Illinois State University
ISU ReD: Research and eData

Master's Theses - Economics

Economics

6-1-2008

Differences Lead to Differences: Diversity and Income Inequality Across Countries

Michael Hotard
Illinois State University

Differences Lead to Differences:
Diversity and Income Inequality Across Countries

Michael Hotard
Master's Student
Department of Economics
Illinois State University
Normal, IL 61790

Under the Supervision of Dr. Oguzhan "Oz" Dincer

Submitted to Illinois State University in Partial Fulfillment of
the Requirements for a Master's Degree in Economics

June 2008

Differences Lead to Differences:
Diversity and Income Inequality Across Countries

Abstract

This paper will test the relationship between income inequality and ethnic heterogeneity. Although previous research has used ethno-linguistic fractionalization as a control variable in inequality regressions, no research has focused primarily on an alternative measure of heterogeneity – polarization. Using Gini coefficients from the from the World Income Inequality Database and polarization data from Montalvo and Reynal-Querol (2005b), a pooled OLS regression is run using data from 58 countries with 205 total observations. The results of these regressions suggest that ethnic polarization does have a positive effect on income inequality, even controlling for country characteristics and allowing for regional differences.

I. Introduction

Income inequality refers to the distribution of income across different populations in a society, specifically the gap between the income levels of the rich and the poor. The level of income inequality differs for countries throughout the world, and researchers have often questioned the causes and effects of these variations. Much research has studied the relationship between income inequality and growth, but the determinants of income distribution are not well discerned. This paper will test the relationship between income inequality and ethnic heterogeneity. Although previous research has used ethno-linguistic fractionalization as a control in inequality studies, no research has focused primarily on the measure of polarization or used the data set this paper employs.

This paper uses the most recent measures for both income inequality found in the World Income Inequality Database and polarization (Montalvo and Reynal-Querol 2005b). A pooled OLS regression is run using data from 58 countries with 205 total observations. The dependent variable used to represent income inequality is the Gini coefficient. The main explanatory variable is the polarization index, and a number of control variables suggested in previous research are included as well. The results of these regressions suggest that countries with higher levels of ethnic polarization also have higher levels of income inequality. Regressions using the more common fractionalization index are also run. The effects of ethnic fractionalization are found to be diminishing, offering further support that polarization may be more important for understanding income inequality.

II. Literature Review

The most well-known study of income inequality was done by Simon Kuznets in the 1950s. Kuznets (1955) finds that a country's level of income inequality is affected by its state of economic development. As an undeveloped country experiences economic growth, income inequality will initially grow as people leave the traditional sector for the industrial sector. But when enough of the population has made this initial transition, additional development will lead to less inequality. Since Kuznets's initial proposition, his hypothesis has been studied by numerous researchers using data from many different countries and many different time periods. And like many topics that are studied extensively with econometric analysis, the results are decisively mixed.

Higgins and Williamson (1999) find a Kuznets curve in their cross-country study of inequality once they control for other variables. Barro (2000) also finds results supporting the inverted-U theorized by Kuznets. However, in a study of industrialized nations since the 1950s, Ram (1997) actually found the opposite of the Kuznets's inverted-U; that is, income inequality fell initially before rising again in the 1970s. For further information on the history of Kuznets's hypothesis, see Moran (2005)

Although the level of economic development may be the most studied determinant of income inequality, it is by no means the only one. In addition to economic development, Kaasa (2005) identifies four other factors that researchers commonly identify as possible determinants of income inequality: demographic factors, macroeconomic factors, cultural and environmental factors, and political factors. Of these, political and cultural factors directly relate to the measurements of heterogeneity this paper investigates. However, demographic and macroeconomic factors will be discussed briefly in order to present a full view of the research on income inequality.

Kaasa (2005) classifies the demographic factors into a number of components such as urbanization, share of children in the population, share of the elderly in the population, composition of the household, education level, education inequality, and education expenditures. He reports that the only factor that has a consistent effect is education inequality, and it is positively related to economic inequality. All of the other factors have been found to have either positive, negative, or insignificant results depending on the study.

Macroeconomic factors include inflation, unemployment, financial development, trade levels, and foreign investments. Similar to the demographic factors, Kaasa (2005) finds mixed results within the literature for most of these determinants. Financial development is found to have a negative effect on income inequality and foreign investments have a positive effect. All of the other macroeconomic variables have been found to be either positive or negative.

While these elements should be considered as control variables, they do not obviously relate to the heterogeneity characteristic this paper attempts to test. Heterogeneity by itself cannot cause differences in a country's income distribution. It is simply a characteristic of the population. However, it can have an effect on other parts of society such as educational equality, government expenditures, and institutional quality. These other factors are included under Kaasa's (2005) classifications of cultural and political factors. Many of these variables have only been included in a few studies, but they have had significantly positive effects on income inequality. These include land concentration, shadow economies, corruption, and natural resource abundance. It is likely that any effect that heterogeneity has on income inequality is caused by one of these factors.

The term heterogeneity has been used frequently in this report, but it is important to define what is actually meant by this term. Within the literature and in this paper as well, heterogeneity will be measured in two distinct ways: fractionalization and polarization. Although fractionalization and polarization are related to one another, they are different concepts that look at different aspects of a population. Fractionalization measures the total diversity within a population. It increases with the number of groups until it reaches a maximum value of one if every individual in a society is from a different group. However, as the number of groups in a society grows, the respective power of each group diminishes. There would be the greatest amount of social conflict when there are groups that are the most powerful, i.e. two equally large opposing groups. Polarization attempts to account for this power and social conflict relationship. It reaches a maximum value of one when there are two groups of equal size before declining with the addition of more groups. Its proponents suggest that accounting for group power in this manner is a better way to capture potential social conflict.

Specifically, the fractionalization index is constructed using the following formula (Alesina et al. (2003):

(1) $$FRAC_i = 1 - \sum_{j=1}^{n} s_{ji}^2$$

where s_{ji} is the population share of group *j* in country *i*. This measurement gives the probability of two randomly selected people selected from country *i* belong to different groups. Thus, a higher value for FRAC indicates a more heterogeneous population. Fractionalization equal to one would be perfect heterogeneity, in which all members of the country were from different groups and zero would be perfect homogeneity, in which all members of society were from the same group. This index can be calculated for various characteristics, but this paper will only examine ethnicity heterogeneity.

Alesina and La Ferrara (2005) briefly outline heterogeneity's possible effects on an economy. They argue that heterogeneity can affect preferences, strategies, and actual production. Furthermore, heterogeneity can lead to a lower provision of public goods because competing groups do not want to assist one another. Empirical evidence of these effects has been shown internationally and within the US. Easterly and Levine (1997) report that the high levels of ethnic diversity in Africa help to explain its low growth rate. They demonstrate that fractionalization affects growth through a number of mechanisms including low schooling and political instability. Alesina et al. (2003) also show that ethnic and linguistic heterogeneity can negatively affect a country's growth rate and the quality of their political institutions. Alesina and Glaeser (2004) also show that the amount of social welfare spending in developed countries is negatively affected by ethnic fractionalization A similar relationship between social spending and fractionalization in the US is found when examining social expenditures across various municipalities (Alesina et al. 2004).

As previously mentioned, there has been some debate in the literature over the use of fractionalization as a representative measure for heterogeneity. Montalvo and Reynal-Querol (2005a) argue that polarization can often be a better variable than fractionalization in determining the social effects of diversity. Rather than measuring the total diversity in a society, polarization measures the difference from a bimodal distribution. It is given by the formula:

(2) $$POL_i = 1 - \sum_{j=1}^{n} \left(\frac{0.5 - s_{ji}}{0.5}\right)^2 s_{ji}$$

where s_{ji} is the population share of group *j* in country *i*. Polarization reaches a peak when there are two groups of the same size in a society, but then its values slowly decrease as more groups enter the society. It attempts to measure the strength of potential conflict, which is greater when large groups are competing against one another. Montalvo and Reynal-Querol (2005a) use their constructed polarization indexes to find that social polarization has a negative effect on growth because it increases public consumption, lowers investment, and increases the likelihood for civil war.

Within the United States, polarization has also been directly linked with income inequality. Dincer and Lambert (2008) provide strong evidence that ethnic and religious heterogeneity do affect income inequality even when controlling for education, corruption, unemployment, and income levels. They test a model that finds polarization has a positive and significant relationship with income inequality. They also find that fractionalization is significant when it is modeled using a quadratic. In further tests, they find that the estimated effects of their heterogeneity change with the inclusion or exclusion of a welfare spending control variable. Therefore, they show that the amount of government transfer payments is one mechanism through which heterogeneity affects income inequality.

No similar study focusing on the relationship between heterogeneity and income inequality has been done with cross country data, although a number of studies have included ethno-linguistic heterogeneity as a control variable. In Barro's (1999) attempt to measure the determinants of income inequality, he includes variables for both ethno-linguistic and religious heterogeneity but finds that neither is significant in the model. Clarke et al. (2003) also include ethno-linguistic variables in their model to measure the effect between financial intermediary development and income inequality. In their regression results, they do find a positive relationship between fractionalization and income inequality. However, they only find fractionalization to be significant in their specifications using the generalized method of moments technique.

III. Model

Despite the numerous studies done on income inequality, there is still not one accepted model to use. Previous research indicates over two dozen variables that could be included in a model (Kaasa, 2005). This paper closely follows a model in a study by Gupta et al. (2002) that studies the effects of corruption on income inequality and poverty. This model is replicated because it is from a relatively recent publication, it features a variety of different control variables, and most of the data used is easily accessible. Thus the model for this paper will be:

(3) $$Inequality_{it} = B_0 + B_1\ Heterogeneity_i + B_2\ (Income\ Level)_{it} + B_3\ (Educational\ Inequality)_{it} + B_4\ (Land\ Inequality)_i + B_5\ (Corruption)_{it} + u$$

Separate regressions using polarization and fractionalization as measures of heterogeneity will be used. A comparison of the results will show which conceptualization is better suited for studying income inequality. If polarization is significant, then it is likely that fractionalization by itself will not be significant, but fractionalization and its quadratic will be significant. Diminishing effects of fractionalization would suggest there is a certain point of fragmentation in society in which income inequality is greatest, supporting the argument that the size of the relative groups competing against one another matters more than the overall number of groups. Based on the results of Dincer and Lambert (2008), it is likely that polarization is significant and has a positive effect on income inequality. Thus, fractionalization is expected to have a positive but diminishing effect.

The other variables are control variables that have been found to be significant in previous studies. Traditionally, income level in the model has been modeled using GDP per capita as well as its quadratic. This method tests for the Kuznets curve that was explained earlier in the paper. However, the existence of a Kuznets curve is a highly contested issue within the study of income inequality. The regressions for this paper were run using both a linear and quadratic representation for GDP per capita. Because the quadratic was not significant in any regression tested, a linear model is used for this study.

Land inequality is another control for income inequality. Because higher land inequality limits the opportunities for the people in a country in terms of employment and collateral, it is predicted to have a positive effect. Education inequality should also have a positive effect on income inequality. Education is theorized to be a determinant of wages, so inequality in this should lead directly to inequality in income. Finally, corruption is also included as a control variable. If higher heterogeneity is related to higher levels of corruption, then this could be a possible mechanism through which polarization affects income inequality. Corruption is expected to lead to more income inequality levels.

IV. Data

Because of the cross-country nature of this study, the data is taken from a variety of sources. The main dependent variable in this study is the Gini coefficient. Although there are other measures of inequality, such as Thiel coefficients, the Gini is the most common measure used in cross-country studies of income inequality. The Gini coefficient is a number ranging from 0 to 1, with 1 being perfect inequality and 0 being perfect income equality.[1] The actual Gini coefficients for this study are obtained from the WIDER World Income Inequality Database (WIID). The database contains over 5000 Gini coefficients for 159 countries spanning from 1960 to 2006. It is a compilation of Gini calculations from a variety of scholarly sources usually using in-country surveys for their calculations.

For Gini coefficients to be calculated, the income shares of parts of the population must be known. However, surveys that inquire about these income shares often use different methods and definitions. For example, the survey may or may not have national coverage; it can use the household or the individual as the unit of analysis; it can be based on income or expenditure; it

[1] The regressions in this paper actually scale the Gini between 0 and 100, but none of the inferences are changed because of this transformation.

can be based on monetary income or income that includes in-kind receipts and services. Before the WIID was developed, most researchers used the Gini estimates from Deininger and Squire (1998). Deininger and Squire outline a number or recommendations for choosing which observations to include. Their guidelines of excluding non-representative surveys are followed in the construction of this dataset. Furthermore, their concerns for variations in gross income compared to net income, monetary income compared to non-monetary income, and income compared to consumption have been accounted for using dummy variables.

For this study, the WIID was first filtered to only those Gini coefficients that were nationally representative in terms of regions, household characteristics, and age. For consistency, it was then filtered to include only the measurements that used households as the income share and a person as the unit of analysis. This filtering resulted in instances where a country had a number of Gini coefficients for one specific year. If they varied due to factors controlled for by dummy variables, then each of the observations remained in the dataset. However, there were instances in which countries had two distinct Gini coefficients for the same income definition for the same year. When this occurred, the highest quality Gini was taken; if they were the same quality, then the Deininger and Squire measure was taken; if there was no Deininger and Squire observation, then the average of the two numbers was taken.[2]

The main independent variable used was ethnic polarization. The method for calculating this polarization is discussed in a previous part of the paper. The specific measures for this index come from the dataset used by Montalvo and Reynal-Querol (2005b), which is derived mainly from the *World Christian Encyclopedia*. Although the exact dates for these values are not given, Montalvo and Reynal-Querol treat them as stable across time for each country, and this study will follow that methodology.

[2] The three observations for which averages were taken are Finland 2000, Germany 2000, and Honduras 1999.

Each of the other variables is a control variable that has been used in previous studies of income inequality. The GDP measure is from the World Bank's World Development Indicators (2006) and corresponds to the year of the Gini estimate. GDP is in terms of Real GDP per capita in terms of purchasing power parity.

The education inequality measure uses values from the Barro and Lee (2000) education dataset commonly used in cross-country studies. The Barro and Lee data set has data for every five years for most countries spanning from 1960 to 2000. Education inequality is calculated using a method described by Gupta (2003) which takes the ratio of those not receiving any schooling to those who have graduated from secondary or post-secondary school by age 15. In order to obtain a more robust account of this measure, an average of education inequality is taken rather than just using data from one time period. The average of two periods is taken, with each being at least the same year as the Gini measure or before it.[3] The natural log of this variable is taken for easier interpretation and to correct for a few outliers.

A third control variable is a land Gini obtained from Frankema (2006). His dataset uses in-country surveys of land ownership to calculate land inequality measures. For some countries, multiple land Ginis are given. For each observation, this study uses the most recent land Gini that is still before the year of that observation's income Gini.

Corruption is the final control variable in the model. The corruption measure is taken from the World Bank's World Governance Indicators, which have been compiled very two years since 1996. Of the six indicators provided in the data, this study uses the "control of corruption" measure which measures "the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as 'capture' of the state by elites and

[3] For example, the measurement of the education inequality in Colombia for the 1998 Gini estimate would be obtained by using the Barro and Lee data from 1995 and 1990.

private interests" (World Bank 2007). These measures are created by combining a number of in-country studies and are then normalized. A positive score represents good governance and a negative score represents poor governance. This study uses the corruption score from the year that the Gini is from or from the year prior to the Gini.

Table 1 provides the summary statistics for the observations included in this study. Although each dataset usually contained over 100 countries, having to match the datasets resulted in a reduction to only 58 countries. Some countries have just one observation but many have more than one. The earliest Gini used is from 1996 and the latest is from 2006. In all, there are 205 observations. Table 2 provides a breakdown of the summary statistics by region. Table 3 provides a correlation matrix for all of the main variables.

V. Analysis and Results

For this analysis, unpooled OLS is conducted using SAS. Although this study does analyze countries over time, the nature of the data makes this method preferred to using panel techniques. Fixed effects cannot be used because the polarization measure does not vary across time for the countries, and doing a transformation would subtract out the variable of interest. Also, it is not logical to do a random effects study with this sample of countries. Attempting random effects is also complicated because some countries have more than one data point for each year (when there are multiple definitions of income inequality available).

When using OLS, one of the concerns for researches is heteroscedasticity. While this will not produce biased results, it does invalidate the standard errors and subsequent significance calculations. The regressions for this study were tested for heteroscedasticity using the White test for heteroscedasticity. These tests showed that heteroscedasticity is present in the data.

Therefore, all of the regressions are calculated with standard errors that are robust to heteroscedasticity.

Table 4 shows the results of the regressions of income inequality on various combinations of the variable of interest (polarization) and the control variables. The first regression shows the effect of polarization on income inequality using income, land inequality, and educational inequality as control variables. Corruption is not included because it could theoretically be a mechanism through which polarization affects inequality. As Table 4 shows, this initial regression is significant and explains approximately 70% of the variation in the Gini coefficients.

Regional factors are often quite important in cross country studies.[4] In many studies, including regional dummies may decrease the significance of certain variables. The results of the third regression demonstrate this because both land inequality and GDP lose their significance when regional dummies are included. The effect of polarization remains significant when regional dummies are introduced, but its effect falls by half. If the average country were to change from having no polarization to having complete polarization, the Gini coefficient would increase by approximately 5 units rather than 10 units. Therefore, some of the effects of polarization in the first two regressions are actually caused by regional differences rather than polarization itself.

The second regression introduces corruption into the model without regional dummies to see if it has an effect on polarization. Corruption is significant at the 0.05 level, but there is little change in polarization when this new variable is added. Similar results are seen in the fourth regression which includes corruption and the regional dummies. There is small fall in the

[4] Here the base for regional dummies is essentially developed countries and one Eastern European country. Eastern Europe is not included a dummy variable because only Poland had data available for each of the variables in the dataset.

estimated polarization coefficient (from 5.05 to 4.78), but it remains significant at the 0.05 level. Corruption is the only control variable other than the regional dummies with significance at the 0.05 level. These results suggest that corruption is not a mechanism through which polarization acts, but rather it has its own unique effect on income inequality.

The strength and significance of the regional dummies in each of the models indicate that there are regional characteristics that are affecting the Gini measure. Because of the unknown nature of the model, it is possible that polarization may also have different effects across different regions. Therefore, a regression including interaction terms is run to test for this possibility. The results of this regression are difficult to fully interpret. Polarization remains significant and increases slightly for the average country. The only region to have a statistically significant interaction term is the Middle East. It actually shows that polarization has a smaller effect on income inequality for this region than in the average country. Also,adding the Middle East's coefficients for polarization and the interaction terms results in a negative number. This suggests that polarization is not a major factor in the income inequality for this region. The regional dummies for Sub-Saharan Africa and for South Asia fall by at least four units indicating that polarization may account for some of the differences in Ginis for these two regions. The small changes for Latin America and East Asia suggest that the differences that cause higher Gini coefficients in these regions are not due to differences in the effect of polarization.

Overall, the results show that polarization does seem to affect income inequality. Controlling for income levels, other levels of inequality, and regional differences did diminish the effect of polarization, but it did not do so entirely. Corruption did not greatly mitigate the effect of polarization. Polarization can probably account for approximately a five unit change in the Gini coefficient. In this dataset, the range of Gini coefficients is from 24.4 to 66.6, but the standard deviation is 11. Thus, going from near complete polarization like Jordan (0.982) to almost no polarization like Norway (0.090), would increase the expected Gini by 5 units. A full swing from non-corrupt to corrupt could have an impact of 10 units on the Gini. Just by being an African or Latin American nation, the expected Gini rises by at least 8 units. Thus, the change due to polarization is significant, but it does not explain a large amount of the variation in income inequality.

In addition to using the polarization figures, an interesting test can be conducted by using fractionalization measures as well. As mentioned earlier, fractionalization is often used as the first indicator of ethnic heterogeneity, but it is very different from polarization. If polarization is linearly related to income inequality, then there should be a quadratic relationship between inequality and fractionalization. Table 5 shows the results of substituting fractionalization for polarization. When fractionalization is modeled as having a linear relationship with income inequality, it is insignificant. However, when a quadratic for fractionalization is also included, both terms are significant at the 0.01 level. These results bolster the claim that there is a relationship between income inequality and polarization.

VI. Conclusions

Income inequality has interested economists for over half a century. Many studies have examined this problem and have found wide range of results. This study should be included in that jumble of findings. Whereas heterogeneity had commonly been assumed to affect inequality, actual empirical results have been mixed. The results from this study show that the use of fractionalization rather than polarization may have been one reason why.

The results from this study show that income inequality is affected by polarization. A complete change from no polarization to complete polarization would on average increase a country's Gini by about 5 units. This is about half a standard deviation for the Gini coefficients employed in this dataset. This effect is robust to the inclusion of control variables and regional dummy variables.

Learning that polarization may affect income inequality is just the first piece of a large puzzle. The most obvious question from this finding is through what mechanisms does polarization affect income inequality. Because it is a demographic characteristic, it is obvious that polarization alone cannot increase income inequality. Corruption was proposed as a possible mechanism, but the results here do not support that hypothesis. More research is needed to find out how polarization is related to other characteristics that may have a direct effect on income inequality.

This study does have limitations that must be acknowledged. The first is the small country sample size, and specifically, the exclusion of most Eastern European countries. As stated earlier, the individual datasets used each had a large number of observations, but many countries fell out because the data did not completely overlap. Using different sources or different control variables that had more overlap may prevent this problem. Employing alternative measures for income inequality or finding other data sources for polarization measures would also further test the robustness of the relationship. Repeating the study with different control variables has its own merit as well, simply because of the lack of consistency in previous research. When most variables have been shown to be significantly positive, significantly negative, or insignificant, then an appropriate model is hard to predict. Perhaps a Bayesian approach similar to studies done for economic growth might be appropriate at determining what the true causes of income inequality.

While there is always room for further research on any problem, the results suggesting that polarization does have a positive effect on income inequality are valid. Furthermore, this relationship appears to be robust to many changes in the model. With more research, the reasons for this can be better discerned and then possible policy recommendations could be crafted. At the least, this study shows that researchers should consider using polarization as acontrol variable for future income inequality studies.

Reference List

Alesina, A. and E. La Ferrara, 2005, “Ethnic Diversity and Economic Performance,” *Journal of Economic Literature*, 63, 762-800.

Alesina, A., R. Baqir, and W. Easterly, 2004, “Public Goods and Ethnic Divisions,” *The Quarterly Journal of Economics*, 114, 1243-1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg, 2003, “Fractionalization,” *Journal of Economic Growth*, 2003, 155-194.

Alesina, A. and E. Glaeser, 2004, *Fighting Poverty in the US and Europe*, Oxford: Oxford University Press.

Barro, R., 2000, “Inequality and Growth in a Panel of Countries,” *Journal of Economic Growth*, 5, 5-32.

Clarke, G, L. Xu, and H. Zou, 2003, Finance and Income Inequality: Test of Alternate Theories, *World Bank Policy Research Working Paper*.

Deininger, K. and L. Squire, 1996, “A New Data Set Measuring Income Inequality,” *World Bank Economic Review*, 10, 565-595.

Dincer, O. and P. Lambert, 2008, “Taking Care of Your Own: Ethnic and Religious Heterogeneity and Income Inequality,” Working paper, available at http://ideas.repec.org/p/inq/inqwps/ecineq2006-48.html

Easterly, W. and R. Levine, 1997, “Africa’s Growth Tragedy,” *The Quarterly Journal of Economics*, 112, 1203-1250.

Gupta, S., H. Davoodi, and R. Alonso-Terme, 2002, “Does Corruption Affect Income Inequality and Poverty?”, *Economics of Government*, 3, 23-45.

Higgins, M. and J. Williamson, 1999, “Explaining Inequality the World Round: Cohort Size, Kuznets Curves, and Openness,” *FRB of New York Staff Report No. 79*.

Kaasa, Anneli, 2005, "Factors of Income Inequality and Their Influence Mechanisms: A Theoretical Overview," University of Tartu Faculty of Economics and Business Administration Working Paper No. 40, available at SSRN: http://ssrn.com/abstract=868491

Kuznets, S., 1955, “Economic Growth and Income Inequality,” *American Economic Review*, 45, 1-28.

Montalvo, J. and Reynal-Querol, M, 2005a, “Ethnic Diversity and Economic Development,” *Journal of Economic Development*, 76, 293-323.

Moran, T., 2005, “Kuznet’s Inverted U-Curve Hypothesis: The Rise, Demise, and Continued Relevance of a Socioeconomic Law,” *Sociological Forum*, 20, 209-244.

Ram, R., 1997, “Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World,” *Southern Economic Journal*, 64, 576-583.

World Bank, 2007, *A Decade of Measuring the Quality of Governance*, available at <http://info.worldbank.org/governance/wgi2007/pdf/booklet_decade_of_measuring_governance.pdf>

Data Sources

Barro, R. J. and J. Lee, "International Data on Educational Attainment: Updates and Implications," CID Working Paper No. 42, April 2000.

Frankema, E, 2006, The Colonial Origins of Inequality: Exploring the Causes and Consequences of Land Distribution, Working paper (University of Gröningen), available at < http://www.ggdc.net/pub/gd81.pdf>

Montalvo, J. and M. Reynal-Querol, 2005b, Ethnic Polarization, Potential Conflict and Civil War, 95, *American Economic Review*.

WIDER, 2008, *World Income Inequality Database V2.0*, available at <http://www.wider.unu.edu/research/Database/en_GB/database/>.

World Bank, 2007, *World Development Indicators 2006*, CD-Rom.

World Bank, 2008, *World Governance Indicators 1996-2006*, available at <http://info.worldbank.org/governance/wgi2007/home.htm>.

Table 1. Summary of Variables

Variable	Mean	Standard Deviation	Minimum	Maximum
Gini	44.73639	10.94003538	24.4	66.6
Dummy for Consumption	0.282927	0.451523803	0	1
Dummy for Monetary	0.24878	0.433364602	0	1
Dummy for Gross Income	0.170732	0.37719547	0	1
SSA Dummy	0.092683	0.290697474	0	1
East Asia Dummy	0.097561	0.29744649	0	1
Latin America Dummy	0.419512	0.494687222	0	1
Middle East Dummy	0.053659	0.225894259	0	1
Ethnic Polarization	0.5312	0.233922402	0.0901551	0.9824263
Ethnic Fractionalization	0.421165	0.261107626	0.045386	0.958587
Land Gini	65.60488	14.2087682	36.8	90.9
GDP	10129.41	9279.043203	540.774963	34207.82
Education Inequality	-0.28371	1.990109736	-5.365976	4.9374671
Corruption	0.189197	1.053535028	-1.431661	2.4619752

Table 2. Summary of Variables by Region

Variable	Mean	Standard Deviation	Minimum	Maximum
SSA (10)				
Gini	51.63511	10.09784859	32.7	66.6
Ethnic Polarization	0.526452	0.153658838	0.2710496	0.7177831
Ethnic Fractionalization	0.755613	0.190457698	0.185026	0.958587
Land Gini	57.30526	14.3781653	36.8	79
GDP	2128.58	2573.041042	540.774963	9419.12207
Education Inequality	2.677625	0.895978951	1.31907806	4.937467051
Corruption	-0.68221	0.552590669	-1.1514638	0.617476102
East Asia (5)				
Gini	44.65208	7.425202587	30.8	58.5
Ethnic Polarization	0.587442	0.071217294	0.4965057	0.761617
Ethnic Fractionalization	0.595782	0.196800703	0.36078	0.842858
Land Gini	47.66	5.91647121	43.8	68
GDP	4885.702	1614.483784	2955.69434	8660.820313
Education Inequality	0.290117	1.083204337	-2.2829666	1.641472139
Corruption	-0.38251	0.346461821	-1.1726894	0.491939764
South Asia (5)				
Gini	40.88889	11.27091882	27.6	61
Ethnic Polarization	0.471468	0.281039664	0.1317742	0.7492799
Ethnic Fractionalization	0.442556	0.310799586	0.0684201	0.901163
Land Gini	51.07778	8.66584932	41.8	62.3
GDP	2009.166	985.794273	1212.63708	3625.534424
Education Inequality	1.716908	1.096715222	-0.0342071	2.665214496
Corruption	-0.46685	0.31615389	-1.0391018	-0.17996227
Middle East (4)				
Gini	38.41429	3.188222374	34.5	44
Ethnic Polarization	0.573082	0.306710817	0.1673397	0.9824263
Ethnic Fractionalization	0.373453	0.229785827	0.087164	0.75625
Land Gini	67.38571	5.337736278	61.6	73
GDP	4249.877	1224.843064	3100.49561	6251.55127
Education Inequality	1.065555	0.294314927	0.67947658	1.345018321
Corruption	-0.1284	0.225414084	-0.4643364	0.118143524
Latin America (18)				
Gini	53.23579	5.323172878	38.47869	63.7
Ethnic Polarization	0.645642	0.16811409	0.2070293	0.9546801
Ethnic Fractionalization	0.444724	0.20789137	0.047643	0.7084
Land Gini	76.50814	7.287882209	46.2	90.9
GDP	5442.079	2145.809046	1798.40564	9115.474609
Education Inequality	0.231196	1.034244235	-1.1354862	2.448298085
Corruption	-0.3679	0.536274144	-1.431661	1.388079309
Other (16)				
Gini	32.52614	3.934037999	24.4	40.8
Ethnic Polarization	0.365072	0.253300074	0.0901551	0.8707317
Ethnic Fractionalization	0.237862	0.210927382	0.045386	0.766822
Land Gini	60.87344	12.71517332	39.2	79.1
GDP	22226.91	7048.801234	6510.35352	34207.82031
Education Inequality	-2.46298	1.388266007	-5.365976	0.988331164
Corruption	1.502143	0.696253533	-0.1870141	2.461975194

Table 3. Correlation Matrix

	Gini	*Ethnic Polarization*	*Ethnic Fractionalization*	*Land Gini*	*GDP*	*Education Inequality*	*Corruption*
Gini	1						
Ethnic Polarization	0.460872579	1					
Ethnic Fractionalization	0.350185968	0.661858552	1				
Land Gini	0.379756662	0.422880185	0.098642227	1			
GDP	-0.638823173	-0.267010584	-0.368285952	-0.090638089	1		
Education Inequality	0.545695868	0.279711836	0.471408816	0.184325379	-0.750126001	1	
Corruption	-0.660149438	-0.307214308	-0.429316095	-0.157116963	0.914036891	-0.777141926	1

Table 4. Regression Results for Polarization

Variable	Inequality	Inequality	Inequality	Inequality	Inequality
Intercept	41.03***	39.32***	36.44***	34.92***	35.53***
	(2.5061)	(2.6227)	(3.3921)	(3.5451)	(3.4873)
Consumption Dummy	-8.59***	-8.25***	-6.47***	-6.41***	-6.52***
	(1.1474)	(1.1321)	(1.0904)	(1.0621)	(1.1007)
Monetary Dummy	0.02	-0.23	-1.18	-1.33*	-1.4*
	(0.9467)	(0.8892)	(0.7852)	(0.7428)	(0.7365)
Gross Income Dummy	4.29***	4.79***	4.38***	4.71***	4.63***
	(1.066)	(1.1064)	(1.0519)	(1.0428)	(1.0278)
Ethnic Polarization	10.4***	9.81***	5.05**	4.78**	5.72**
	(2.208)	(2.1049)	(1.9134)	(1.877)	(2.6295)
GDP	-0.00064***	-0.00038***	-0.00012	0.000063	0.000054
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education Inequality	0.81**	0.49	0.54	0.23	0.3
	(0.3168)	(0.3474)	(0.4092)	(0.4304)	(0.4287)
Land Gini	0.1***	0.1***	-0.01	-0.01	-0.02
	(0.0356)	(0.0351)	(0.0498)	(0.0488)	(0.048)
Corruption	--	-2.86**	--	-2.15**	-2.19**
	--	(1.3023)	--	(1.0247)	(1.0692)
SSA	--	--	15.21***	15.57***	11.17
	--	--	(3.1257)	(3.157)	(7.4701)
East Asia	--	--	8.07***	7.88***	8.52
	--	--	(2.4097)	(2.4135)	(10.4535)
South Asia	--	--	4.76	5.29	1.67
	--	--	(3.5391)	(3.6083)	(3.8983)
Middle East	--	--	6.17**	6.9***	11.73***
	--	--	(2.4742)	(2.4455)	(2.4894)
Latin America	--	--	15.2***	14.96***	16.23***
	--	--	(1.8956)	(1.9124)	(2.1455)
*Polarization * SSA*	--	--	--	--	6.86
	--	--	--	--	(12.5237)
*Polarization * East Asia*	--	--	--	--	-2.31
	--	--	--	--	(16.72)
*Polarization * South Asia*	--	--	--	--	6.18
	--	--	--	--	(9.248)
*Polarization * Middle East*	--	--	--	--	-9.46***
	--	--	--	--	(3.0493)
*Polarization * Latin America*	--	--	--	--	-2.86
	--	--	--	--	(3.3255)
Observations	205	205	205	205	205
R-squared	0.6962	0.7054	0.7848	0.7895	0.7887
F stat	67.79	62.06	62.99	59.87	43.31

Standard errors are reported below the coefficient estimates in parenthesis

*** 0.01 significance, ** 0.05 significance, * 0.10 significance

Table 5. Regression Results for Fractionalization

Variable	Inequality	Inequality
Intercept	33.58*** (3.1062)	32.86*** (3.3786)
Consumption Dummy	-5.53*** (1.0742)	-6.17*** (1.0665)
Monetary Dummy	-1.33 (0.941)	-1.17 (0.7851)
Gross Income Dummy	5.15*** (1.1329)	4.94*** (1.0927)
Ethnic Fractionalization	-0.97 (1.8317)	13.31* (7.2579)
Ethnic Fractionalization Squared	-- --	-17.25* (8.6114)
GDP	0.000091 (0.0001)	0.00012 (0.0001)
Education Inequality	0.085 (0.3882)	0.05 (0.434)
Land Gini	0.02 (1.8327)	0.02 (0.0457)
Corruption	-2.21** 0.9557	-2.81** (1.0658)
SSA	17.75*** (2.7197)	18.76*** (3.3171)
East Asia	10.28*** (2.2423)	9.71*** (2.4204)
South Asia	6.98** (2.7388)	6.93* (3.7203)
Middle East	6.54*** (2.04)	7.84*** (2.4499)
Latin America	16.79*** (1.8871)	15.75*** (1.8956)
Observations	205	205
R-squared	0.7830	0.7885
F stat	57.63	55.32

Standard errors are reported below the coefficient estimates in parenthesis
*** 0.01 significance, ** 0.05 significance, * 0.10 significance