# Factors that Influence Income Inequality Across the Globe

Emma Francis

Stephanie Gaysowski

Nathan Moen

**Abstract**

This paper initially sought to analyze the relationship between income inequality, as measured by the Gini coefficient, and the multidimensional poverty index, which measures the levels of deprivation related to health, education, and living standards within a developing nation. However, in the pursuit identifying the factors that affect income inequality, we discovered that other population demographics such as life expectancy and median age are more accurate predictors of levels of income inequality across the globe.

## I. Introduction

In recent years, the global perception and actual structure of poverty around the world has shifted from a country with a poor, underserved population to a society with drastically polarized socioeconomic classes. An impoverished country in the past might have been seen as a country with an altogether poor population, whereas now the issue has become the distribution of wealth within these developing nations. Within these countries, it is common to see an extremely wealthy, albeit narrow, group of people collecting the majority of the benefits from the country. However, in these same countries, there is also a substantial population living a life of poverty. Countries with high inequality suffer from a large divide between the classes within the nation, inhibiting potential growth and economic success.

Understandably, economists have shifted their focus to this issue due to inequality's notable effect upon the success of developing nations. Experts everywhere have begun searching for the root causes of inequality within nations who suffer from these large socioeconomic divides, and one of the most interesting relationships uncovered relates to the levels of inequality and overall poverty found within the country. As stated above, within nations with high levels of inequality, there are groups of extraordinarily wealthy people; however, the majority of population are still penniless and struggling to make ends meet. This paper aims to analyze whether or not this level of impoverishment exacerbates the gap between the classes.

The Gini index is used to measure the overall deviation of a country's economy from a perfectly equal distribution of wealth across the population. Since the Multidimensional Poverty Index (MPI) is defined as the product of the average intensity of the deprivation indicators and the percent of the population experiencing these poverty levels, we expect income inequality to increase along with the different indicators which constitute the MPI in our simple regression analysis.

**II. Literature Review**

Much of the literature surrounding inequality seeks to understand the underlying factors that can predict income inequality across nations throughout the world. Some studies attempt to explore the relationship between poverty, growth, and inequality in developing nations, sometimes stratifying sample countries based on economic or political regimes. However, other studies emphasize how population statistics, such as median age and life expectancy, may be more accurate indicators of income inequality levels.

One article presents the strong positive correlation found between unemployment and income inequality in a diverse range of economies across the globe (Cysne & Turchick, 2012). However, the article clarifies that this relationship is only observed in situations where the unemployment rate is no larger than 15%. Furthermore, the article explains how unemployment is relatively higher amongst low-skilled workers who often endure longer spells of unemployment due to technological progress, etc.

The following study also explores the relationship between demographic variables and the distribution of income within the United States (Lam 1997). The study particularly examines how a changing population composition may alter income inequality, focusing specifically on age distribution, fertility, marriage, migration, and mortality. Some analyses state that the age distribution of a population may alter the overall levels of income inequality within a country, without actually altering the levels of income inequality between age groups. However, further analysis concluded that in some cases, a larger, younger workforce may actually decrease wages. The effects of fertility on inequality were also deemed ambiguous as much of it could similarly be attributed to changes in the composition of the population over time. Overall, the article concedes that although demographic factors produce significant changes in the distribution of income, much of this change could be due to changes in labor demand.

The next article studies the relationship between poverty, growth, and inequality in developing nations and the poverty-reduction performance of the recent wave of global economic growth occurring since the early 1990s (Kwasi Fosu 2016). The article distinguishes between the various decreasing rates of poverty and the resulting both increasing and decreasing rates of income inequality. However, the article recognizes that generalities exist. For example, more than 75 percent of the countries demonstrated decreasing income inequality, although most of these seemingly decreasing levels of income inequality can be attributed to income growth rather than an actual redistribution of income within the country. Furthermore, the main force behind the increases or decreases in poverty is primarily related to average income growth in these countries.

The following study investigates the forces that affect carbon emissions patterns and changes in economic growth, inequality, and poverty in Pakistan in the period 1980-201, utilizing a multivariate cointegration approach (Hassan, Zaman, & Gul 2015). The results demonstrated a positive relationship between economic growth and income inequality as well as poverty and income inequality in the short run; in the long run, the relationship holds true even when adding carbon emissions. However, it is important to note a negative relationship between carbon emissions and income inequality. Ultimately, the study is limited by the fact that it only focuses on Pakistan; however, it also incorporates the Kuznets curve hypothesis into the standard exploration of growth, poverty, and inequality.

Our initial model focused primarily upon the effects of poverty on income inequality in developing countries, which has also been studied in an attempt to test the validity of Kuznets hypothesis. However, over the course of our analysis and the expansion of our model, we have explored the significance of a variety of other variables indicative of the population's overall health and wellbeing. Although our paper originally sought to offer a simple analysis of the relationship between nonmaterial aspects of poverty and inequality, our revised analysis allows for a better understanding of the ambiguous effects of many of these variables on income inequality and the overall complexity of the issue.

**III. Data**

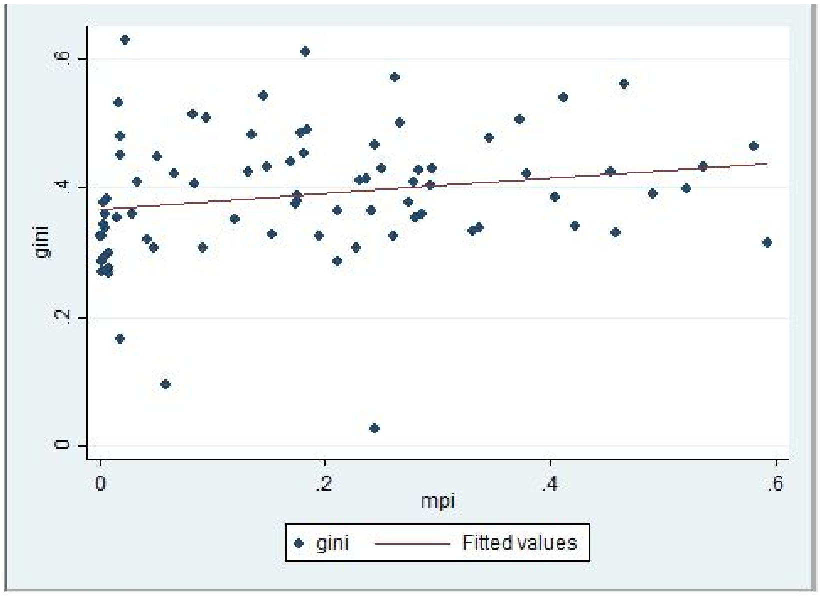
The variables used in our analysis include the Gini index and the ten indicators used to make up the Multidimensional Poverty Index (MPI). The MPI is divided into three categories corresponding to the three main dimensions of poverty: health, education, and living standard. These three categories are then broken up further into indicator variables that are measured using surveys. Health is divided into nutrition and child mortality; education is divided into years of schooling and school attendance; living standard is divided into cooking fuel, sanitation, drinking water, electricity, housing, and assets. Each of the indicator variables is measured using surveys and each one has different survey criteria in order for someone to be considered deprived of this particular indicator.

The original data set used was provided by the United Nations Development Programme. Alleviating poverty is one of the principal goals of the United Nations, so they have also sought to track and understand the relationship between inequality and the multidimensional aspects of poverty that extend beyond simple economic deprivation. The MPI data was collected through yearly surveys. The original data set contains preliminary 2018 survey results of 105 developing nations, which covers about 74% of the global population (Sabire & Kanagaratnam 2018). After dropping countries with missing observations, we analyzed the 79 countries that remained.

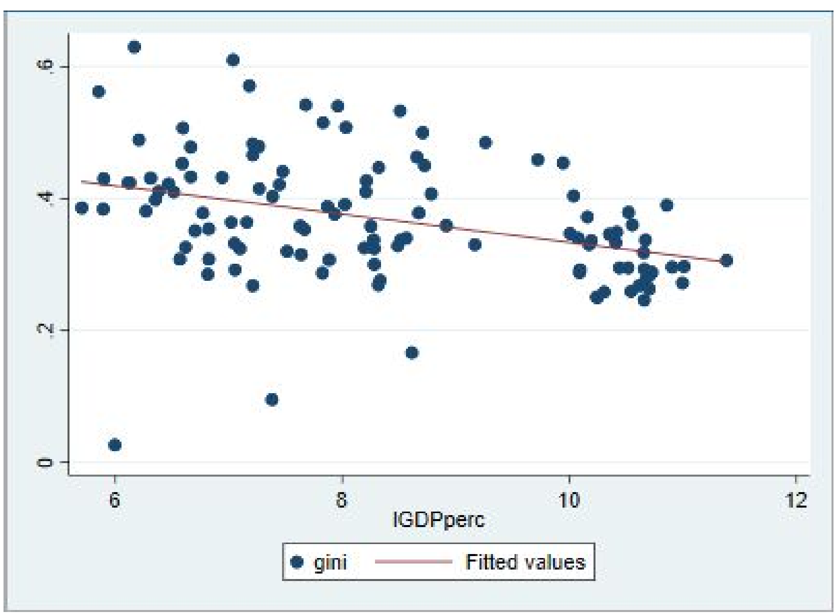
To add more variation to the dataset we added the 36 OECD countries along with the 79 developing countries to better understand the global impacts that economic and demographic data have on income inequality, bringing the total number of countries analyzed to 115. The variables we chose to focus on in the analysis are fertility rates, life expectancy, median age, unemployment rates, and the natural log of GDP per capita and their effects on the Gini Index of the 115 countries included in our data set. We decided to investigate life expectancy and fertility rates since these variables are indicators of a growing or declining population. We hypothesized that a high fertility rate would lead to higher inequality, since the expenses on children would be much higher. Along those same lines, we expected life expectancy to have a negative effect on inequality, since longer life span hints at more welfare programs within a country. Furthermore, we chose median age and the unemployment rate to emphasize the relationship between age distribution, workforce demographics, and income inequality. As stated in our literature review, in some cases a large, young workforce decreases wages and exacerbates income inequality in a country. From this we expected median age to have a negative impact on Gini, meaning as the median age decreases, the income inequality increases. We also presumed unemployment would have a positive relationship, which would imply the more people out of a job, the higher the income inequality in a country. Finally, we needed a measure of income within these countries, since Gini is a measure of income inequality. We hypothesized that as income levels (GDP per capita) fall, the measure of income inequality (Gini) will rise. These values were collected for the 79 developing countries through The World Bank's World Development Indicators Database using the year 2015. Data for the 36 OECD countries was collected through the OECD's Databases on Main Economic Indicators and Demographics. Median age for all countries was found using the CIA World Factbook 2017 estimates.

**Descriptive Statistics of the Variables**

**Scatterplots of MPI vs Gini and lGDPperc vs Gini**




The scatterplot above illustrates the weak positive relationship between MPI and Gini in the sample of 79 developing countries.




The scatterplot above shows the weak negative relationship between Gini and the log of GDP per capita in the larger sample of 115 countries.

**Summary Table of all Variables**

The table below shows the summary statistics for each variable in the regression. Each variable has 115 observations; MPI is not measured in OECD countries, therefore it only has 79 observations.

| **Variable** | **Observations** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|
| Gini | 115 | 0.3684522 | 0.0950944 | 0.026 | 0.63 |
| mpi | 79 | 0.1933344 | 0.160913 | 0.0006754 | 0.5914328 |
| fertrate | 115 | 3.104078 | 1.511595 | 1.2 | 7.29 |
| lifeexp | 115 | 70.20357 | 8.988621 | 51.41 | 85.3 |
| medage | 115 | 28.59652 | 9.823521 | 15.4 | 47.3 |
| unemploy | 115 | 7.945445 | 5.933121 | 0.35 | 27.33 |
| GDPperc | 115 | 13308.29 | 17990.15 | 300.6766 | 87842 |
| lGDPperc | 115 | 8.370789 | 1.61699 | 5.706035 | 11.3833 |

**Gauss Markov Assumptions**

**Linear in parameters**

The regression equation is linear in parameters, because we are using the STATA regression command to calculate our results. Therefore, our equation will be:

$$\widehat{gini} = \widehat{\beta}_0 + lGDPperc(\widehat{\beta}_1) + fertrate(\widehat{\beta}_2) + lifeexp(\widehat{\beta}_3) + unemploy(\widehat{\beta}_4) + medage(\widehat{\beta}_5) + \widehat{u}$$

As can be seen by the equation above, our regression is linear in parameters and meets the first Gauss Markov assumption.

**Random sampling**

The data meets the random sampling assumption because the original simple regression and multiple regression models include 79 developing countries with data provided by the UNDP; none of the countries chosen are from any particular region or economic background. Where some datasets might just include OECD or Asian nations, this dataset includes a diverse mix of countries. The variation in the dataset ensures the randomness of the sample and eliminates worries of bias in the sampling. For example, we have data ranging from Mongolia to France. Our revised multiple regression model also includes 36 OECD nations from all around the world.

**No perfect collinearity**

If a variable was perfectly collinear with another variable, an increase in one of the variables would result in a perfectly linear increase in the other. For example, if one of our variables was "total mortality", and it was measured by adding child mortality and adult mortality, then it would be perfectly collinear with the variable "cmort" or child mortality. Some of our variables such as fertility rate and median age show a strong negative correlation with each other, but since none of our variables are correlated this heavily with one another based on the measurements in which they were collected, we can assume that there is no perfect collinearity among our independent variables. *See Appendix Tables 2A and 2B for correlation coefficients of the independent variables.*

**Zero conditional mean: $E(u \mid x_1, x_2, \ldots, x_k) = 0$**

This assumption states that there are no omitted variables that have an effect on the independent variable. This assumption would be violated if a pertinent variable was omitted or left out of the regression due to insufficient data. After extensive research, we can conclude that we are not omitting any variables that would have a significant impact on our dependent variable, Gini.

**Homoscedasticity: $Var(u \mid x_1, x_2, \ldots, x_k) = \sigma^2$**

This assumption states that the variance for error term u is the same for all combinations of the independent variables. For example, the variance of u does not depend on the median age of the population or the unemployment rate in a country.

**III. Results**

**Simple Regression Model 1:**

Equation: $\widehat{gini} = \widehat{\beta}_0 + mpi(\widehat{\beta}_1) + \widehat{u}$

After regression: $\widehat{gini} = 0.3674 + mpi(0.1209)$

N=79    $R^2$= 0.0375

| Variable | Coefficient (Std. Error) | T-value | P > \|t\| | $H_0$: $B_j$=0 $H_1$: $B_j \neq 0$ |
|---|---|---|---|---|
| mpi | 0.1209452* (0.0693933) | 1.74 | 0.085 | Reject at 10% |
| constant | 0.3674288*** (0.0173006) | 21.24 | 0.000 | Reject at 1% |

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at 1%)

*See Appendix Output 1 for STATA Output.*

**Simple Regression Model 2 (Developing Countries + OECD):**

Equation: $\widehat{gini} = \widehat{\beta}_0 + lGDPperc(\widehat{\beta}_1) + \widehat{u}$

After Regression: $\widehat{gini} = 0.547853 + lGDPperc(-0.0214368)$

N= 115 $R^2$=0.1329

| Variable | Coefficient (Std. Error) | T-value | P > \|t\| | $H_0$: $B_j$=0 $H_1$: $B_j \neq 0$ |
|---|---|---|---|---|
| lGDPperc | -0.0214368*** (0.0051517) | -4.16 | 0.000 | Reject at 1% |
| constant | 0.547853*** (0.0439141) | 12.48 | 0.000 | Reject at 1% |

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at 1%) *See Appendix Output 2 for STATA output.*

**Multiple Regression Model 1 (Developing Countries):**

Equation:

$$\widehat{gini} = \widehat{\beta}_0 + mpi(\widehat{\beta}_1) + lGDPperc(\widehat{\beta}_2) + fertrate(\widehat{\beta}_3) + lifeexp(\widehat{\beta}_4) + medage\widehat{\beta}_5) + unemploy(\widehat{\beta}_6) + \widehat{u}$$

After regression:

$$\widehat{gini} = 1.09 + mpi(.089) + lGDPperc(.0014) + fertrate(-.047) + lifeexp(-.0063) + medage(-.007) + unemploy(.003)$$

N=79 $R^2$=0.2517

| **Variable** | **Coefficient (Std. Error)** | **T-value** | **P>\|t\|** | **$H_0$: $B_j$=0 $H_1$: $B_j \neq 0$** |
|---|---|---|---|---|
| mpi | 0.0885083 (0.1252115) | 0.71 | 0.482 | Fail to reject at 10% |
| lGDPperc | 0.0013614 (0.0122026) | 0.11 | 0.911 | Fail to reject at 10% |
| fertrate | -0.0474059*** (0.0178446) | -2.66 | 0.010 | Reject at 1% |
| lifeexp | -0.006266** (0.0027535) | -2.28 | 0.026 | Reject at 5% |
| medage | -0.0070032* (0.0035985) | -1.95 | 0.056 | Reject at 10% |
| unemploy | 0.0034234* (0.0017608) | 1.94 | 0.056 | Reject at 10% |
| constant | 1.087997*** (0.2313891) | 4.70 | 0.000 | Reject at 1% |

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at 1%) *See Appendix output 3 for STATA output.*

**Multiple Regression Model 2 (Developing + OECD countries)**

Equation:

$$\widehat{gini} = \widehat{\beta}_0 + lGDPperc(\widehat{\beta}_1) + fertrate(\widehat{\beta}_2) + lifeexp\ (\widehat{\beta}_3) + unemploy(\widehat{\beta}_4) + medage(\widehat{\beta}_5) + \widehat{u}$$

After regression:

$$\widehat{gini} = .948 + lGDPperc(.0098) + fertrate(-.033) + lifeexp\ (-.006) + unemploy(.0027) + medage(-.0053)$$

N=115 $R^2$= 0.3156

| **Variable** | **Coefficient (Std. Error)** | **T-value** | **P > \|t\|** | **$H_0$: $B_j$=0 $H_1$: $B_j \neq 0$** |
|---|---|---|---|---|
| lGDPperc | 0.0098492 (0.0083859) | 1.17 | 0.243 | Fail to reject at 10% |
| fertrate | -0.0334391*** (0.0123131) | -2.72 | 0.008 | Reject at 1% |
| lifeexp | -0.0060967*** (0.0019783) | -3.08 | 0.003 | Reject at 1% |
| unemploy | 0.0027103** (0.0013204) | 2.05 | 0.043 | Reject at 5% |
| medage | -0.0053165** (0.0021232) | -2.50 | 0.014 | Reject at 5% |
| constant | 0.9483155*** (0.1463884) | 6.48 | 0.000 | Reject at 1% |

(*Statistically Significant at 10%, **Statistically Significant at 5%, ***Statistically Significant at 1%) *See Appendix Output 4 for STATA output.*

**Interpretation:**

In all of the above regression models we conducted a two-tailed hypothesis test in order to determine the significance of the relationships. In most cases, the authors found scattered results in their relationships with income inequality, so we decided to simply test if our independent variables had any significant relationship with Gini rather than test for a specific type of relationship.

The output from the Simple Regression Model 1 proves that MPI and the Gini index are slightly positively correlated, however the result was only statistically significant at 10%. It also produced a low R-squared value of 0.0375, meaning only 3.75% of the variation in Gini could be explained by the MPI as a whole. We then analyzed the statistical significance of the individual MPI indicators, but they also did not demonstrate significant correlations with the Gini index. The lack of statistical significance of the MPI is likely do to the fact that it is an aggregated index of 10 indicators, so countries could have vast differences in their scores for drinking water, child mortality, or electricity, but still have similar scores for the MPI overall, limiting the significance of their impact on Gini.

Following these attempts, we added data on GDP per capita from the 36 OECD countries and tested the relationship between the natural log of GDP per capita and Gini (see Simple Regression Model 2), which produced a coefficient of -0.021 that was statistically significant at 1%, with a higher R-squared of 0.1329.

We then expanded upon the original simple regression model and formed our first multiple regression model, which included MPI, lGDP per capita, fertility rate, life expectancy, median age, and unemployment. We included MPI and lGDP per capita in order to test the significance of factors directly related to poverty against what could be considered more indirect factors related to poverty. Multiple Regression Model 1 displayed that fertility rate was significant at the 1% level with a coefficient of -0.047 and life expectancy was statistically significant at the 5% level with a coefficient of -0.006. Median age and unemployment rate were both statistically significant at the 10% level with coefficients of -0.007 and 0.003, respectively; however MPI and the natural log of GDP per capita were not statistically significant at even the 10% level, so they were removed from the model to conduct an F-test for joint significance. Neither MPI nor the natural log of GDP per capita were jointly significant to Multiple Regression Model 1 at 10% significance, with an F-stat of just 0.27, compared with the critical value of $F_{2,72}$= 2.37 for 10% significance. *See Extensions for the F-test results.* Multiple Regression Model 1 demonstrated a weak but statistically significant negative correlation between the Gini index and the fertility rates as well as life expectancy in the developing world and a weak but significant positive correlation between Gini index and a country's unemployment rate. Due to the lack of significant correlations found between the Multidimensional Poverty Index and Gini, we decided to remove MPI as an independent variable. This allowed us to expand our sample to include 36 OECD countries, which effectively increases the variation in the data set for Multiple Regression Model 2.

Multiple Regression Model 2 provided statistically significant coefficients for fertility rate and life expectancy at the 1% significance level, with coefficients -0.033 and -0.006, respectively. Unemployment and median age were statistically significant at the 5% level of significance, with coefficients of 0.0027 and -0.005, respectively, while the natural log of GDP per capita remained insignificant even at the 10% level with a p-value of 0.243. In Multiple Regression Model 2, fertility rate and life expectancy continued to have weak but statistically significant negative correlations with the Gini index. Unemployment rate continued to have a weak but statistically significant positive correlation with Gini index.

To see how the variables' relationships with the Gini index change between developing and developed countries, we ran our Multiple Regression Model 2 again with only the 36 OECD countries, and surprisingly the coefficient for the natural log of GDP per capita changed from a statistically insignificant positive coefficient in Multiple Regression Models 1 and 2 to a statistically significant, strongly negative coefficient, similar to what was shown in our Simple Regression Model 2 between Gini and the natural log of GDP per capita.We suspect the changes in this relationship are due to the stages of development within the country. Similar to Kuznets' theory which states that economic inequality will increase during the beginning stages of development and eventually decrease as the country becomes more developed, we found the relationship between developing countries' income inequality and the natural log of GDP per capita to be positive, while the relationship between developed countries' income inequality and natural log of GDP per capita to be negative. *See Appendix Output 8 for STATA output of Multiple Regression Model 2 for only the 36 OECD countries.*

**IV. Extensions**

Since the variables MPI and lGDPperc were statistically insignificant in Multiple Regression Model 1, we conducted an F-Test to determine if these two variables are jointly significant. *See Appendix Output 5 for STATA output of the Restricted Model.*

$H_0$: $\widehat{\beta}_1 = \widehat{\beta}_2 = 0$ $H_1$: $H_0$ not true

Unrestricted Model (Multiple Regression Model 1):

$$\widehat{gini} = \widehat{\beta}_0 + mpi(\widehat{\beta}_1) + lGDPperc(\widehat{\beta}_2) + fertrate(\widehat{\beta}_3) + lifeexp(\widehat{\beta}_4) + medage(\widehat{\beta}_5) + unemploy(\widehat{\beta}_6)$$

Restricted Model:

$$\widehat{gini} = \widehat{\beta}_0 + fertrate(\widehat{\beta}_3) + lifeexp(\widehat{\beta}_4) + medage(\widehat{\beta}_5) + unemploy(\widehat{\beta}_6)$$

Critical Value at 10% significance $F_{2,72}$= 2.37 Model F-Stat: 0.2729

Therefore we fail to reject the Null Hypothesis at 10% and the variables MPI and lGDPperc are not jointly statistically significant among the 79 developing countries.

To be sure there were no errors in our functional form, the natural log of Gini was also used as the dependent variable instead of Gini. The natural log of gini did not produce coefficients that were more statistically significant. *See Appendix Output 6 for STATA output using Natural log of Gini as the dependent variable.*

Due to the insignificance of GDP per capita in our model, we also tried to change the functional form of the model by regressing GDP per capita squared, as well as the natural log of GDP per capita, along with GDP per capita. GDP per capita squared did not produce statistically significant results; however, we decided to primarily use the natural log of GDP per capita. The natural log of GDP per capita more clearly captured the relationship between Gini and income level in the country, because it shows percent change of GDP per capita rather than the effect of a dollar difference on inequality. *See Appendix Output 7A and 7B for STATA Outputs using GDP per capita squared and the natural log of GDP per capita.*

**V. Conclusions**

Our final model included the natural log of GDP per capita, fertility rate, life expectancy, unemployment, and median age. However, in the process of developing our final model, we also explored the relationship between the Gini Index and many other factors related to development, poverty, and economic or political freedom. For instance, we tested the significance of the multiple factors that constitute the MPI as well as the individual categories themselves such as education, health, and living standards. We also explored the significance of the freedom index, literacy rate, carbon dioxide emissions, urbanization, imports as a percent of GDP, and the number of cellular subscriptions per 100 people within a country. Although previous literature pointed to a relationship between these factors and income inequality, we did not find such a relationship to exist.

After testing the significance of these factors in our sample of developing nations, we expanded our sample to include OECD nations as well in an effort to increase the variation and sample size of our model. As expected, this analysis found that fertility rates, life expectancy, unemployment, and median age were factors that could be used to predict levels of income inequality. These factors are typically representative of the overall health and welfare of a nation. For example, in more developed countries with less inequality, life expectancy is typically higher because better healthcare and government services allows for the population to live longer. Similarly, a higher median age indicates an aging population, which is a common occurrence in many developed nations today, unlike in developing countries which tend to have lower median ages. These two variables demonstrated a negative relationship with Gini. Fertility rates also showed a strongly negative relationship with Gini, which as a surprising outcome, since we expected fertility rates to have a positive relationship with income inequality. On the other hand, as expected, greater unemployment predicts greater levels of income inequality because unemployment typically affects low or unskilled labor, further exacerbating the income divide.

As a final note, after performing the final multiple regression analysis on the smaller sample of the OECD countries, the natural log of GDP per capita become the only statistically significant factor to predict inequality. The relationship between natural log of GDP per capita and Gini also becomes negative. This seems to support Kuznets Hypothesis, which would be interesting to analyze in the future as well.

## References


Alkire, Sabina, and Gisela Robles. *Global Multidimensional Poverty 2017*. Oxford Poverty and Human Development Initiative, 2017, www.ophi.org.uk/wp-content/uploads/B47_Global_MPI_2017.pdf.

Alkire, S., Kanagaratnam, U. and Suppa, N. (2018). 'The Global Multidimensional Poverty Index (MPI): 2018 revision', OPHI MPI Methodological Notes 46, Oxford Poverty and Human Development Initiative, University of Oxford.

Cysne, Eubens Penha, and David Turchick. "Equilibrium Unemployment-Inequality Correlation." *Journal of Macroeconomics*, vol. 34, no. 2, June 2012, pp. 454–469. *ScienceDirect*, doi: 10.1016/j.jmacro.2011.12.009.

Fosu, Augustin. "Growth, Inequality, and Poverty Reduction in Developing Countries: Recent Global Evidence." *SSRN Electronic Journal*, vol. 71, no. 2, May 2011, pp. 306–336., doi:10.2139/ssrn.1813968.

Hassan, Syeda Anam, et al. "The Relationship between Growth-Inequality-Poverty Triangle and Environmental Degradation: Unveiling the Reality." *Arab Economic and Business Journal*, vol. 10, no. 1, June 2015, pp. 57–71., doi:10.1016/j.aebj.2014.05.007.

Lam, David. "Chapter 18 Demographic Variables and Income Inequality." *Handbook of Population and Family Economics Handbook of Population and Family Economics Volume 1*, 1997, pp. 1015–1059. *ScienceDirect*, doi:10.1016/s1574-003x(97)80010-4.

"Median Age Country Comparison." The World Factbook, CIA, 2017, www.cia.gov/library/publications/the-world-factbook/fields/2177.html.

OECD (2018), Fertility rates (indicator). doi: 10.1787/8272fb01-en (Accessed on 19 November 2018)

OECD (2018), "Health status", *OECD Health Statistics* (database), https://doi.org/10.1787/data-00540-en (accessed on 19 November 2018)

OECD (2018), "Table 1 - Gross domestic product, volume, 2015=100: Seasonally adjusted", in *Main Economic Indicators, Volume 2018 Issue 11*, OECD Publishing, Paris, https://doi.org/10.1787/mei-v2018-11-table1-en.

*World Development Indicators*, The World Bank, 14 Nov. 2018, datacatalog.worldbank.org/dataset/world-development-indicators.

## Appendix

**Table 1: List of Countries included in Dataset (*denotes an OECD member country)**

| | | | | | |
|---|---|---|---|---|---|
| **Albania** | **Comoros** | **Guinea** | **Luxembourg*** | **Paraguay** | **Tanzania** |
| **Algeria** | **Congo** | **Guinea-Bissau** | **Madagascar** | **Peru** | **Thailand** |
| **Angola** | **Congo, Democratic Republic of** | **Haiti** | **Malawi** | **Poland*** | **The Netherlands*** |
| **Armenia** | **Côte d'Ivoire** | **Hungary*** | **Maldives** | **Portugal*** | **Timor-Leste** |
| **Australia*** | **Czech Republic*** | **Iceland*** | **Mali** | **Rwanda** | **Togo** |
| **Austria*** | **Denmark*** | **India** | **Mauritania** | **Sao Tome and Principe** | **Tunisia** |
| **Azerbaijan** | **Djibouti** | **Iraq** | **Mexico*** | **Senegal** | **Turkey*** |
| **Bangladesh** | **Ecuador** | **Ireland*** | **Moldova** | **Serbia** | **Uganda** |
| **Belgium*** | **El Salvador** | **Israel*** | **Mongolia** | **Sierra Leone** | **United Kingdom*** |
| **Belize** | **Estonia*** | **Italy*** | **Montenegro** | **Slovakia*** | **United States*** |
| **Benin** | **eSwatini** | **Japan*** | **Morocco** | **Slovenia*** | **Uzbekistan** |
| **Bhutan** | **Ethiopia** | **Jordan** | **Mozambique** | **South Africa** | **Vanuatu** |
| **Bolivia** | **Finland*** | **Kazakhstan** | **Myanmar** | **South Korea*** | **Yemen** |
| **Burkina Faso** | **France*** | **Kenya** | **Namibia** | **South Sudan** | **Zambia** |
| **Burundi** | **Gabon** | **Kyrgyzstan** | **Nepal** | **Spain*** | **Zimbabwe** |
| **Cameroon** | **Gambia** | **Laos** | **New Zealand*** | **Sudan** | |
| **Canada*** | **Germany*** | **Latvia*** | **Niger** | **Sweden*** | |
| **Central African Republic** | **Ghana** | **Lesotho** | **Nigeria** | **Switzerland*** | |
| **Chad** | **Greece*** | **Liberia** | **Norway*** | **Syria** | |
| **Chile*** | **Guatemala** | **Lithuania*** | **Pakistan** | **Tajikistan** | |

**Table 2A: Correlation Coefficients of Independent Variables (115 observations):**

```
. correlate lGDPperc fertrate medage lifeexp unemploy
(obs=115)
```

| | lGDPperc | fertrate | medage | lifeexp | unemploy |
|---|---|---|---|---|---|
| lGDPperc | 1.0000 | | | | |
| fertrate | -0.6404 | 1.0000 | | | |
| medage | 0.7877 | -0.8843 | 1.0000 | | |
| lifeexp | 0.7694 | -0.8462 | 0.8623 | 1.0000 | |
| unemploy | 0.0217 | -0.1233 | 0.0870 | -0.0193 | 1.0000 |

**Table 2B: Correlation Coefficients of Independent Variables (MPI included: 79 Observations)**

```
. correlate lGDPperc fertrate medage lifeexp unemploy mpi
(obs=79)
```

| | lGDPperc | fertrate | medage | lifeexp | unemploy | mpi |
|---|---|---|---|---|---|---|
| lGDPperc | 1.0000 | | | | | |
| fertrate | -0.2264 | 1.0000 | | | | |
| medage | 0.2397 | -0.8719 | 1.0000 | | | |
| lifeexp | 0.3284 | -0.7536 | 0.7206 | 1.0000 | | |
| unemploy | 0.0782 | -0.1661 | 0.1371 | -0.0229 | 1.0000 | |
| mpi | -0.1850 | 0.8067 | -0.7311 | -0.7583 | -0.2456 | 1.0000 |

**Appendix Output 1:** Simple Regression Model of Gini vs. MPI

```
. regress gini mpi
```

| Source | SS | df | MS |
|---|---|---|---|
| Model | .030062319 | 1 | .030062319 |
| Residual | .771925631 | 78 | .009896482 |
| Total | .80198795 | 79 | .010151746 |

| | | |
|---|---|---|
| Number of obs | = | 80 |
| F(1, 78) | = | 3.04 |
| Prob > F | = | 0.0853 |
| R-squared | = | 0.0375 |
| Adj R-squared | = | 0.0251 |
| Root MSE | = | .09948 |

| gini | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| mpi | .1209452 | .0693933 | 1.74 | 0.085 | -.0172063 | .2590967 |
| _cons | .3674288 | .0173006 | 21.24 | 0.000 | .3329859 | .4018717 |

**Appendix Output 2:** Simple Regression Model 2 of Gini vs. lGDPperc

```
. regress gini lGDPperc

      Source |       SS           df       MS      Number of obs   =       115
-------------+----------------------------------   F(1, 113)       =     17.31
       Model |  .136974836         1  .136974836   Prob > F        =    0.0001
    Residual |  .893921651       113  .007910811   R-squared       =    0.1329
-------------+----------------------------------   Adj R-squared   =    0.1252
       Total |  1.03089649       114  .009042952   Root MSE        =    .08894

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lGDPperc |  -.0214368   .0051517    -4.16   0.000    -.0316433   -.0112304
       _cons |   .5478953   .0439141    12.48   0.000     .4608935    .6348972
------------------------------------------------------------------------------
```

**Appendix Output 3:** Multiple Regression Model 1 STATA output

```
. regress gini mpi lGDPperc fertrate lifeexp medage unemploy

      Source |       SS           df       MS      Number of obs   =        79
-------------+----------------------------------   F(6, 72)        =      4.04
       Model |  .201287076         6  .033547846   Prob > F        =    0.0015
    Residual |  .598508899        72  .008312624   R-squared       =    0.2517
-------------+----------------------------------   Adj R-squared   =    0.1893
       Total |  .799795975        78  .010253795   Root MSE        =    .09117

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mpi |   .0885083   .1252115     0.71   0.482    -.1610963    .3381129
    lGDPperc |   .0013614   .0122026     0.11   0.911    -.0229641    .0256868
    fertrate |  -.0474059   .0178446    -2.66   0.010    -.0829786   -.0118333
     lifeexp |   -.006266   .0027535    -2.28   0.026     -.011755    -.000777
      medage |  -.0070032   .0035985    -1.95   0.056    -.0141767    .0001703
    unemploy |   .0034234   .0017608     1.94   0.056    -.0000867    .0069336
       _cons |   1.087997   .2313891     4.70   0.000     .6267309    1.549263
------------------------------------------------------------------------------
```

**Appendix Output 4:** Multiple Regression Model 2 STATA output

```
. regress gini lGDPperc fertrate lifeexp unemploy medage

      Source |       SS           df       MS      Number of obs   =       115
-------------+----------------------------------   F(5, 109)       =     10.05
       Model |  .325333058         5  .065066612   Prob > F        =    0.0000
    Residual |  .705563429       109  .006473059   R-squared       =    0.3156
-------------+----------------------------------   Adj R-squared   =    0.2842
       Total |  1.03089649       114  .009042952   Root MSE        =    .08046

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lGDPperc |   .0098492   .0083859     1.17   0.243    -.0067714    .0264698
    fertrate |  -.0334391   .0123131    -2.72   0.008    -.0578432   -.009035
     lifeexp |  -.0060967   .0019783    -3.08   0.003    -.0100176   -.0021758
    unemploy |   .0027103   .0013204     2.05   0.043     .0000932    .0053273
      medage |  -.0053165   .0021232    -2.50   0.014    -.0095246   -.0011083
       _cons |   .9483155   .1463884     6.48   0.000     .6581785    1.238453
------------------------------------------------------------------------------
```

**Appendix Output 5:** Restricted Multiple Regression Model 1 to perform F-test:

```
. regress gini fertrate lifeexp medage unemploy

      Source |       SS           df       MS      Number of obs   =        79
-------------+----------------------------------   F(4, 74)        =      6.04
       Model |  .196749139         4  .049187285   Prob > F        =    0.0003
    Residual |  .603046836        74  .008149282   R-squared       =    0.2460
-------------+----------------------------------   Adj R-squared   =    0.2052
       Total |  .799795975        78  .010253795   Root MSE        =    .09027

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    fertrate |  -.0425676   .0164192    -2.59   0.011    -.0752835   -.0098517
     lifeexp |   -.007029    .002343    -3.00   0.004    -.0116974   -.0023605
      medage |  -.0070202   .0035596    -1.97   0.052    -.0141129    .0000724
    unemploy |   .0030553   .0016397     1.86   0.066    -.0002118    .0063224
       _cons |   1.150472   .2126561     5.41   0.000      .726746    1.574199
------------------------------------------------------------------------------
```

**Appendix Output 6:** Testing the different functional form of our Multiple Regression model using natural log of Gini as the dependent variable. (lgini represents the natural log of the Gini index)

```
. regress lgini fertrate medage unemploy lifeexp GDPperc

      Source |       SS           df       MS      Number of obs   =       115
-------------+----------------------------------   F(5, 109)       =      3.11
       Model |  1.81218376         5  .362436752   Prob > F        =    0.0116
    Residual |  12.7076845       109  .116584261   R-squared       =    0.1248
-------------+----------------------------------   Adj R-squared   =    0.0847
       Total |  14.5198683       114  .127367265   Root MSE        =    .34144

------------------------------------------------------------------------------
       lgini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    fertrate |  -.0850818   .0532412    -1.60   0.113    -.1906041    .0204404
      medage |  -.0093326   .0092239    -1.01   0.314    -.0276141    .0089489
    unemploy |    .006272   .0056105     1.12   0.266    -.0048479    .0173919
     lifeexp |  -.0174309   .0082214    -2.12   0.036    -.0337254   -.0011365
     GDPperc |   1.64e-06   3.06e-06     0.54   0.593    -4.42e-06    7.70e-06
       _cons |   .6386927    .705386     0.91   0.367    -.7593595    2.036745
------------------------------------------------------------------------------
```

**Appendix Output 7A:** Testing different functional forms of GDP per capita (GDPperc2 represents GDP per capita squared, lGDPperc represents the natural log of GDP per capita)

```
. regress gini fertrate medage unemploy lifeexp GDPperc GDPperc2

      Source |       SS           df       MS      Number of obs   =       115
-------------+----------------------------------   F(6, 108)       =      8.63
       Model |  .334211788         6  .055701965   Prob > F        =    0.0000
    Residual |  .696684699       108  .006450784   R-squared       =    0.3242
-------------+----------------------------------   Adj R-squared   =    0.2867
       Total |  1.03089649       114  .009042952   Root MSE        =    .08032

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    fertrate |  -.0388999   .0131764    -2.95   0.004    -.0650177    -.012782
      medage |  -.0067499    .002447    -2.76   0.007    -.0116003   -.0018996
    unemploy |    .002817   .0013202     2.13   0.035     .0002003    .0054338
     lifeexp |  -.0063528   .0019673    -3.23   0.002    -.0102524   -.0024532
     GDPperc |   2.85e-06   1.79e-06     1.59   0.114    -6.92e-07    6.39e-06
    GDPperc2 |  -2.94e-11   2.31e-11    -1.27   0.206    -7.53e-11    1.64e-11
       _cons |   1.082555    .176587     6.13   0.000     .7325285    1.432581
------------------------------------------------------------------------------
```

**Appendix Output 7B:**

```
. regress gini fertrate medage unemploy lifeexp GDPperc lGDPperc

      Source |       SS           df       MS      Number of obs   =       115
-------------+----------------------------------   F(6, 108)       =      8.34
       Model |  .326416246         6  .054402708   Prob > F        =    0.0000
    Residual |  .704480241       108  .006522965   R-squared       =    0.3166
-------------+----------------------------------   Adj R-squared   =    0.2787
       Total |  1.03089649       114  .009042952   Root MSE        =    .08076

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    fertrate |  -.0345735     .01267    -2.73   0.007    -.0596876   -.0094593
      medage |  -.0055681   .0022191    -2.51   0.014    -.0099667   -.0011696
    unemploy |    .002772   .0013342     2.08   0.040     .0001274    .0054165
     lifeexp |   -.006184   .0019974    -3.10   0.002    -.0101432   -.0022248
     GDPperc |   3.83e-07   9.41e-07     0.41   0.684    -1.48e-06    2.25e-06
    lGDPperc |   .0069911   .0109571     0.64   0.525    -.0147278      .02871
       _cons |   .9834877   .1704245     5.77   0.000     .6456767    1.321299
------------------------------------------------------------------------------
```

**Appendix Output 8:** Multiple Regression Model 2 with only OECD countries

```
. regress gini lGDPperc fertrate lifeexp unemploy medage

      Source |       SS           df       MS      Number of obs   =        36
-------------+----------------------------------   F(5, 30)        =      3.41
       Model |  .036799464         5  .007359893   Prob > F        =    0.0147
    Residual |  .064699759        30  .002156659   R-squared       =    0.3626
-------------+----------------------------------   Adj R-squared   =    0.2563
       Total |  .101499222        35  .002899978   Root MSE        =    .04644

------------------------------------------------------------------------------
        gini |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lGDPperc |   -.069907   .0317882    -2.20   0.036    -.1348272   -.0049869
    fertrate |   .0148932   .0338932     0.44   0.664    -.0543259    .0841123
     lifeexp |   .0023673   .0041354     0.57   0.571    -.0060784     .010813
    unemploy |   .0001923   .0018738     0.10   0.919    -.0036345    .0040192
      medage |  -.0041528   .0027819    -1.49   0.146    -.0098341    .0015286
       _cons |   1.002281   .2906091     3.45   0.002     .4087784    1.595784
------------------------------------------------------------------------------
```