An Analysis of Inequality Measures and Common Datasets

Chandini Sankaran – University of Wisconsin – Stevens Point

Abstract
Results from studies of within-country income inequality are inconclusive and sometimes contradictory. This paper emphasizes that these varied findings may be due to differences in the datasets used. There are four major indices to measure inequality within societies, income shares, income ratios, the gini index, and the theil index. Common datasets include the World Development Indicators Database (World Bank 2003b), Deininger and Squire's (1996) database, and the theil index compiled through the University of Texas Inequality Project (UTIP 2005). This paper discusses the strengths and weaknesses of these measures, and investigates the comparability of these datasets. Findings indicate that UTIP's (2005) theil dataset is not strongly correlated with the inequality measures from the World Bank data base (World Bank 2003b) and the Deininger and Squire's inequality measures. Since their theil dataset is based on manufacturing wage and employment data, it does not accurately reflect society-wide income inequality. There is indication that wages within the manufacturing sector are less spread out than society-wide wages. Findings also indicate inconsistency in Deininger and Squire's (1996) income inequality database. Finding a uniform and comprehensive dataset of income inequality still remains a challenge. Until then, researchers and policymakers should proceed with caution when interpreting the results of studies on inequality, keeping in mind the drawbacks of the existing measures.

Introduction

Simon Kuznets (1955) in a study of income inequality within and between the agricultural and non-agricultural sectors of three countries over the process of development, hypothesized that economic progress initially causes rising inequality within a country, but as the country develops further this inequality falls. Visualizing a graph with inequality on the vertical axis and GDP per capita on the horizontal axis, Kuznets projected an inverted-U curve. This curve illustrates a positive relationship between income inequality and per capita output for a country in the early development phase, and a negative relationship for a country in the later stages of industrialization. Empirical studies that followed Kuznets' seminal analysis in the next three decades claimed to have evidence that supported the inverted-U hypothesis. Kravis (1960), Paukert (1973), Ahluwalia (1976), and Tsakloglou (1988), analyzing cross-country data of income shares were able to support the Kuznets' hypothesis that "with economic development income inequality tends to increase, then becomes stable, and then decreases" (Paukert (1973), p. 121).

Over the past decade, there has been renewed interest in within-country income inequality.

> Up to the 1970s, the Kuznets' hypothesis seemed to account for the experience not only of the US but also of most of the OECD countries… However, the downward trend in inequality experienced by these economies during the twentieth century has reversed sharply in recent times. In

particular, the past fifteen years have witnessed a significant increase in wage inequality (Aghion, Caroli, and Garcia-Penalosa (1999), p. 1616).

Because the majority of OECD countries are at the later stages of development, Kuznets' hypothesis would lead us to anticipate falling income inequality in these countries. As observed by Gottschalk (1997), income inequality in OECD countries has increased over the past few decades, with the United States and the United Kingdom as leaders of this present trend. In these countries, "the gains to the rich have exceeded the gains to the poor" (Gottschalk (1997), p. 24). Foster and Pearson (2000) find the percentage of wealth held by the rich increased from the mid-1980s to the mid-1990s in all 19 OECD countries included in their study. For 18 out of the 19 OECD countries, income shares of the bottom and middle of the population fell, indicating widening inequality.[1]

This widening gap between the rich and the poor within the mature OECD economies is contrary to Kuznets' hypothesis. Kuznets' hypothesis has also been challenged for failing to explain the experiences of countries currently or recently embarking on the process of industrialization. According to his proposition, because these countries are just in the beginning of industrialization, they should see increases in income inequality. However, the high-performing countries of East Asia (South Korea, Taiwan, Malaysia, Thailand, and Indonesia) have had declines in inequality after 1965 (World Bank (1993)). By examining the gini coefficients of each of these East Asian countries individually,[2] Jomo (2000) finds a similar picture whereby "income inequality declined or did not worsen in South Korea, Taiwan, Malaysia and Indonesia over the 1976-85 period… there were significant reductions (almost 20 percent or more) in the gini coefficients of the five economies by the 1980s" (pp. 8-9).

The consensus thus far is that "the relationship postulated by Kuznets generally is present in the cross-section data…. and the relationship does not hold up when time-series data are examined (Perkins, Radelet, Snodgrass, Gillis and Roemer (2001) pp. 129-130). Deininger and Squire (1996, 1998) and Lundberg and Squire (2003) using a large-scale, cross-country and time-series data set, test relationships between inequality, the level of development, and growth. They fail to discover any evidence of a Kuznets' curve and conclude that development is not an important determinant of inequality. However, Lindert and Williamson (1985) using time-series data declare that "the British experience since 1688 looks like an excellent advertisement for the Kuznets' curve, with income inequality rising across the Industrial Revolution, followed by a prolonged leveling in the last quarter of the nineteenth century" (p. 344).

[1] Income shares are explained in Section 3.1.
[2] The gini index is explained in Section 3.2.

TABLE 1: AVERAGES OF DATA FROM 1968-1999

	All Countries	Low Income	Lower-Middle Income	Upper-Middle Income	High Income	High Income Non-OECD	High Income OECD
GDP per capita (constant 1995 US$)	5,749.78	432.17	1,448.04	4,065.13	17,053.76	14,736.37	19,371.15
Income Share Highest 10%	32.24	33.32	34.76	33.50	27.36	29.73	24.99
Income Share Lowest 10%	2.36	2.51	2.13	2.16	2.65	2.55	2.75
Income Ratio (highest 10% / lowest 10%)	13.66	13.28	16.32	15.51	10.33	11.66	9.09
Income Share Highest 20%	47.69	48.58	50.44	48.94	42.80	45.45	40.16
Income Share Lowest 20%	6.04	6.18	5.40	5.61	6.98	6.58	7.39
Income Ratio (highest 20% / lowest 20%)	7.90	7.86	9.34	8.72	6.13	6.91	5.43
Gini Index	40.72	41.44	44.05	42.60	34.78	37.46	32.10
Theil Index	0.065	0.089	0.058	0.056	0.055	0.082	0.021

Note: Low income economies are countries with $745 or less in annual per capita income.
Lower-middle income economies are countries with $746-$2,975 in annual per capita income.

Upper-middle income economies are countries with $2,976-$9,205 in annual per capita income.

High-income economies are countries with $9,206 or more in annual per capita income.

Source: World Bank Classifications (2003b).

Data on the income ratios and Gini index are from the World Development Indicators (World Bank 2003b). Theil index is from UTIP (2004).

This paper partly attributes these inconclusive and sometimes contradictory findings to differences in the datasets used. Currently, there are four major indices used to measure inequality within societies, income shares, income ratios, the gini index, and the theil index. Table 1 shows the averages of the GDP per capita and various measures of income inequality for the entire sample divided into groups. The GDP per capita (measured in constant 1995 US $) is intended to gauge the level of development of a country. This cross-sectional analysis indicates an explicit relationship between the level of development and the three measures of inequality from the World Development Indicators (World Bank 2003b). In low-income countries, 33 percent of all income is held by the richest 10 percent of the population, whereas the income share of the richest in OECD high income countries is 25 percent. All measures of income inequality show that developed economies, classified as high-income, have a more egalitarian distribution of income than any of the other economies. The income shares of the highest 10 percent of the population first increases and then decreases as we progress from low income, to lower-middle, upper-middle, and high income countries, whereas the income shares of the lowest 10 percent of the population follows an opposite pattern. Moreover, the income ratios and the gini index show an inverted-U relationship between the level of development and inequality (i.e. inequality first increases and then falls as we advance across the columns of the table progressing from low, lower-middle, higher-middle and high income economies). These patterns show evidence that there is a cross-sectional relationship between the level of development and income inequality when measured using the World Bank's income data, and indicate a Kuznets' curve. The theil index, on the other hand, shows an inverse relation between the level of development and inequality, and does not indicate an inverted-U pattern. It instead shows that inequality continuously falls as we progress through the process of industrialization.

The purpose of this study is to provide a comparison of these measures of inequality that are currently used in various papers. Results from this study may bring some light as to why researchers with different datasets arrive at different conclusions. Of particular interest to policymakers is whether inequality is inevitable, a direct result of the process of development, or whether there are other underlying sources that can be influenced by government policies. Before conclusions can be drawn based on these vast results, there needs to be a comprehensive investigation of the comparability of datasets currently used.

Data and Methodology

Section 3 briefly investigates the four most commonly used measures of within-country income inequality- income shares, income ratios, the gini index, and the theil index.[3]

[3] See Ray ((1998), pp. 187-192) for a comprehensive discussion of the other measures of inequality.

After a brief investigation of the methods used to calculate these indices, the strengths and weaknesses of each index is presented. Section 4 discusses three widely used datasets of inequality- the World Bank (2003b) database (income shares and gini), the Deininger and Squire (1996, 1998) dataset (income shares and gini), and UTIP's (2005) manufacturing theil dataset. Section 5 of this paper conducts comparability tests between the different measures of inequality within a society.

Income share data and data on the gini index are obtained from the World Bank (2003b) database and the Deininger and Squire (1996, 1998) dataset.[4] The theil dataset compiled by Galbraith and Kumh (2001) is available online from the University of Texas Inequality Project (2005).[5] The full sample includes 138 countries for the years 1968-1999, but missing data restricts the analysis to the available years and countries. Countries included in each group are listed in Appendix A, and the definitions of variables are listed in Appendix B. Appendix C provides summary statistics for the data used.

Measures of Inequality

An early measure of income inequality used by Kuznets (1955) and Ahluwalia (1976) is the income shares of various percentile groups in total income. Table 2 provides a simple illustration of the concept of income shares.

TABLE 2: CALCULATION OF INCOME SHARES FOR GREECE USING 1998 WORLD BANK (2003B) DATA

Quintiles *Col. 1*	**Share of Population** Col. 2	**Income (in millions of dollars) held by each quintile** *Col. 3*	**Income share of each quintile in total income** *Col. 4*	**Income share of each quintile in total income (%)** *Col. 5*
5	20% (lowest)	7,670	0.0706	7.06%
4	20%	12,439	0.1145	11.45%
3	20%	17,198	0.1583	15.83%
2	20%	23,922	0.2202	22.02%
1	20% (highest)	47,410	0.4364	43.64%
Total	*100%*	*108,639*	1.000	*100%*

After obtaining income information by surveying the population of Greece in 1998, arrange all the incomes from the lowest to the highest. Then, the units are divided into five equal size groups called quintiles. Column 1 shows the five quintiles. The segments ranked lowest by income, receive the smallest shares of total income. Column 5 reports the percentage of total income for each quintile. This is calculated by dividing the income held by each quintile with the total income of the country and multiplying by 100. Another

[4] The Deininger and Squire dataset is available online at http://www.worldbank.org/.
[5] The theil index is available at http://utip.gov.utexas.edu.

approach is to divide the population into ten deciles, with the top-most decile consisting of the richest 10 percent of the population and the lowest quintile, the poorest 10 percent of the population. Kuznets (1955) uses the share of the lowest (poorest 20 percent of the population) and top (wealthiest 20 percent of the population) quintiles, whereas Ahluwalia (1976) utilizes the cumulative income share of the top 20 percent (wealthiest), middle 40 percent, lowest 60 percent, and lowest (poorest) 40 percent of the population. Income shares are also sometimes reported as income ratios, which are usually the ratios of income shares of the top and lowest quintiles.[6] This measure is used in more recent papers such as by Tsakloglou (1988), Persson and Tabellini (1994), Perotti (1996), as well as Aghion, Caroli, and Garcia-Penalosa (1999).

Although this measurement is easy to understand, researchers using income share data should be cautioned that simply comparing a few points on the income distribution curve could be misleading. Focusing on just the richest or the poorest of the population may disregard changes in the dispersal of income inequality within the entire population. Tsakloglou (1988) states that "these indices have the disadvantage that--unlike summary measures of inequality--they refer to a single point of the Lorenz curve, disregarding the rest of the distribution" (p. 512).[7] On the other hand, the focus of this data set on specific population groups is beneficial because it enables researchers to examine "the behavior of the income shares of specific population groups during the process of economic development" (Tsakloglou (1988), p. 512).

Gini Index

The gini coefficient "measures the extent to which the distribution of income among individuals or households within a country deviates from a perfectly equal distribution" (World Bank (2003a)). It is one of the most commonly used measures of income inequality and has an advantage over income shares data since it measures income distribution within the entire population. It was developed by the Italian statistician Corrado Gini in 1955.

The first step towards calculating the gini coefficient is to construct the Lorenz curve.[8] To construct a Lorenz curve, the cumulative distributions for each quintile must be calculated and then plotted on a graph. Figure 1 shows a Lorenz curve. The closer the Lorenz curve is to the line of perfect equality (the 45^0 diagonal line from the origin which represents a uniform distribution), the lesser the degree of inequality in income, and the smaller the gini coefficient. The more bowed out the Lorenz Curve, the greater the degree of inequality and the bigger the gini coefficient. This ratio, when expressed as a percentage, is referred to as the gini index. A gini index of 0 percent corresponds with perfect equality (everyone has the same income) and 100 percent corresponds with perfect inequality (one person has all the income, and everyone else has zero income) (Perkins, Radelet, Snodgrass, Gillis and Roemer (2001)). The gini index can be derived from income, consumption, wage, or employment data.[9]

---

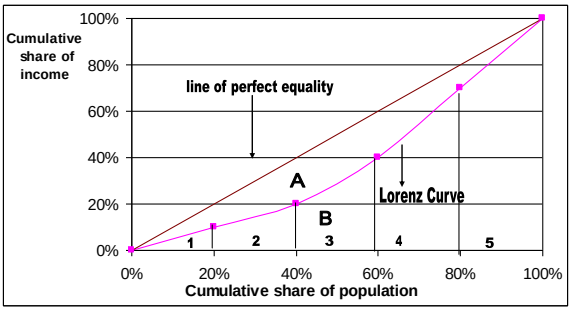
[6] Income ratios for Greece in Table 2 above is 6.18.

[7] The Lorenz curve is discussed in Section 3.3.

[8] The Lorenz curve is a cumulative frequency curve that gives the distribution of a specific variable, in this case, income. It was developed by Max O. Lorenz in 1905. It portrays observed income distributions and compares this to a state of perfect income equality.

[9] Nafzinger ((1997), pp. 132-134) explains the Lorenz curve and the gini index in more detail.

Figure 1: The Lorenz Curve



The gini coefficient is calculated as a ratio of areas on the Lorenz curve diagram. If the area between the line of perfect equality and the Lorenz curve is A, and the area beneath the Lorenz curve is B, then the gini coefficient is A/(A+B) (Perkins, Radelet, Snodgrass, Gillis and Roemer (2001). The gini coefficient for the 1998 Greek income distribution is given by the calculations in Table 3 below. Column 1 shows the share of population in each quintile in Greece or $X_{i(j)}$ (where $i=1\ldots n$ is the number of quintiles, and j indexes nation). In Column 2, $\sigma X_{i(j)}$ represents the cumulative share of population. Column 3 shows the income shares that goes to each quintile within Greece ($Y_{i(j)}$), and Column 4 the cumulative income shares ($\sigma Y_{i(j)}$). For example, $\sigma Y_{4(j)}$ in Column 4 is the cumulative income share of the fourth quintile and is calculated as the sum of 0.0706 and 0.1145. The triangle that represents Area A + Area B in Figure 1 can be calculated for Greece by ½ x 100(base) x 100(height) = 5000. Area B is found by adding up areas 1,2,3,4, and 5 on Figure 1. Area A is found by taking the difference between 5000 and Area B.

TABLE 3: CALCULATION OF THE GINI COEFFICIENT FOR GREECE USING A LORENZ CURVE WITH 1998 INCOME SHARE DATA FROM THE WORLD BANK (2003B)
(ASSUMING LINEAR SEGMENTS)

Quintile	$X_{i(j)}$ (i=1...5) **Share of Population in each quintile (%)** *Col. 1*	$\sigma X_{i(j)}$ **Cumulative Share of Population (%)** *Col. 2*	$Y_{i(j)}$ **Income share of each quintile in total income** *Col. 3*	$\sigma Y_{i(j)}$ **Cumulative Share of Income** *Col. 4*	$(\sigma Y_{i(j)})x100$ **Cumulative Income Shares (%)** *Col. 5*	**Calculating Areas Below the Lorenz Curve** *Col. 6*
5	20	20	0.0706	0.0706	7.06	20*7.06/2= 70.6
4	20	40	0.1145	0.1851	18.51	20*(7.06+18.51)/2= 255.7
3	20	60	0.1583	0.3434	34.34	20*(18.51+34.34)/2 = 528.5
2	20	80	0.2202	0.5636	56.36	20*(34.34+56.36)/2 = 907
1	20	100	0.4364	1	100	20*(56.36+100)/2= 1563.6

Total	100	1.000	Area B = 3325 Area (A + B) = 5000 Area A = 5000 – 3325 =1647
Gini			*A/(A+B)= 1647/5000 = 0.33*

Most studies calculate the gini coefficient using the less complicated Brown (1994) formula as:

$$G_{(j)} = \left| 1 - \sum_{i=1}^{n} (\sigma X_{i-1(j)} - \sigma X_{i(j)})(\sigma Y_{i-1(j)} + \sigma Y_{i(j)}) \right|$$

where $G_{(j)}$ is the gini coefficient within nation *j*, *j* indexes nation, *n* is the number of observations, which are the 5 quintiles, *X* is the share of population in each quintile calculated as the reciprocal of the number of observations or data points ($X=1/n$), *i* is the rank of values in ascending order, $Y_{i(j)}$ is the income share of each quintile in total income, $\sigma X_{i(j)}$ is the cumulative proportion of the population variable, and $\sigma Y_{i(j)}$ is the cumulative proportion of the income variable. Table 4 presents an example of the gini coefficient calculated with the Brown formula for Greece in 1998.

TABLE 4: CALCULATION OF THE GINI COEFFICIENT USING THE BROWN FORMULA FOR GREECE WITH 1998 INCOME SHARE DATA FROM THE WORLD BANK (2003B)

Quintile	$Y_{i(j)}$ **Income share of each quintile in total income** *Col. 1*	$X_{i(j)}=1/n$ **Share of Population** *Col. 2*	$\sigma X_{i(j)}$ **Cumulative Share of Population** *Col. 3*	$\sigma Y_{i(j)}$ **Cumulative Share of Income** *Col. 4*	$\sigma Y_{i-1(j)}+\sigma Y_{i(j)}$ *Col. 5*	$\sigma X_{i-1(j)}-\sigma X_{i(j)}$ *Col. 6*	Col 5*6 *Col. 7*
Lowest 20%	0.0706	0.2	0.2	0.0706	0.0706	0.2	0.014
Second 20%	0.1145	0.2	0.4	0.1851	0.2557	0.2	0.051
Third 20%	0.1583	0.2	0.6	0.3434	0.5285	0.2	0.106
Fourth 20%	0.2202	0.2	0.8	0.5636	0.9070	0.2	0.181
Highest 20%	0.4364	0.2	1	1	1.5636	0.2	0.312
Total	1.000	1.0					0.664
Gini						*1-0.665*	*= 0.33*

Theil's Statistic

The theil index, a measure developed by Henri Theil (1967) measures how the income distribution of a country differs from the population distribution of that country. "It compares the income and population distribution structures by summing, across groups, the weighted logarithm of the ratio between each group's income and population shares. When this ratio is one for some group, then this group's contribution to inequality is zero. When all the groups have a share of income equal to their population share, the overall theil measure is zero" (Conceicao (2001), p. 13). The basic formula used to calculate the theil index with income share data is:

$$T_{(j)}=\Sigma_{i(j)}y_{i(j)}ln(y_{i(j)}/p_{i(j)})$$

Using quintile data, since $i=1,2,3,4,5$ this equation can be expressed as:

$$T_{(j)}=\Sigma_{i(j)}y_{i(j)}ln(5y_{i(j)}).$$

where j indexes nation, i indexes quintile within nation, $T_{(j)}$ is the within-nation theil for the j^{th} nation, $y_{i(j)}$ is income share for the i^{th} quintile within nation j, and $p_{i(j)}$ is the population share of the i^{th} quintile within nation j.

Table 5 presents a simple example of calculating the theil index of inequality using 1998 Greek income share data. Because the richest group's income share is higher than its population share and the poorest group's income share is lower than its population share, the richest group's contribution to inequality is always positive and the poorest group's contribution is always negative (Conceicao (2001)).

TABLE 5: CALCULATION OF THE THEIL INDEX FOR GREECE WITH 1998 INCOME SHARE DATA FROM THE WORLD BANK (2003B)

P_{ij}= POPULATION SHARES	**Y_{ij}= INCOME SHARES**
0.2	0.0706
0.2	0.1145
0.2	0.1583
0.2	0.2202
0.2	0.4364
Calculation of the Log of the Ratio of Shares:	.0706*ln(.0706/.2) .1145*ln(.1145/.2) .1583*ln(.1583/.2) .2202*ln(.2202/.2) .4364*ln(.4364/.2)
Log of the Ratio of Shares/Contribution to Inequality:	=-0.073515 =-0.063862 =-0.037015 = 0.021187 = 0.340498
Theil	*0.187298*

Datasets on Inequality

Three data sets on inequality are discussed in this section. The World Development Indicators database (World Bank 2003b), Deininger and Squire's (1996, 1998) database, and UTIP's (2005) theil database. The World Income Inequality Database (WIDER 2007) a recent addition to inequality databases, is briefly presented.

The World Development Indicators Database (World Bank, 2003b) provides data on income shares and gini coefficients. This data, however, is limited in coverage especially for less developed countries. Deininger and Squire (1996, 1998) claim to have advanced the quality of data available and expanded the coverage of data with a "new and improved cross-country data set on inequality" ((1996), p. 3). They compute gini coefficients using a "high quality" sample based on income and consumption for a panel of approximately 700 country/year observations, from 1950 for 58 nations. These statistics can be obtained for free at the World Bank's website.[10]

The data set of Deininger and Squire (1996) has been used widely as a measure of income inequality in the last decade. The publication of the Deininger and Squire data set benefited the study of income inequality since it provides expanded income inequality data, enabling time-series and panel data analysis. However, the results of empirical analysis using this data set are varied, or even contradictory, when testing a relationship between income inequality and the level of development or growth. Using the Deininger and Squire data set, Barro (2000) finds an inverted-U curve, whereas Ram (1997) finds an upright-U relationship between income inequality and *the level of GDP*. Deininger and Squire (1996, 1998) and Quah (2001) fail to find a significant relationship between inequality and *GDP growth*. In contrast, Forbes (2000) finds a positive relationship between inequality and *growth* while Alesina and Rodrick (1994) find a negative relationship. The reason why these studies arrive at different conclusions using the same dataset may be attributed to sample selection.

Such authors as Szekely and Hilgert (1999), Atkinson and Brandolini (2001), Galbraith and Kumh (2002), and Panizza (2002) attribute these conflicting results to the quality and comparability of the data set. They believe that the data set has unbalanced coverage and is inaccurate. The inaccuracy of the data set, due to the different definitions of income inequality used for different countries, yields incomparable results for cross-sectional studies depending on the data selected for the study. This may be because "some (countries) use income and others expenditure data, some are based on gross and others on net income" (Teulings and van Rens (2002), p. 21).

Deininger and Squire (1998) suggest several remedies such as the inclusion of dummy variables and/or additional adjustments that may remedy this problem. However, researchers such as Atkinson and Brandolini (2001) state "that differences in definitions may be quantitatively important, but we doubt whether a single additional or multiplicative adjustment is a satisfactory solution to the heterogeneity of the available statistics. Our preference is for the alternative approach of using a data set where the observations are as fully consistent as possible" (p. 773).

In addition to not being consistent, the coverage of the Deininger and Squire data set is unbalanced. Data for western nations and Asia is pretty comprehensive, but data for

[10] Deininger and Squire (1996) specify data as "high quality" if they are based on a national household survey which is representative of the population and in which all sources of income have been counted.

Africa, Eastern Europe, and Latin America is limited. Firebaugh (2005) mentions that only 28 percent of African nations are covered in this data set. Africa, however, is an important region with trends in inequality that need to be understood better.

The World Income Inequality Database (WIDER 2007) consists of different estimates of income inequality data that have been compiled by different sources. The first version of this dataset was published in September of 2000 (WIDER 2007). This dataset which is also freely available on the internet provides useful inequality data for cross-country analysis. Income shares in deciles, quintiles and the gini index is reported for selected countries. All data reported are divided into three groups each group indicative of the quality of the data. The most recent compilation encompasses data from 1867-2004 (WIDER 2007). However, data before 1950 is sparse. Years after 1960 include data for more than 20 countries each year. The 1970s include on average data for 33 countries each year, the 1980s report on average data for 46 countries each year, the 1990s include an average count of 69 countries each year, and the 2000's report on average data for 56 countries each year. The largest number of countries reported is 77 in 1997. This dataset appears to be the most comprehensive dataset of income shares available but has similar problems of inaccuracy and imbalance faced by Deininger and Squire's data.

Ever since Henri Theil's publication of *Economics and Information Theory* in 1967, the theil index has been used widely to calculate inequality in various fields. More recently, Galbraith and Kumh (2001), under the University of Texas Inequality Project (UTIP 2005), have assembled a comprehensive data set of inequality through a compilation of the theil's t-statistic with approximately 3000 country/year observations. Their theil statistic is based on manufacturing wage and employment data and is compiled from the United Nations International Development Organization (UNIDO) Industrial Statistics.[11] According to Galbraith and Kumh (2001), their data is "more stable, more reliable and more comparable across countries than those of Deininger and Squire's" (p. 1).

The following formula specifies the method by which the theil index can be calculated given manufacturing pay data:

$$T_{(j)} = \sum_{i=1}^{m} w_{i(j)} \ln \frac{w_{i(j)}}{n_{i(j)}} = \sum_{i=1}^{m} \frac{Y_{i(j)}}{Y_{(j)}} \ln \left(\frac{Y_{i(j)} / Y_{(j)}}{N_{i(j)} / N_{(j)}} \right)$$

$$\text{where} w_{i(j)} = \frac{Y_{i(j)}}{Y_{(j)}}, n_{i(j)} = \frac{N_{i(j)}}{N_{(j)}},$$

$T_{(j)}$ is the theil index within country *j* measuring inequality between *i* segments, $w_{i(j)}$ is segment *i*'s income share within country *j*, $n_{i(j)}$ is segment *i*'s population share within country *j*, $N_{(j)}$ is the total employment in all segments in country *j*, $N_{i(j)}$ is the number of individuals in segment *i* of country *j*, $Y_{(j)}$ is total income from all segments in country *j*, and $Y_{i(j)}$ is income for segment *i*.

An illustration of measuring inequality using manufacturing pay data is presented in Table 6 where *i=1* represents employment in segment 1 and so on. People in each industry are classified into different income groups from the richest to the poorest. *Group a* represents the richest of the workers and *group c* the poorest. C_a represents income received

[11] See Conceicao (2001) for a detailed discussion of the theil's t-statistic of inequality.

by the richest group of workers, C_c is income received by the poorest group of workers. The first row in Table 6 below can be interpreted as: there are 23 people in the top wage bracket in segment 1 and their compensation is $120, $80 is the income received by the middle 15 wage-earners in this segment, and the poorest group consists of 12 people who make $40.

TABLE 6: EXAMPLE OF CALCULATING THE THEIL INDEX USING MANUFACTURING DATA

	Employment				**Compensation**			
	E_a=poor	E_b	E_c=rich	**sum**	C_a=poor	C_b	C_c=rich	**sum**
i=1	12	15	23	**50**	40	80	120	**240**
i=2	10	15	20	**45**	30	60	85	**175**
i=3	15	20	20	**55**	20	40	80	**140**
i=4	20	20	20	**60**	10	20	40	**70**
	$n_{i(j)}$ **POPULATION SHARE CALCULATION:**				$w_{i(j)}$ **INCOME SHARE CALCULATION:**			
	12/50 =.2400	15/50 =.3000	23/50 =.4600		40/240 =.1667	80/240 =.3333	120/240 =.5000	
	10/45 =.2222	15/45 =.3333	20/45 =.4444		30/175 =.1714	60/175 =.3429	85/175 =.4857	
	15/55 =.2679	20/55 =.3636	20/55 =.3636		20/140 =.1429	40/140 =.2857	80/140 =.5714	
	20/60 =.3333	20/60 =.3333	20/60 =.3333		10/70 =.1429	20/70 =.2857	40/70 =.5714	
	CALCULATION OF THE LOG OF THE RATIO OF SHARES:							
	.1667*ln(.1667/.2400)	.3333*ln(.3333/.3000)	.5000*ln(.5000/.4600)					
	.1714*ln(.1714/.2222)	.3429*ln(.3429/.3333)	.4857*ln(.4857/.4444)					
	.1429*ln(.1429/.2679)	.2857*ln(.2857/.3636)	.5714*ln(.5714/.3636)					
	.1429*ln(.1429/.3333)	.2857*ln(.2857/.3333)	.5714*ln(.5714/.3333)					
	LOG OF THE RATIO OF			**SUMMATI**				

	SHARES/CONTRIBUTION TO THEIL:			ON:	
	=-.0608	=.0351	=.0417	-.0608+.0351+.0417=.0160	
	=-.0445	=.0097	=.0432	-.0445+.0097+.0432=.0084	
	=-.0898	=-.0689	=.2583	-.0898-.0689+.2583=.0996	
	=-.1210	=-.0440	=.3080	-.1210-.0440+.3080=.1430	
The il				**0.2670**	

Since this compilation of the theil data set is relatively recent, there has not been much published research that makes use of this measure. Galbraith and Kumh (2001) run several model specifications and find a U-curve. With their findings, they conclude that the less industrialized countries experience a more egalitarian distribution of income accompany development, whereas industrialized nations observe increases in inequality accompany increases in GDP. They attribute this to perhaps what they call an augmented Kuznets' curve.

Compared to other existing databases on inequality, UTIP's (2005) data set covers more countries for longer periods, with fewer missing values, enabling a comprehensive empirical investigation of issues regarding wage inequality within manufacturing. Furthermore, data is compiled from a consistent source, the UNIDO Statistical Database. It is also available online for free.

Galbraith and Kumh (2001) claim that their data set is a "good proxy for income inequality" because inequality of pay is an important component of total income inequality. However, their theil index is derived not from inequality statistics of the entire economy, but from inequality in manufacturing pay which just represents a small proportion of the economy.

The different measures of income inequality presented in this section are each associated with certain strengths and weaknesses. Researchers using income shares and ratios are able to conduct a detailed analysis on the evolution of income between different population groups, but lack data to conduct a comprehensive time-series study. The benefit of Deininger and Squire's gini dataset and UTIP's (2005) theil dataset is the extensive availability of data enabling researchers with more flexibility in examining the distribution of income within countries. The Deininger and Squire data set however is inconsistent and incomparable across countries. A problem with the theil data set of income inequality is that the data set focuses narrowly on manufacturing wages. As illustrated by Table 7, employment in the manufacturing sector only accounts for 26% of total employment for all

countries, and only 16% of employment in low-income countries, causing some concern. Since movements in inequality within manufacturing may not be similar to movements in inequality within services and agriculture, it is necessary to assess the ability of the theil to proxy economy-wide inequality.

TABLE 7: THEIL INDEX AND EMPLOYMENT IN AGRICULTURE, INDUSTRY AND SERVICES AVERAGES OF DATA (1968-1999)

	ALL COUNTRIES *Col. 1*	**LOW INCOME** *Col. 2*	**LOWER MIDDLE INCOME** *Col. 3*	**UPPER MIDDLE INCOME** *Col. 4*	**HIGH INCOME** *Col. 5*	**OECD HIGH INCOME** *Col. 6*
Theil Index	0.065	0.089	0.058	0.056	0.055	0.021
Agriculture Employment (% of total employment)	21.50	51.35	26.53	16.45	7.77	8.60
Employment in Industry (% of total employment)	26.40	15.79	26.47	28.08	30.13	29.59
Employment in Services (% of total employment)	50.78	28.03	46.00	54.69	61.82	61.58

Comparability Tests

The purpose of this section is to compare the theil data set with four different data sets of inequality, income ratios and the gini indexes compiled by the World Development Indicators Database (World Bank 2003b) and Deininger and Squire (1996).

TABLE 8: DESCRIPTIVE STATISTICS FOR ALL COUNTRIES LISTED IN APPENDIX A

	Mean	Std.Dev.	Skewness	Kurtosis	Minimum	Maximum	Num Cases
Income Ratio	10.96	9.52	2.60	10.84	3.37	57.64	93
Gini	40.93	10.60	0.37	2.20	24.44	63.01	93
Theil	0.054	0.060	4.98	52.54	0.001	1.03	2787

Note: Countries All results based on non-missing observations.

Data on the income ratios and Gini index are from the World Development Indicators (World Bank 2003b).

Theil index is from UTIP (2004).

TABLE 9: CORRELATION MATRIX FOR ALL COUNTRIES LISTED IN APPENDIX A

	Income Ratio	Gini	Theil
Income Ratio	1.00	.90	.31
Gini	.90	1.00	.45
Theil	.31	.45	1.00

Note: All results based on non-missing observations.
Data on the income ratios and Gini index are from the World Development Indicators (World Bank 2003b).
Theil index is from UTIP (2004).

Table 8 describes the income ratio (income share held by the highest 20% of the population divided by the income shares of the lowest 20% of the population) and gini index from the World Development Indicators database (World Bank 2003b), as well as the theil index from UTIP (2005). As can be seen, the availability of the World Bank inequality data is only 93, mostly 1 data point for each country. The theil index, on the other hand, has 2787 data points.[12] The data compiled for the theil index is more skewed than the gini and income ratios. Kurtosis is the degree of peakedness of a distribution. The closer the kurtosis measure is to 0, the more normal the distribution. All three measures exhibit a distribution with a high peak, which is called leptokurtic, with the theil index being the most leptokurtic. The abovementioned differences between the theil and other inequality measures may be an indication that wages within the manufacturing sector are less spread out than society-wide wages.

Table 9 shows the correlation matrix for these three measures of inequality. It is clear that the income ratio and the gini index have a strong positive association at +.90. This strong correlation is expected since both measures are calculated based on income share data. The theil index does not exhibit high correlation with the World Bank income ratios and the gini index.

TABLE 10: DESCRIPTIVE STATISTICS FOR HIGH-INCOME OECD COUNTRIES

	Mean	Std.Dev.	Skewness	Kurtosis	Minimum	Maximum	NumCases
Income Ratio	5.71	1.57	0.26	2.13	3.37	9.01	22
Gini	32.06	5.03	-0.19	1.79	24.7	40.81	22
Theil	0.02	0.014	2.00	8.36	0.003	0.094	682

Note: All results based on non-missing observations.
Data on the income ratios and Gini index are from the World Development Indicators (World Bank 2003b). Theil index is from UTIP (2004).

[12] The original file compiled by Galbraith and Kumh and available on UTIP's website contains more data than this but certain countries were excluded from my analysis since they were not included in the World Bank database.

TABLE 11: DESCRIPTIVE STATISTICS FOR HIGH-INCOME OECD COUNTRIES

	Mean	Std.Dev.	Skewness	Kurtosis	Minimum	Maximum	NumCases
D&S Income Ratio	9.11	1.70	0.17	2.29	5.86	12.73	197
D&S Gini	32.39	4.24	-0.13	2.71	22.9	44	197
Theil	0.02	0.014	2.00	8.36	0.003	0.094	682

Note: All results based on non-missing observations.

Data on the income ratios and Gini index are from the Deininger and Squire Dataset (1996). Theil index is from UTIP (2004).

Table 10 shows descriptive statistics for 23 OECD high-income countries. As can be seen, the World Bank database only has one data point for each of the 23 high-income OECD countries, with an exception of Iceland which doesn't have any data on inequality. For most countries, this data point is between the years 1995-1998, with four exceptions: the data for Netherlands is from 1994, Japan 1993, Spain 1990, and Ireland 1987. The theil index, on the other hand, has 682 data points. Table 11 shows that income ratios (income share held by the highest top quintile divided by the income shares of the lowest quintile) and gini data availability from the Deininger and Squire high-quality data set is 197 for all 23 OECD countries, more than the World Bank's database.

TABLE 12: CORRELATION MATRIX FOR HIGH-INCOME OECD COUNTRIES

	Income Ratio	Gini	Theil
Income Ratio	1.00	.96	.45
Gini	.96	1.00	.56
Theil	.45	.55	1.00

Note: All results based on non-missing observations.

Data on the income ratios and Gini index are from the World Development Indicators (2003b).

Theil index is from UTIP (2004).

TABLE 13: CORRELATION MATRIX FOR HIGH-INCOME OECD COUNTRIES

	Income Ratio	Gini	Theil
D&S Income Ratio	1.00	.78	.16
D&S Gini	.78	1.00	.28
Theil	.16	.28	1.00

Note: All results based on non-missing observations.

Data on the income ratios and Gini index are from the Deininger and Squire Dataset (1996).

Theil index is from UTIP (2004).

Tables 12 and 13 report the correlation matrixes of inequality measures for OECD high-income countries. Income ratios and the gini indexes calculated by the World Bank (2003b) still exhibit positive associations with each other. Even though the Deininger and Squire's (1996) income ratios and gini indexes are highly correlated, this correlation is not as strong as it should be considering that both measures are calculated using income share data. This might be an indication of a problem with this data set. Results from the correlation

matrices in both tables indicate that the theil index is not strongly correlated with the inequality measures from the World Bank data base and the Deininger and Squire inequality measures.

Summary and Conclusions

All three of the widely employed datasets on inequality are associated with certain strengths and weaknesses. Even though the World Development Indicators Database (World Bank 2003b) covers years from 1960-2001, time-series data on income shares and gini are scarce. In general, each country only has income shares and gini data for 1 year, out of the 40 years supposedly covered, limiting time-series analysis. The strength of this dataset is the availability of inequality data for over 90 countries, which may prove useful for cross-section analysis.

The Deininger and Squire's (1996) data base also consists of income share and gini data. It is available for free online. Even though their data set covers approximately 120 countries, the coverage is imbalanced, with some countries having only one data point whereas others, such as the United States, having as many as 70 data points. This data set consists of data collected from various sources-various authors and different coverage (national, urban etc). Then, the reported data is sorted into acceptable and unacceptable groups- data in the acceptable group include only those based on national coverage and from a consistent source. This technique is useful since it yields more consistent data, but it also results in the loss of data coverage. Results in Section 5 indicate that the data from this compilation may not be reliable due to the lack of uniformity between two basic measures of inequality, the income ratios and the gini index.

The manufacturing theil data set compiled by Galbraith and Kumh (2001) through the UTIP project (2005) is also available for free online, and is more comprehensive and complete than the rest, covering approximately 3000 country/year observations with fewer missing data. In addition, the data reported is from a consistent source, the United Nations International Development Organization (UNIDO) Industrial Statistics. However, this theil statistic is based on manufacturing wage and employment data and does not accurately reflect society-wide inequality. The analysis in Section 5 shows that UTIP's (2005) theil index is not strongly correlated with any of the other commonly used datasets on within-country income inequality, and should therefore, be interpreted carefully. However, it is a useful depiction of the distribution of wages within the manufacturing sector.

Income inequality within countries leads to wide disparities in infant mortality rates for all births and mortality rates for children less than five years old (World Bank 2003b). "This is due to differential access to basic health services, safe drinking-water, adequate nutrition, and safe motherhood and child initiatives" (World Bank2003b) between the rich and the poor. Finding a uniform and comprehensive dataset of income inequality still remains a challenge. Until then, researchers and policymakers should proceed with caution when interpreting the results of studies on inequality, keeping in mind the drawbacks of the existing measures. The World Income Inequality Database (WIDER 2007) which is yet to be fully explored, offers some progress towards a more comprehensive income share dataset for cross-sectional analysis.

References

Aghion, Philippe, Eve Caroli and Cecilia Garcia-Penalosa (1999). "Inequality and Economic Growth: The Perspective of the New Growth Theories," *Journal of Economic Literature*, XXXVII (December): 1615-1660.

Ahluwalia, Montek S. (1976). "Income Distribution and Development: Some Stylized Facts," *American Economic Review*, 66(2): 128-135.

Alesina, Alberto and Roberto Perotti (1994). "The Political Economy of Growth: A Critical Survey of the Recent Literature," *World Bank Review*, 8(3): 351-371.

Atkinson, Anthony and Andrea Brandolini (2001). "Promise and Pitfalls in the Use of Secondary Data-Sets: Income Inequality in OECD Countries as a Case Study," *Journal of Economic Literature*, 34: 771-799.

Barro, Robert J. (2000). "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, 5: 5-32.

Brown, Malcolm (1994). "Using Gini-Style Indices to Evaluate the Spatial Patterns of Health Practioners: Theoretical Considerations and an Application Based on Alberta Data, *Social Science Medicine*, 38(9): 1243-1256.

Conceicao, Pedro (2001). "Inequality Dynamics: A Note on the Time-wise Decompositions of Entropy-based Measures of Inequality with a Special Focus on Theil Measures," UTIP Working Paper19, UTIP and Center for Innovation, Technology and Policy Research, Lisbon, Portugal, April.

Deininger, Klaus and Lyn Squire (1996). "A New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10: 565-591.

_____ (1998). "New Ways of Looking at Old Issues: Inequality and Growth," *Journal of Development Economics*, 2: 318-334.

Firebaugh, Glenn (2005). "The New Geography of Income Inequality: How World Inequality Changed in the Latter Part of the 20th Century," Working Paper, Session: Trends in Income Inequalities, Pennsylvania State University, USA.

Forbes, Kristin J. (2000). "A Reassessment of the Relationship Between Inequality and Growth," *American Economic Review*, 90: 869-887.

Foster, Michael and Mark Pearson (2000). "Income Distribution in OECD Countries," Poverty and Income Inequality in Developing Countries: A Policy Dialogue on the Effects of Globalisation, OECD Development Centre, Paris.

Galbraith, James and Hyunsub Kumh (2001). "Inequality and Growth Reconsidered Once Again: Some New Evidence from Old Data," Association for Comparative Economic Studies (ACES), Ne:w Orleans, January.

Gottschalk, Peter (1997). "Inequality, Income Growth, and Mobility: The Basic Facts," *Journal of Economic Perspectives*, 11(2): 21-40.

Jomo, Kwame S. (2000). "Globalisation, Liberalisation, Poverty and Income Inequality in Southeast Asia," Poverty and Income Inequality in Developing Countries: A Policy Dialogue on the Effects of Globalisation, OECD Development Centre, Paris.

Kravis, Irving (1960). "International Differences in the Distribution of Income," *Review of Economics and Statistics*, 42 (November): 408-416.

Kuznets, Simon (1955). "Economic Growth and Income Inequality," *American Economic Review* 45(1): 1-28.

Lindert, Peter and Jefferey Williamson (1985). "Growth, Equality, and History," *Explorations Economic History*, 22 (October): 341-377.

Lorenz, Max C. (1905). "Methods of Measuring the Concentration of Wealth," *Publication of American Statistical Association*, 9: 209-219.

Lundberg, Mattias and Lyn Squire (2003). "The Simultaneous Evolution of Growth and Inequality," *Economic Journal*, 113 (April): 326-344.

Mukhopadhaya, Pundarik (2004). "World Income Inequality Data Base (WIID) Review," *Journal of Economic Inequality*, 2(3): 229-234.

Nafziger, Wayne E. (1997). "The Economics of Developing Countries," 3rd edition, Upper Saddle River, N.J.: Prentice Hall.

Panizza, Ugo (2002). "Income Inequality and Economic Growth: Evidence from American Data," *Journal of Economic Growth*, 7(1): 25-41.

Paukert, Felix (1973). "Income Distribution at Different Levels of Development: A Survey of Evidence," *International Labor Review* 108 (August- September): 97-125.

Perkins, Dwight, Steven Radelet, Donald R. Snodgrass, Malcolm Gillis and Michael Roemer (2001). "Economics of Development," 5th edition, W.W. Norton & Company, Inc, New York.

Perotti, Roberto (1996). "Growth, Income Distribution, and Democracy: What the Data Say," *Journal of Economic Growth*, 1(2): 149-187.

Persson, Torsten and Guido Tabellini (1994). "Is Inequality Harmful for Growth?" *American Economic Review*, 84(3): 600-621.

Quah, Danny (2001). "Some Simple Arithmetic on How Income Inequality and Economic Growth Matter," *Journal of Applied Econometrics Lecture*, Econometric Society Meetings, Auckland NZ, 1(July): 1-42.

Ram, Rati (1997). "Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World," *Southern Economic Journal,* 64: 576–583.

Ray, Debraj (1998). *Development Economics*, Princeton University Press, Princeton, New Jersey.

Szekely, Miguel and Marianne Hilgert (1999). "What's behind the Inequality We Measure: an Investigation using Latin American Data," Washington, D.C. IADB, Working Paper 409.

Theil, Henri (1967). *Economics and Information Theory,* Chicato: Rand McNally.

Tsakloglou, Panos (1988). "Development and Inequality Revisited," *Applied Economics*, 20:509-531.

United Nations International Development Organization, UNIDO (2001), Industrial Statistics Database.

UTIP (2005). "University of Texas Inequality Project," Home page. Aug. 31 2005. Oct. 20 2005. <http://utip.gov.utexas.edu>

World Bank (1993). "The East Asian Miracle," Oxford University Press, New York.

_____ (2003a). "Human Development Indicators 2003," Home page. May 15 2005. June 20 2005. <http://www.undp.org/hdr2003/indicator/indic_126_1_1.html>

_____ (2003b). "World Development Indicators CD-Rom," the World Bank, Washington D.C.

WIDER (2007). "World Income Inequality Database V 2.0b," Home page. Sept. 17 2007. Sept. 21 2007. <http://www.wider.unu.edu/wiid>

APPENDIX A: Countries Included in this Study

TABLE 14: COUNTRIES INCLUDED IN THE STUDY BY INCOME CLASSIFICATIONS

LOW INCOME	LOWER MIDDLE INCOME	UPPER MIDDLE INCOME	HIGH INCOME	OECD HIGH INCOME
Afghanistan	Albania	Argentina	Australia	Australia
Angola	Algeria	Barbados	Austria	Austria
Armenia	Belize	Botswana	Belgium	Belgium
Azerbaijan	Bolivia	Brazil	Canada	Canada
Bangladesh	Bosnia Herzegovina	Chile	Cyprus	Denmark
Benin	Bulgaria	Costa Rica	Denmark	Finland
Burkina Faso	Cape Verde	Croatia	Finland	France
Burundi	China	Czech Republic	France	Gremany
Cameroon	Colombia	Gabon	Germany	Greece
Central African Republic	Cuba	Hungary	Greece	Iceland
Congo, Rep.	Dominican Republic	Latvia	Hong Kong, China	Ireland
Cote D'ivoire	Ecuador	Libya	Iceland	Italy
Eritrea	Egypt, Arab Rep.	Lithuania	Ireland	Japan
Ethiopia	El Salvador	Malaysia	Israel	Korea, Rep
Gambia, The	Fiji	Malta	Italy	Luxembourg
Ghana	Guatemala	Mauritius	Japan	Netherlands
Haiti	Honduras	Mexico	Korea, Rep.	New Zealand
India	Iran, Islamic Rep.	Oman	Kuwait	Norway
Indonesia	Iraq	Panama	Luxembourg	Portugal
Kenya	Jamaica	Poland	Macao, China	Spain
Kyrgyz Republic	Jordan	Puerto Rico	Netherlands	Sweden
Lesotho	Macedonia, FYR	Seychelles	New Zealand	United Kingdom
Madagascar	Morocco	Slovak Republic	Norway	United States
Malawi	Peru	Trinidad And Tobago	Portugal	
Moldova	Philippines	Uruguay	Qatar	
Mongolia	Romania	Venezuela, RB	Singapore	
Mozambique	Russian Federation		Spain	
Myanmar	South Africa		Sweden	
Nepal	Sri Lanka		United Arab Emirates	
Nicaragua	Suriname		United Kingdom	
Nigeria	Swaziland		United States	
Pakistan	Syrian Arab Republic			
Papua New Guinea	Thailand			
Rwanda	Tonga			
Senegal	Tunisia			
Sierra Leone	Turkey			
Somalia	Yugoslavia, Fed. Rep.			
Tanzania				
Togo				
Uganda				
Ukraine				
Yemen, Rep.				
Zambia				
Zimbabwe				

APPENDIX B: Definition of Explanatory Variables

From the World Development Indicators, 2003

Employment in agriculture (% of total employment): "is the ratio of people employed in agriculture to the total number of people working. Data on employment are drawn from labor force surveys, establishment censuses and surveys, administrative records of social insurance schemes, and official national estimates. The concept of employment generally refers to people above a certain age who worked, or who held a job, during a reference period. Agriculture corresponds to division 1 (ISIC revision 2) or tabulation categories A and B (ISIC revision 3) and includes hunting, forestry, and fishing."

Employment in industry (% of total employment): "is the ratio of people employed in industry to the total number of people working. Industry corresponds to divisions 2-5 (ISIC revision 2) or tabulation categories C-F (ISIC revision 3) and includes mining and quarrying (including oil production), manufacturing, construction, electricity, gas, and water."

Employment in services (% of total employment): "is the ratio of people employed in industry to the total number of people working. Services correspond to divisions 6-9 (ISIC revision 2) or tabulation categories G-P (ISIC revision 3) and include wholesale and retail trade and restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services."

APPENDIX C: Comparison of Inequality Datasets

The following provide a detailed analysis of the Income Share and Gini indexes from the World Bank(2003b) dataset and the theil index..

TABLE 15: PARTITION OF RANGE: MINIMUM TO MAXIMUM FOR ALL COUNTRIES

Range of X	Income Ratios	GINI	Theil
Minimum	3.3696	24.440	.14034E-02
1st.Qrtl	16.936	34.082	.25748
Midpoint	30.503	43.725	.51356
3rd.Qrtl	44.070	53.367	.76964
Maximum	57.636	63.010	1.0257

TABLE 16: ORDER STATISTICS FOR ALL COUNTRIES

PERCENTILE	INCOME RATIOS	GINI	THEIL
Min.	3.3696	24.440	.14034E-02
10th	4.0309	25.880	.86123E-02
20th	5.0767	31.600	.14815E-01
25th	5.4008	32.740	.17850E-01
30th	5.6003	34.360	.22104E-01
40th	6.6981	36.420	.29204E-01
Med.	7.4961	39.500	.39398E-01
60th	9.1381	42.998	.49553E-01
70th	11.452	46.240	.61823E-01
75th	12.275	48.220	.69662E-01
80th	13.801	50.560	.78125E-01
90th	20.297	57.140	.10426
Max.	57.636	63.010	1.0257

TABLE 17: PARTITION OF RANGE: MINIMUM TO MAXIMUM FOR HIGH-INCOME OECD

Range of X	Income Ratios	GINI	THEIL
Minimum	3.3696	24.700	.28545E-02
1st.Qrtl	4.7801	28.728	.25631E-01
Midpoint	6.1906	32.755	.48408E-01
3rd.Qrtl	7.6011	36.783	.71185E-01
Maximum	9.0117	40.810	.93962E-01

TABLE 18: ORDER STATISTICS FOR HIGH-INCOME OECD COUNTRIES

Percentile	Income Ratios	GINI	THEIL
Min.	3.3696	24.700	.28545E-02
10th	3.5708	24.925	.67317E-02
20th	4.0333	25.300	.93003E-02
25th	4.4058	25.695	.10456E-01

30th	4.6289	28.140	.11803E-01
40th	5.1578	31.130	.14550E-01
Med.	5.4965	32.550	.17176E-01
60th	5.8897	33.965	.20298E-01
70th	6.5940	35.637	.24913E-01
75th	6.8942	35.937	.26417E-01
80th	7.0498	36.003	.27423E-01
90th	7.5021	37.194	.34764E-01
Maximum	9.0117	40.810	.93962E-01

The following provide a detailed analysis of the Income Share and Gini indexes from the Deininger and Squire dataset the theil index.

TABLE 19: PARTITION OF RANGE: MINIMUM TO MAXIMUM FOR HIGH-INCOME OECD

Range of X	D&S Income Ratio	D&S Gini	Theil
Minimum	5.8633	22.900	.33266E-02
1st.Qrtl	7.5805	28.175	.14288E-01
Midpoint	9.2977	33.450	.25249E-01
3rd.Qrtl	11.015	38.725	.36210E-01
Maximum	12.732	44.000	.47171E-01

TABLE 20: ORDER STATISTICS FOR HIGH-INCOME OECD COUNTRIES

	D&S Income Ratio	D&S Gini	Theil
Min.	5.86326	22.9	0.0033266
10th	6.98315	26	0.0063523
20th	7.49704	28.39	0.0092312
25th	7.71395	30.04	0.0099598
30th	8.08861	30.8	0.0109448
40th	8.76144	31.83	0.013617
Med.	9.10469	32.54	0.0153819
60th	9.42679	33.58	0.0185245
70th	9.77966	34.6	0.0227727
75th	10.3545	35.06	0.0248927
80th	10.7091	35.7	0.0260747
90th	11.7255	37.56	0.0274547
Max.	12.7322	44	0.0471712