Income Inequality in IKO Counties

C. Boone, R. Khan, B. Welker

Introduction:

The topic of income inequality is one that has captured the attention and inflamed the passions of the general public throughout history, especially in recent years. The Occupy Wall Street movement of 2011 demonstrated the rage that many Americans felt about an economic system perceived as rigged to benefit the top 1% of wealthy Americans. Great and growing differences in earnings and wealth between members of society has the ability to incite resentment, fear, and premature death. Indeed, the case was made in the American Journal of Public Health that low income and income instability can be linked to higher mortality rates (McDonough, et al. 2011).

Policymakers have high incentive to pay close attention to trends of and attitudes about extreme income disparities, as the consequences of rampant inequality can be a danger to society, whether locally, nationally, or globally (Kochhar, K., et al., 2015). This paper will focus on causes of income inequality at the local level in the states of Indiana, Kentucky and Ohio. This analysis examines the effects to which certain social statistics that are postulated to drive variance in economic equality affect GINI at a county level within the scope of the model.

The GINI coefficient, developed by Italian statistician Corrado Gini in 1912, will function as the dependent variable. The coefficient, which ranges from zero (perfect equality in within a population) and one (all wealth or income in a population held by one individual). Using our county-level data gleaned from the U.S. Census Bureau, we will attempt to demonstrate the relationship between economic growth and income inequality. The Lorenz Curve is a graphical representation of the cumulative income earned by a given percentile of people sorted by income from lowest to highest. (US Census Bureau) The GINI coefficient describes the degree of difference from the cumulative distribution of income in the Lorenz Curve from that of perfect equality of income, or linear equality. Using our county-level data gleaned from the U.S. Census bureau and other sources, we will attempt to demonstrate the relationship between economic growth and income inequality.

In the United States, the study of income inequality can be traced back to the 20th century. Early works by Edwin Cannan's "Division of Income" (1905) and H.J. Davenport's "Value and Distribution" (1908) mark the earliest mentions of the issue of income distribution. More recently, the conversation has shifted towards examining the economic, political, and societal forces driving the mostly consistent growth in inequality. Nationally, a study out of UC Berkeley (Saez, 2016) tracked the rise in income share of the top 10% and top 1% of incomes over nearly the last century. As can be seen in Figure 1 below, the top 10% of earners in the U.S. have never accounted for less than 30% of all income nationally. Further, the top earning decile recently accumulated as much as half of all income.

Levernier, Rickman, and Partridge (1995) examined economic data for the contiguous 48 states in the U.S. and studied inequality trends at a regional level. Key findings of their research were that some notable causes of inequality increase during the 1980s were due to heightened international immigration and the rise of households headed by a single female. Additionally, their research supported arguments that factors such as greater high school attainment, labor force participation, goods-producing employment share, and transfer payments aided in reducing or mitigating the rise of income inequality. Both of these later values were included as independent variables in all regressions.

The work of with Bennet and Vedder (2013) examined this relationship using state-level data. Their research postulates that increasing degrees of economic freedom is associated with lower degrees of income inequality. Building on the work Ashby and Sobel (2008) argued that the relationship between economic freedom and inequality was represented by an inverse U-curve. At low levels of economic freedom, a marginal increase will yield higher inequality. However, as economic freedom continues to rise, marginal increases will begin to lower inequality. Figure 1 below illustrates this relationship.

Figure **1**. Average Economic Freedom.

Figure **2.** The Poverty, Inequality, and Growth Triangle.

The relationship between income inequality, growth and poverty is conceptualized via Figure **2**. The interactions between poverty, income inequality, and economic growth show that when these values experience growth or decline, there is a ripple effect that reaches across all income groups (Naschold, 2003). Data regarding these values were also included in our regressions to see how they affected results.

Data:

The dependent variable, $GINI_{ij}$, (GINI) is a value calculated by the US Census that is specific to each county in the sample population. Where, *i*, represents the respective county in the analysis, and *j,* represents the respective census year for which the value was calculated. The analysis was composed of data collected from American Community Survey, collected by the US Census, which aggregates survey data from households in a certain geographic to confidently assign aggregate values for calculations regarding income inequality. The values are calculated using the varying income levels respective to their share of the population area. (US Census)

Data was collected at a county-wide level for Indiana, Kentucky, and Ohio, which comprises some 300 independent county locales. All data was compiled by variable and by year, with the time and state variables being controlled by factored dummy time and dummy state variables. These were controls used by Levernier and Rickman to control for regional and time effects in their pooled OLS regression. The observations were collected in 1990, 2000, and 2010, using the ACS 5-year estimate values.

Table **1.** Descriptive Statistics.

The explanatory variables, unemployment, per capita income, population, and percent of people over age 25 with a high school degree (education), were common factors used in the OLS regression suggested by Levernier and Rickman in 2013, amongst other more specific economic and demographic variables. To investigate the model using quantile and finite mixture models, data for median income per quintile was added to further characterize the counties. This data was unused in the OLS and Panel models. The percent of people below the federal poverty line, $Poverty_{ij}$, (poverty), was added to see how it would affect the regression results. These variables are hypothesized to be indicators of the robustness of wealth distribution across a populace and will be tested to see how they affect GINI coefficients at a county-level.

When collecting data regarding multiple geographic across a common area, Levernier and Rickman, in their analysis of a state by state income inequality, discuss how multicollinearity affected their results. They note that single period regressions have fewer degrees of freedom by nature, and that further, measures that affect both measures that labor markets and demographic variables, like education, could affect each other. Controls for these interactions will be discussed in the empirical approach section.

The models for regression were adapted from a thorough literature review which identified two major trends in this type of analysis, pooled OLS and panel regressions. The data was collected into a panel frame for the specified variables and time periods, *j*. Pooled OLS was thought to aid in the control of regional effects on results and could aid in predicting how the variables are able to affect GINI values (Levernier and Rickman, 1995).

The data for the GINI coefficient, income, education rates and poverty rates were arranged for the 3 states in a single panel data format. Quantitative data analysis such as the ADF and KPSS tests were conducted to dictate the stationarity of the data. Descriptive statistics were performed to determine the nature of the data, which is indicated in Table **1**.

The dataset was tested for several OLS assumptions to address issues such as multicollinearity, heteroskedasticity, and stationarity. It was hypothesized that the effects of time would have a correlation on results, and this was identified once graphing poverty vs. time, which is displayed in Figure **3**.

Figure **3**. Plot of Poverty vs. time.

Literature addressed for issues arising between using single period regressions by adding dummy variables for time. Multicollinearity was addressed by adding dummy variables, and by adding more defining demographic statistics regarding both economic effects such as income per sector, or income of goods producing sectors, and demographics including ethnicity and gender. These data could possibly positively affect the validity of the results and need to be investigated further. These specific regressions used in this paper used the rates of change for the entire, unspecified population.

Figure **4**. GINI values per county for 1990, 2000, and 2010.

Empirical Approach:

The variables income and population follow a wide distribution which ranges over several orders of magnitude thus it is crucial to take the logarithm of those variables. This transformation helps to normalize the skewness of the data. (Croissant and Milo, 2008)

The explanation behind each test was driven by the Journal of Statistical Software, Panel Data Econometrics in R: The plm Package, and Chapter 15 from Principles of Econometrics with R, Panel Data Models. The Lagrange Multiplier Test (Gouriéroux, Monfort, 1989) was ran for two-way fixed effects (county and year). We reject the null hypothesis at 99.9% significance level that there are no significant two-way fixed effects with a p-value < 2.2 e^{-16}. Thus, there are significant two-way fixed effects. With respect to the null hypothesis of trend stationarity, the KPSS test indicates that the poverty and unemployment variables were non-stationary, while affirming that the other variables were stationary. (Kwiatoski et al, 1992)

Serial correlation was tested using the Breusch-Godfrey/Woolridge test with the null hypothesis being that there is no serial correlation. P-value of 0.003454 was obtained, which means we reject the null at 99.9% level of confidence. The Breusch-Pagan test was used to check for heteroskedasticity. With a p-value of 5.609e^{-7}, the null hypothesis of homoscedasticity has to be rejected at 99.9% confidence level. Thus, the data is heteroscedastic.

Using a pooled fixed effects OLS model was supported by Levernier and Rickman. They use this type of regression to mask the bias caused by controlling for year and location of where the samples were collected by means of dummy variables. The authors also note that, the fixed effects used in their model (which was a regional analysis), could have been incomplete in their goal of controlling for state fixed effects in such a cross-section specification, which could bias results. In order to address heteroskedasticity and stationarity for the OLS regression the variables were differenced to an order of I(1).

With regards to the panel data analysis, a large sample size and small number of time periods fixed effects model is said to the more efficient than the first difference model. (Advanced Panel Data Methods)

The first difference (FD) model is a solution to tackle the problem of unobserved time constant variant heterogeneity of the data. When serial correlation and heteroskedasticity is present in the data the FD model is a more efficient than the fixed effects model (Wooldridge, 2006). We decided to base the panel data analysis was established upon fixed effects and first difference models. The fixed effects also were verified as valid in the pooled OLS regressions by using ANOVA. The results from this analysis are displayed in Table **2**.

Table **2**. ANOVA.

The ANOVA results showed that the added dummy variables were significant. The added fixed effects also tried to address the possibility of omitted variable bias, and sought to mitigate the effects that regional, state, and time bias were affecting the results.

The regressions were tested via VIF testing for multicollinearity. The VIF values in the undifferenced regression had unacceptably high VIF values for two variables, log per capita income, 12.545916, and the factored dummy year variable, with a VIF value equaling 31.280846. This problem was sought to be mitigated by differencing values that literature suggesting could be causing the discrepancies, and the protocol used to correct this is listed in results. Multicollinearity in regressions using data such as single locales over specific time periods is a problem is common across literature (Levernier and Rickman, 1995), especially among education and unemployment values. Post-differencing, VIF values for all variables decreased significantly.

The inclusion of a quantile model was hypothesized to better understand the interactions between independent variables and GINI at the 25th and 75th

percentile. Distributions does not only differ by their means, but they also might differ based on their upper or lower quantiles. Using OLS regression to model the mean may disregard the association between the dependent and the independent variables especially when the dependent variable is biased by different groups that exist within it. Thus, the quantile regression helps us to model the 25th and 75th percentile of the dependent variable (GINI coefficient) (Beyerlein, 2014).

Upon examining Figure **4.**, counties with high income inequality can be easily identified, and aggregated. The finite mixture model analyzed the probabilistic tendencies of grouping counties into separate clusters to analyze common trends among the sample population. This model was first suggested to be used for analyzing trends among clusters by McLachlan in 2000. Although the clusters will be identified in this investigation, the regressions indicated below will only account for these clusters by means of dummy variables, where indicated.

Models

Fixed-Effects Pooled OLS Model (and) Quantile Model (without interaction)

$$GINI = \beta_0 + \beta_1 Population\ (log)_{ij} + \beta_2 Education_{ij} + \beta_3 Poverty_{ij} + \beta_4 Percap\ Income(log)_{ij} + \beta_5 Unemployment_{ij} + \phi_6 KYdummy_{ij} + \phi_7 OHdummy_{ij} + \phi_8 2000dummy_{ij} + \phi_9 2010dummy_{ij} + u$$

Pooled OLS Model (with interaction)

$$GINI = \beta_0 + \beta_1 Population\ (log)_{ij} + \beta_2 Education_{ij} + \beta_3 Poverty_{ij} + \beta_4 Percap\ Income(log)_{ij} + \beta_5 Unemployment_{ij} + \phi_6 KYdummy_{ij} + \phi_7 OHdummy_{ij} + \beta_8 (Education * Poverty)_{ij} + u$$

Panel Model

$$GINI = \beta_0 + \beta_1 Education_{ij} + \beta_2 Poverty_{ij} + \beta_3 Percap\ Income(log)_{ij} + \beta_4 Unemployment_{ij} + \phi_5 KYdummy_{ij} + \phi_6 OHdummy_{ij} + u$$

Finite Mixture Model Classification Model

$$P(GINI) = \beta_0 + \beta_1 Poplation_i + \beta_2 Education_i + \beta_3 Poverty_i + \beta_4 Percap\ Income_i + \beta_5 Unemployment_i + \phi_6 Percap1Q_{ij} + \cdots + \phi_{10} Percap5Q_{ij} + u$$

Results:

Tables **3**. and **4.** Fixed-Effects Pooled OLS Regressions;
Table **5.** Panel Regressions;
Table **6.** Quantile Regressions;
Table **7**. Finite Mixture Probabilities.

The results from each analysis will be described below in their respective section. For detailed information regarding individual elasticities and results, refer to the tables indicated above which are listed in the supplementary section at the end of the investigation.

Fixed Effects-Pooled OLS

Empirical estimates of the pooled OLS models are shown in tables **3** and **4**. The regression, without control for interaction terms (in table **3**.), estimates show factors that impact interstate state level of income inequality at a particular point in time. All variables, explicitly excluding per capita income, unemployment and the state dummy for Ohio, are significant with p-values < 0.01. The estimates along with the value of the adjusted R^2 values show that the investigated variables on average explain a considerable share of interstate differences in state income inequality between the years 1990-2010.

Upon examining the values of the regression that did not include an interaction term, the time dummies become less negative over time. This specifically is in relation to the dropped time control, 1990. The regression showed that simply being in the year 2010 (as opposed to the control year), had a negative effect of -0.0218 units on GINI. The magnitude for the coefficient of the state dummy for Kentucky is higher than the state control dummy (Indiana), indicating that Kentucky's marginal effect on the average levels of the GINI coefficient is the higher than that of Indiana, which supports the empirical findings by the Center of Budget and Policy Priorities indicating that Kentucky has higher levels of income inequality than Indiana.

Further, we will examine the effects on the model by adding an interaction term which examined the effects between poverty and education. The interaction term was highly significant, with a p value of $1.21e^{-9}$. With this interaction term, the effect of the time dummy became highly multicollinear, and for that reason it was removed from the regression to stabilize the other variables. The joint elastic effect of poverty and education shows that for each one unit increase in education and poverty, there is a negative effect to the magnitude of -0.14 units on GINI. Poverty and income inequality share a negative relationship. This negative relationship is demonstrated by the poverty-growth-income triangle. In an area with high income inequality, economic growth leads to a reduction in poverty rates and exacerbates the level of income inequality. (Bourguignon, 2004)

The effect of the time dummies on the model with an interaction term was significantly diminished, and further, VIF values taken after this regression was shown proved to that the time dummies added multicollinearity to the regression, and for this reason they were removed. The results from this regression also diminished the confidence of the state fixed effects, but this will be discussed further in the finite mixture results.

The OLS regression that was deemed to be the best fit model was that for which the nonstationary variables, poverty, and unemployment were differenced to the first order, I(1), and for which an interaction term was added to control for the synergy between education and poverty. Both models were fixed effects models after ANOVA showed its significance. As discussed previously, KPSS testing and visual inspection (figure **3**.) showed these variables to be the most affected by stationarity. Other values were trimmed to 899 observations to avoid variable length errors. Although both the R^2 and AIC/BIC statistics for this regression were lower than that of the non-differenced model, VIF testing confirmed that multicollinearity was significantly decreased by differencing the nonstationary variables, and in the case of the model with an interaction term, multicollinearity was further aided by dropping the state fixed-effects. This led to the conclusion that differencing the model was overall a more accurate representation of the results.

Both regressions showed negative elasticities in the educational category, which was expected, with a calculated value of -0.24 percentage points for the differenced model without an interaction term. This value, however, must be discounted due to, as noted earlier, the significant interactions between poverty and education in the way to which these characteristics affect GINI values, ceteris paribus. Per the model, educational attainment is positively related to GINI, but as poverty increases this value become less positive and can turn negative when poverty raises to a certain extent. Higher educational attainment rates were expected to decrease per unit GINI, and under inspection this relationship has validity. As noted in previous research, as this value increases there is a negative effect on income inequality (Sale, 1974).

The first differenced OLS model without an interaction term shows that all the variables are significant other than the unemployment rate, log income and the state dummy variable for Ohio, when compared to the control for state fixed-effects, Indiana. Upon addition of the interaction term, all values excluding the fixed effects variables are highly statistically significant.

The log of income per capita was found to be significant in the final OLS regression, with a positive elastic effect of 0.0002 on GINI for a one percentage point increase of per capita income, all else constant. However, the depth to which income per capita reaches could be a cause for biased results. Research into this topic typically calculates their own statistics for analyzing real share of income in a geographic region (Bennet, Vedder, 2013). Considering this, further investigation into the GINI coefficients when considering a dataset like ours would need to consider the pros and cons for determining which value to use for a true median income value.

Panel

The dependent variable in all the fixed effects regressions is the GINI coefficient. We started with all the variables with the fixed effects model. It was determined that all the variables other than the education variable were statistically significant. The coefficient of the independent variable indicates how much the GINI coefficient changes over time on average per county when the independent variable increases by one unit. A change in one percentage point of the logged independent variables would cause a responsive change in GINI, which is determined by the beta coefficient being divided by 100 percentage points. An increase in one percentage point of income would lead to an increase in 0.0063 percentage points of the GINI.

An interesting finding of this analysis is that income inequality is positively correlated with the county level incomes. Our results align with the findings of Partridge (1997) who used a panel of 48 US states to show that state level incomes were positively correlated with income inequality. His hypothesis was the first one to contest the idea that income inequality was harmful for economic growth. Income inequality increases in times of high economic growth.

The first differenced model indicates that the log of per capita income, educational attainment, poverty and the state dummies are statistically significant. Compared to the fixed-effects estimation, all the variables of the first differenced model have a higher magnitude after removing the unobserved time heterogeneity. The education rate has a negative coefficient indicating that a unit increase in the education rate would decrease the GINI coefficient by 0.000729 percentage points on average, ceteris paribus. As economic growth is more robust the gap between the rich and poor is increasing. This leads to higher education being accessible to families with higher incomes while students from low income households are less likely to go to college and to graduate. (Hill, 2015)

Quantile Model

A quantile regression was run for cross section of the data consisting of data specific to the year 2010. The dependent variable, GINI was tested at the 25th and 75th percentiles while the independent variables being population, education, poverty, unemployment and income. The variable education was insignificant while poverty, population and unemployment were significant at both the 25th and 75th percentiles of GINI. The coefficients for poverty and population are same at both the percentiles. This shows that the effect of both poverty and population does not differ at the 25th and the 75th percentile. The effect of income differs considerably with the effect being significant on GINI only at the 25th percentile. At lower levels of income inequality (GINI at the 25th percentile), a one percentage point increase in income would decrease GINI by 0.0001 percentage points.

Finite Mixture Classification Model

This model characterized the data set into two categorically significant groups, with distinctive characteristics between them. As seen by the results of this characteristic model, the first group of counties, Group 1 (G1), which comprises some 22 observations out of the 900, contains counties only in Kentucky. The second, Group (G2), comprising 878 observations, had significantly different characteristics which affected GINI. G1 had a mean GINI value of 0.503 compared to G1's value of 0.422.

Their median per capita incomes differed by $9,525.50, with G1's average per capita income amounting to a shockingly low sum of $8,354.46. This model shows that there were two distinct groups of counties being analyzed in the collective dataset, and that the group with exceedingly higher GINI values (G1) were related in a distinct way from that of G2. Poverty values also were significantly higher in G1, were some 39.9% of the population was below the federal poverty line, and where G2's average was only 14.9%. The fact that G1 was comprised of only Kentucky counties further agrees with the OLS results indicating that Kentucky, compared to the control state of Indiana, had a significant effect on the regressions that were run.

Conclusion:

The paper examined the relationship between the commonly used measure of county income inequality, the GINI coefficient and numerous county specific economic variables such as population, income and unemployment along with time and state wise dummies. Interpreting the factors that can affect income inequality in highly developed countries is difficult and can sometimes show that higher levels of values such as per capita income and educational attainment can thus lead to lower levels of income inequality, such as suggested by Kutznets' U-Curve. The results of the previous studies were updated as it specifically examined the county wise income inequality. The OLS results show that the factors that contribute to changes in income inequality in our neighboring states and Ohio were education and poverty. Furthermore, it also shows that a county is more likely to have higher income inequality if it has lower education rates, marginally higher levels of poverty and is situated in Kentucky, this is visualized for our population set in Figure **4**.

The panel results show income inequality within a county over time is attributed to the log per capita income levels, poverty rates, education rates and the state that it is located in. Clustering by years is also apparent in the data under visual inspection, and further the finite mixture model showed two distinct groups in the entire dataset. To account for this, further investigations into this topic should make note of running different type of regressions to control for this bias. One such regression could use a spatial dependence model to account for the clustering. More so, as shown by the quantile model, that different variables have different impacts on GINI at the 25th and 75th percentiles, with income only being significant at the 25th percentile of GINI. The different magnitudes and significance of the coefficients of unemployment and income between the 25th and 75th percentile of GINI reiterates our justification of using a quantile regression.

This has several policy implications. Economic growth enhancing reforms should be implemented so that lower income households have access to higher levels of disposable income. Higher levels of disposable income would give households access to better education, healthcare and lower rates of poverty. This would narrow the income inequality in disposable income. (OECD, 2015)

Works Cited:

Ashby, N. J., & Sobel, R. S. (2008). Income inequality and economic freedom in the US states. Public Choice, 134(3-4), 329-346.

Advanced Panel Data Methods [Scholarly project]. (n.d.). In Montana State University. Retrieved from http://www.montana.edu/cstoddard/562/Chapter14studentnotes.pdf

Bennett, D. L., & Vedder, R. K. (2012). A Dynamic Analysis of Economic Freedom and Income Inequality in the 50 U.S. States: Empirical Evidence of a Parabolic Relationship. SSRN Electronic Journal. doi:10.2139/ssrn.2134650

Beyerlein, A. (2014). Quantile regression—opportunities and challenges from a user's perspective. American journal of epidemiology, 180(3), 330-331.

Bourguignon, F. (2004). The Poverty-growth-inequality triangle (Working paper No. 125). New Delhi: Indian Council for Research on International Economic Relations (ICRIER).

Cannan, E. (1905). The division of income. The Quarterly Journal of Economics, 19(3), 341-369.

Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. Journal of Statistical Software, 27(2), 1 - 43. doi:http://dx.doi.org/10.18637/jss.v027.i02

Davenport, H. J. (1908). Value and Distribution: A critical and constructive study. University of Chicago Press.

Frank, M. W. (2009). Inequality and Growth in The United States: Evidence from A New State-Level Panel of Income Inequality Measures. Economic Inquiry,47(1), 55-68. doi:10.1111/j.1465-7295.2008.00122.x

Galili, T. (2013, May 27). Log Transformations for Skewed and Wide Distributions. Retrieved from https://www.r-statistics.com/2013/05/log-transformations-for-skewed-and-wide-distributions-from-practical-data-science-with-r/

Glassman, B. (2017). Income Inequality among Regions and Metropolitan Statistical Areas: 2005 to 2015. Retrieved from https://www.census.gov/content/dam/Census/library/working-papers/2017/demo/SEHSD-WP2017-41.pdf

Gouriéroux, C., & Monfort, A. (1989). A general framework for testing a null hypothesis in a "mixed" form. Econometric Theory, 5(1), 63-82.

Grundy, P. (2017). Pulling Apart. Color and Character. doi:10.5149/northcarolina/9781469636078.003.0006

Hill, C. B. (2015). Income Inequality and Higher Education. Retrieved from https://www.acenet.edu/the-presidency/columns-and-features/Pages/Income-Inequality-and-Higher-Education.aspx

Iii, T. S. (1974). Interstate Analysis of the Size Distribution of Family Income, 1950-1970. Southern Economic Journal,40(3), 434. doi:10.2307/1056017

INCOME INEQUALITY HAS GROWN IN KENTUCKY(Rep.). (n.d.). Retrieved https://www.cbpp.org/sites/default/files/atoms/files/Kentucky.pdf

Kochhar, K., Suphaphiphat, N., Ricka, F., & Tsounta, E. (2015, June). Causes and Consequences of Income Inequality: A Global Perspective. Retrieved March 1, 2019, from https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2016/12/31/Causes-and-Consequences-of-Income-Inequality-A-Global-Perspective-42986

Kuznets, S. (1955). Economic Growth and Income Inequality. The American Economic Review,45(1). Retrieved from https://www.jstor.org/stable/pdf/1811581.pdf?refreqid=excelsior:5a1b7a97a79b79320f9062cb058a5a8e&seq=1#page_scan_tab_contents.

Levernier, William, Dan S. Rickman, and Mark D. Partridge. "Variation in US state income inequality: 1960-1990." International Regional Science Review 18.3 (1995): 355-378.

McDonough, P., PhD, Duncan, G. J., PhD, Williams, D., PhD, & House, J., PhD. (1997). Income Dynamics and Adult Mortality in the United States, 1972 through 1989. American Journal of Public Health,87(9). Retrieved from https://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.87.9.1476.

McLachlan, G.J., and Peel, D. (2000). Finite Mixture Models. New York: John Wiley & Sons.

Naschold, F. (n.d.). Why Inequality Matters for Poverty (Issue brief).

Saez, E. (2016). Striking it Richer: The Evolution of Top Incomes in the United States (Updated with 2015 preliminary estimates) (Tech.). University of California, Berkeley.

Sewell, W. H. (1971). Inequality of opportunity for higher education. American Sociological Review, 36(5), 793-809.

US Census Bureau. (2016, January 25). Income Inequality. Retrieved from https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html

Variation in U.S. State Income Inequality: 1960-1990. (n.d.). Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/016001769501800305

Variation in U.S. State Income Inequality: 1960-1990. (n.d.). Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/016001769501800305?casa_token=408gLj3fP-8AAAAA:j0EGJguJj_u2pd3HGQQhDDDRsYuqqf1JhnblR4eylm4_9Ifwqk-OtZARhD5jv-b-nti2wmYd5w

Wooldridge. (2006). Introductory Econometrics (3rd ed). Cincinnati, OH: South-Western, Cengage Learning.

Supplementary Data and Figures:

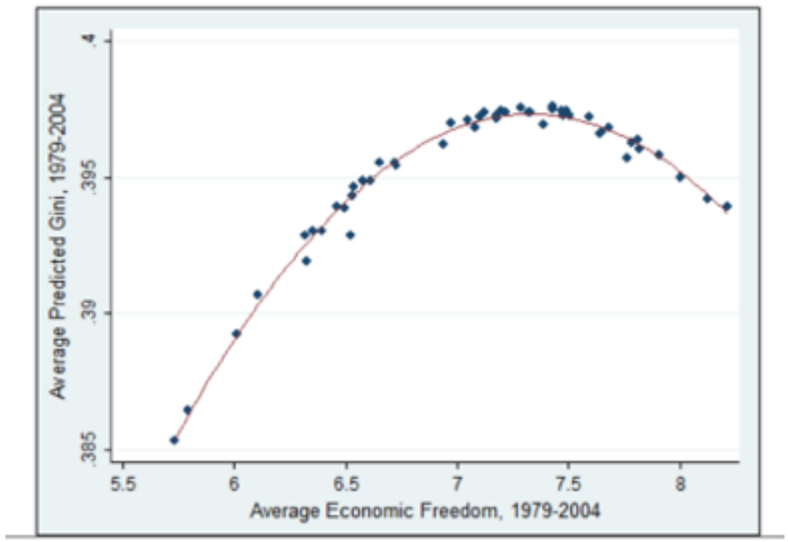

Figure **1**.

GINI coefficients as a function of Average Economic Freedom (1979-2004).

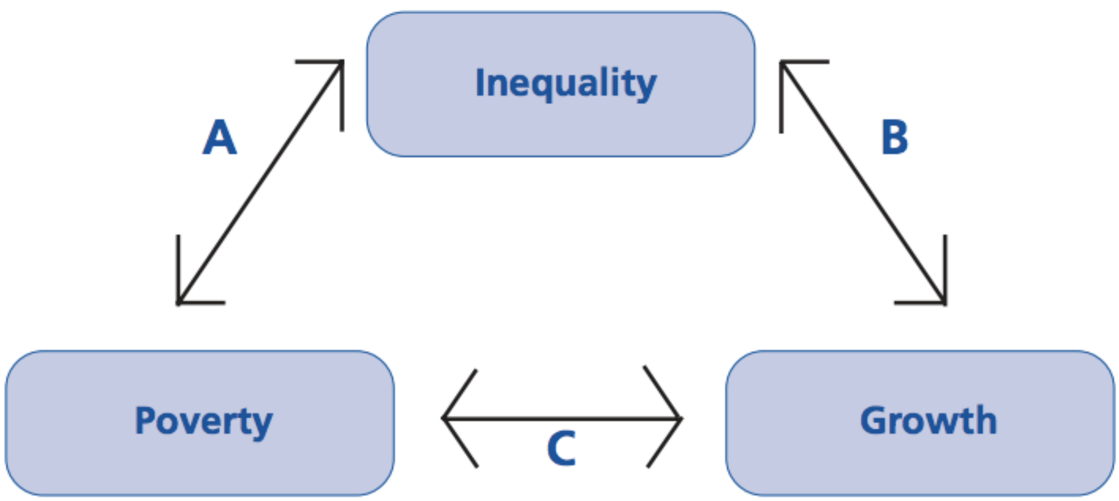

Figure **2**, The Poverty, Inequality, and Growth Triangle.

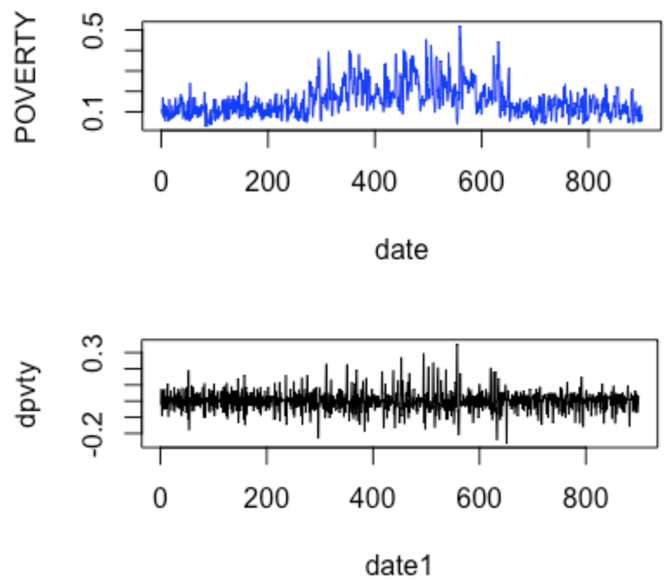

Figure **3**. Plot of Poverty vs. time

A plot of poverty versus time, where observations (date) 1-300 represent values in 1990, 301-600 represent values in 2000, and 601-900 represent values in 2010. There are three visible cluster which represent nonstationarity and spatial dependence. The second chart (date1) represents the poverty values post-differencing.

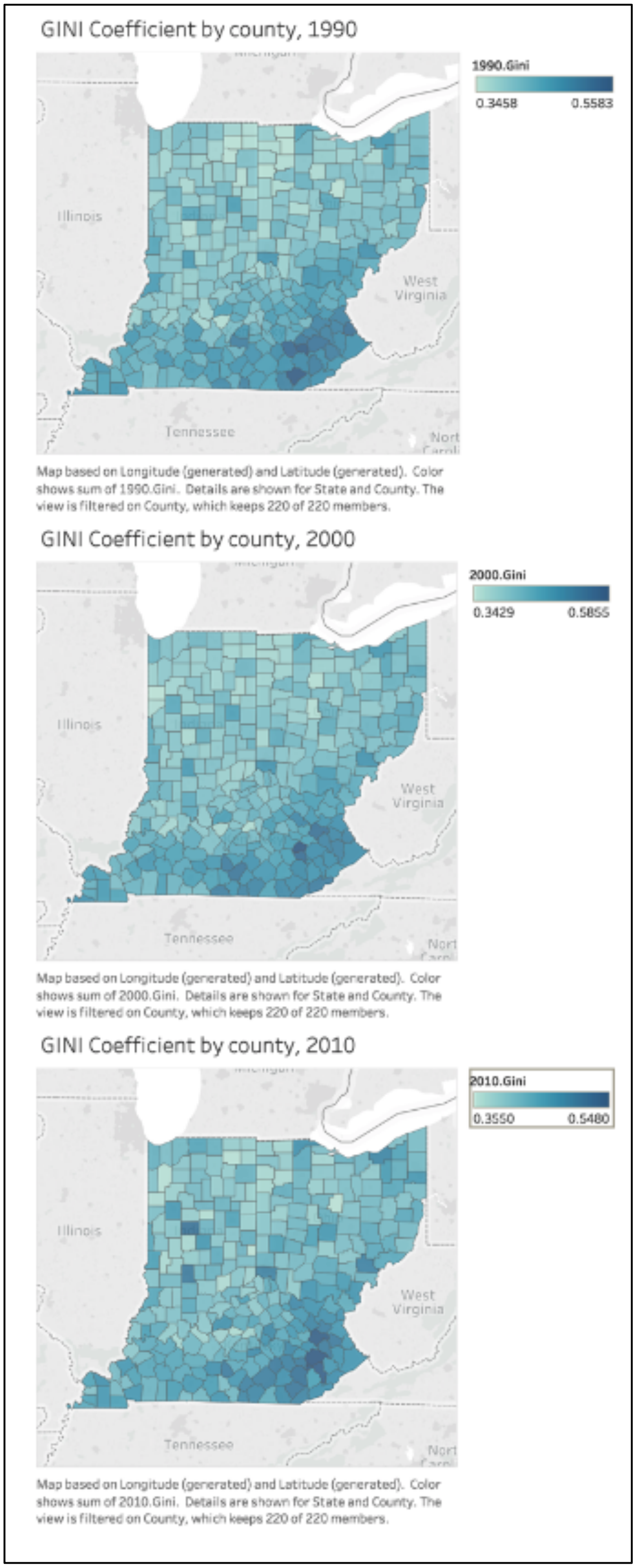

Figure **4.** GINI by county for 1990, 2000, and 2010. Kentucky has a disproportionate number of counties with high GINI compared to that of Indiana and Ohio.

Descriptive Statistics								
Variable Name	Units	Year	Observations	Mean	Min	Max	Std Dev	SOURCE
		1990	300	0.4184	0.3458	0.5583	0.4184	US Census
GINI	Index: 0 to 1	2000	300	0.4259	0.3429	0.5855	0.4259	US Census
		2010	300	0.4266	0.3550	0.5480	0.4266	US Census
		1990	300	6.7%	2.2%	15.3%	6.7%	US Census
Unemployment	Percent	2000	300	4.4%	1.9%	10.5%	4.4%	US Census
		2010	300	11.2%	6.5%	20.5%	11.2%	US Census
		1990	300	$ 10,975	$ 5,153	$ 20,426	$ 10,975	US Census
Income	USD	2000	300	$ 17,566	$ 9,716	$ 33,109	$ 17,566	US Census
		2010	300	$ 24,400	$ 10,767	$ 53,680	$ 24,400	US Census
		1990	300	66,922	2,124	1,412,140	66,922	US Census
Population	Number of People	2000	300	71,585	2,266	1,393,978	71,585	US Census
		2010	300	74,532	2,282	1,280,122	74,532	US Census
		1990	300	66.7%	35.5%	88.7%	66.7%	US Census
Educational Attainment	Percent	2000	300	75.6%	49.2%	94.2%	75.6%	US Census
		2010	300	81.5%	56.3%	96.1%	81.5%	US Census
		1990	300	15.5%	3.6%	52.1%	15.5%	US Census
Poverty Rate	Percent	2000	300	13.3%	2.9%	45.4%	13.3%	US Census
		2010	300	16.9%	4.6%	40.1%	16.9%	US Census

Table **1.**

This table contains summary statistics for the dataset which was used in the regressions. The data is was arranged vertically, as opposed to its horizontal representation above, and was used with dummy variables to distinguish between state and time. The 300 observations above include summary values for each county, collectively. The data was obtained by using the US Census' Factfinder tool.

Model	RSS	Df Sum of Sq	F	p-value
1 (without FE)	0.47033			
2 (with FE)	0.42184	0.048485	25.573	<2.2e-16***

P-value significance codes: " ` " p<.1 " * " p < .05 " ** " p < .01 " *** " p < .001

Table **2**. ANOVA

The above table was the results of an ANOVA analysis used to test the significance of the added fixed effects (FE), time period (1990, 2000, and 2010) and state (Indiana, Kentucky, and Ohio).

The results show that there is a significant correlation between the the model with FE to that of the control model which excludes them.

Undifferenced	Estimate (Std. Error)	**Differenced**	Estimate (Std. Error)
(Intercept)	-0.1229750* (0.0578)	Intercept	0.488638*** (0.0604)
Population (log)$_{ij}$	0.0089506*** (0.0009782)	Population (log)$_{ij}$	0.015857*** (0.00109)
Education$_{ij}$	-0.0434780* (0.0182)	Education$_{ij}$	-0.24261*** (0.0180)
Poverty$_{ij}$	0.4433866*** (0.0220)	dPoverty$_{ij}$	-0.071487*** (0.0170)
Per Capita Income$_{ij}$ (log)	0.0434665*** (0.00646)	Per Capita Income$_{ij}$ (log)	-0.004367 (0.00700)
Unemployment$_{ij}$	-0.0977200 . (0.0501)	dUnemployment$_{ij}$	-0.003608 (0.0447)
factor(YEAR)2000	-0.0025127 (0.00325)	factor(YEAR)2000	-0.046488*** (0.00493)
factor(YEAR)2010	-0.0218075 *** (0.00639)	factor(YEAR)2010	-0.012010* (0.00565)
factor(STATE)KY	0.0225679*** (0.00231)	factor(STATE)KY	0.028897*** (0.00261)
factor(STATE)OH	0.0096398*** (0.0022326)	factor(STATE)OH	0.002076 (0.00261)
Adjusted R^2	0.7043		0.5756
AIC	-4322.873		-3993.622
BIC	-4270.047		-3940.808

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table **3.** OLS Regressions, differenced and undifferenced models, without interaction control. The above table shows regression results for the two OLS restricted and unrestricted models that were analyzed during the investigation. Differenced variables are indicated with a d-prefix.

Descriptive statistics

	Dependent variable:	
	GINI (1)	dgini (2)
logpop	0.01***	0.002***
	(0.001)	(0.001)
EDU	-0.1***	0.1***
	(0.02)	(0.02)
POVERTY	0.2***	0.3***
	(0.05)	(0.02)
loginc	0.05***	0.02***
	(0.01)	(0.01)
unemp	-0.1**	-0.2***
	(0.05)	(0.03)
factor(YEAR)2000	-0.005	
	(0.003)	
factor(YEAR)2010	-0.03***	
	(0.01)	
factor(STATE)KY	0.02***	0.002
	(0.002)	(0.003)
factor(STATE)OH	0.01***	0.001
	(0.002)	(0.003)
EDU:POVERTY	0.4***	-0.6***
	(0.1)	(0.1)
Constant	-0.1**	-0.2***
	(0.1)	(0.05)
Observations	900	899
R2	0.7	0.3
Adjusted R2	0.7	0.3
AIC	-4346.065	-3880.982
BIC	-4288.436	-3832.969
Residual Std. Error	0.02 (df = 889)	0.03 (df = 892)
F Statistic	223.4*** (df = 10; 889)	55.9*** (df = 6; 892)

Note: *p<0.1; **p<0.05; ***p<0.01

Table **4.** OLS Regressions, differenced and undifferenced models, with interaction control. All variables under the dgini subsection have been differenced to order I(1). Upon adding the interaction term, the effect directly between GINI and education is significantly altered. This is because education acts through poverty when affecting GINI values, to an elastic effect of -0.14 units, ceteris paribus.

First Differenced Model	Estimate (Std. Error)	**Fixed Effects Model**	Estimate (Std. Error)
		(Intercept)	-0.1462255** (0.0517)
Education$_{ij}$	0.0097028 (0.0181)	Education$_{ij}$	-0.07286946*** (0.0194)
Poverty$_{ij}$	0.5173889*** (0.0212)	Poverty$_{ij}$	0.562566094*** (0.0208)
Per Capita Income$_{ij}$ (log)	0.6325697*** (0.0629)	Per Capita Income$_{ij}$ (log)	1.22995853*** (0.0876)
Unemployment$_{ij}$	-0.146225** (0.0517)	Unemployment$_{ij}$	-0.02091237 (0.0506)
factor(STATE)KY	0.0236551*** (0.00239)	factor(STATE)KY	0.0678401*** (0.0180)
factor(STATE)OH	0.0174029*** (0.00214)	factor(STATE)OH	0.10465427*** (0.0255)
Adjusted R^2	0.56638		0.6791

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table **5**. PLM Regressions

The above table shows regression results for the fixed effects and first difference estimations. The two-way fixed effects capture marginal effects of county and time.

Quantile Regression for the year 2010

	Dependent variable:	
	GINI	
	(25th)	(75th)
educ	0.02	-0.1
	(0.04)	(0.04)
povert	0.5***	0.5***
	(0.04)	(0.05)
Unemp	-0.4***	-0.5***
	(0.1)	(0.1)
logincome	-0.01*	0.01
	(0.01)	(0.01)
logpopulation	0.01***	0.01***
	(0.001)	(0.002)
Constant	0.4***	0.3**
	(0.1)	(0.1)
Observations	300	300
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table **6**. Quantile Regression Results

The above table shows regression results for the fixed effects and first difference estimations. The two-way fixed effects capture marginal effects of county and time.

Variable	Obs	Mean	Std. Dev.	Min	Max
state	0				
county	0				
CNTY_NUM	22	161.0909	31.43839	105	211
year	22	1995.455	7.385489	1990	2010
gini	22	.5036364	.0293656	.45	.56
unemp	22	9.631818	2.843939	5.7	15.3
percapin	22	8354.455	2376.351	5153	12983
pop	22	14156.09	7858.387	4755	31795
edu	22	.4737273	.0757277	.355	.653
poverty	22	.3990909	.0396519	.348	.521

Variable	Obs	Mean	Std. Dev.	Min	Max
state	0				
county	0				
CNTY_NUM	878	150.2346	87.57886	1	300
year	878	2000.114	8.159514	1990	2010
gini	878	.4218907	.0381927	.34	.59
unemp	878	7.378929	3.453651	1.9	20.5
percapin	878	17879.96	7177.796	6823	53680
pop	878	72437.59	144275.4	2124	1412140
edu	878	.7529754	.1071782	.395	.961
poverty	878	.1462128	.0690858	.029	.38

Tables **7a** and **7b**. Finite Mixture Model Results

Table **7a** shows results for Group 1 of counties and Table **7b** shows results for Group 2. Their identifying characteristics show that there a significantly different types of people living within each group. Results for income per quintile was excluded from results.