RELATIONSHIPS BETWEEN HAPPINESS AND INCOME INEQUALITY

by

Mats Karlsen

A Bachelor's Thesis

Handelshøgskolen UiS

Supervisor: Sara Helene Rønningstad

BØKBAO 2024

Abstract

This thesis explores the relationship between income inequality and societal happiness. Using the Gini coefficient, and Life Ladder scores from the World Happiness Report, it investigates the impact of economic differences on a national level of well-being. The study employs ordinary least squares regression analyses to examine existing theories such as the Easterlin Paradox, considering GDP per capita as well as time and country-fixed effects.

The results reveal an inverse correlation between income inequality and happiness. As income inequality rises, happiness tends to fall. Unexpectedly this pattern did not hold true when accounting for a nation's overall wealth. Happiness increased alongside income inequality when the analysis included GDP per capita. This thesis emphasizes the importance of fair wealth distribution in enhancing societal happiness. Lastly, it suggests that further research is needed to deepen the understanding of these dynamics.

Table of Contents

1 Introduction

In an era with stark contrasts between the wealthy and the less fortunate, the question of how economic inequalities impact our well-being is crucial. This thesis seeks to examine the relationship between income inequality and how it affects happiness in various countries, a subject at the core of economic and social sciences. This subject is important, as the pursuit of happiness is a fundamental human goal, and understanding how income inequality may influence this pursuit, can help shape policies that intend to improve life satisfaction across societies.

Using the Gini coefficient as a measure of income inequality, and happiness ratings from the World Happiness Report, this study provides statistical analyses to get empirical insights into the relationship that links economic conditions with psychological states. Why do countries with similar economic statuses report different levels of happiness? Can higher levels of GDP compensate for happiness lost to income inequality? Have countries become happier over time? Do cultural differences affect happiness? This thesis presents a complex picture of this relationship, influenced by a blend of economic, cultural, social, and policy-driven factors. More specifically, this thesis will investigate the following hypotheses:

Hypothesis 1: There is a significant relationship between income inequality and reported levels of happiness. Specifically, as income inequality increases, happiness levels decrease.

Hypothesis 2: The negative relationship between income inequality and happiness remains significant even after controlling for the GDP per capita. Confirmation of this relationship would suggest that the impact of income inequality on happiness does not depend on a country's overall economic wealth.

Hypothesis 3: The negative relationship between income inequality and happiness remains significant after controlling for year, indicating that the observed relationship is not driven by underlying time trends in the variables.

Hypothesis 4: The negative relationship between income inequality and happiness remains significant after controlling for country-specific effects, suggesting that the relationship is not driven by countries' unique characteristics.

The first chapter after the introduction lays the foundation for this thesis by outlining the theories from existing literature. Informing about our current understandings of happiness in the context of economic disparities and inequality, the thesis builds upon the works on the Easterlin Paradox (Easterlin & O'Connor, 2022) and surrounding studies on the same topics.

The method chapter focuses on the methodological choices used to perform the statistical analyses in this thesis. It describes the ordinary least squares regression analyses utilized to examine the correspondence between the Gini coefficient, a measure of income inequality, and the self-reported Life Ladder values, measuring happiness.

Chapter 4 covers the empirical backbone of this thesis, the datasets. This chapter introduces the reader to the Life Ladder dataset and the Gini coefficient dataset. It discusses their origin, their applications, and their implications of usage in this thesis, as well as descriptive statistics.

The analysis chapter is the analytical core of the thesis. Here the datasets are utilized for regression analyses, controlling for different variables. Both linear and multiple regression models are examined to discern trends and patterns, providing empirical results to the theoretical material discussed earlier.

At last, the discussion seeks to reflect and interpret the essence of the thesis in conclusion. Discussing the key findings, comparing these to the existing literature discussed earlier, as well as interpreting these results. This chapter continues by discussing the limitations to this thesis and gives recommendations for future research as well as reflecting on the methodological choices made, before finally concluding the thesis.

2 Theory

Richard A. Easterlin and Kelsey J. O’Connor’s article (2022), describes a phenomenon now known as the Easterlin Paradox. Originally formulated in 1974, the Easterlin Paradox states that economic growth and happiness varies directly with each other at one point in time, but over long-term, economic growth does not necessarily result in increased happiness. Easterlin and O’Connor primarily attributes this to social comparison, in which a person will assess their own income relative to others. In short-term individuals with higher income will be happier when comparing their income to those who are less fortunate, and reversely for individuals with lower income. However, in the long-term overall income rises due to economic growth and the positive effects of the individual’s income growth are neutralized by the simultaneous income growth, in the group they were comparing themselves to. This has been consistent over nearly five decades of research, when considering the trends of GDP and happiness. Still, Easterlin says that happiness could be increased by introducing policies promoting strong social safety nets and full employment, even at low levels of GDP per capita.

A previous study by Shigehiro Oishi and Selin Kesebir (2015) try to provide evidence that the Easterlin paradox can be partly explained by its relationship with income inequality. Using Veenhoven’s World Database of happiness, an extensive archive researching happiness and life satisfaction, and Latinobarómetro data, an annual public survey conducted in 18 Latin American countries, and Spain, as their datasets. Both datasets measuring life satisfaction on a 4-point scale, covering a total of 34 countries. Veenhoven’s World Database of happiness used data spanning from 1959 to 2006 and the Latinobarómetro were available from 2001 and 2003 to 2009. The authors then used a weighted regression analysis and examined the within-nation correlations between GDP per capita and life satisfaction across the 34 countries, while considering the means of the Gini coefficient and log-transformed GDP per capita. Their results showed that economic growth was not associated with increases in happiness when it at the same time was accompanied by growing income inequality. They propose these findings are significant considering the rising income inequality observed in most parts of the world, implying that even more instances of the Easterlin Paradox may be seen in the future. They conclude with stating that these findings may suggest that to raise nationwide happiness, a more even distribution of growth in national wealth may be in order.

The 2011 study “Income Inequality and Happiness” (Oishi et al., 2011) used General Social Survey to examine connections between income inequality and happiness in the United States. Using data from 1972 to 2008, the authors found that in America individuals were on average happier in the years where the national income inequality was lower than the years where the national income inequality were high. They further try to explain this inverse relationship between income inequality and happiness by the people’s perceived fairness and general trust. During the years with higher national income inequality Americans trusted other people less and perceived other people to be less fair, than in the years with lower national income inequality. This negative association was strong for the lower-income samples but did not hold for the higher-income samples. The link between lower-income households and income inequality, was not found to be explained by lower income, but by the perceived unfairness and lack of trust. Therefore, policies trying to reduce income inequality could contribute to higher happiness, especially in the lower-income class.

Bjørnskov. et.al. (2013) also argues that the perceived fairness of income generation affects subjective well-being. Their study extends previous research by adding individuals’ subjective perceptions of fairness in the income generation process and varying levels of actual fairness across countries in their analysis. The study uses a pseudo micro panel, with data collected from different individuals at different time points, from the World Values Survey. It measures income inequality using Gini coefficients and individuals’ fairness perception concerning the income generation process. They argue that there are systematic differences regarding this, between countries, that are characterized by a high or low level of actual fairness. According to Bjørnskov.et.al. the empirical literature on the relationship between happiness and income inequality, spanning from 1990 to 2008, has yielded ambiguous results. They suggest that the cause of this confusion might be because people evaluate the fairness of income distribution differently, and that eventually these subjective evaluations can affect their subjective well-being. Their findings challenge the standard argument that more redistribution and less income inequality unambiguously led to an increase in welfare of the average person. Their findings leave out several questions for future research, such as looking at perceived fairness in the long run. Their results also suggest that to foster a subjective well-being, it would be preferable with a society that offers equal opportunities, compared to an overly redistributive welfare state.

This literature shows that the relationship between income inequality and happiness is complex with many aspects to take into consideration. It suggests that outcomes depend on factors such as national contexts and distribution of economic gains. While some findings are consistent, including that higher inequality often correlates with lower happiness, both the mechanisms measuring the relationship, and the significance of the relationship varies. Influenced by both objective levels of inequality and subjective perceptions of that inequality, policies for economic growth and redistribution needs to consider the psychological and social dimensions to effectively increase society’s well-being. Based on this, moving forward it’s expected to find a negative relationship between happiness and income inequality. However, a possibly complex and nuanced relationship that varies greatly from psychological and social factors as well as economic factors, is expected.

3 Method

In the following paragraph, the method of this paper will be described. First ordinary least squares regression analysis is used to explore the relationships between happiness and income inequality. Then the key assumptions for regression analysis will be reviewed, such as linearity, independence, homoscedasticity, and normality. Finally important concepts for interpreting the regression results, including coefficients, significance levels and explanatory power, will be described.

The simple linear regression is an approach for predicting a quantitative response (Y), based on a predictor variable (X). This model assumes that there is an approximately linear relationship between the variables (James et al., 2017, p. 61). The intercept and the slope for this model are both unknowns, this means that observations to settle the values is needed. In most cases the observations tend to be scattered and subjects to errors. The observations then do not fit a straight line. To find the line that best fits the observed data points, the method called ordinary least squares (OLS) is used. OLS finds this line by minimizing the sum of the squared differences between the model’s predicted values and the observed values. The line should then be in such a way that the line has the minimum distance between each point and the line (Alto, 2023).

In this simple linear regression model, one explanatory variable to explain the development of the dependent variable is used. Continuing, it is favourably to factor in other variables that may influence the dependent variable. Multiple regression models allow for this. A multiple regression model is simply put a regression model that includes more than one independent variable (Hayes, 2020). The model extends the simple linear regression model so that it can accommodate multiple variables directly. This can be done by giving each variable a separate slope coefficient in the same model. Following the same steps as previously, the same ordinary least squares approach to estimate the parameters is used (Uboe, 2017, p. 263). The reason for using multiple regressions is that it extends the concept of simple linear regression by adding more dimensions that allow for a more comprehensive analysis of how the various factors can influence the outcome simultaneously. This makes it more applicable to real-world scenarios where outcomes are often affected by more than one factor.

When performing the regression, there are some classical linear model assumptions that form the foundation for the analysis. These assumptions exist for the validity of the statistical interference. These six assumptions are respectfully: linear in parameters, random sampling, no perfect collinearity, zero conditional mean of error terms (residuals), homoskedasticity, normality of error terms (Onozaka, 2024). In a linear regression model, it is assumed that the relationship between the independent and dependent variables are a straight-line relationship. If the relationship is not linear, then all conclusions drawn from the regression are suspect. Random sampling implies that the data points used in the analysis are drawn from the relevant population randomly. This is to ensure that the sample is representative of the population and helps to reduce selection bias. Collinearity refers to when two or more independent variables in the model are highly correlated to each other. This should be avoided to ensure that each independent variable provides its own unique information to the model. The assumption of zero conditional mean of error terms states that the expected value of the error term is zero, given the values of the independent variables. Homoskedasticity means that the spread of the residuals around the regression model line remains constant. This ensures that the coefficients are unbiased and accurate, opposed to heteroskedasticity where the variance of the error terms varies across the observations, which leads to inaccurate coefficients and biased standard errors. The normality of error terms assumption states that the residuals are normally distributed with a mean of zero (‘Assumptions of OLS’, 2016).

To understand the results, it is important to correctly interpret the coefficients and p-values from the regression models. This involves understanding the relationship between the independent variables (predictors) and the dependent variable (outcome). The value of the coefficient represents the size of the impact the independent variable has on the dependent variable. A positive coefficient means that as the value of the independent variable increases, the mean of the dependent variable also increases, and reversely for a negative coefficient (Frost, 2017). For example, if an independent variable has the coefficient value of 2.0, for each one unit increase in that variable, the dependent variable is expected to increase by 2.0 units, assuming all other variables are held constant.

The p-value is measuring the probability that the results given could have occurred by chance if there is actually no association between the independent and dependent variables in the population. The threshold used for declaring statistical significance is $p < 0.05$. When the p-value is less than 0.05 (5%), the sample data provide enough evidence to reject the null hypothesis of no effect and conclude that there is a statistically significant effect. Oppositely if the p-value is greater than the significance level at 0.05, there is insufficient evidence in the data sample to conclude a non-zero correlation (Frost, 2017).

The coefficient of determination (R^2) is used to determine how much of the variance in the dependent variable is explained by the independent variables included in the regression model. It always takes on a value between 0 and 1, where 0 means none of the independent variables explains any of the variance in the dependent variable. A value of 1 indicates that all the variance in the dependent variable is explained by the independent variables. Simply put R2 measures the proportion of variability in Y (dependent variable) that can be explained using X (independent variable) (Uboe, 2017, p. 252).

To evaluate the goodness of fits of a regression model, the graphical tool called residuals vs. fitted is used. In the residual plot the residuals are plotted on the y-axis and the fitted (predicted) values on the x-axis. (*Applied Regression Analysis*, 2018). The horizontal line where residual equals 0 represents where the residuals would be if a model made perfect predictions. Ideally the residuals should be symmetrically distributed around this line, without clear patterns. If the residual spread changes across the fitted values, this could mean the model shows heteroscedasticity, meaning the error variance is not constant, and that there may be a non-linearity the model don’t account for. There should be no patterns in the residual plot, it should not be in the shape or a trend of curves, waves, cones, or systematic structures. Points that are far away from the rest of the data can be outliers, if outliers have large residuals they can have a big impact on the regression model and should be further investigated (Frost, 2017).

4 Data

To perform the regression analyses, the datasets used are quantitative. These datasets are The Life Ladder measurement from The World Happiness Report 2023 (Helliwell et al., 2023), downloaded from their official website, and the Gini Coefficient Index (after tax), downloaded from Our World in Data (Hasell et al., 2024). The datasets are panel data, also known as longitudinal data. Panel data combines cross-sectional and time series data by following the same subjects over a period of time. This means it contains data observations collected at a regular, chronologically frequency and contains observations across multiple collections of individuals (2019).

4.1 World Happiness Report

The first dataset used is the Life Ladder which is countries own self-measured level of happiness. The dataset is reported by The World Happiness Report, a publication of the Sustainable Development Solutions Network, a global initiative of the United Nations (Helliwell et al., 2023). The report includes 137 countries with a year range from 2005 to 2022. The World Happiness Report use data from the Gallup World Poll surveys. The results are based on answers to the Cantril Ladder: it asks participants to imagine a ladder where the top of the ladder, being numbered 10, represents the best possible life for them, and the bottom, numbered 0, represents the worst possible life. The World Happiness Report first used this Gallup World Poll metric in their 2012 report and have been publishing reports annually ever since 2016. The median level of happiness measured by the Life Ladder variable in this dataset is 5.43 while the mean is 5.48. The highest measured Life Ladder score was 8.02 in Denmark 2005, while the lowest recorded Life Ladder score belonged to Afghanistan in 2022 at 1.28. Each year the number of countries and people surveyed varies somewhat, but approximately more than 100,000 people in 130 countries participate in the survey each year. For each country the typical annual sample size is 1000 people. The World Happiness Report states that this sample size is adequate to give a good estimate at a national level, as confirmed by their 95% confidence intervals for each country in the dataset.

4.2 Gini Coefficient

The Gini Coefficient is intended to measure the extent of which the distribution of income among households or individuals within an economy deviates from a perfectly equal distribution (*Gini Index | DataBank*, 2024). It measures the income inequality on a scale from 0 to 1, where the 0 represents a perfect equality, where everyone would have the exact same income, while the 1 (or 100%) represents a perfect inequality, where one individual receives all income and everyone else receives nothing. The reason for using the Gini coefficient after tax instead of before tax, is that it better captures the impact of taxes and transfers on income distribution, and government interventions. This provides a more comprehensive view of the income the people actually receive. The median Gini coefficient in this dataset was 0.36, while the mean was 0.38. The highest measured value at 0.658 belonged to Malawi in 1997, oppositely the lowest value belonged to China in 1984 at 0.178. The dataset contains 184 different countries with a year range of 1967 to 2021. Countries not included are some known for small populations such as Andorra, Monaco, Liechtenstein, and San Marino. There are also some high-income countries missing such as Singapore, Qatar, Kuwait, Oman, and Saudi Arabia. Especially the wealthy middle eastern countries with significant oil wealth often have different income distribution characteristics. A significant number of island nations, such as the Bahamas, Barbados, Fiji, and the Maldives, are also missing. Island nations often have unique economic structures and are therefore often excluded from global datasets.

The calculation of the Gini coefficient involves several steps, usually visualised through the Lorenz curve. Individuals in a population are arranged in ascending order by their levels of income, from lowest to highest. Then the cumulative share of income by each percentile of the population is calculated. The Lorenz curve is plotted with the cumulative percentage of the population on the horizontal axis, and the cumulative percentage of the income on the vertical axis (Damgaard, 2024). A line at 45 degrees would mean a perfect equality of income distribution, called the line of equality. The more the Lorenz curve bows away from this line, the greater the inequality. The Gini coefficient is then calculated as the ratio of the area between the line of equality and the Lorenz curve, over the total area under the line of equality. If A is the area between the Lorenz curve and the line of equality, and B is the area under the Lorenz curve, the Gini coefficient G is calculated as $G = A/(A+B)$. The Gini coefficient can also be calculated directly from the Lorenz curve, as $G = 1-2B$ since $A+B=0.5$ (Hasell & Roser, 2023).

It should be noted that there are some issues in interpreting a Gini coefficient, as a single value may result from many different distribution curves. This means that two countries with the same Gini coefficient might have very different underlying income distributions among their populations. One should therefore take into account the demographic structure. A country with an aging population would have a larger proportion of their population being retired, and no longer earning wages. This demographic would then tend to have lower income levels than working adults, which can skew the income distribution. This could suggest a growing inequality even if the income distribution among the working adults has not changed. It would simply be the demographic that shifted, rather than a change in income inequality, which increased the Gini coefficient. Another example of this is countries with increased birth rates. With more people in the population not earning income, the Gini coefficient might increase, without any real change in income inequality among those working (Sung, 2010).

5 Analysis

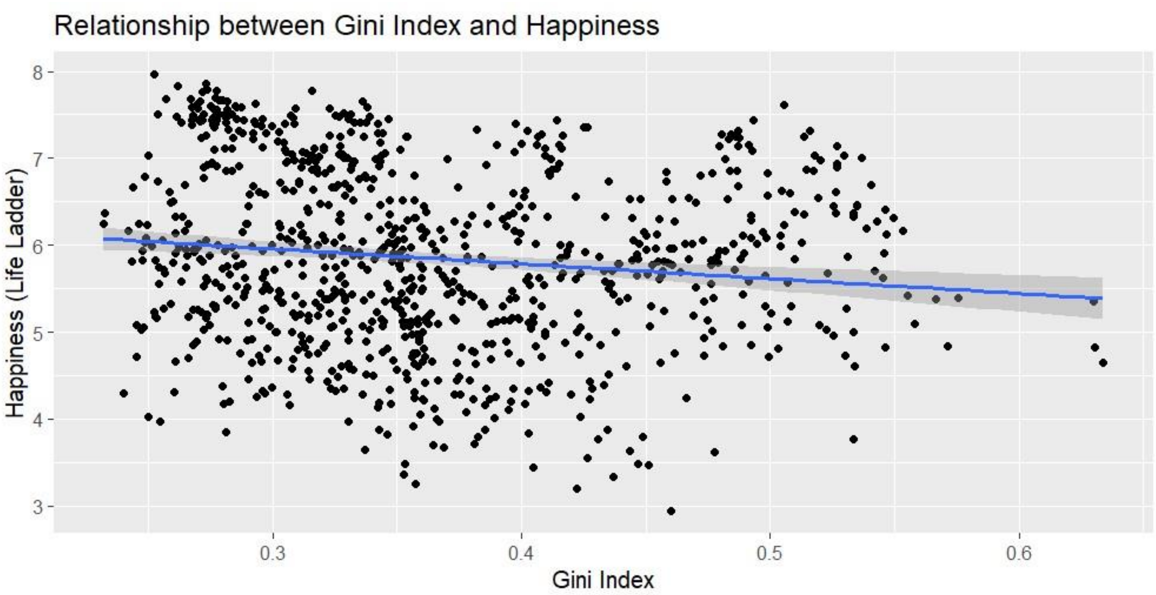
In the following paragraphs I will show results from four different regression models. Using both linear and multiple regression models factoring in different variables, the regression models will examine trends and patterns in countries' levels of happiness and its relationship with income inequality. The first model being a simple linear regression model including the Gini's coefficients effect on the dependent variable Life Ladder, which measures happiness. Secondly, adding the variable GDP per capita to the model to see what effects this would have on the Life Ladder variable as well. Thirdly, another regression model including both the effects of Gini and Year. Lastly, the model included the independent variables Gini and Country and their effect on the dependent variable Life Ladder.

5.1 Life Ladder ~ Gini

Table 1: Regression Analysis of Gini Coefficient's impact on Life Ladder

	Dependent variable:
	LifeLadder
Gini	-1.708^{***}
	(0.430)
Constant	6.470^{***}
	(0.160)
Observations	901
R^2	0.017
Adjusted R^2	0.016
Residual Std. Error	1.018 (df = 899)
F Statistic	15.759^{***} (df = 1; 899)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 1: Relationship between Gini Coefficient and Happiness



There is a statistically significant inverse relationship between countries' happiness (LifeLadder) and income inequality (Gini). From the regression analysis performed, the coefficient associated with the Gini Index, is -1.708, with a standard error of 0.43. It is statistically significant with a p value less than 0.01, which means it's unlikely that this relationship happened by pure chance. This result indicates that, on average, for each unit increase in the Gini coefficient, the Life ladder variable decreases by 1.708 units. The constant has a value of 6.470, indicating the expected value of the Life Ladder score when the Gini Index is 0. In this case this is a theoretical scenario, as the Gini Index would never actually be 0. The R^2 is unfortunately low with a value of 0.017, which means that only around 1.7% of the variability in the Life Ladder scores is explained by the model. This suggests that the Gini Index alone does not account for that much of the variation of happiness across observations. The purpose of the adjusted R^2 is to adjust for the number of predictors in the model, as in this case is just one, which gives it an almost identical value at 0.016.

Figure 1 visualises this negative relationship between happiness and income inequality, indicated by the regression model. The Gini Index is shown on the x-axis and the Life Ladder on the y-axis. The fitted regression line shows the downward trend, indicating that higher income inequality is associated with lower happiness. It is important to note that this model does not include for other control factors that might affect happiness. The scatter plot should also be viewed with some caution since the R^2 value is so low. Therefore, even though the plot shows a trend, it does not account for the complexity of all the factors contributing to happiness.

Figure 1. 1: Goodness-of-fit model for Life Ladder Regression with Gini Coefficient

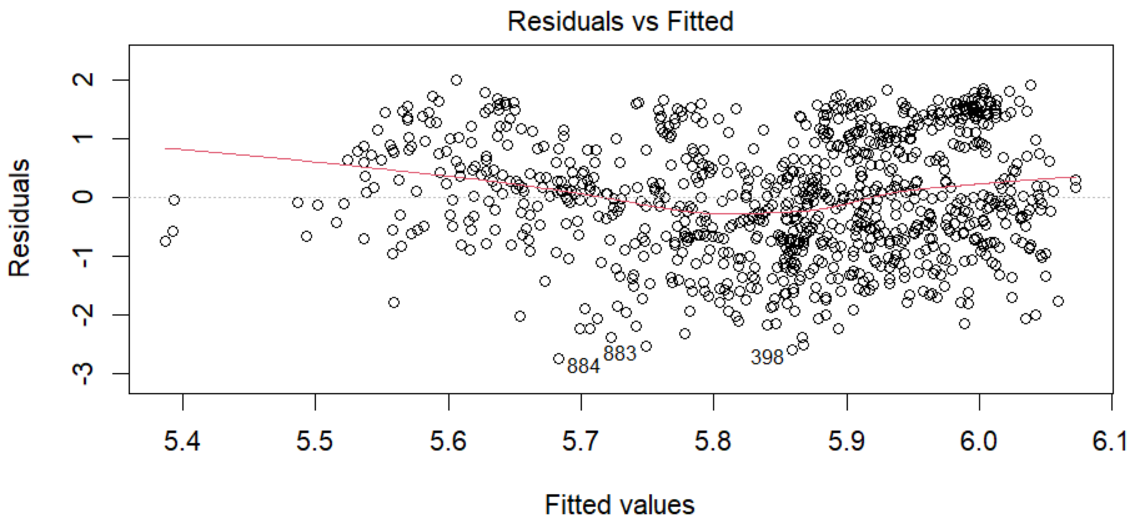


Figure 1. 1 reveals some more characteristics of this model's performance. The plot shows a non-constant spread of residuals across the range of fitted values, suggesting some heteroskedasticity. The pattern shows residuals tend to fan out as the fitted values increase, suggesting the presence of non-linearity. The red line, meant to smooth out the residuals, also shows a slight curve, which should be flat if the relationship between the variables

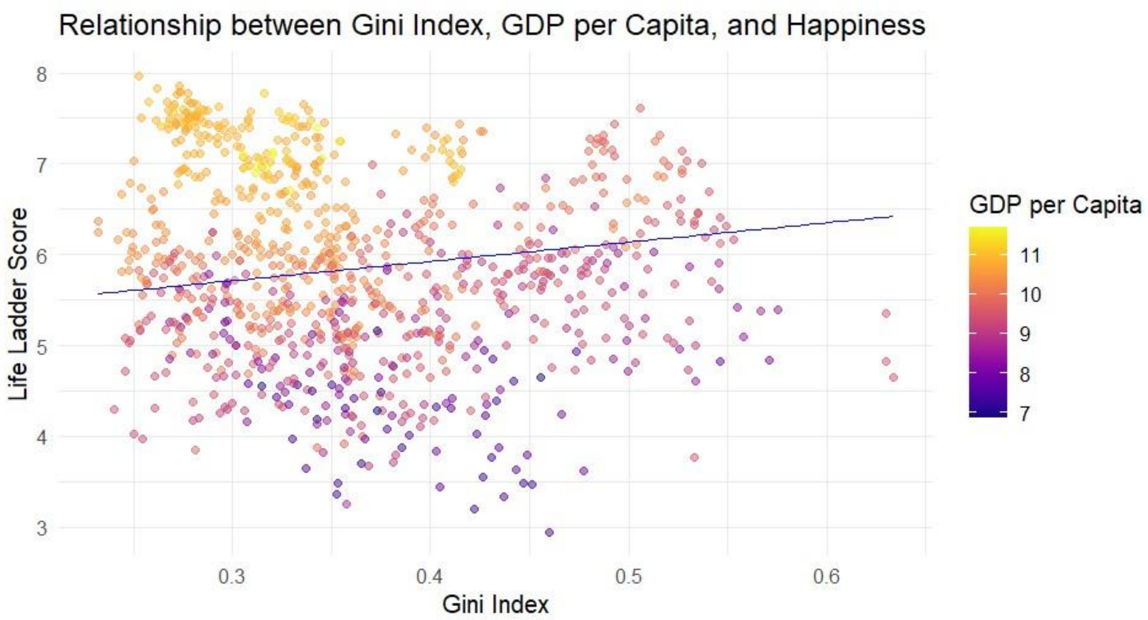
were linear. This curve indicates that a linear model may not be capturing all the aspects of the relationship. There do not appear to be any obvious outliers.

5.2 Life Ladder ~ Gini + GDP per Capita

Table 2: Regression Analysis of Gini Coefficient's impact on Life Ladder, controlling for GDP per capita

	Dependent variable:
	LifeLadder
Gini	2.126***
	(0.303)
GDPpercap	0.898***
	(0.026)
Constant	−3.719***
	(0.314)
Observations	901
R^2	0.577
Adjusted R^2	0.576
Residual Std. Error	0.668 (df = 898)
F Statistic	612.816*** (df = 2; 898)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2: Relationship between Gini Coefficient and Life Ladder, controlling for GDP per capita



By including GDP per capita as a control variable, the analysis accounts for the overall economic prosperity of a country, isolating the effect that income distribution has on happiness. From the multiple regression model including both the Gini coefficient and GDP per capita, this result can be interpreted as countries with higher GDP per capita usually scores higher on the Life Ladder than the countries with a lower GDP per capita. Holding the GDP per capita variable constant, the Gini coefficient is 2.126. This indicates that for each unit increase in the Gini coefficient, the LifeLadder increases with 2.126 units. Here the R^2 has a value of 0.577 (57.7%), this leaves approximately 42.3% of the variability to be explained by other factors not included in the model. This is a moderate fit and a more acceptable determination coefficient than previously. This suggests that, within the context of this model, higher income inequality is associated with higher reported happiness. This seems counterintuitive as it might be expected that higher inequality leads to lower happiness. However, this positive relationship could be influenced by various factors, such as cultural norms, societal values or other unmeasured variables that correlate with both happiness and income inequality.

The coefficient for GDP per capita is 0.898. This aligns with the assumption that countries with a higher GDP per capita is associated with higher levels of happiness. As shown in Figure 2, it can be seen trends that the countries with the highest GDP per capita tend to also be the ones with the lowest score on the Gini Index, and highest score on the Life Ladder, with exceptions. Both the Gini variable and the GDP per capita variable are very significant with p-values less than 0.01.

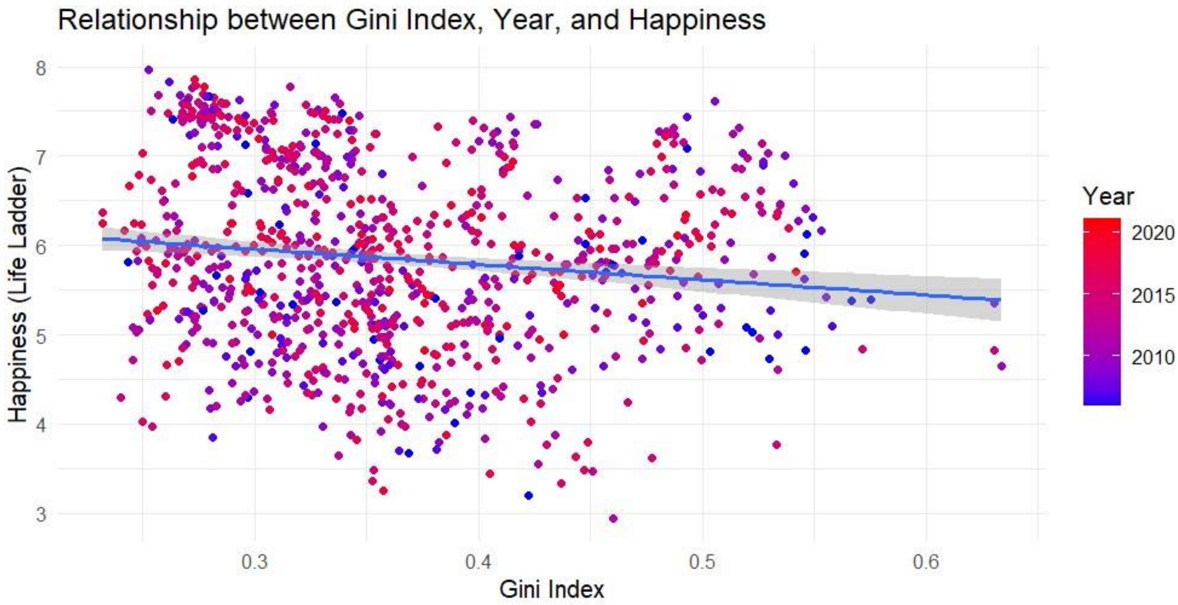
Visually, Figure 2 supports these results, showing the spread of Life Ladder scores across different Gini Index values with the colour indicating different values of GDP per capita. The slope of the regression line is positive, indicating a positive relationship between the Gini coefficient and happiness when controlling for GDP per capita.

5.3 Life Ladder ~ Gini + Year

Table 3: Regression Analysis of Gini Coefficient's impact on Life Ladder, controlling for time

	Dependent variable:
	LifeLadder
Gini	-1.637^{***}
	(0.432)
Year	0.015^{*}
	(0.009)
Constant	-24.666
	(17.611)
Observations	901
R^2	0.021
Adjusted R^2	0.018
Residual Std. Error	1.017 (df = 898)
F Statistic	9.461^{***} (df = 2; 898)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 3: Relationship between Gini Coefficient and Life Ladder, controlling for time



This model is looking at the impact of the Gini coefficient on Lide Ladder, controlling for the "Year" variable to account for time trends. From Figure 3, although hard to immediately spot, there seems to be more red plots at the higher Life Ladder score and more blue towards the lower scores. This indicates that the happiness tends to be somewhat higher in recent years.

The Gini coefficient is relatively similar to that of the first regression model (-1.708), at -1.637, with a standard error of 0.432, and a p-value less than 0.01, which makes it statistically significant. This means that everything else kept constant, the Life Ladder variable decreases with approximately 1.637 units for every unit increase of the Gini variable. The variable for Year is at 0.015, which means for every unit increase in year, the Life Ladder variable increases with 0.015. However, the Year variable is not significant enough, with a p value of less than 0.1, but higher than 0.05. The constant is quite large with a value of -24.666, but it is not statistically significant. The R^2 value is 0.021, which means the model explains 2.1% of the variability of the Life Ladder scores. This is a slight improvement from the first regression model without the "Year" control variable, but it is still a low figure. This indicates that other factors not included in the model are responsible for the vast majority of the variation in happiness.

The model indicates that while there is a negative relationship between income inequality and happiness, there may also be a slight positive trend in happiness over time. However, the models explanatory power remains low, and this topic may need more research in the future.

5.4 Life Ladder ~ Gini + Countries

Table 4: Regression Analysis of Gini Coefficient's impact on Life Ladder, controlling for country-specific effects

	Dependent variable:
	LifeLadder
Gini	-3.049^{***}
	(0.591)
Constant	5.845^{***}
	(0.229)
Observations	901
R^2	0.904
Adjusted R^2	0.890
Residual Std. Error	0.341 (df = 780)
F Statistic	61.447^{***} (df = 120; 780)
Note:	*p<0.1; **p<0.05; ***p<0.01

This regression model shows the effect of Gini on the Life Ladder score while controlling for country fixed effects. It uses dummy variables for each country, which accounts for any unobserved heterogeneity across countries – differences that don't change over time within each different country. These differences might include the country's culture, its historical trends or legal-, political-, and social-institutions that have persisted over time.

Holding all the countries' coefficients constant, the Gini coefficient is sizeable at -3.049, with a statistically significant negative impact ($p < 0.01$). This means that increasing the Gini coefficient with one unit, the Life Ladder approximately decreases with 3.049 units. Compared to the first regression model, only including the Gini as an independent variable, the coefficient was -1.708, which is remarkably different from the new coefficient of -3.049 when including all the countries coefficients in the regression model.

In this model the R^2 value is very high at 0.904, indicating that this model explains about 90.4% of the variation in the Life Ladder scores. The adjusted R^2 is also similarly high at 0.890, adjusting for the number of predictors in the model. Comparing this to previous models, this is a substantial increase and suggests that country-specific factors are important in explaining happiness.

5.5 Summary

Examining the relationship between happiness and income inequality, a negative correlation emerged, indicating that higher income inequality is usually associated with lower happiness. Adversely, when accounting for GDP per capita, this negative correlation did not hold true, with happiness increasing with income inequality. When controlling for time, the results suggested a slight increase in happiness over the years. Most notably, using dummy variables to account for country fixed effects, further increased the explanatory power of the inverse relationship between happiness and income inequality. This relationship might still be heavily influenced by many different factors not taken into account for in this thesis and may create interest for deeper dives in future research.

6 Discussion

6.1 Key Findings

Throughout the empirical analyses conducted in this thesis, examining the relationship between happiness and income inequality, there is a somewhat consistent inverse relationship between the Gini coefficient and the Life Ladder scores. The relationship was seen through most of the analyses, adjusting for factors such as temporal, and societal changes. As seen in Table 1, the regression model's explanatory power was pretty low, indicating that income inequality alone is not enough to understand the happiness levels. Including GDP per capita into the regression model showed that countries with higher GDP per capita tends to correlate with higher levels of happiness. The model accounting for GDP per capita also showed a positive relationship between income inequality and happiness, suggesting that while income inequality tend to affect happiness negatively, overall economic growth can offset this effect. The models including Year (time) as a variable indicated a slight positive trend in happiness over the years, suggesting that over time, regardless of income inequality, there might be other factors that contribute to gradually increasing happiness. The last model, controlled for country-specific fixed effects, showed that income inequality's negative impact on happiness is more significant when holding constant unobserved cultural diversity across countries. With a high explaining power, this model suggests that there is a stronger relationship between income inequality and happiness within countries, than between countries. Overall, this relationship underscores a finding that higher income inequality tends to correlate with lower levels of happiness.

6.2 Compared to existing literature

The Easterlin Paradox, states that over time, long-term economic growth and levels of happiness and income are not significantly related (Easterlin & O'Connor, 2022). This thesis' findings somewhat align with this theory, as it shows how relative income disparities can impact happiness, as well as it focuses on changes within countries. On the other hand, when controlling for GDP per capita, as shown in Table 2, the results show a positive relationship. Otherwise, the negative correlation between the Gini coefficient and the Life Ladder scores, could reinforce Easterlin's argument that social comparison and relative income levels determines well-being.

These findings are backed by earlier studies such as Oishi & Kesebir (2015), who found similar trends covering different countries, also controlling for country fixed effects, reinforcing the theory that increased income inequality tend to worsen happiness. Their argument is that without a more equal distribution, growth leads to a stagnation in happiness growth. However, the surprising finding in this thesis indicates that sometimes higher GDP per capita can mitigate some of the negative effects of income inequality on happiness.

This thesis' findings almost echo the findings of Oishi, Kesebir & Diener (2011). They found that in the United States higher happiness correlates with lower income inequality. This thesis found the same consensus but on a more global scale, instead of within the United States. However, this relationship disappeared once GDP per capita was controlled for. Broadening the application of their U.S. findings, this thesis once again reinforces the notion that perceived social justice significantly influence happiness.

This notion is the focus of Bjørnskov et al.'s (2013) study. Focusing on the importance of perceived fairness of income distribution, suggesting that individuals' subjective perceived fairness can mediate the effects of income inequality on happiness. In contrast, this thesis explores the direct relationship between income inequality and reported levels of happiness. This study utilizes regression models for analysis, in contrast to Bjørnskov et al.'s use of pseudo micro-panel data to explore the psychological implications of income inequality. By combining the psychological insights of Bjørnskov et al. with this thesis' economic data analysis, it highlights the complex interplay between the subjective perceptions and economic realities that shapes the level of happiness across different countries.

6.3 Interpreting results

As previously mentioned, except for model 2, controlling for GDP per capita, our findings show a negative relationship between income inequality and happiness. Although with a low explanatory power for model 1 and 3, this supports our hypotheses 1, 3, and 4. This may likely be linked to increased social comparison, and less social cohesion. In more unequal societies this may also be linked to feelings of stress and anxiety associated with economic and social uncertainty.

What was not expected to find in this thesis, was that a higher GDP per capita could mitigate the negative effects of income inequality on happiness. Contrary to the typical negative assumption of inequality, and our assumption in Hypothesis 2, Table 2, controlling for GDP per capita revealed a positive association between the Gini coefficient and the Life Ladder score. This may suggest that while inequality reduces happiness, higher levels of overall wealth can provide a buffer by improving general living conditions and more resources for public services, which can improve the individual's quality of life. Additionally, in wealthy countries, even the lower classes might live above a certain standard of living, which may reduce the impact felt from inequality. This gives GDP per capita sort of a dual role and further highlights the complex interplay between growth and distribution. Suggesting that economic growth alone is not sufficient, but how the benefits of said growth is distributed and used plays a crucial role on happiness.

Even though Table 3, examining Gini's impact on Life Ladder while controlling for time, gave a negative result, it was unexpected to find a positive trend in happiness over time, although minor. This could be caused by improvements in non-economic factors such as better healthcare, education, technology, etc. and a general shift toward higher standards of living over the decades.

Including country-specific fixed effects in the regression model, the results from Table 4 yielded a substantially high explanatory power. This indicates that local factors significantly influence happiness. Examples of such factors may include cultural norms, strength of social safety nets, quality of government, and other policies regarding wealth distribution. These results show the importance of policies tailored to specific national contexts, where one approach for enhancing happiness through economic policy may work for some country but then again not for others.

6.4 Limitations

This thesis has some limitations, the main limitations being found in the datasets, the variables, variability in the application of findings and time. Also important to note is that this study only examines correlations and the findings do not imply causation.

Both datasets have some limitations that weaken the interpretation of the findings in this thesis. First of all, both the Gini dataset and the World Happiness Report dataset are missing some countries, specifically small population countries and island nations, such as Andorra, Monaco, Fiji and the Maldives. To be able to give a more holistically view, it would be preferable with all countries included in the analysis. The same goes for the timeframe of the datasets. Though 2005-2021 is 17 years, this is not as long a time-period as desirable when examining economic growth, income inequality and happiness on a global scale. World Happiness Report's dataset for the Life Ladder scores also used a self-report measurement, and since happiness is subjective it can be difficult to measure consistently across different cultures, as well as it might not encapsulate its broader more nuanced implications for well-being. Self-measured statistics also often consists of personal biases, which may skew results and give an unclear picture of countries' true happiness.

By choosing to use just the variables Gini and Life Ladder, it may oversimplify the whole picture. Although controlling for other variables such as GDP per capita, temporal changes, and country specific fixed effects, it risks that the relationship between income inequality and happiness is being ascribed to only these variables, even though other factors may contribute significantly. Other variables that are not considered could be employment rates, health factors, education, political stability, and social support. The Gini variable measures income inequality but does not account for other forms of inequality, such as wealth, opportunity or access to resources, all factors that may affect happiness.

At last, a big limitation is time. As this thesis had a time limit to be completed within, it restricts further examination of results. Referring to the results from Table 2, which was unexpected, it would be preferable to dive deeper into the analysis and continue to examine the model for further understanding the results. It would also be advantageous to further investigate the cause of the residuals shown in Figure 1. 1, as well as perform residual vs. fitted analyses for every regression model. Due to time management and prioritization this was not possible but would be interesting to examine in future research.

6.5 Future research

The limitations of this thesis raise some open questions that could be addressed in future research. To get a better understanding of income inequality's effect on happiness, future studies could use more longitudinal studies to better track changes over time. Future research should include more variables to better understand why and how income inequality impacts happiness. Other studies than this thesis might also use qualitative data to provide in-depth insights of the psychological aspects of income- and wealth inequality, opposed to just the economic aspects. As mentioned in limitations, although difficult, a study including more, or all countries would provide an even more holistic analysis. Considering the unexpected results from including GDP per capita, a study focusing more on the relationships between GDP per capita, income inequality and happiness would be interesting. A study researching the effectiveness of specific policies aimed at reducing income inequality could also provide more practical insights.

6.6 Reflecting on methodological choices

The choice to use regression analyses, specifically OLS, was to provide a robust framework examining relationships between variables. Even though OLS assumes a linear relationship and does not capture the more complex aspects of the relationship, this method was a more feasible analysis for this thesis. The variable and dataset selection provided the Gini coefficient and Life Ladder scores as key variables for representing income inequality and happiness. This choice simplified a usually more complex construct but sought to provide the fundamentals of the relationship between happiness and income inequality.

7 Conclusion

This thesis, examining how income inequality affects happiness in various countries, illustrate that while economic factors like GDP per capita and the Gini coefficient are significant, their effect on happiness are nuanced and deeply influenced by social, cultural and policy factors. The relationship between income inequality and happiness remains somewhat ambiguous from our analyses, with some models showing low explanatory power, and some showing correlations in opposite directions. Although certain findings support existing literature and contribute to academic research, they underscore the need for future research. A deeper dive into these relationships is required to develop a more comprehensive understanding and clearer results.

References

4.2—Residuals vs. Fits Plot | STAT 462. (2018). [The Pennsylvania State University]. PennState Eberly College of Science. https://online.stat.psu.edu/stat462/node/117/

Alto, V. (2023, February 14). *Understanding Ordinary Least Squares (OLS) Regression | Built In*. https://builtin.com/data-science/ols-regression

Assumptions of OLS: Econometrics Review | Albert.io. (2016, November 25). *Albert Resources*. https://www.albert.io/blog/key-assumptions-of-ols-econometrics-review/

Bjørnskov, C., Dreher, A., Fischer, J. A. V., Schnellenbach, J., & Gehring, K. (2013). Inequality and happiness: When perceived social mobility and economic reality do not match. *Journal of Economic Behavior & Organization*, *91*, 75–92. https://doi.org/10.1016/j.jebo.2013.03.017

Damgaard, C. (2024). *Lorenz Curve*. Wolfram Research, Inc. https://mathworld.wolfram.com/

Easterlin, R. A., & O'Connor, K. J. (2022). Explaining happiness trends in Europe. *Proceedings of the National Academy of Sciences*, *119*(37), e2210639119. https://doi.org/10.1073/pnas.2210639119

Eric. (2019, November 29). *Introduction to the Fundamentals of Panel Data | Aptech*. https://www.aptech.com/blog/introduction-to-the-fundamentals-of-panel-data/

Frost, J. (2017a, April 12). *How to Interpret P-values and Coefficients in Regression Analysis*. Statistics By Jim. http://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/

Frost, J. (2017b, August 13). *Heteroscedasticity in Regression Analysis*. Statistics By Jim. http://statisticsbyjim.com/regression/heteroscedasticity-regression/

Gini Index | DataBank. (2024). The World Bank. https://databank.worldbank.org/metadataglossary/gender-statistics/series/SI.POV.GINI

Hasell, J., Arriagada, P., Ortiz-Ospina, E., & Roser, M. (2024). Economic Inequality. *Our World in Data*. https://ourworldindata.org/economic-inequality

Hasell, J., & Roser, M. (2023). Measuring inequality: What is the Gini coefficient? *Our World in Data*. https://ourworldindata.org/what-is-the-gini-coefficient

Hayes, A. (2020). *Multiple Linear Regression (MLR) Definition, Formula, and Example*. Investopedia. https://www.investopedia.com/terms/m/mlr.asp

Helliwell, J. F., Layard, R., Sachs, J. D., Neve, J.-E. D., Aknin, L. B., & Wang, S. (2023, March 20). *World Happiness Report 2023*. https://worldhappiness.report/ed/2023/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning: With applications in R* (Corrected at 8th printing). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Oishi, S., & Kesebir, S. (2015). Income Inequality Explains Why Economic Growth Does Not Always Translate to an Increase in Happiness. *Psychological Science, 26*(10), 1630–1638. https://doi.org/10.1177/0956797615596713

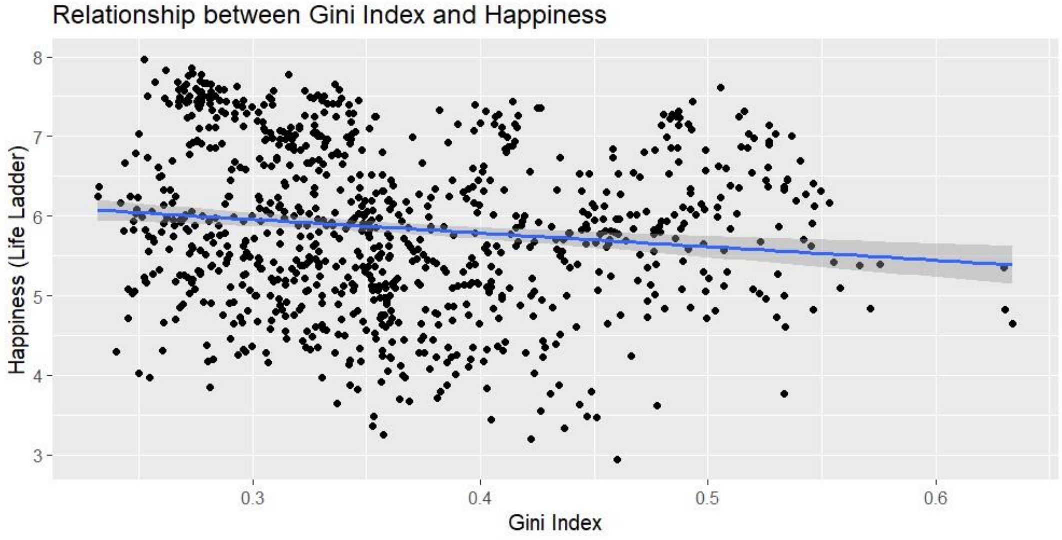
Oishi, S., Kesebir, S., & Diener, E. (2011). Income Inequality and Happiness. *Psychological Science, 22*(9), 1095–1100. https://doi.org/10.1177/0956797611417262

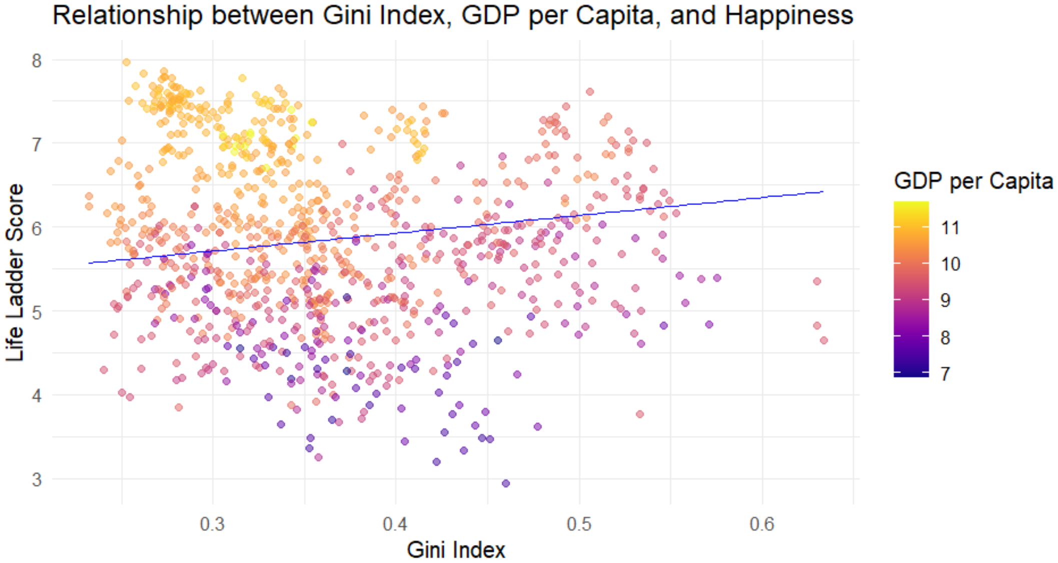
Onozaka, Y. (2024). *Chapter 11 Regression Analysis (3)*.

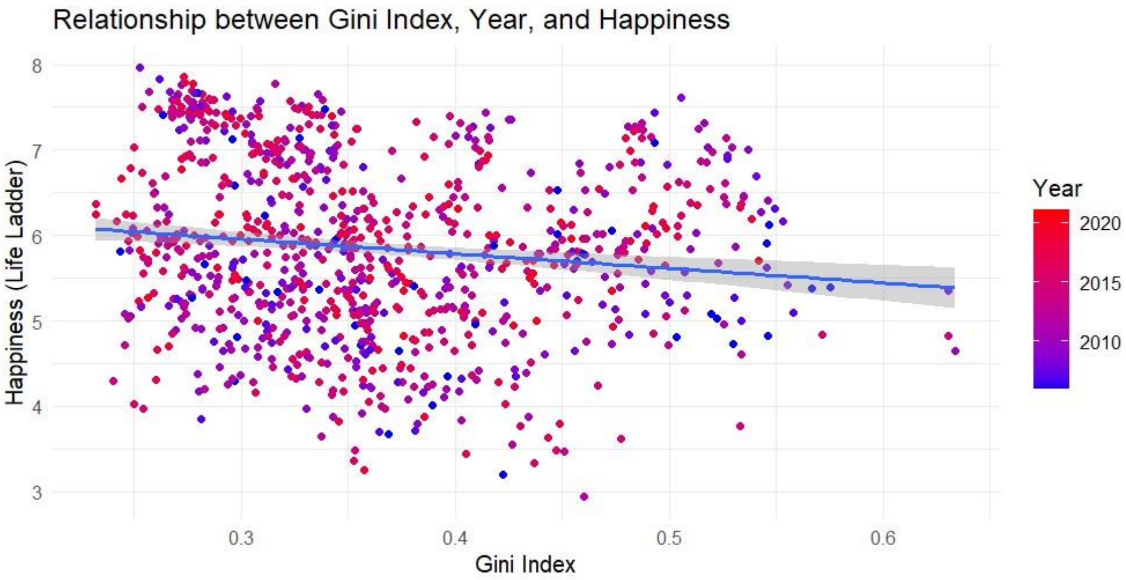
Sung, M. J. (2010). *Population Aging, Mobility of Quarterly Incomes, and Annual Income Inequality: Theoretical Discussion and Empirical Findings*.

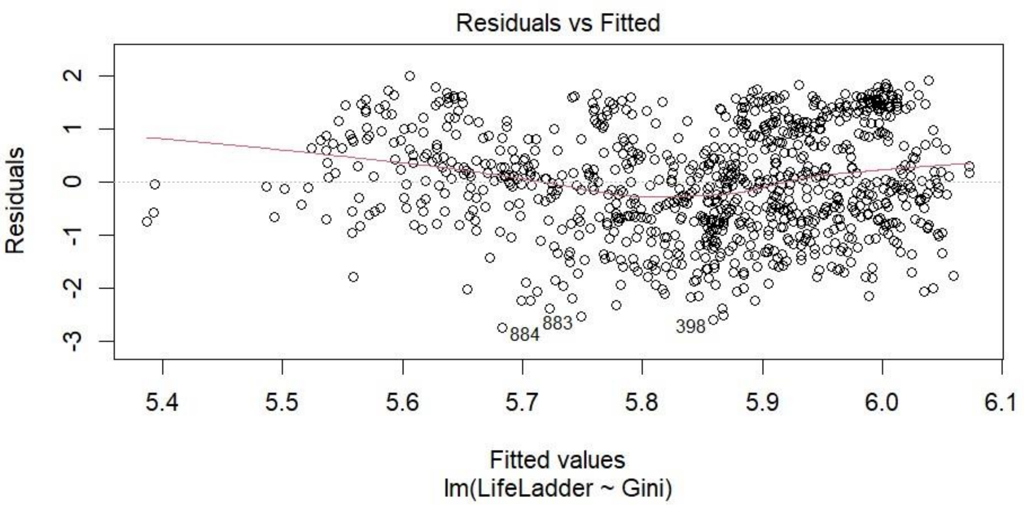
Uboe, J. (2017). *Introductory statistics for business and economics: Theory, exercises and solutions*. Springer Berlin Heidelberg.

Table of Figures






Relationship between Gini Index, Year, and Happiness
Happiness (Life Ladder)
Gini Index
Year
2020
2015
2010
8
7
6
5
4
3
0.3
0.4
0.5
0.6


Residuals vs Fitted
Residuals
Fitted values
lm(LifeLadder ~ Gini)
884
883
398
2
1
0
-1
-2
-3
5.4
5.5
5.6
5.7
5.8
5.9
6.0
6.1

	Dependent variable:
	LifeLadder
Gini	-1.708^{***}
	(0.430)
Constant	6.470^{***}
	(0.160)
Observations	901
R^2	0.017
Adjusted R^2	0.016
Residual Std. Error	1.018 (df = 899)
F Statistic	15.759^{***} (df = 1; 899)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

	Dependent variable:
	LifeLadder
Gini	2.126^{***}
	(0.303)
GDPpercap	0.898^{***}
	(0.026)
Constant	-3.719^{***}
	(0.314)
Observations	901
R^2	0.577
Adjusted R^2	0.576
Residual Std. Error	0.668 (df = 898)
F Statistic	612.816^{***} (df = 2; 898)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

	Dependent variable:
	LifeLadder
Gini	−1.637***
	(0.432)
Year	0.015*
	(0.009)
Constant	−24.666
	(17.611)
Observations	901
R^2	0.021
Adjusted R^2	0.018
Residual Std. Error	1.017 (df = 898)
F Statistic	9.461*** (df = 2; 898)
Note:	*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	LifeLadder
Gini	−3.049***
	(0.591)
Constant	5.845***
	(0.229)
Observations	901
R^2	0.904
Adjusted R^2	0.890
Residual Std. Error	0.341 (df = 780)
F Statistic	61.447*** (df = 120; 780)
Note:	*p<0.1; **p<0.05; ***p<0.01