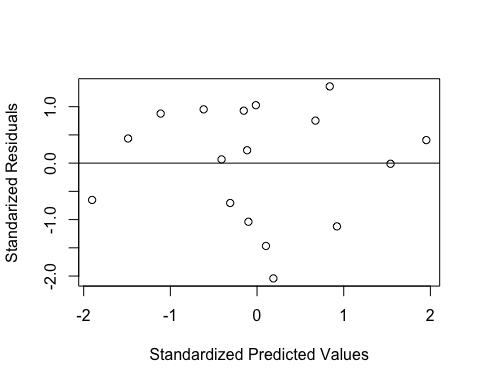
## Shapiro-Wilk normality test  
## data: resid.1  
## W = 0.92044, p-value = 0.1502

The test output shows that the p-value is 0.1502, which is greater than the significance level of 0.05. Therefore, we fail to reject the null hypothesis, indicating that the residuals are normally distributed.

Based on these tests, we can conclude that the normality assumption is satisfied.

# 2. Heteroscedasticity:

To check for heteroscedasticity, we can plot the standardized residuals against the standardized predicted values. If there is a pattern in the plot, then it indicates that the variance of the residuals is not constant across the range of predicted values.



From the plot, we can observe that there is no clear pattern in the scatterplot, which suggests that there is no significant heteroscedasticity present in the model.

We can also perform a statistical test to confirm the absence of heteroscedasticity. The studentized Breusch-Pagan test is a commonly used test for heteroscedasticity in linear regression. The null hypothesis is that the variance of the errors is constant across predicted values. A significant p-value (less than 0.05) would indicate that the null hypothesis can be rejected, and that heteroscedasticity is present.

## studentized Breusch-Pagan test  
## data: model1  
## BP = 5.2818, df = 3, p-value = 0.1523

The studentized Breusch-Pagan test output shows a p-value of 0.1523, which is greater than the significance level of 0.05. Thus, we fail to reject the null hypothesis, suggesting that the variance of the errors is constant across the predicted values and that there is no significant heteroscedasticity present in the model.

# 3. Collinearity:

To check for collinearity in our model, we can use the variance inflation factor (VIF). The VIF measures the extent to which the variance of the estimated regression coefficient is increased due to multicollinearity in the predictors.The variables are treated as potentially-collinear if their VIFs approach or exceed 4. If a VIF is greater than 10, it is probably a serious problem.

**Table 3: VIF Test**

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | tot.fert | mleave | unempl.y |
| VIF | 1.05 | 1.01 | 1.04 |

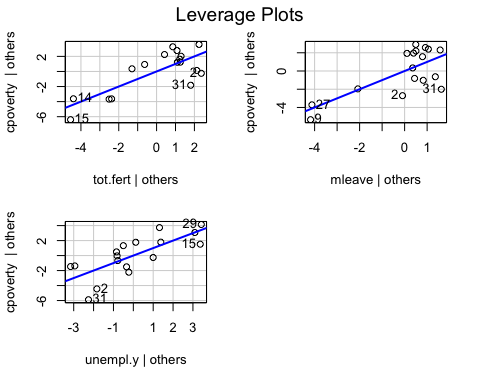
The VIF values are all close to 1, indicating that the predictors are not highly correlated with each other. The test suggests that collinearity is probably not a problem. Therefore, we can conclude that collinearity is probably not a problem in this model.

# 4. Outliers :

To detect outliers, we can use leverage plots and the Bonferroni outlier test.

### Leverage Plots:

One way to ascertain potential influential observations is through leverage plots. These plots give the partial relationship between the model outcome and individual predictors



From the graph we can conclude that there is no obvious outlier problems.

### Bonferroni Outlier Test:

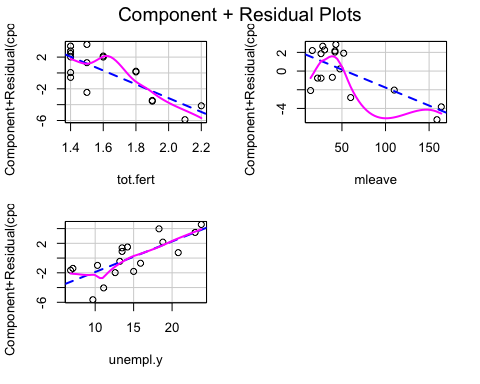
The Bonferroni outlier test is a statistical test that detects outliers by comparing the p-value of each observation’s standardized residuals to a modified significance level. If the p-value is less than the modified significance level, we can conclude that the observation is an outlier.

## No Studentized residuals with Bonferroni p < 0.05  
## Largest |rstudent|:  
## rstudent unadjusted p-value Bonferroni p  
## 31 -2.415938 0.032558 0.55349

The output of the Bonferroni outlier test shows that there are no significant outliers in the data.

# Linearity:

# Linearity is one of the key assumptions of linear regression, and it means that the relationship between the predictor variables and the outcome variable should be linear.



Here, the predictor variable and the outcome variable are roughly linear.

# Conclusion:

Based on the regression diagnostics conducted on our model, we can conclude that the model is accurate and reliable.The analysis conducted in this study demonstrated that total fertility rate, maternity leave duration, and youth unemployment rate all have a significant impact on the percentage of children living in poverty. These findings highlight the need for policymakers to focus on increasing maternity leave duration and decreasing youth unemployment rates while also considering the potential implications of total fertility rate on resource allocation and population growth. Overall, this study provides valuable insights into the intricate relationships between social, economic, and demographic factors that contribute to child poverty rates.