Re-examining Kuznets Hypothesis: Does Data Matter?

Jalil, Mohammad Muaz

North South University

August 2009

Re-examining Kuznets Hypothesis: Does Data Matter?

Mohammad Muaz Jalil
Student ID# 071 889553

Department of Economics
North South University
Dhaka, Bangladesh

August 2009

Re-examining Kuznets Hypothesis: Does Data Matter?

Submitted By
Mohammad Muaz Jalil
Student ID# 071 889553

Thesis Supervisor:
Dr. A.K.M Atiqur Rahman
Professor, Department of Economics
North South University
Dhaka, Bangladesh

*A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Economics, Department of Economics, North South University, Summer 2009

Thesis Examination Committee

Chairman

Dr. A. K. M. Atiqur Rahman
Professor, Department of Economics
North South University

Members

Dr. Mohammad Ali Rashid
Professor, Department of Economics &
Dean, School of Arts and Social Sciences
North South University

Dr. Amirul Islam Chowdhury
Professor, Department of Economics
Dean, School of Business
North South University

Dr. Gour Gobinda Goswami
Associate Professor & Proctor,
Department of Economics
North South University

ACKNOWLEDGEMENT

I would like to extend my gratitude to several people who have made this thesis a reality.

First of all I wish to thank my thesis supervisor Dr. A. K. M. Atiqur Rahman Professor, Department of Economics, for his time, patience, support, guideline and encouragement every step of the way. Working with him was a great learning experience and I hope to utilize this in my future endeavors.

I am really grateful to my family for encouraging me to write this thesis and for their continuous concern and support.

My thanks also go out to Dr. Gour Gobinda Goswami, Dr. Mohammad Ali Rashid, Dr. Amirul Islam Chowdhury, for having agreed to be a part of the thesis examination committee. I also wish to thank them for their valuable comments and insights which I have tried to incorporate to the best of my abilities.

Table of Content

List of Tables

List of Figures

ABSTRACT

Kuznets Hypothesis has been in debate since Simon Kuznets published his seminal paper "Economic Growth and Income Inequality" hypothesizing that inequality follows an inverted U shaped curve. He suggested that inequality rises as an economy develops, due to urbanization and industrialization, which later is abated as leveling forces gradually reduces inequality. Extensive research has been done in this arena but consensus is yet to be reached. In this paper we test the robustness of Kuznets hypothesis by employing newly available EHII and UTIP manufacturing pay inequality dataset which has been developed by University of Texas Inequality Project. Panel unit root tests are undertaken to develop the parametric equation for testing the hypothesis. We also use various econometric methods (Fixed Effect, dynamic panel analysis, fixed effect with autoregressive term) to analyze the effect of economic model on existence or absence of Kuznets Curve. Data are also segregated to Global dataset and OECD dataset, one including all the countries in the world and the other only OECD countries respectively. The objective is to see if there exists a different inequality dynamics for highly developed economy. The paper found that income variables should be in log formed and not level form while testing the hypothesis, as otherwise they are not stationary. It was also found that gender segregated cohort size has an implication for inequality, with matured male cohort having negative and matured female cohort size having positive relation with inequality In relation to Kuznets hypothesis it was found that inverted U shaped curve appears in case of D&S and WIID2 data but U shape curve appears in case of EHII and UTIP dataset. In case of OECD countries the difference is enhanced. The primary reason rests on rising manufacturing pay inequality which EHII captures. It is hypothesized that inequality follows a zigzag pattern with inequality rising and falling as economy develops and moves from an agrarian to an urbanized industrialized economy. This is in line with Kuznets hypothesis. After that, the economy faces major technological innovations which on onset increase the inequality within the manufacturing sector first but given the industrialized nature of the economy, this translates quickly to overall rise in inequality. Hence it seems that absence or presence of Kuznets curve greatly depends on the usage of inequality dataset. It is suggested that existing EHII dataset may be augmented by basing it on WIID2 instead of D & S and by making estimates which are gross of individual income resulting in data harmonization.

I. Introduction

Five decades ago, Simon Kuznets (1955) expressed the important hypothesis that income inequality first increases, but after a turning point it decreases in the course of economic development. This premise, usually termed Kuznets's hypothesis or Kuznets's inverted-U, has been widely investigated, but the results of that research are far from well established. Kuznets' original hypothesis relied on historical data for the first half of the nineteenth century from only three developed countries, the US, the UK and Germany, and he cautiously concluded that the data appeared to *'justify a tentative impression of constancy in the relative distribution of income before taxes, followed by some narrowing of relative income inequality after the first world war — or earlier'*[1]. Kuznets (1955) did not set out a formal theory of the relationship between the degree of income inequality within a country and its level of economic development; but he drew an argument.

Here is how Kuznets curve is supposed to work: in early stage of development investment opportunities for those who have money multiply, while wages are kept low due to influx of cheap labor from rural to urban areas. In Kuznets own words "An invariable accompaniment of growth in developed countries is the shift away from agriculture, a process usually referred to as industrialization and urbanization." With industrialization concomitantly inequality increases. Hence you get in to a situation where there are many business moguls coexisting with large body of impoverish day laborers. But gradually this urbanization or rural urban migration flattens out, hence wages begin to rise. At the same time education, enhanced social and political consciousness forces government or people in power to undertake redistributive efforts. These forces combine together to reduce inequality.

As of date the hypothesis has found many supporters, to the point of being considered 'fully confirmed' by Oshima (1970), a 'stylized fact' by Ahluwalia (1976a), and an 'economic law' by Robinson (1976). Recent literature nonetheless has been more cautious in their conclusions. They note that the statistical significance of the income variables of the basic

[1]Kuznets, S. (1955) Economic Growth and Income inequality, American Economic Review, page. 5

Kuznet model[2] tend to get eliminated with addition of other right-side variables such as education (Bourguignon and Morrison, 1990). Many studies however studies go on supporting empirically the hypothesis, as is the case of Dawson (1997), Li et al. (1998), Barro (2000), Thornton (2001), and Huang (2004). Similarly there are those who question the hypothesis, as did Adelman and Morris, (1973); Saith, (1983), Papanek and Kyn, (1986). More recently other skeptical authors have joined this group, who challenge the hypothesis, as for example Hsing and Smith (1994), Deininger and Squire (1998), or Mátyás et al. (1998) who labeled the hypothesis as a 'myth'. So, the hypothesis remains a theme of substantial debate in development literature.

In order to rigorously test the Kuznets hypothesis it is necessary to at least use longitudinal data although panel data structure is even better. Kuznets himself, as mentioned before, used time series data for three countries to formulate his hypothesis. Since at that time panel data analysis did not exists and neither did adequate level of inequality data, it was impossible for Kuznets to go beyond his conjecture. Even thought panel data analysis has existed for quite sometime, lack of adequate data on inequality forbade its use and hence most early researchers had to employ dataset which were almost entirely cross-sectional in nature, with typically one[3] observations per country. With these data, a number of studies found support for the Kuznets curve (Ahluwalia, 1976a, 1976b; Campano & Salvatore, 1988; Chenery & Syrquin, 1975; Dawson,1997; Eusufzai, 1997; Jha, 1996; Kravis, 1960; Mbaku, 1997; Papanek & Kyn, 1987; Paukert, 1973; Randolph & Lott, 1993; Tsakloglou, 1988; Bourguignon, 1994; Milanovic, 1995; Jha, 1996).

In so far as the lack of inequality data is concerned, Deininger and Squire's effort (hereafter D&S, 1996) is monumental. D&S collected many different surveys of income inequality, and compiled those meeting certain criteria of process[4] into a single "high-quality" panel, offering 693 country/year observations since 1950. Although this dataset allows for undertaking panel data analysis, but when one tries to undertake analysis with all countries

[2] With inequality measure as dependent and income variable (with quadratic term) as regressor
[3] Sometimes a few observations per country were also available
[4] Three main criteria are that observations should be (1) drawn from a published household survey, (2) based on the whole population, and (3) based on a comprehensive measure of income or expenditure.

included, degrees of freedom is significantly reduced and then there may not be sufficient data points. Even with this limitation, in absence of alternatives, this is now a standard reference, on which dozens of papers have been based. Deininger and Squire (1998) using their own dataset rejected the presence of the Kuznets curve for the fixed-effects case. They do find it present in the pooled case for their functional form (namely real GDP per capita and 1/(real GDP per capita)). Barro (2000) uses a different functional form (log y and its square) and finds the inverted-U shape present in both the crosssectional pooled and fixed-effects cases. Anand and Kanbur (1993) found that the functional form chosen to test the inverted-U hypothesis could have considerable impact on the ‘turning point’, of the curve, where inequality begins to decline. They also found that the U-shape is significant for some functional forms and not for others.

More recent studies have adopted a panel data approach by using the Deininger and Squire (1996) data set and have obtained different forms of the inequality-growth relationship (Ram, 1997; Barro, 2000; Forbes, 2000). However, the D & S data set has been criticized for not generating an accurate outcome since many of its observations are not consistent and comparable, even after applying “high quality filters”, and because its coverage is limited and unbalanced (Atkinson et al., 2001; Galbraith and Kum, 2002). Still other recent papers have used the Deininger and Squire (1996) dataset (updated, with more observations) and also used country-specific fixed effects. Higgins and Williamson (1999) examine the impact of openness and cohort size on inequality in addition to the Kuznets process and found that Kuznets Curve comes out of hiding when the inequality relationship is conditioned by the cohort size. Munir and Muaz (2004) used new datasets introduced by University of Texas inequality project, UTIP. The study used 24 countries purposely selected to develop a balance panel covering LDCs, developing countries and developed countries, over a time period of 37 years from 1963 to 1999. The results were negative for both level and log quadratic formed of equations that were tested for. Time series analysis was also performed on individual countries and the results were still negative.

The studies, so far, that have explored the relationship between inequality and the level of development have broadly differed in terms of: inequality dataset employed, parametric form used, conditionality imposed (independent variables), and the econometric model used. Also one of the major limitations of the studies has been comparability of the data across countries. The present study will try to address all of these issues in a systematic manner. The ultimate objective is to bring a reasonable consensus in relation to Kuznets hypothesis. In order to achieve this objective the paper will undertake the following:

1. Four different types of inequality dataset will be used as dependent variable, namely World income inequality database (WIID2), UTIP UNIDO Manufacturing Pay inequality dataset (UTIP), Estimated Household Income Inequality Data Set (EHII) and D&S, 1996[5].

2. Gamut of explanatory variables will be used, taken from existing literature, to see:
    a. The affect of such variables on the Kuznets relationship. Whether presence of certain explanatory variables remove the significance or alters the sign of the income variables, as some research have shown.
    b. The relationship that exists between inequality and such explanatory variables and to analyze the stability of such relationship. Whether such relationship varies across the types of inequality dataset used or the structure of the dataset or the econometric method employed

3. Stationarity test will be used to ascertain the correct functional form of the econometric model, whether one has to use variables in their log or level form. Anand and Kanbur (1993b) found that the functional form chosen to test the Kuznets hypothesis can have considerable impact. They found that the inverted U-shape is significant for some functional forms and not for others.

4. Analysis will be carried out on three separate dataset, namely
    a. Annualized global dataset of 188 countries
    b. 4 years average of the global dataset
    c. A dataset including only OECD countries

[5] In the following the detail descriptions of the inequality datasets will be provided.

The objective of employing these 3 types of dataset is to assess if the structure of the data itself has any impact on the Kuznet hypothesis and the relationship between inequality and the other explanatory variables. Also many studies like Higgins and Williamson (1999), Ram (1991 and 1997), Alderson and Nielsen (2002) etc, only researched on OECD countries. Hence it is worth noting whether this has any impact on the Kuznets hypothesis. 4 years average is used to reduce serial correlation and also because researchers have often suggested that inequality is likely to be a stable across time. It is worth noting whether smoothing the dataset has any impact in determining the presence or absence of Kuznets curve.

5. Various econometric methods will be used in line with current literature and as the data demands. Literature suggests that most often used models are pooled regression and fixed effect model. However some recent researchers like Galbraith and Kum (2004), Meschi and Vivarelli (2007) employ dynamic panel model, specifically Arellano and Bond GMM methodology. While earlier researches have used cross sectional data in absence of adequate panel data, since we do not face this constraint it seems inappropriate to use this method; hence this will not be included in this paper.

Although it is improbable to answer all queries but it is hoped that present study will go a long way in testing the robustness of the elusive Kuznets curve and possibly provide key reasons for existence or absence of Kuznets curve under different circumstances.

Section III will discuss the four different types of inequality dataset and their interrelationship will also be explored in this section. In section IV the conditioning or independent variables that will be employed in this study will be analyzed; the rationale for their choice and sources of dataset will be discussed. Section V will focus on developing the functional form of the econometric model to be tested by undertaking battery of panel unit root tests on the aforesaid variables, in their log transformed and level form. Once the variables and the functional formed is identified Section VI will focus on estimating the econometric relationship by using various tests and models. Section VII will present the findings of the study and Section VIII will offer the concluding remarks.

II. The Inequality Datasets

Deininger and Squire collected many disparate surveys of income and expenditure inequality, and compiled those meeting certain criteria of process1 into a single “high-quality” panel, offering about 693 country/year observations since 1947. The database uses different sources to compute Gini coefficients, depending on the data available in each country. There are three major differences. The first is whether the unit of analysis is a household or an individual. If, as is usually the case, poor households have more members, the distribution of income at a household level will be more equal than when computed at the individual level. Therefore, one would expect to find that the Gini coefficients are greater (more unequal income distribution) in countries that report data at the individual level. The second issue is whether income data refers to income before or after tax. Provided the tax system is progressive, countries that collect data on gross (before-tax) income will probably have a higher Gini coefficient than countries that report data on net income. Finally, some countries measure the distribution of income, while others measure the distribution of expenditure, which is measured on the basis of net income. In addition, given that high income households presumably save a bigger proportion of their income than poor households, it is expected that countries that use income rather than expenditure will have higher Gini coefficients.

Table 1: The distribution of inequality measures by different definitions in D & S data

	Reference unit*									
	Household		Household equivalent		Person		Person equivalent		Total	
Source***	Gross**	Net	Gross	Net	Gross	Net	Gross	Net	Gross	Net
Expenditure		23				104		1		128
Income	254	72		12	108	46		34	362	164

* Indicates whether the recipient unit is the household, the household equivalent (weighted by number of persons), the person, or the person equivalent (wherein the effective number of household members is assumed to be the square root of the actual number). – ** Indicates whether the measure of income is gross or net of taxes. – *** Indicates whether the measure of source is expenditure or income.

From the above table we see that 50% of the data in D&S is based on household estimates, gross and net inclusive. Similarly 54% of the data are based on are gross estimate, that is before tax deduction. Income based inequality measures accounts roughly 80% of the total data. Hence it is very likely that the gini measured by D&S is likely to be an overestimation of the actual underlying inequality scenario.

Despite the large number of observations, the coverage of the D&S data set remains limited and unbalanced. Serious questions have been raised as to whether the data points are in fact comparable either across countries or through time. As Atkinson and Brandolini (2001) especially argue, the D&S inequality measures are based on various income definitions, reference units and processing procedures that cannot be wholly reconciled to each other, even with "high-quality" filtering. Even within individual countries, the range of fluctuation in the D&S data is occasionally far too wide. For instance, the measure of inequality in Sri Lanka plummets by 16 Gini points during three years from 1987 to 1990. And there is an increase of almost 10 Gini points in Venezuela in just one year, 1989-1990. D&S suggest adding 6.6 Gini points to measures of inequality in expenditure data, in order to make the figures comparable to measures of income inequality. However Atkinson and Brandolini reject this methodology, that whether a simple additional or multiplicative adjustment is a satisfactory solution to the heterogeneity of the available statistics. Instead they suggest that one should be using a data-set where the observations are as fully consistent as possible. UTIP UNIDO and EHII dataset were developed by University of Texas inequality project (UTIP) as an answer to this criticism.

At the initial stage UTIP developed the UTIP UNIDO manufacturing pay inequality dataset. The strategy followed in constructing this dataset was to narrow the focus of overall inequality to the measures of inequality in manufacturing pay. While this may seem an extreme concession, it was motivated by several considerations. First, pay is a major source of total income. Thus, changes in pay inequality are reflected in income inequality. For example, Williamson (1982) argues that the "wage differential and its development seems to parallel broader trends in income distribution;" Second, while Kuznets' hypothesis was based mainly on between-sector inequalities in a two-sector (agriculture-industry) model of the economy, the role of inequality within each of these sectors is surely substantial. According to Fields (1980), the largest share of overall inequality can be accounted for by inequality within sectors, and the inequality in modern, industrial and urban sector rather than in the traditional and agricultural sectors is the driving force behind the evolution of inequality.

Third, manufacturing pay has been measured with reasonable accuracy as a matter of official routine in most countries around the world for nearly forty years. Berman (2000) has recently endorsed the coverage and accuracy of the United Nations International Development Organization's (UNIDO) compilation of these measures. Moreover, UNIDO's measures are comparable and consistent across countries, since they are based on a two or three digit code of the International Standard Industrial Classification (ISIC), a single systematic accounting framework. The measure of inequality using the UNIDO data is the between-groups component of Theil's T statistic, an entropy measure whose functional form is defined as

$$T = \sum \left(\frac{Y_i}{Y}\right) T_i + \sum \frac{Y_i}{Y} \log\left(\frac{Y_i / Y}{N_i / N}\right) = T^W + T^B$$

where T^w and T^B indicate within-group and between-group inequality measures respectively. N and Y stand for total employment and total pay respectively, and subscript i denote group identity. T^B is used as the inequality measure, where groups are defined as categories within the UNIDO industrial classification codes. Theil (1972) has shown that T^B is a consistent lower-bound inequality measure, where the within-groups component is unobserved The UNIDO source permits calculation of inequality measures for nearly 3200 country/year observations, covering over 150 countries during the period 1963 to 1999. These measures were computed for the University of Texas Inequality Project.

Galbraith and Kum (2004), part of the UTIP, developed a second dataset called the Estimated household income inequality (EHII) dataset. Basically they have regressed Deininger and Squire's Gini coefficients on the values of explanatory variables, which include the different income measures of Deininger and Squire's data set, the set of measures of the dispersion of pay in the manufacturing sector, and the manufacturing share of the population. Unexplained variations in Deininger and Squire's income measures are treated as inexplicable, and they are discarded from the calculations of EHII Gini coefficients. According to Galbraith and Kum (2004) EHII Gini has three clear advantages over the Deininger and Squire's Gini index. First, with more than 3,000 estimates, the coverage basically matches that of the UTIP-UNIDO exercise, providing substantially annual estimates of household income inequality for most countries, including developing countries that are badly under-represented in D&S. Second, this data set borrows accuracy from the UTIP-UNIDO pay dispersion measures. Thus, changes over time and differences across countries in pay dispersion are reflected in income inequality, in proportion to their historical importance with due adjustment for the different employment weight of manufacturing in different economies. Third, all estimates are adjusted to household gross income hence data comparability across countries is greatly enhanced. Previously in this section we discussed the potential comparability problem that arises in D&S dataset as the unit of analysis varies between household and individuals, income used in calculating the gini is either in gross or net of value and lastly gini is calculated based either on income or expenditure. EHII dataset avoids this comparability issue by providing estimates which are adjusted to household gross income. However as mentioned in Section II, estimates based on gross household income may significantly overestimate the underlying inequality. Thus data comparability or homogenous method of reporting may come at a price.

As mentioned before the index is calculated from OLS estimates with conditioning variables, just two exogenous variables: pay inequality and manufacturing share, plus dummies for data type as described below.

In its log form the "EHII Gini" is simply:

$$EG = \alpha + \beta * T + \gamma * X$$

where EG stands for estimated household income inequality, T is for UTIP-UNIDO pay inequality, and X is a matrix of conditioning variables, including the three types of data source (H,G and I), manufacturing employment share to population (mfgpop).

Our last inequality dataset is the latest United Nation’s World Income Inequality Database WIID2. The WIID2 is a data set collected by the United Nation University and the World Institute for Development Economics Research (WIDER). The data set covers about 159 countries. The sources for the inequality measures in the different countries are central statistical offices of the countries involved as well as the Transmonee database of UNICEF/ICDC, the unit record data of the Luxembourg Income Study, the World Bank Poverty Monitoring database, the Socio-Economic database for Latin America and the Caribbeans, and various research studies (in particular, Deininger and Squire 2004). The measures of inequality included in the data set are: the Gini coefficient, quintile/decile group shares, income shares of the poorest 5% and richest 95% of the population, survey means, and medians. However, for most countries only a subset of these measures is available. In the present study the gini data was used (not reported gini) and when multiple sources of data was available for a particular country in a particular year, the average of all the available gini inequality was taken. This was done partly because there was no specific reason to choose any specific source of gini over the other and second in literature, unlike in case of D&S, there was no prescribed method to deal with the heterogeneity of the source of data.

III. Income Inequality and Explanatory Variables

A major part of this study is the presence of the conditioning variables. Major studies have shown that conditioning variables play a critical role in determining or unearthing the presence of Kuznets curve. Literature abounds with plethora of independent variables suspected to have influence on income inequality. Kuzents in his seminal paper did not specifically mention any independent variables as such and hence much of the one present in the literature are derived from an intuitive understanding of the way inequality works. However Kuznets did indicate few areas which will affect inequality and one may derive few variables in this regards. For instance he mentions that *'An invariable accompaniment of growth in developed countries is the shift away from agriculture, a process usually referred to as industrialization and urbanization'* and in another place he states that '*particularly so during the periods when industrialization and urbanization were proceeding apace and the urban population was being swelled',* therefore urban population or share of manufacturing sector in labor force may be a plausible independent variable and it has been used in Galbraith and Kum (2004).

One can also use share of labor in agriculture as proxy for degree of agricultural dependence. Some may question this view stating that in current economy beside agriculture there exist two other sectors, service and manufacturing hence lack of agricultural dependence doesn't necessarily imply industrialization as it might be a service driven economy like Hong Kong or Singapore. But the author believes such criticism is misplaced as Kuznets (1955) state in his paper that *'....let there be two sectors: agriculture (A) and all others (B). The basic assumptions used throughout are that the per capita income of sector B (nonagricultural) is always higher than that of sector A......'*. Therefore it is not required that the other sector be manufacturing; so long the per capita income of the non agricultural sector, service or manufacturing, is higher than that of agriculture the forces mentioned by Kuznets curve should still operate. Therefore Kuznets hypothesis should also apply to service driven economy.

We see similar conclusion being reached by Alderson and Nielsen (2002). As they suggest that inequality is attributable to differences in average income between sectors, which are called sector dualism and that sector dualism, is a function of the difference in average income between sectors and the relative size of the sectors. They use percentage of labor force in agriculture as a conditioning variable to capture this affect. In this paper we will also use the same variable. Data on the total agricultural labor force are estimated by FAO based on the close relationship existing between the ratio of economically active population in agriculture to the total economically active population and the ratio of agricultural population to total population. Annual figures are obtained through interpolating and extrapolating from the ILO decennial series[6].

Since EHII dataset is developed with ratio of manufacturing employment to population as a conditioning variable, it is very likely that multicollinearity will exist if we use percentage of labor force in agriculture as the other conditioning variable. Therefore incase of EHII dataset we will use share of urban population instead of labor force in agriculture in order to capture the sector dualism, or industrialization as such. Estimates of the proportion of the population living in urban areas are obtained from national sources, such as censuses or population registers. Variations between countries make it nearly impossible to adopt uniform criteria for distinguishing urban from rural areas. As such, national statistical offices are often in the best position to establish appropriate criteria to characterize urban areas in their respective countries. [7]

The relation between education and inequality still remains much unexplored. One of the primary reasons for this is the lack of global dataset. However, recently Barro and Lee (1993, 1995, 2001) have developed a global dataset for International measures of schooling years and schooling quality. In Alderson and Nielsen (2002), Gregorio and Lee (1999) they use secondary school enrollment ratio as a conditioning variable while analyzing the interrelationship between education and inequality. Also Barro and Lee (2001) themselves

[6] For more information, please see the FAO's Annual Series of Demographic Estimates explanatory notes.

[7] For further information on country-specific definitions of urban areas, please refer to the data sources listed in World Urbanization Prospects: The 2005 Revision.

suggest that the over-15 age group corresponds better to the labor force for many developing countries. The data set comprises at least one observation for 142 economies, of which 109 have complete information at five-year intervals from 1960 to 2000. Since the dataset are on a five year interval, linear interpolation was used to fill in the gap in the years in-between. Although this might result in serial correlation, since analysis will also be done on five year average dataset as well, this criticism may not be that severe. In this paper Percentage of "secondary school complete" in the total population was used as proxy for educational attainment, inline with current literature. However in this paper we will not look in to the relationship between educational inequality and income inequality although recent panel educational inequality dataset developed by Castelló and Doménech (2001) makes it possible. The author believes that further research can be carried out in this arena.

Demography has also been shown to have implications when it comes to inequality. Higgins and Williamson (2002) used cohort size as a conditioning variable and found that it has a significant impact on inequality. The cohort-size hypothesis is simple enough: fat cohorts tend to get low rewards. When those fat cohorts lie in the middle of the age-earnings curve, where life-cycle income is highest, this labor market glut lowers their income, thus tending to flatten the age earnings curve. Earnings inequality is moderated. When instead the fat cohorts are young or old adults, this kind of labor market glut lowers incomes at the two tails of the age-earnings curve, thus tending to heighten the slope of the upside and the downside of the age-earnings curve. Earnings inequality is augmented. In their paper they used the variable MATURE to capture this effect and the variable was defined to be the proportion of the adult population who are 40-59. In the current paper the variable is further segregated to include the proportion of adult female and male population who are 40-59, thus bringing an additional gender dimension to the study. The data is taken from United Nations Statistics Division, UNSD, Demographic Statistics.

Another variable that finds itself much in the spot light is the ‘trade openness’ of a country. Much research has been done in trying to ascertain the impact of openness on inequality. It is customarily believed that globalization has resulted in greater economic integrations between nations and that trade flow has significantly increased ergo trade openness of countries. However the impact of trade openness on poverty and income inequality remains much contestable area. At the same time, increasing opportunities to trade are likely to affect income distribution and whether or not increasing openness to trade is accompanied by a reduction or an increase inequality is highly controversial. The usual hypothesis is developing countries have an abundant supply of unskilled labor relative to skilled labor and developed countries have an abundant supply of skilled labor relative to unskilled labor. Hence increased openness in developing countries is assumed to boost the relative demand for skilled labor which in turn will increase overall inequality, all else being equal. However results remain much less clear cut. In the following page there is a snapshot of major studies done so far on inequality and openness relationship. What becomes clear is that results are far from conclusive. In this paper we use total trade as a percentage of GDP as defined by openk variable in Penn World Table 6.2.

Table 2: Studies on openness and inequality[8]

Study on Gini	Sample	Measure of openness	Effect of openness on inequality
Edwards 1997	43 countries in 1970 and 1980 by decade averages First difference	Tariffs, Sachs - Warner, Adjusted Trade	=0 for developed countries =0 for developing countries
Savvides 1998	34 countries on 1978-1994 in two periods First difference	Tariffs and NTBs, Sachs -Warner	=0 for developed countries >0 for developing countries
Li, Squire and Zou 1998	49 countries on 1960-1990 5 years period average OLS	X/GDP	=0
Higgins & Williamson 1999	85 countries on 1960-1990 Decades averages OLS and Fixed Effect	Tariffs, NTBs, Sachs-Warner, Adjusted Trade	<0 for developed countries in OLS <0 for developing countries in OLS =0 for developed countries in FE =0 for developing countries in FE
Barro 2000	84 countries on 1960-1990 OLS and Fixed Effect	Adjusted Trade	<0 for developed countries in OLS >0 for developing countries in OLS >0 for countries in FE
Calderon and Chong 2001	102 countries on 1960-1995 5 years period average GMM	Trade to Gdp ratio, Sachs-Warner,	<0 for developing countries =0 for developed countries
Ravallion 2001	50 countries on 1947-1994 5 years period average OLS	X/GDP	<0 for developed countries >0 for developing countries
Rama 2001	97 countries on 1960-1990 period average OLS	X+M/PIB	>0 for countries <0 for skill intensive countries
Dollar and Kraay 2002	92 countries on 1950-1999 Fixed Effect	Trade to Gdp ratio, Adjusted Trade, Sachs-Warner, Tax on imports	=0 for developed countries =0 for developing countries
Milanovic 2002	83 countries in 1988, 1993 and 1998 OLS and GMM	Trade to Gdp ratio	>0 for poor countries <0 for middle income countries
Lundberg et Squire 2003	38 countries on 1960-1994 5 years period average OLS and TSLS	Trade to Gdp ratio, Sachs-Warner	>0

8 Gourdon, Julien *Openness and Inequality in Developing Countries: A New Look at the Evidence*, MPRA Paper No. 4176, posted 07. November 2007

The final variable is the standard real GDP per capita/worker which is the most widely used indicator for capturing the Kuznets effect Deininger and Squire (1998), Barro (2000), RAM (1991, 1997), Frazer (2006), Higgins and Williamson (2002) etc . Although whether to use its level or log form remains to be decided. In this paper we will use real GDP per capita/worker (in constant 1996 US$) Penn World Tables dataset, version 6.2. In some paper, particularly Higgins and Williamson (2002) use per worker instead of the usual per capita and hence both will be used in this study.

In the literature we also find other interesting variables which have been shown to influence income inequality. For instance in Bahmani-Oskooee, Goswami, Mebratu (2006) it was shown that income inequality is higher in countries that have black market for foreign exchange. While in Chong (2004) it was found that democracy has non-monotonic link with income inequality. However in this paper we will restricts ourselves to conditioning variables discussed in Higgins and Williamson (2002), since the paper was found to be significantly broad in scope in terms of its coverage of conditioning variables. In some cases slight variation will be used, for instance we will employ gender segregated matured cohort size instead of the mature adult cohort size, which was not gender segregated , as was used in Higgins and Williamson (2002). In case of EHII we will use share of urban population, which was not mentioned in Higgins and Williamson (2002) but this is done to avoid multicollinearity issue. But nonetheless the current paper draws heavily, in terms of choice of explanatory variables, from Higgins and Williamson (2002) specifically their extended regression model.

But we will not employ financial depth and political freedom variable, as it was found to be insignificant in Higgins and Williamson (2002).Therefore in this study the independent/conditioning variables that will be included are: real GDP per capita/worker, mature, trade openness, secondary school enrollment ratio, percentage of labor force in agriculture or proportion of urban population. In the following page it is given in tabular form:

Table 3: Relevant Variables and Labels

Variables	Label
gdp_pc	Real GDP per capita
gdp_wc	Real GDP per worker
recgdp_pc	Reciprocal of Real GDP per capita
recgdp_wc	Reciprocal of Real GDP per worker
lngdpc	Ln of Real GDP per capita
lngdpc2	Square of Ln of Real GDP per capita
lngdpw	Ln of Real GDP per worker
lngdpw2	Square of Ln of Real GDP per worker
openk	Trade Openness
labor_agri	Percentage of labor force in agriculture
U_Pop	Share of urban population
Edu_Sec_15	Percentage of secondary school complete
lnedu	Ln of Percentage of secondary school complete
male4059	Proportion of the male population who are 40-59
female4059	Proportion of the female population who are 40-59
wiid2	World Income Inequality dataset
ehii	Estimated Household income inequality dataset
utip	UTIP UNIDO Manufacturing Pay inequality
ds96	Deininger and Squire 1996

IV. Econometric issues Panel Unit root tests

At one time, conventional wisdom was that in order to apply standard inference procedures in such studies, the variables in the system needed to be stationary since the vast majority of econometric theory is built upon the assumption of stationarity. Consequently, for many years econometricians proceeded as if stationarity could be achieved by simply removing deterministic components (e.g., drifts and trends) from the data. However, stationary series should at least have constant unconditional mean and variance over time, a condition which hardly appears to be satisfied in economics, even after removing those deterministic terms. Yule (1926) pointed out that spurious correlation may persist in large sample despite the absence of any connection between the underlying series.

Those problems were somehow ignored in applied work until important papers by Granger and Newbold (1974) and Nelson and Plosser (1982) alerted many to the econometric implications of non-stationarity. It was established that the stationarity or otherwise non stationarity of a series can strongly influence its behavior and properties (e.g. persistence of shocks will be infinite for nonstationary series). *Nonsense* or *spurious* regressions was also possible if the variables used in the analysis were not stationary, Granger and Newbold (1974), Granger (1981). For instance if one were to regress two variables that are trending over time, a regression of one on the other could have a high R^2 even though the two variables might be totally unrelated. It was also shown that if the variables in the regression model are not stationary, then the standard assumptions for asymptotic analysis will not be valid. In other words, the usual "t-ratios" will not follow a t-distribution, so one cannot validly undertake hypothesis testing about the regression parameters. In their seminal paper Engle and Granger (1987) defined a series to be integrated of order one if it became stationary after being differenced and it was denoted I(1). If a series is integrated of degree one then it is also said to poses unit root. In general, a series which is stationary after being differenced d times is said to be integrated of order d and denoted I(d). A series which is stationary without differencing is said to be I(0).

Therefore before we undertake any econometric analysis we must first need to ascertain the order of integration for all the variables, dependent and independent. If all the variables are stationary then we can perform econometric analysis without being concerned about the possibility of spurious regressions and inappropriate standards errors. However, some variables may possess unit root and therefore be mean or variance non stationary. In such cases either we need to difference the variable (mean non stationary) and/or log transformed the series (variance non stationary). In order to address the issue of stationarity of the variables, we undertake panel unit root tests

All the tests are primarily based on the following ADF specification:

$$\Delta X_{it} = \alpha_i + \beta_i t + \rho_i + \varepsilon_{it}$$

Here, $\alpha_i and\ \beta_i$ represents the country specific fixed effects and unit specific linear time trends respectively. For LLC and BR tests, it is assumed that $\rho_i = \rho$. Hence, the null hypothesis of a unit root translates to $\rho = 0$. The IPS, PP and ADF tests allow the autoregressive coefficient to vary across countries which entails the alternative hypothesis as $H_a\!: \rho_i < 0\ for\ i = 1 \dots K\ and\ \rho_i = 0\ for\ i = k \dots N$. Therefore, the reported t-statistic is the sample-weighted average of the t-statistics for the individual countries.

The following table summarizes the findings of the panel unit root tests.

Table 4: Panel Unit Root test

	Assumes Common Unit root process		Assumes Individual Unit root process		
Variables	LLC	BR	IPS	ADF	PP
gdp_pc	8.14290 1.00	-4.19777*** 0.00	6.04848 1.0000	422.247** 0.03	368.892 0.4771
gdp_wc	5.11711 1.00	-3.46151*** 0.00	3.58881 1.00	397.594*** 0.02	406.697** 0.01
recgdp_pc	-24.7066*** 0.00	3.32759 1.00	-10.9679*** 0.00	926.450*** 0.00	1085.24*** 0.00
recgdp_wc	-24.1748*** 0.00	3.31715 1.00	-11.4652*** 0.00	919.625*** 0.00	1062.57*** 0.00
lngdpc	-8.7147*** 0.00	1.16797 0.88	-1.97892** 0.02	531.808*** 0.00	511.570*** 0.00
lngdpc2	-6.53514*** 0.00	0.53263 0.70	-0.88189 0.19	503.329*** 0.00	472.16*** 0.00
lngdpw	-9.31405*** 0.00	1.10156 0.86	-3.74212*** 0.00	556.554*** 0.00	555.671*** 0.00
lngdpw2	-7.74178*** 0.00	0.63578 0.74	-2.67462*** 0.00	519.968*** 0.00	523.791*** 0.00
openk	11.6819 1.00	-6.45575*** 0.00	4.17227 1.00	460.190*** 0.00	438.624*** 0.00
labor_agri	-12.4096*** 0.00	-0.84754 0.20	0.40563 0.66	739.324*** 0.00	1395.59*** 0.00
U_Pop	-1.18706 0.12	-1.03590 0.15	5.96336 1.00	480.972*** 0.00	1360.99*** 0.00
Edu_Sec_15	-2.26410** 0.01	0.44246 0.67	-0.09550 0.46	207.711 0.24	154.233 1.00
lnedu	-19.7140*** 0.00	0.91261 0.82	-7.44115*** 0.00	375.985*** 0.00	611.797*** 0.00
male4059	-5.29458*** 0.00	0.29392 0.62	4.05010 1.00	414.592** 0.02	232.989 1.00
female4059	-3.90801*** 0.00	-0.19542 0.42	3.79569 1.00	429.598*** 0.00	233.254 1.0000
wiid2	-18.2890*** 0.00	-3.71722*** 0.00	-11.3200*** 0.00	237.169*** 0.00	302.776*** 0.00
ehii	-3.25311*** 0.00	0.28422 0.61	0.50871 0.69	300.238** 0.04	339.427*** 0.00
utip	-14.5080*** 0.00	-2.99645*** 0.00	-0.90191 0.18	368.280*** 0.00	379.442*** 0.00
ds96	1.02856 0.85	0.14671 0.56	-0.19307 0.42	41.4044* 0.08	77.2453*** 0.00

From the above table one thing becomes very clear and that is in case of real GDP per capita and real GDP per worker, the log transformed performs much better than level form. Although we see in both cases the reciprocal of level form performs very well in the unit root test. However we must remember that in order to avoid spurious regression, all variables ought to be stationary and so even though the reciprocal of Real GDP per worker/capita is stationary, since Real GDP per worker/capita itself is not stationary hence the functional form their functional form used by Deininger and Squire (1998) (namely real GDP per capita and 1/(real GDP per capita)) is not justified and neither is the form real GDP per capita and square of real GDP per capita, as used by others. Unless of course one can show that a cointegrating relationship exists between these variables and the inequality variable, in which case spurious relationship can be avoided. Since the log transformed form of both real GDP per capita and real GDP per worker, along with their quadratic form, perform consistently better than their level form, we will use them in this paper, in line with Barro (2000).

Other than the real GDP per capita, the rest of variables give a mixed result but in most cases 3 of the tests at least show the other variables to be stationary. Only in case of Percentage of secondary school complete, Edu_Sec_15, do we see gross violation of unit root test. However we see that the log transformed form performs much superiorly. Hence we can see that in its level form, like real GDP per capita and real GDP per worker, the variable is variance non-stationary. In this paper we are going to use the log transformed form.

Since the objective of the study is testing the robustness of the Kuznets hypothesis, much greater emphasis will be given on the real GDP per capita and real GDP per worker. Therefore in reference to conditioning variables, even at the chance of receiving criticism, the author believes that the variables are satisfactorily stationary and hence no further transformation will be carried out. Thus the functional form we end at is

$$\mathbf{INEQ_{it} = \alpha_i + \beta_1 (\ln Y_{it}) + \beta_2 (\ln Y_{it})^2 + \beta_i X_{it} + \varepsilon_{it}} \qquad (1).$$

Where $\mathbf{INEQ_{it}}$ is the inequality measure (WIID2, UTIP, EHII, DS) , $\boldsymbol{\alpha_i}$ is the country specific fixed effect[9] , $\mathbf{\ln Y_{it}}$ is Real GDP per capita or worker and $\mathbf{X_{it}}$ is the constellation of conditioning variables, namely Trade Openness, Percentage of labor force in agriculture or Share of urban population, Ln of Percentage of secondary school complete, Proportion of the male population who are 40-59 and Proportion of the female population who are 40-59.

[9] Most research done after early 90s use Fixed effect modeling, Deininger and Squire (1998), Ram (1991 1997), Barro (2000), Higgins Williamson (2002) etc . LM test and Hausman test confirm the assertion also but they are not shown in the paper.

V. Alternative Estimation of Inequality Relationship

Previous studies on inequality and development studies used crosssection or pooled datasets. Naturally, what we want to understand is how inequality changes over time, or with level of development, within a country, and yet, because of previous data limitations, the empirical tests were forced to draw conclusions largely from cross-sectional (or pooled) datasets. But with the advent of D&S, 1996 this problem has been significantly mitigated, if not completely so. In this paper we will use D&S, 1996 inequality dataset, along with three other datasets[10] , which also have panel structure. In line with existing literature[11] we will initially undertake pooled regression on the three types of dataset, namely Annualized global dataset of 188 countries, 4 year average of global dataset and OECD datasets. This is the simplest form of analysis and after that we will undertake specification tests to see whether pooled or random effect or fixed effect modeling is appropriate. We will also try to refine our modeling to ensure that there is no misspecification error or serial correlation. Therefore pooled regression is done to enhance the comparability of present research with earlier research.

- <u>Pooled Regression</u>

In this section we will try to develop the econometric model in order to investigate the shape and existence of Kuznets curve using the explanatory variables mentioned in the previous section. We will run regression on the annualized global dataset, 4 years average dataset and finally on the OECD section of the annualized dataset, as much research has been done on investigating inequality relationship for OECD countries. The following table gives the result of pooled regression for the three types of dataset and the 4 different inequality measures.

[10] World income inequality database (WIID2), UTIP UNIDO Manufacturing Pay inequality dataset (UTIP), and Estimated Household Income Inequality Data Set (EHII)
[11] Deininger and Squire (1998), Galbraith and Kum (2004)

Table 5: Pooled regression on annualized dataset, all countries

	Dependent Variable				
Regressor	DS96	WIID2	EHII	UTIP	UTIP-W
c	-86.271*** 0.00	-38.562* 0.08	83.077*** 0.00	0.612* 0.00	0.756* 0.00
lngdpc	26.610*** 0.00	20.645*** 0.00	-5.676*** 0.00	-0.111** 0.00	
lngdpc2	-1.033** 0.01	-0.914*** 0.00	0.328*** 0.00	0.007*** 0.00	
openk	0.0158** 0.05	0.012*** 0.01	0.007*** 0.00	8.14E-05*** 0.00	7.36E-05*** 0.00
labor_agri	0.173*** 0.00	0.001 0.97		-8.70E-05*** 0.39	-0.00036*** 0.00
u_pop			-0.053*** 0.00		
lnedu	-1.346** 0.00	-2.867*** 0.00	-0.324** 0.02	-0.001 0.42	-0.001 0.49
male4059	-3.007*** 0.00	-4.577*** 0.00	-0.634*** 0.01	0.012*** 0.00	0.011*** 0.00
female4059	0.693 0.11	1.773*** 0.00	-0.105 0.42	-0.010*** 0.00	-0.01*** 0.00
lngdpw					-0.123** 0.00
lngdpw2					0.006*** 0.00
Cross section	74	87	93	93	93
N	469	1094	2307	2284	2281
Adj R-squared	0.57	0.49	0.57	0.22	0.22
DW stat	0.52	0.56	0.14	0.33	0.34
F-statistic	89.01	152.26	438.50	92.18	95.47
Prob(F-stat)	0.00	0.00	0.00	0.00	0.00

For Kuznets hypothesis to be true we would expect $\beta_1 > 0$ and $\beta_2 < 0$. From the above table we see that Kuznets curve appear in case of D&S and WIID2 while in case of EHII and UTIP we see the un-inverted U shaped curve. The coefficients are also significant to 1% level. In case of the independent variables we see that openness seems to have a positive relationship with inequality, irrespective of dataset. Proportion of labor in agriculture is rather a troublesome variable. In case of DS96 and WIID2 it has a positive coefficient which is counter intuitive on the ground that one would assume that agriculture has less inequality and hence it should have a negative coefficient. We see in case of UTIP the coefficient value is negative. In case of EHII we used urban population instead of Proportion of labor in agriculture to avoid multicollinearity issue and we see that it has negative coefficient. This is also counter intuitive as increased urbanization should increase inequality and not decrease it. However we see in some cases it is not significant. Once fixed effect model is run, one needs to monitor the effect it has on this variable. In case of Galbraith Kum we see that fixed effect may wash away the significance of urbanization or such variables. Hence further analysis should be deferred till that time. Education variable is negative in all cases, although in case of UTIP it is not significant. It is in line with literature as education is considered to be one of the key leveling factors which reduce inequality.

Gender segregated cohort size show a very interesting result. In case of D&S and WIID2, male cohort has a negative coefficient while positive for female. In accordance with Williamson Higgins (2002) finding large mature working-age cohorts are associated with lower aggregate inequality, which seems to hold for male cohort size and goes in opposite direction when it comes to females' cohort size. It is very interesting as one over the last 20-30 years female participation in the labor force has increased significantly. Hence in coming decade, worldwide, there will be a significant proportion of matured female labor force. In EHII both coefficient are negative but in case of female it is insignificant. But in case of UTIP the signs are reversed with Male cohort size being positive and negative in case of female. It could be attributed to the fact that the dynamics of inequality within manufacturing sector may differ from that of overall country inequality dynamics but it is an area which surely needs further investigation.

The pooled regression has adjusted R-squared running between 22% to 57%, which suggest that although the aforesaid variables are very important, there are individual heterogeneity that cannot be captured by a single intercept as is done in case of pooled regression. Also we see that there is high auto correlation, with such low DW statistics. In the following table we will see the result of pooled regression on OECD countries and 4 year average dataset.

Table 6: Pooled regression on Average dataset

	Dependent Variable			
Regressor	**WIID2**	**EHII**	**UTIP**	**UTIP-W**
c	-23.519 0.44	85.614*** 0	0.611*** 0	0.781*** 0
lngdpc	17.789*** 0.01	-6.222** 0.05	-0.108*** 0	
lngdpc2	-0.790** 0.05	0.363* 0.06	0.006*** 0	
openk	0.019** 0.02	0.008** 0.04	0.0001*** 0.01	0.0001*** 0.01
labor_agri	0.006 0.88		0.0002 0.4	-0.0005** 0.02
u_pop		-0.052*** 0		
lnedu	-3.003*** 0	-0.266 0.28	-0.001 0.71	-0.0006 0.82
male4059	-2.752*** 0	-0.618 0.16	0.013*** 0	0.015*** 0
female4059	0.745 0.14	-0.133 0.59	-0.010*** 0	-0.0113*** 0
lngdpw				-0.122*** 0
lngdpw2				0.006*** 0
Cross section	88	93	93	93
N	492	646	640	640
Adj R-squared	0.48	0.57	0.21	0.22
DW stat	0.59	0.3	0.5	0.5
F-statistic	65.53	124.62	25.47	27.21
Prob(F-stat)	0	0	0	0

Due to lack of data points after averaging for 4 years, it was not possible to carryout regression for D&S dataset. What becomes clear from above tables is that the Kuznets curve still does not appear for UTIP or EHII dataset while it is clearly seen in case of WIID2 and D&S (OECD). This is an important finding because in Galbraith and Kum (2004) it is mentioned that *'For the OECD countries (Western Europe and North America) where the direct measurement of household income inequality is likely to be most advanced and most consistent, there is not much systematic divergence between the two data sets'*- EHII and D&S. this might be true in case of descriptive statistics but in case regression analysis the results are still divergent. Similarly results are divergent with WIID2 dataset which is a much improved version of D&S 1996. Further analysis on this disparity will be carried out in the later part. The coefficient values of the quadratic and the linear term of per capita income significantly increases for all datasets.

In terms of independent variables we see divergence emerge among the findings of dataset for OECD countries. In case of openness, there is strong negative relationship with inequality in case of WIID2 and D&S dataset, while it is just the opposite in case of EHII and UTIP. In case of labor participation in agriculture, the coefficient value remains positive for all dataset. It might be hypothesized that because of the modernization and mechanization of agriculture in OECD countries the increase in labor force participation in the sector may actually increase inequality. Also the sector may not be as unionized as the manufacturing sector and hence leveling effect may be missing.

Table 7: Pooled regression on OECD dataset

	Dependent Variable				
Regressor	**DS96**	**WIID2**	**EHII**	**UTIP**	**UTIP-W**
c	-84.798	-265.475***	247.760***	0.2511**	0.4171**
	0.14	0	0	0.03	0.02
lngdpc	13.7518	57.819***	-44.899***	-0.0659***	
	0.26	0	0	0.01	
lngdpc2	-0.1129	-2.633***	2.455***	0.0044***	
	0.87	0	0	0	
openk	-0.0552***	-0.058***	0.011**	0.0001***	0.0001***
	0	0	0.04	0	0
labor_agri	0.4556***	0.320***		0.0008***	0.0007***
	0	0		0	0
u_pop			-0.086***		
			0		
lnedu	0.6352	0.165***	-0.835***	-0.0046***	-0.0039***
	0.17	0.74	0	0	0
male4059	-1.5573***	-2.980***	-1.528***	-0.0049***	-0.0055***
	0.01	0	0	0	0
female4059	0.6517*	1.339***	0.847***	0.0027***	0.0031***
	0.07	0	0	0	0
lngdpw					-0.0954***
					0.01
lngdpw2					0.0055***
					0
Cross section	21	23	22	23	23
N	238	495	744	748	748
Adj R-squared	0.45	0.26	0.24	0.21	0.2
DW stat	0.17	0.53	0.05	0.17	0.17
F-statistic	28.83	25.23	35.2	28.8	27.74
Prob(F-stat)	0	0	0	0	0

Similarly in case of urbanization we see a negative sign which is also difficult to explain as with a large urban population and an un-inverted U shape relationship between per capita income and inequality, the only way one can explain the finding is if one were to assume that inequality in developed countries is higher in areas outside urban locale. Now this may be due to the fact that most of the factories/manufacturing units are outside urban areas due to high real estate cost, hence urbanization may not necessarily imply industrialization. It may also mean that in developed countries agriculture has higher inequality. This may explain the positive relationship between agricultural labor participation, urbanization and inequality.

In case of cohort size we see a consensus between all datasets and the gender dimension becomes even more pronounced. As the size of matures female working population increases inequality decreases and vise versa for male. This may actually stem from the fact that incase of female the wage/income differential is not as high as in the case for male. This egalitarian income differentiation among female may actually be a result of gender discrimination rather than homogenous skill sets among female which thus fetches similar wages/income. Since the current matured population entered the labor force when gender discrimination may still have been prevalent, it is very likely that employment opportunity for woman back then were less in comparison to male. Hence wage/income for the current female population is similar and hence the coefficient value is negative. If this is indeed the case then in future this difference between male and female mature working population will diminish and may even disappear. In case of OECD countries the sign reverses and it seems now increasing size of female population increases inequality while decreases it in the case of male. Could it be due to the fact that OECD countries have tried to mainstream gender and in the process have selectively focused on high skilled female labors in order to set up example for future generation or could it be that the limited opportunities available to female covered opposite areas of the income spectrum. Hence a female could either enter as secretary or business executive resulting in current high inequality among current female matured workforce. This area requires further investigation and may provide some interesting findings in future research.

There seems to be broadly an agreement between the coefficient signs and in most cases the values between the finding of Pooled regression on annualized and 4 year average dataset. Adjusted R-squared value does not improve either. The disparity in finding for different datasets in reference to cohort size and education still holds. Counter intuitive finding in terms of labor participation in agriculture and urbanization is still prevalent. The only significant difference arises in the DW statistics which improves slightly, indicating that serial correlation may not necessarily stem only from the linear interpolation used in case of some independent variables. Hence different econometric method must be employed to remove the problem of serial correlation, dynamic panel or inclusion of AR1 may be used.

In case of UTIP in one case we use per capita income and its square, while in the other case, UTIP-W, we used per worker and its square. From the above tables we see that the sign and the coefficient value for the per capita and per worker do not change significantly to warrant running separate set of regressions for per worker. The sign of the coefficient and values for independent variables also do not change much. Therefore in the following sections we will limit our study to per capita income, excluding per worker, and will only focus on annualized and OECD datasets, excluding average dataset.

- Fixed effect Regression

After running pooled regression, two formal specification tests are performed. One is Breusch and Pagan's LM test (1980), to see the relevance of random-effects specification; If the test statistic, based on chi square distribution, rejects the null hypothesis (which it does in this case), then a random effects model is regarded as preferable. The other test is a Hausman test for specification (1978). The null hypothesis in this test is that country-specific effects are not correlated with any regressor in the model equation, implying that the estimates are efficient. If this null is rejected, the random effects model estimates are inconsistent and fixed effects model specification would be preferred. Test results show that a random-effects model provides inconsistent estimates in equation (1).

Based on these test results, the estimates from fixed-effects model appear more robust in present circumstances. Two way fixed effect is not to be performed here as our objective is to precisely capture the time element and the possibility of quadratic relationship between per capita income and inequality over time; time specific dummies in two way effect will wash away any such relationship and therefore will defeat the objective of our research. In this section we will initially test unconditional Kuznets hypothesis, before proceeding in to testing the conditional Kuznets hypothesis with all the explanatory variables.

Table 8: Fixed effect regression for Unconditional Kuznets curve on Annualized all countries

Global Dataset								
	Dependent Variable							
Regressor	DS96		WIID2		EHII		UTIP	
		Prob		Prob		Prob		Prob
c	18.389	0.42	-12.996	0.50	87.986	0	-0.126	0.28
lngdpc	4.596	0.39	13.799	0.00	-10.622	0	0.043	0.11
lngdpc2	-0.277	0.38	-0.881	0.00	0.603	0	-0.002	0.11
Cross section	107		139		145		146	
N	575		1442		2892		2901	
Adj R-squared	0.9		0.78		0.86		0.62	
DW stat	0.93		0.9		0.42		0.91	
F-statistic	51.4		37.99		121.67		32.59	
Prob(F-stat)	0		0		0		0	

In the above table we see that the difference between WIID2 and EHII continues, with Kuznets hypothesis being confirmed in case of first measure of inequality while being rejected in case of the second. In case of D&S and UTIP although both confirm Kuznets hypothesis however the coefficient values are not significant. The following table shows the result for OECD countries.

Table 9: Fixed effect regression for Unconditional Kuznets curve on OECD countries

OECD								
	Dependent Variable							
Regressor	DS96		WIID2		EHII		UTIP	
		Prob		Prob		Prob		Prob
C	10.556	0.82	-125.642	0.02	309.048	0	0.658	0
Lngdpc	4.162	0.68	39.944	0	-60.931	0	-0.143	0
lngdpc2	-0.199	0.72	-2.430	0	3.365	0	0.008	0
Cross section	24		26		25		26	
N	242		525		784		793	
Adj R-squared	0.72		0.55		0.79		0.53	
DW stat	0.69		0.7		0.18		0.29	
F-statistic	26.23		24.87		111.39		33.56	
Prob(F-stat)	0		0		0		0	

In case of OECD dataset we see that the difference in finding continues to persist between WIID2 inequality measure and EHII measure. However in case of OECD countries, coefficient values become significant for both UTIP and D&S but the difference remains, with D&S confirming Kuznets curve while UTIP rejects it. The finding are in consensus with Munir and Muaz (2004) where an un-inverted U shaped curve was found, while for testing unconditional Kuznets curve, in case of both UTIP and EHII dataset. Munir and Muaz (2004) carried out the study on a balanced panel of 24 countries. In the following page we run fixed effect model on conditional Kuznets curve to further analyze the issue.

The fixed effect modeling on conditional Kuznets curve, still does not bring any consensus among the findings. Under both WIID2 and D&S dataset we see the Kuznets inverted U curve but un-inverted curve for EHII and UTIP. This is applicable for both OECD and annualized Global Dataset. In case of OECD, the coefficient value of the linear and quadratic term of per capita income for D&S and EHII are almost equivalent in value but exactly in opposite in 'sign', showing the stark difference between the findings of the two sets of dataset. So instead of convergence of findings between the datasets in case of OECD countries, we see that the divergence actually becomes more pronounced, a finding very much contrary to usual interpretation given to the heterogeneity of D&S data. Literature usually suggest that because D&S includes data that are derived through various means (consumption, household, personal, income, expenditure etc), intra country comparability is near to impossible. Hence EHII is constructed by taking these differences in to account through use of dummy variables for source of data. Galbraith Kum suggests that disparity should decrease between D&S and EHII for OECD countries, as method of collection among OECD countries should be more or less consistent. However the above result suggests that this is not to the case and as a matter of the fact the results are just polar opposite, which is indeed a matter of concern.

Table 10: Fixed effect regression on Annualized all countries and OECD dataset

	Global Dataset								OECD							
	Dependent Variable								Dependent Variable							
Regressor	DS96		WIID2		EHII		UTIP		DS96		WIID2		EHII		UTIP	
		Prob		Prob		Prob		Prob		Prob		Prob		Prob		Prob
c	-75.098	0.01	-79.475	0.00	129.777	0.00	0.399	0.00	-292.041	0.00	-397.167	0.00	293.272	0.00	0.705	0.00
lngdpc	18.453	0.01	27.286	0.00	-20.276	0.00	-0.050	0.11	55.630	0.00	85.738	0.00	-55.650	0.00	-0.139	0.00
lngdpc2	-0.822	0.03	-1.738	0.00	1.027	0.00	0.001	0.56	-2.493	0.00	-4.554	0.00	2.904	0.00	0.007	0.00
openk	0.042	0.00	-0.014	0.17	-0.010	0.00	0.000	0.00	0.079	0.05	-0.075	0.00	0.093	0.00	0.000	0.00
labor_agri	0.291	0.00	0.054	0.17			-0.001	0.00	0.754	0.00	0.499	0.00			0.000	0.07
u_pop					0.088	0.00							0.027	0.43		
lnedu	-0.185	0.66	-0.312	0.48	0.732	0.00	0.010	0.00	0.170	0.77	1.822	0.01	-0.604	0.00	-0.005	0.00
male4059	-0.249	0.67	-1.986	0.00	-0.192	0.37	-0.001	0.64	0.105	0.87	-1.681	0.01	-0.275	0.15	-0.001	0.34
female4059	0.263	0.44	1.567	0.00	0.245	0.04	0.002	0.27	0.167	0.65	1.470	0.00	0.232	0.03	0.001	0.32
Cross section	74		87		93		93		21		23		22		23	
N	469		1094		2307		2284		238		495		744		748	
Adj R-squared	0.92		0.80		0.86		0.61		0.78		0.57		0.83		0.59	
DW stat	0.90		0.94		0.36		0.62		0.80		0.82		0.23		0.34	
F-statistic	68.03		49.47		136.37		37.06		31.23		23.48		134.59		38.08	
Prob(F-stat)	0.00		0.00		0.00		0.00		0.00		0.00		0.00		0.00	

In case of openness the results are rather mixed and are no longer as straightforward as before. To begin with we see that in case of D&S for Global Dataset the coefficient is positive, insignificant for WIID2 and negative for EHII . But in case of OECD we see that there is a disagreement between WIID2 and D&S dataset, which may be of interest. For labor participation in agriculture the result remains consistent with the previous counter intuitive result. Education is significant mostly for EHII and UTIP dataset. In the annualized dataset it has a positive coefficient while in case of OECD it is negative. One may attribute this to the fact that annualized Global Dataset, especially EHII and UTIP, has a greater representation of developing LDC countries and in these countries return to education may be very high. Hence at the initial stages of development higher secondary level education attainment may increase inequality as the skill set demanded in the developing labor market may not be that high. In OECD countries education may play the role of leveling effect and return to education might be lower, which reduces inequality. In case of cohort size the gender dimension with negative sign for male cohort still persists although in some cases, especially in OECD dataset, it looses significance. But this time we see across both OECD and Global Dataset that mature male population tend to reduce inequality while female tend to increase it.

- Fixed effect Regression with AR1 error

It is also seen that there is significant improvement in Adjusted R-squared value and DW statistics for both regressions under both dataset. However this is very likely the result of country specific constant which is boosting the explanatory power of the regression without enhancing the interpretive power of the regression in any significant way. There is still serial correlation as the DW statistics is still very low. In equation (1) the error term (ε_{it}) is naively supposed to be white noise, satisfying the standard I.I.D.~$(0,\sigma^2)$ assumption. However, his is not so reasonable in longitudinal data. If the assumption of zero serial correlation is not correct, then standard errors of the estimates are biased, leading to biased test statistics. Autoregressive specification, usually AR (1), is recommended to cope with this problem. We apply the AR(1) procedure to fixed effects models following Baltagi and Wu's method

(1999), which can deal with unbalanced panel structure of our data. Then the equation (1) is modified as

$$\mathbf{INEQ_{it} = \alpha_i + \beta_1 (ln Y_{it}) + \beta_2 (ln Y_{it})^2 + \beta_i X_{it} + \varepsilon_{it}} \quad (2)$$

Where $\varepsilon_{it} = \rho\varepsilon_{it-1} + \eta_{it}$

where ρ is a correlation coefficient among (εit , $\varepsilon it\text{-}1$) and η_{it} is again conventional white noise satisfying the I.I.D.~$(0,\sigma^2)$ assumption.

Based on the aforesaid discussion we incorporated AR1 and tested for unconditional Kuznets curve. The results are given in Appendix 2. The findings are more or less consistent with previous findings, with D&S and WIID2 confirming Kuznets curve while EHII and UTIP rejecting it. In case of OECD countries the difference is even more significant. This one again validates finding made by Munir and Muaz (2004). We then incorporate AR1 to our conditional Kuznets curve framework. In the following we show the findings the augmented fixed effect equation , in case of OECD and standard annualized dataset.

Table 11: Fixed effect regression on annualized all countries and OECD dataset with AR1

	Global Dataset				OECD			
	Dependent Variable				Dependent Variable			
Regressor	**DS96**	**WIID2**	**EHII**	**UTIP**	**DS96**	**WIID2**	**EHII**	**UTIP**
C	-99.572*** 0.01	-89.841*** 0.00	135.737*** 0.00	0.3379*** 0.00	-256.428*** 0.01	-445.902*** 0.00	308.995*** 0.00	0.8923*** 0.00
Lngdpc	20.395*** 0.01	28.306*** 0.00	-21.766*** 0.00	-0.0334 0.16	47.056*** 0.01	94.081*** 0.00	-59.214*** 0.00	-0.1778*** 0.00
Lngdpc2	-0.797** 0.05	-1.813*** 0.00	1.108*** 0.00	-0.0001 0.94	-2.010** 0.03	-4.984*** 0.00	3.111*** 0.00	0.0091*** 0.00
Openk	0.077*** 0.01	-0.020*** 0.09	-0.010*** 0.00	-0.0001*** 0.00	0.041 0.44	-0.080*** 0.00	0.085*** 0.00	0.0004*** 0.00
Labor_Agri	0.410*** 0.00	0.124*** 0.02		-0.0011*** 0.00	0.727*** 0.00	0.627*** 0.00		-0.0005*** 0.00
U_Pop			0.093*** 0.00				0.021 0.22	
Lnedu	-0.028 0.96	0.272*** 0.63	0.785*** 0.00	0.0107*** 0.00	1.678* 0.10	2.919*** 0.00	-0.651*** 0.00	-0.0069*** 0.00
Male4059	-0.488 0.44	-2.314*** 0.00	-0.049*** 0.69	0.0004 0.83	-0.883 0.32	-1.903*** 0.00	-0.110 0.22	-0.0012*** 0.08
Female4059	0.352 0.32	1.847*** 0.00	0.174*** 0.01	0.0009 0.43	0.741 0.13	1.707*** 0.00	0.147*** 0.00	0.0007*** 0.06
Ar1	0.544*** 0.00	0.499*** 0.00	0.852*** 0.00	0.7217*** 0.00	0.615*** 0.00	0.592*** 0.00	0.918*** 0.00	0.8865*** 0.00
Cross section	31	62	89	89	15	23	22	23
N	217	658	2156	2133	142	366	720	723
Adj R-squared	0.96	0.88	0.96	0.81	0.92	0.75	0.96	0.87
DW stat	1.75	2.12	2.11	2.23	1.65	2.32	2.00	1.96
F-statistic	127.62	69.71	490.10	94.14	77.94	37.10	665.47	165.31
Prob(F-stat)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

In aforesaid table we see that divergence between the findings of the dataset continues. We see the existence of inverted U shaped curve in case of WIID2 and D&S dataset but un inverted U shaped curve in case of EHII dataset. Although in case of standard annualized dataset we see Inverted U shaped curve for UTIP but the coefficient values are insignificant even at 10% confidence level. In case of OECD, as before the divergence is even more pronounced, D&S dataset and EHII have exactly opposite signed coefficient value and UTIP once again shows the un –inverted U shaped curve.

In case of openness, although the coefficient value changes somewhat but the sign remains more or less consistent with previous fixed effect regression result. For the standard annualized dataset the signs are positive for D&S dataset but negative for the rest. While in case of OECD, the coefficient value and sign becomes more robust for WIID2 but changes significantly for UTIP and EHII, both in terms of sign and value.

For participation of labor force in agriculture, for both OECD and annualized dataset the coefficient is positive and highly significant. However UTIP shows a negative sign in case of Global Dataset and OECD datasets. So the disparity in findings for the dataset continues. For urbanization, coefficient is positive but insignificant in case of OECD.

The variable of secondary education attainment shows positive coefficient for EHII and UTIP in case of Global Dataset but significant and negative for OECD countries, which may be due to the fact, as mentioned before, that Global Dataset has greater proportion of LDCs and developing countries where the return to education is higher than OECD, hence education tend to increase inequality. For WIID2 and D&S dataset, the variable is insignificant in Global Dataset but becomes significant and positive in case of OECD countries and this finding obviously puts to question the aforementioned reason for negative sign in case of EHII and UTIP dataset in OECD countries. Which result is valid can only be answered if one can suggest or choose a dataset over the other thereby invalidating the findings of the other dataset

For Cohort size we see that Female factor remains significant in case of WIID2 for both standard and OECD dataset and it is positive, while male cohort size is negative and is also significant. Female cohort size is significant in case of EHII for both dataset and is also positive. But one thing that becomes clear is that after the inclusion of AR1, which is significant at 1% level in all cases, apart from WIID2 in all other cases it is negative but insignificant[12]. The DW statistics shows significant improvement after the addition of AR1 term, as was expected.

[12] For UTIP it is significant at 10% confidence level for OECD dataset

However serial correlation may arise in residuals (ε_{it}) from another source, that is, from some influence of omitted lagged dependent variables, then not only could standard errors of the estimates but also coefficient estimates be biased. This is a plausible suspicion, because the previous year's inequality could have some persistency in determining the current year's inequality. If this were the case, the previous remedy focused on only the error term would not generate a reliable result. To address this problem, a lagged dependent variable (LDV) specification is adopted. Then equation (1) can be modified as

$$\mathbf{INEQ_{it} = \alpha_i + \gamma_1 * INEQ_{i(t-1)} + \beta_1 (lnY_{it}) + \beta_2 (ln\,Y_{it})^2 + \beta_i X_{it} + \varepsilon_{it}} \tag{3}$$

estimates, the lagged dependent variable [$INEQ_{i(t-1)}$] should not be correlated with current error term: $E(INEQ_{i(t-1)}, \varepsilon_{it}) = 0$ and the time dimension (t) should be expanded to infinity, which is particularly not feasible in this study. To deal with this problem the popular method suggested by Arellano and Bond (1991) is adapted, which corrects the lagged dependent variable bias as well as permits a certain degree of endogeneity in the other regressors. This Generalized Method of Moment (GMM) estimator modifies our model by specifying a first-difference form, eliminating country-specific effects first, and uses the lagged value of each differenced term as instruments. Model (4) can be rewritten as

$$\mathbf{[INEQ_{it} - INEQ_{i(t-1)}] = \gamma_1 * [INEQ_{i(t-1)} - INEQ_{i(t-2)}] + \beta_1 * [LnY_{it} - LnY_{it-1}] + \beta_2 * [(ln\,Y_{it})^2 - (ln\,Y_{it-1})^2] + \beta_i * [X_{it} - X_{it-1}] + [\varepsilon_{it} - \varepsilon_{it-1}]} \tag{4}$$

The result of the estimation is given in the following page for the OECD and annualized Global Dataset.

Table 12: Arellano-Bond on annualized all countries and OECD dataset

	Global Dataset				OECD			
	Dependent Variable				Dependent Variable			
Regressor	DS96	WIID2	EHII	UTIP	DS96	WIID2	EHII	UTIP
Lag	0.586*** 0.00	0.141*** 0.00	0.665*** 0.00	0.5469*** 0.00	0.711*** 0.00	0.405*** 0.00	0.941*** 0.00	0.8741*** 0.00
Lngdpc	15.328** 0.05	36.153*** 0.00	-5.184 0.18	0.1065* 0.06	43.008 0.21	129.269* 0.06	-5.726 0.14	-0.0640** 0.02
lngdpc2	-0.714* 0.09	-2.259*** 0.00	0.237 0.29	-0.0077** 0.02	-1.907 0.26	-6.526* 0.06	0.288 0.15	0.0032** 0.02
Openk	0.033 0.27	0.002 0.87	-0.003 0.30	0.0000 0.63	-0.033 0.62	-0.019 0.52	0.010** 0.03	0.0001*** 0.00
labor_agri	-0.105 0.52	-0.151 0.19		-0.0007*** 0.01	0.470* 0.06	1.330*** 0.00		-0.0002 0.18
U_Pop			0.057*** 0.01				0.015 0.50	
Lnedu	0.693 0.41	-0.692 0.47	0.355* 0.06	0.0055* 0.07	0.791 0.42	0.339 0.77	0.011 0.93	- 0.0023*** 0.01
male4059	-1.637* 0.06	-2.547*** 0.01	-0.145 0.60	0.0035 0.44	0.472 0.63	0.424 0.74	0.302** 0.02	0.0008 0.42
female4059	0.802* 0.07	1.798*** 0.00	0.196 0.18	-0.0006 0.81	-0.082 0.88	0.425 0.51	-0.126* 0.06	-0.0004 0.46
C	-61.679* 0.08	-116.003*** 0.00	33.744** 0.05	-0.3422 0.18	-240.379 0.17	-659.198** 0.05	27.228 0.14	0.3178** 0.02
Cross section	23	49	88	88	12	22	22	23
N	169	529	2038	2015	115	313	699	701
Number of instruments	166	465	629	627	116	295	510	512

In the above table we see the result of the dynamic panel model. It becomes evident that Kuznets curve appears only in case of WIID2 and D&S dataset but un-inverted U shaped curve in case of EHII dataset but they are insignificant. Although in case of standard annualized dataset we see Inverted U shaped curve for UTIP but it disappears in case of OECD dataset. In both standard and OECD dataset, trade openness becomes insignificant and this may be due to lack of data points as Arellano-Bond requires usage of many instruments which may diminish degrees of freedom. In case of labor participation in agriculture, for WIID2 and D&S in Global Dataset they are negative, contrary to previous findings, but insignificant. However it is very much significant in OECD dataset and is positive in congruence with previous finding. Urbanization has positive coefficient in both standard and OECD dataset but it is insignificant in case of OECD dataset. Education loses much significance in all the datasets as we can see that in most cases the coefficient is insignificant. Cohort size is only significant in case of Global Dataset for WIID2 and we find that male mature cohort population has a negative relationship with inequality while for female it is opposite. The result is opposite in case of EHII for OECD dataset, which is in line with previous findings.

VI. Analysis

After going through the above procedure some of the key findings are mentioned below –

1. Kuznets curve is evident when dependent variable is WIID2 and D&S and an un-inverted curve is found when inequality is EHII. This finding is independent of controlling variable and the econometric model used (pool, fixed, autoregressive fixed effect, dynamic panel)

2. Relationship between openness and inequality is not clear cut and varies based not only on choice of dependent variable but also on the econometric model used

3. Labor participation in agriculture seems to have a positive relationship with inequality especially in case of OECD countries. The finding is in stark opposition to that found in current literature and what Kuznets hypothesized where agriculture was assumed to have lower level of income inequality.

4. Other than pooled regression, in other cases (fixed, autoregressive fixed effect, dynamic panel) urbanization has a significant, especially for Global Dataset, positive relationship with inequality. This is pretty much in line with findings in literature and Kuznets hypothesis, which states that greater urbanization should lead to higher inequality.

5. Cohort size provides a rather interesting finding. This paper used male and female mature cohort size, a novelty. It is found that in most cases, male cohort , age group 40-59, are associated with lower aggregate inequality but in case of female cohort the result is opposite that is they are associated with higher aggregate inequality. The result is more robust for OECD countries in comparison standard annualized dataset. This finding is of very much interest and provides an avenue for further research.

6. Very little additional information can be extracted from 4 year average dataset in comparison annualized Global Dataset. The findings are more or less similar, although average dataset reduces serial correlation but does so imperfectly; AR1 or Arellano-Bond provides a much better way to mitigate the problem of serial correlation.

7. Per capita worker and per capita income more or less provide similar results.

Therefore what we conclude is that although choice of econometric modeling does play a role in determining the coefficient value and the sign of the independent variables hence the relationship between inequality and the other controlling variables, the primary determining factor is the choice of the measure of inequality (D&S, WIID2, UTIP, EHII). Absence or presence of Kuznets inverted curve seems to be strongly contingent upon the usage of particular inequality dataset. The curve emerges almost always when the measure of inequality is D&S and WIID2, while the opposite does so when we use UTIP and EHII. In most cases the coefficient value and the sign of the controlling variables are also opposite for the aforesaid set of inequality measures. This disparity is accentuated and not abated as suggested by Galbraith and Kum (2004) in case of OECD countries. It is of importance to understand why this disparity exists between the findings.

It is worth recalling that WIID2 and EHII are in a sense derived dataset build on the primary dataset of D&S[13] and UTIP manufacturing pay inequality dataset respectively. EHII is also constructed by taking the fitted value of a regression between D&S and UTIP[14] . Hence the starting point of the research will be to analyze the descriptive statistics of the three datasets, namely D&S, WIID2, and EHII.[15] At the onset we will use the annualized global dataset and then we will focus on the subsection of it, namely the OECD dataset.

[13] It is built on D&S 2004, which is an updated version of D&S 1996 used in this study

[14] Dummies for the three types of data source (G, H, I) are also used as regressor. G=0 if measure is based on gross, otherwise 1, H=0 if measure is based on household, otherwise 1, I=0 if measure is based on income, otherwise=1. The information is extracted from the D&S data

[15] UTIP is excluded from the descriptive analysis as it is based on Theil index and hence comparability will be difficult.

Table 13 : Descriptive Statistics for inequality measures

Common Sample			
	EHII	D&S	WIID2
Mean	37.569	35.062	35.537
Median	35.999	33.325	33.755
Maximum	55.255	62.300	62.200
Minimum	24.155	17.830	17.650
Std. Dev.	7.172	8.736	9.093
Observations	458	458	458

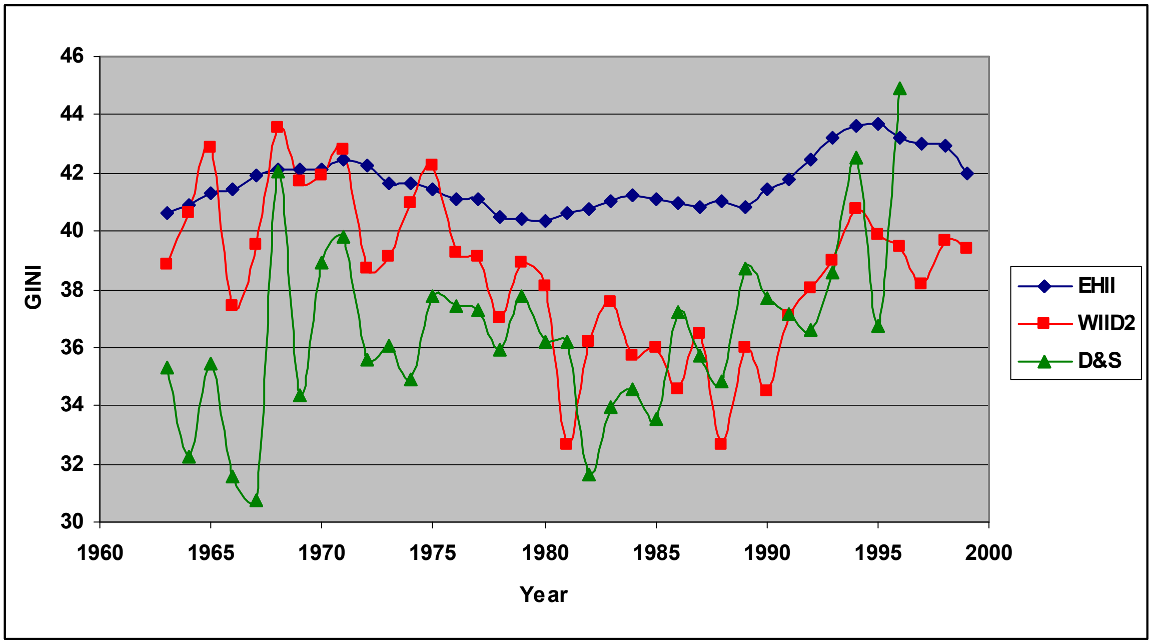
From the table we see that the values of the three inequality series are very much homogenous. The mean and median of the two dataset are very close for all three dataset and so is the standard deviation. But in terms of maximum and minimum value there is similarity between D&S and WIID2 but not with EHII. The results so far does not answer our disparity in findings, however we must note this is drawn from common sample of the three dataset. Now let us look at the situation for overall datasets and not just common sample.

Table 14: Descriptive Statistics for inequality measures

Individual Sample			
	EHII	D&S	WIID2
Mean	41.611	36.459	38.294
Median	43.009	34.420	36.500
Maximum	64.751	63.180	73.900
Minimum	20.075	17.830	4.950
Std. Dev.	7.380	9.424	10.832
Observations	3112	617	1593

Now we see that for EHII dataset the mean and median is roughly 5-10 points higher than that of D&S and WIID2. The difference is higher for maximum and minimum values too. However standard deviation is higher for both D&S and WIID2 in comparison to EHII. This disparity in individual sample descriptive statistics may explain the difference in results obtained between EHII and the other two inequality measure-D&S, WIID2 in annualized global dataset; considering the fact that homogeneity between the inequality measures is high in case of common sample, which then again only accounts for 15% and 29% of the total data points for EHII and WIID2 inequality measures.

Figure 1: Average Global Inequality over the years



In the above Figure we see the average annual world gini for each of the inequality measures. The estimate of mean global inequality index is developed by averaging the inequality measure for all the countries on a yearly basis. The estimate, albeit crude, does raise some interesting question.

- EHII is measured in terms of gross household income, and therefore because it is based on gross income its value can be higher than measures that use net income. But then again since its unit of analysis is household it should drive down inequality, as discussed in section II. Therefore there might be another reason why EHII value is higher than the other two inequality measures[16]
- There seems to be more or less congruence between WIID2 and D&S series. It is important for us to ascertain why there exists such correspondence, as it may explain the similarity in results that we found in the preceding section. In the following table we present some information on measurement aspect of WIID2 dataset, in order for

[16] In its log form the "EHII Gini" is simply:

$EG = \alpha + \beta * T + \gamma * X$

where *EG* stands for estimated household income inequality, *T* is for UTIP-UNIDO pay inequality, and *X* is a matrix of conditioning variables, including the three types of data source (H,G and I), manufacturing employment share to population (mfgpop). The intercept (α) and coefficients (β and γ) are deterministic parts extracted from OLS estimation us to compare it with D&S dataset. This comparison may help us resolve the issue of similarity in findings during regression analysis for the two inequality measure

Table 15: Distribution of inequality measures by different definitions in WIID2 data[17]

Unit of analysis	Obs.
Family	70
Household	259
Person	985

	Net	Gross	Total
Income	917	317	1234
Expenditure/ Consumption	59		59

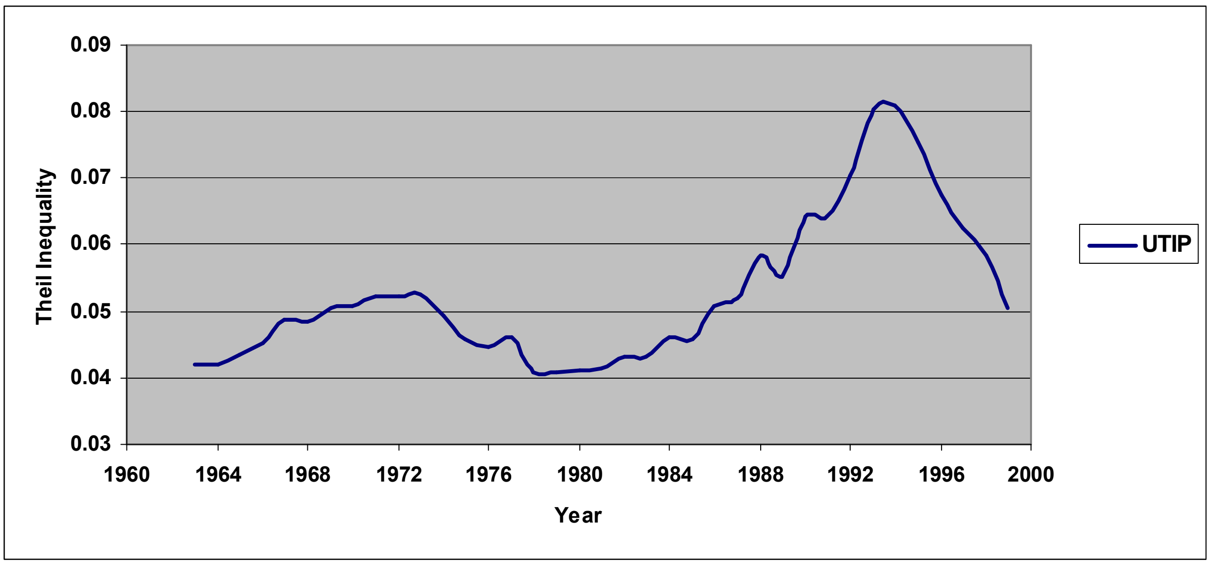
From the above table we see that WIID2 has a significant number of measures where unit of analysis is person or individual, nearly 75 %, while in D&S it was nearly 50%. Similarly values measure in gross accounts for 24%, which was 54% in D&S. Therefore while presence of greater number of individual based unit of analysis may overestimate inequality in case of WIID2, in comparison to D&S, but because most data is based on net of income it may bring down this overestimation significantly. So the overall structure of the series depends on this resultant of these two opposing forces and because they may have nullified each other's opposing effect, probably that is why we see so strong consensus among results between WIID2 and D&S.

Our next question is why there seems to be the large disparity between findings using EHII inequality measures and WIID2/ D&S inequality measures. It is very unlikely that this is because EHII is measured in terms of household income, as we have seen that majority of WIID2 data are based on personal income. Hence if this was the crucial issue then WIID2 would have had higher values than EHII and not lower as our Figure shows. One could suggest that it is because EHII is calculated on gross value whereas WIID2 is substantially based on net of income, but it seems to be a weak reason for such strong divergence in results.

[17] For simplicity and clarity, some of the classifications have been clubbed under broader concepts, e.g. family and family + unrelated individuals fall under the heading of family. The data with original classifications is provided in appendix.

Another possible reason for divergence may stem from UTIP UNIDO manufacturing pay inequality dataset. EHII is based on a regression analysis of UTIP inequality on D&S dataset. Now if manufacturing pay inequality increases over the year, which may have been the case, than that might cause EHII estimates to be higher than WIID2 and D&S inequality measures. The following Figure shows the average global manufacturing pay inequality over 1963-1999 periods. The estimate of mean global manufacturing inequality index is developed by averaging the inequality measure for all the countries on a yearly basis.

Figure 2: Average Global Manufacturing Pay Inequality over the years



We see there has been a rapid rise of manufacturing pay inequality from late 1970s onward, although post 1993 there has been some sharp decline. Manufacturing pay inequality in between this period has almost doubled. This result is consistent with Galbraith and Kum (2004) findings. The rise in manufacturing pay inequality coincides with the economic ascendancy of the Asian tigers, namely Hong Kong, South Korea, Singapore and Taiwan. This is also the time when China and India began rapidly expanding. Apart from these countries other developing countries, especially those in Asia, saw rapid economic growth coupled with increased industrialization. This is probably the reason why we observe the increase in average global inequality in the manufacturing sector, as shown in the graph. However after 1993, the average inequality begins to fall rapidly and by 1999 it drops to the level it was in 1986. This inverted U shape structure of inequality is consistent with Kuznets hypothesis, where he theorized that at the initial stages of economic development, and concomitant industrialization, inequality would increase but at the later stages stabilizing forces would reduce it significantly. So we see that Figure 2 does seem to justify Kuznets hypothesis, however further analysis is required before we reach a conclusion.

Now if we look at the dependent variable used in developing EHII, namely D&S inequality measures, we find that over 60% of its high quality data points lies in between the period 1979-1993, when manufacturing pay inequality (regressor) was rapidly rising . Hence EHII estimate is bound to capture this rise of within sector inequality and therefore its values are very likely to be higher than those reported by WIID2 and D&S inequality measures.

In our previous section we saw that in case of OECD dataset, these differences or varying findings where magnified or became more robust. In order to investigate this strengthening of disparity between findings of different inequality measures in case of OECD countries, we will follow the same route of analysis as we did in the preceding case of annualized global dataset.

Table 16 : Descriptive Statistics for inequality measures OECD

Common Sample			
	EHII	**D&S**	**WIID2**
Mean	34.267	32.111	32.975
Median	34.464	32.855	32.404
Maximum	43.571	41.720	42.850
Minimum	26.154	19.490	20.700
Std. Dev.	3.559	4.455	5.402
Observations	214	214	214

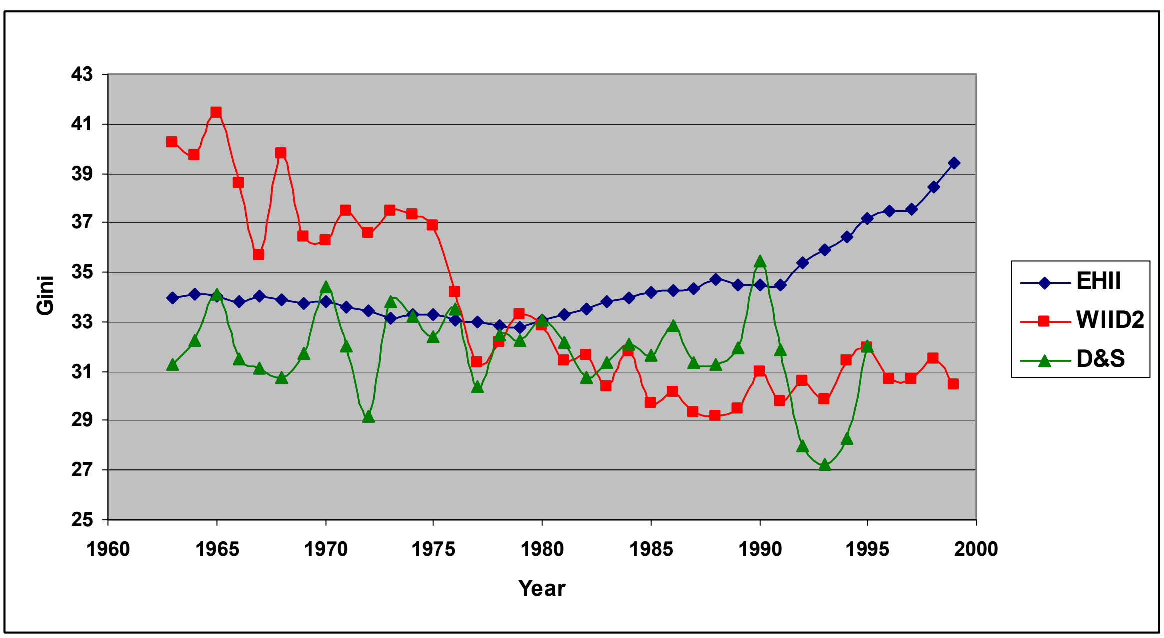
For common sample in case of OECD countries we see high degree of congruence between the three different measure of inequality and this was mentioned by Galbraith and Kum (2004) and was attributed to the fact that in OECD countries the method of data collection for measuring inequality must be consistent and hence all dataset will have similar structure. Now let us look at the descriptive statistics for individual sample rather the common sample, as we did for Global Dataset.

Table 17 : Descriptive Statistics for inequality measures OECD

Individual Sample			
	EHII	**D&S**	**WIID2**
Mean	34.343	32.052	32.742
Median	34.460	32.500	31.700
Maximum	46.177	47.000	58.200
Minimum	20.075	19.490	16.630
Std. Dev.	4.634	4.572	6.886
Observations	818	245	535

The result is indeed interesting because unlike in the case of Global Dataset, we see in this case that for OECD countries even in case of individual sample the descriptive statistics are still very homogenous and even more so than their common sample, which is indeed striking. This is very much contrary to the expected result as we would have expected a greater degree of difference between the EHII and the other two inequality measures -D&S, WIID2 in their individual sample in order to explain our previous regression findings. May be descriptive statistics do not capture the hidden difference. A better approach may be is to see the average annual OECD gini for each of the inequality measures. The estimate of mean OECD inequality index is developed by averaging the inequality measure for all the OECD countries on a yearly basis.

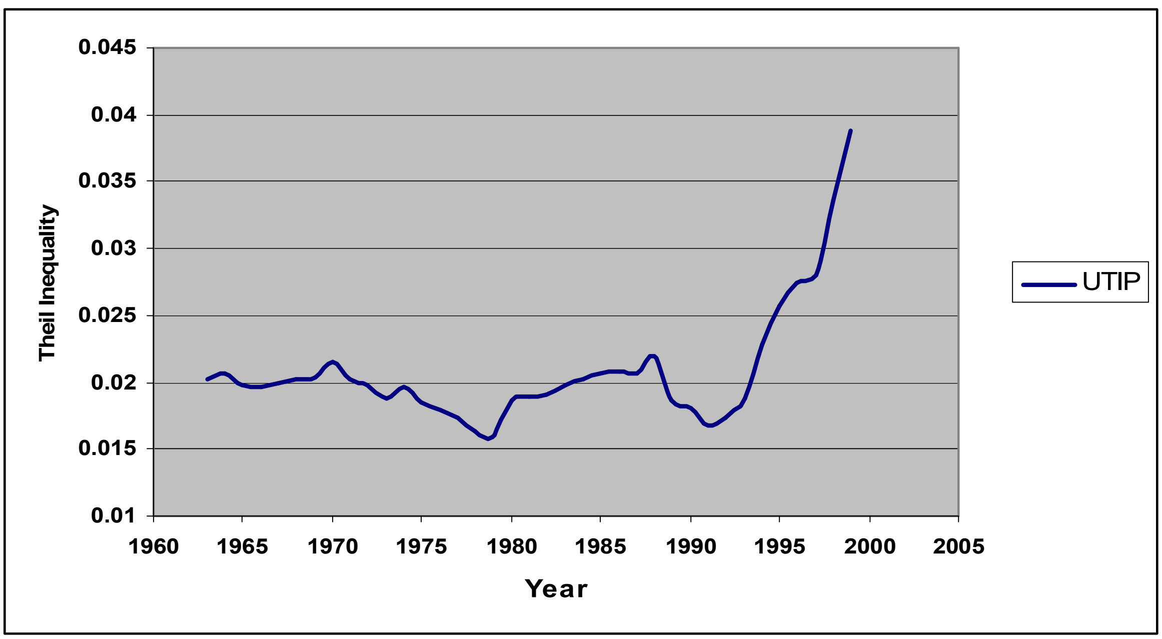
Figure 3: Average OECD Inequality over the years



From the above Figure we can see something that goes missing in descriptive statistics and that is post mid 1970s EHII has consistently higher values than both D&S and WIID2 measures. While D&S and WIID2 show a distinct downward trend, EHII has an equally distinct upward trend. It is interesting to note that the rise in inequality that EHII seems to capture starts from the era of OPEC oil embargo, Yom Kippur conflict and stagflation. During the 1980s USA and other major developed world went through major recessions and rise in inequality. If this is indeed true then we should expect a rapid rise of manufacturing pay inequality during this period and in line with previous argument, significant number of D&S data should lie within this period, as EHII is based on regressing manufacturing pay inequality on D&S data.

In the following Figure we see the average OECD manufacturing pay inequality for the period 1963-1999, using the same methodology mentioned before.

Figure 4: Average OECD Manufacturing Pay Inequality over the years



We see that inequality did rise between the late 1970s till 1999, with brief drop during 1989-1993 periods. However after 1990s inequality actually began to rise rapidly, within the OECD countries, while prior to late 1970s inequality actually diminished by few points. The

OECD countries began rapidly expanding after World War II, with funds being injected under the auspice of Marshall Plan. Reconstruction and massive industrialization took place just after the war. It might be the case that inequality within the manufacturing sector began to rise during that period but gradually, leveling forces stabilized it and by 1970s it has reached a low level equilibrium. This movement of inequality seems to be consistent with the one we saw in Figure 2. However in Figure 2 we saw that inequality began to fall post 1990s, while here we are seeing the exact opposite with inequality rapidly rising after 1990s.

This contradictory finding may be attributed to the fact that after 1990s OECD countries began to experience increasing inequality as they were at a later stage of development than the rest of the world. It may so happen that after the IT boom, which began in early 1990s, and fall of Soviet Union, greater level of technological innovation were spurred within the non military industrial complex of OECD countries. This rapid technological development could have resulted in increasing inequality within the OECD countries. If this hypothesis is indeed true then inequality may follow a zigzag pattern rather than a simple inverted U shaped pattern as predicted by Kuznets. Since major technological innovations are not always incremental and breakthrough (like advent of internet and computers) comes in blocks, it may be hypothesized that: inequality follows an inverted U shaped curve for a given level of technology and economic development but then there is a paradigm shift[18] or cumulative effect of multiple technologies. This may happen only after a country has substantially become economically developed and has extensive R & D spending within the economy. This technological leap radically changes the industrial or manufacturing landscape which in turn increases the inequality within the manufacturing sector. The rise in inequality within the manufacturing sector coupled with already industrialized nature of the developed economy,, resultantly increases overall inequality.

The hypothesis needs further investigation but if found to be true, then we might expect global inequality to rise in the near future as more and more economy develops. Since in this

[18] The idea promulgated here may be consistent with the concept of creative destruction as proposed by Joseph Schumpeter and formalized by Philippe Aghion and Peter Howitt in their paper "A Model of Growth through Creative Destruction," Econometrica 60:2 (1992), pp. 323-51.

paper we are using data till 1999, it might be interesting to see the inequality scenario in the recent 10 years (1999-2009)

However this increased inequality after 1990s within the manufacturing sector of OECD countries does not explain fully the disparity in findings between D&S WIID2 measures and UTIP, EHII measures. In order to analyze this disparity further we look at the D&S inequality measure, which is used as dependent variable in developing EHII. We find that over 72% of its high quality data points for OECD countries lies in between the period 1975-1994 periods. However we can clearly see that the rise in manufacturing pay inequality is highest post 1990s and hence this argument cannot alone account for the difference between findings using EHII and D&S/WIID2 inequality measures for OECD countries. To this end we will focus on correlation between four measures of inequality, including manufacturing pay inequality. The objective is to capture overall disparity or similarity that may exists between the datasets.

Table 18 : Correlation in OECD dataset

	EHII	D&S	WIID2
UTIP	0.854	0.334	0.420
EHII	1.000	0.401	0.414
D&S		1.000	0.745
WIID2			1.000

The above table clearly explains the difference between the findings in the regression analysis. We see that correlation between EHII and D&S, WIID2 is very low while between D&S and WIID2 it is very high as would be expected. Similarly we see that correlation between UTIP manufacturing pay inequality and EHII is very high but with D&S, WIID2 it is very low. This explains the difference between findings using the two sets (D&S, WIID2 and UTIP, EHII) of inequality measures, in case of regressions run on OECD datasets. If the above logic is indeed correct then one would assume that the correlation will be better, if not significantly better, in case of Global Datasets for the two sets of inequality measures since the disparity between the regression findings in case of Global Dataset was lower than for OECD countries. This is shown below –

Table 19 : Correlation in Global Dataset

	EHII	D&S	WIID2
UTIP	0.777	0.413	0.448
EHII	1.000	0.624	0.625
D&S		1.000	0.907
WIID2			1.000

Apart from UTIP and EHII correlation, which is not in our current focus of analysis, all other correlation values significantly become higher. The correlation between EHII and WIID2 jumps by 51% while with D&S it improves by 55%. This explains why the result between the two sets of dataset where opposite in reference to Kuznets curve and other regressor and it also goes in to explaining why the difference heightens in case of OECD countries; it is primarily because of the construction of the dataset.

In Galbraith and Kum (2004), we see that their manufacturing inequality based regression model captures 68% of the variation of the D&S 1996, one would assume that this would be even lesser for D&S 2004, based on which WIID2 is developed. Hence disparity is very likely to creep up between any analysis that uses WIID2 and EHII data and the difference has very little to do with econometric model or the controlling variable used, this paper has clearly demonstrated that. Another features that Galbraith and Kum (2004) mention in regards to EHII is its stability in comparison to D&S. According to the paper EHII coefficients are more narrowly spaced over time than those reported by D&S, which indicates the changes of inequality in the OECD countries are much smaller or stabilized than those of D&S.

However low variance in data may not be a boon in itself and may be due to some other factor than attributed in the paper. For instance dummies[19], used in developing EHII, accounts for 35% of the variation in D&S data while the combine regression with manufacturing inequality explains 68%. Therefore significant amount of variation in data is explained by dummies. The coefficient of the Ln Theil used to estimate EHII was 0.134 and therefore we can understand the magnitude of the impact this variable can have on Ln EHII (the inequality measure was estimated in the log form in the regression) by comparing the

[19] Dummies for the three types of data source (Gross, Household, Individual)

descriptive statistics of the two variables in their log form and also by multiplying the coefficient with Ln Theil, are

Table 20: Descriptive Statistics Ln Theil and Ln D &S

	Ln (Theil)	B x Ln (Theil)	Ln (D & S)
Mean	-3.3563	-0.4497	3.5962
Maximum	0.0254	0.0034	4.1460
Minimum	-6.5689	-0.8802	2.8809

As the above table clearly shows that even if UTIP manufacturing pay inequality is robust and explains 26 % of D&S variation, in terms of magnitude it does not have significant impact on the value of EHII data point. As a matter of fact the constant term[20] for the regression is 4.212, if we take the anti log of the term it becomes 67.50. therefore major share of the EHII data points comes from the Constant and the dummy values, the manufacturing pay inequality, although significant, does not introduce that much variation. It is because of the presence of dummy variables and not necessarily the nature of underlying inequality, that the EHII data are so less dispersed. Hence interpretation based on this dataset may be subject to criticism as it may misrepresent underlying inequality situation.

In the preceding discussion we see that EHII has higher value of gini than WIID2 and D &S for both global and OECD dataset. The primary reason seems to be the rising manufacturing pay inequality within the late 1970s and 1990s period, based on which EHII is constructed. From the pattern seen in case of manufacturing pay inequality in Figure 2 (global dataset) and Figure 4 (OECD), it has been hypothesized that inequality may follow a zigzag pattern rather than a simple inverted U shaped structure. It is believed that inequality may initially rise and fall much like Kuznets had predicted but after inequality stabilizes it may begin to rise once again because of increasing inequality within the manufacturing sector, as it did in case of OECD countries. This upward movement of inequality within the manufacturing sector is believed to be caused by major technological shift within the economy. Therefore one may suggest that inequality movement can be categorized in to two phases:

[20] Although from theoretical aspect the constant term may offer little interpretive power

- Phase 1: In this stage the economy gradually develops from a lowly industrial, agrarian economy to a more developed industrialized urbanized economy, In this phase the usual Kuznets hypothesis works and as predicted inequality will rise and then gradually fall.
- Phase 2: In this stage due to major technological shifts developed economy begin to experience rising inequality within their manufacturing sector as structural changes takes place. Since the economy is already industrialized, a rise in inequality within manufacturing sector results in rise in overall inequality. Once the economy adjusts to this technological shift then inequality may once again begin to stabilize like Kuznets hypothesis.

VII. Conclusion

The paper has shown that the income variables, used in estimating Kuznets hypothesis, should be used in their log form as otherwise the variables are non stationary thus resolving a long standing issue regarding estimating Kuznets curve. The paper also showed that gender discriminated cohort size does seem to have significant impact on inequality, with matured female cohort increasing inequality while matured male cohort reducing it. However the major finding in this paper seems to that "inequality data" matters. Discussion in Section VI clearly pins down the difference in findings in terms of presence of absence of Kuznets curve, to the data structure of the different inequality measures. None of the inequality measures seems to be beyond reproach and when researchers use them, they must be aware of the caveats. In previous researches much has been said about the D & S measures and its limitations. In case of WIID2, although much improved version of original D & S, some of the same criticism leveled against D & S can be used here. e.g. difficulties in data comparability, varied unit of analysis, gross or net of income etc. But WIID2 has much higher number of data points in comparison to D & S, which is a mark improvement. On the other hand EHII series provides a much dense dataset in comparison to D & S and WIID2 and the dataset also shows stability across time. However the stability of the dataset may be attributed to the nature of construction of the dataset. Despite this fact, one significant advantage of EHII is that it captures the rising manufacturing pay inequality and as more developing countries and LDCs become industrialized, this variable is of paramount interest.

In order to ensure comparability of findings using different measures of inequality, data harmonization is in the order. Improving econometric methodology or changing the parametric nature of the equation may affect the results but the magnitude of this affect may not be as significant as in case of usage of different measures of inequality. There may also be further room for improvement of EHII measure. The dataset may be augmented so that it measures inequalities which are adjusted to personal/individual net of income. Also it is suggested in developing this dataset, instead of using D & S, 1996, inequality measure it will be more appropriate to use the latest WIID2 dataset. Kuznets hypothesis can be better explored with the modified dataset.

It is also likely that Kuznets overestimated the leveling effect that was supposed to reduce inequality at the later stage of development. Although urbanization stabilizes with economic development and wages tend to rise within agriculture but the redistributive effort on the part of the government to reduce inequality may no be sufficient. Rapid technological development may greatly increase the inequality scenario within the manufacturing sector. As developing countries and LDCs become more industrializes, inequality within the manufacturing sectors of these countries may increase. Therefore inequality may not be decreasing as Sala-I-Martin (2002) found, but it might so happen that for a time it stabilizes as forces driving the inequality up and down cancel out each other. However with technological advances coupled with rising industrialization, overall inequality may significantly increase via increasing manufacturing pay inequality. This conjecture needs to be investigated and one may do so by developing an improved inequality dataset, as mentioned before. if the zigzag pattern of inequality is confirmed then Kuznet hypothesis may be extended in to cubic functional form adding a cubic income term to the usual quadratic income variables used in testing Kuznets hypothesis.

VIII. References

Adelman, I. and Morris C. (1973). Economic Growth and Social Equity in Developing Countries, Stanford University Press.

Ahluwalia, M. (1976a). Income Distribution and Development: Some Stylized Facts. American Economic Review 66(2), pp. 128-135.

Ahluwalia, M. (1976b). Inequality, Poverty and Development. Journal of Development Economics, 3, 307-342.

Alderson, Arthur S, Nielsen, Francois (2002), Globalization and the Great U-Turn: Income Inequality Trends in 16 OECD Countries, American Journal of Sociology Volume 107 Number 5 (March 2002): 1244–99

Angeles-Castro, Gerardo, (2006) The relationship between economic growth and inequality: evidence from the age of market liberalism No 2, Proceedings of the German Development, Economics Conference, Berlin

Anand, S. and Kanbur, S.M.R. (1993). Inequality and development: A critique. Journal of Development Economics, 41, 19–43.

Atkinson, A. B. and Bourguignon, F. (2000). Introduction: Income Distribution and Economics, in A. B. Atkinson and F. Bourguignon (eds.) Handbook of Income Distribution: Volume 1. Amsterdam: Elsevier.

Atkinson, A.B. and Brandolini, A. (2001). Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries as a Case Study, Journal of Economic Literature, 39(3): 771-799.

Augustin Kwasi Fosu (1993), Kuznets's Inverted-U Hypothesis: Comment, Southern Economic Journal, Vol. 59, No. 3. (Jan., 1993), pp. 523-527

Bahmani-Oskooee, Mohsen, Goswami,Gour Gobinda Mebratu (2006), Solomon. *Black Market Premium and Income Distribution,* The Journal of Developing Areas - Volume 39, Number 2, Spring 2006, pp. 17-28

Barro, R.J. (2000). Inequality and growth in a panel of countries; Journal of Economic Growth, 5, 5–32.

Bourguignon, F. and Morrison C. (1990). Income Distribution, Development and Foreign Trade: A Cross-sectional Analysis, European Economic Review, 34, 1113-1132.

Braulke, M. (1983). A Note on Kuznets' U. Review of Economics and Statistics. 65, 135-139.

Campano, F. and Salvatore, D. (1988). Economic Development, Income Distribution and the Kuznets' U-Shaped Hypothesis, Journal of Policy Modelling, 10, 265-280.

Chen, Been-Lon (2003). An inverted-U relationship between inequality and long-run growth, Economics Letters 78, 205–212

Chong Alberto (2004). Inequality, Democracy, and Persistence: Is There a Political Kuznets Curve?, No 4253, RES Working Papers from Inter-American Development Bank, Research Department

Clarke, G. (1995). More evidence on income distribution and growth. Journal of Development Economics, 47, 403-428.

Dawson, P.J. (1997). On testing Kuznets' economic growth hypothesis, Applied Economics Letters, 4, 409-410.

Deaton, A. and Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. LSMS Working Paper No. 135, Washington, D. C. The World Bank.

Deininger, K. and Squire, L. (1996). A New Data Set Measuring Income Inequality, The World Bank Economic Review, 10, 565-92.

Deininger, K., and Squire, L. (1998). New ways of looking at old issues: inequality and growth. Journal of Development Economics, 57, 259–287.

Fields, G.S. (1979). A Welfare Economic Approach to Growth and Distribution in the Dual Economy, Quarterly Journal of Economics, 93, 325-353.

Forbes, K. (2000). A reassessment of the relationship between inequality and growth. American Economic Review 90, 869–887.

Frazer, G. (2006). Inequality and Development Across and Within Countries. World Development, 34(9), 1459–1481.

Galbraith, J. K. and Kum, H. (2003) Estimating the Inequality of Household Incomes: Filling Gaps and Correcting Errors in Deininger & Squire, Inequality Project, UTIP Working Paper 22, The University of Texas at Austin

Galbraith, J.K. and Hyunsub Kum (2003), Inequality and Economic Growth: A Global View Based on Measures of Pay, CESifo Economic Studies, Vol. 49, 4/2003, 527-556

Galbraith, J.K. and Hyunsub Kum (2004), Estimating the Inequality of Household Incomes: A Statistical Approach to the Creation of a Dense and Consistent Global Data Set, UTIP Working Paper 22 (revised version V), The University of Texas at Austin

Gourdon, Julien (2006), Openness and Inequality in Developing Countries: A New Look at the Evidence, Munich Personal RePEc Archive

Gujarati, D. N. (2003) Basic Econometrics. New York: Mc Graw-Hill, 4th Ed.

Hayashi Fumio (2000): Econometrics, Princeton University Press

Hsing, Y. and Smith, D. (1994). Kuznets' inverted-U hypothesis revisited, Applied Economic Letters, 1, 111-113.

Huang, H-C (River) (2004). A flexible nonlinear inference to the Kuznets hypothesis, Economics Letters, 84, 289–296.

Jalil, Mohammad Muaz and Khasru, Munir (2004), Revisiting Kuznets Hypothesis: A Time Series and Panel Data Approach, Bangladesh Development Studies, Vol. 30, No. 3-4, pp. 89-112

Kuznets, S. (1955). Economic Growth and Income Inequality, American Economic Review, 45(1), 1-28.

Lecaillon, J., Paukert, F., Morrison, C. and Germidis, D. (1984). Income Distribution and Economic Development: An Analytical Survey. Geneva: International Labor Office.

Li, H., Zou, H. (1998). Income inequality is not harmful for growth: theory and evidence. Review of Development Economics 2 (3), 318–334.

Li, H., Squire, L. and Zou, H. (1998). Explaining international and intertemporal variations in income inequality, Economic Journal, 108, 26–43.

Mátyás, L., Konya, L. and Macquarie, L. (1998). The Kuznets U-curve hypothesis: some panel data Evidence Applied Economics Letters, 5, 693-697.

Oshima, H. (1970). Income Inequality and Economic Growth: The post-war experience of Asian Countries, Malayan Economic Review, 15; 7-41.

Papanek, G.F., and Kyn, O. (1986). The Effect on Income Distribution of Development, the Growth Rate, and Economic Strategy, Journal of Development Economics, 23; 55-65.

Paukert, F. (1973). Income Distribution at Different Levells of Development: A Survey of the Evidence, International Labour Review, 108; 97-125.

Pessoa, Argentino (2008), Kuznets Hypothesis and the Data constraints, FEP Working Papers, Faculdade De Economia, Universidade Do Porto

Robinson, S. (1976). A Note on the U Hypothesis Relating Income Inequality and Economic Development. American Economic Review, 66(3); 437-440.

Saith, A. (1983). Development and Distribution: A Critique of the Cross-country U Hypothesis, Journal of Development Economics, 13, 367-382.

Savvides Andreas (1999), On the Turning Point in the Inequality-Development Relationship: Evidence from Six Inequality Measures, Journal Of Economic Development, Volume 24, Number 1

Ram, Rati. (1991). Kuznets's inverted-U hypothesis: Evidence from a highly developed country. Southern Economic Journal 57: 11 12-23.

Ram, R. (1997), Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World, Southern Economic Journal, 64, 576-8.

Thornton, J. (2001). The Kuznets inverted-U hypothesis: panel data evidence from 96 countries. Applied Economics Letters, 8, 15-16.

UNU-WIDER (2005). World Income Inequality Database V 2.0a, User Guide and Data Sources, June 2005, World Institute for Economics Development, United Nations University.

Vicente, J. and Borge, L. (2000). Inequality and growth: inverted and uninverted Ushapes, Applied Economics Letters, 2000, 7, 497-500.

Williamson,Jeffrey G , Mathew Higgins (2002), Explaining Inequality the World Round: Cohort Size, Kuznets Curves, and Openness, Southeast Asian Studies, Vol. 40, No. 3

World Bank (2006). World Development Indicators, data set available on CD-Rom.

Wooldridge Jeffrey (2001) Introductory Econometrics: A Modern Approach: South-Western College Publication, 2nd Ed

Xavier Sala-i-Martin, (2002), “The Disturbing Rise of Global Income Inequality” NBER working paper 8904, http://www.nber.org/papers/w8904

IX. Appendix 1 : Descriptive Statistics Inequality Dataset

Descriptive Statistics for EHII
Categorized by values of COUNTRY
Sample: 1963 1999
Included observations: 3112

COUNTRY	Mean	Obs.
Afghanistan	42.43406	15
Albania	41.14862	8
Algeria	38.52092	28
Angola	53.93387	6
Argentina	43.95437	11
Armenia	52.90648	5
Australia	33.06127	35
Austria	34.39779	37
Azerbaijan	40.87731	5
Bahamas	49.99731	3
Bahrain	53.16349	1
Bangladesh	42.86694	26
Barbados	44.0035	28
Belgium	35.05254	30
Belize	47.23917	2
Benin	49.11796	7
Bhutan	49.86242	1
Bolivia	47.39933	30
Bosnia and Herzegovina	36.951	2
Botswana	46.52021	15
Brazil	47.02352	5
Bulgaria	30.75181	36
Burkina Faso	45.08608	10
Burundi	49.59471	17
Cameroon	50.96366	24
Canada	35.65279	37
Cape Verde	35.51124	2
Central African Republic	47.96143	19
Chile	45.28274	37
China	30.98742	10
Colombia	44.02432	37
Congo, Republic of	52.05451	14
Costa Rica	41.43543	18
Cote d`Ivoire	47.78859	22
Croatia	33.63925	11
Cuba	31.05662	13
Cyprus	41.4484	37
Czech Republic	21.1536	29
Denmark	30.61666	35
Dominican Republic	46.74249	23

Descriptive Statistics for WIID2
Categorized by values of COUNTRY
Sample: 1963 1999
Included observations: 1593

COUNTRY	Mean	Obs.
Albania	29.3	1
Algeria	37.65	2
Argentina	42.20247	30
Armenia	34.55091	11
Australia	36.16778	15
Austria	27.75	14
Azerbaijan	35.7359	13
Bahamas	45.60081	11
Bangladesh	36.64455	14
Barbados	34.68148	9
Belarus	28.47156	12
Belgium	32.11893	16
Bolivia	53.37618	10
Bosnia and Herzegovina	32.88	1
Botswana	52.6625	4
Brazil	58.63173	24
Bulgaria	25.1762	35
Burkina Faso	54.83305	3
Burundi	37.55755	2
Cambodia	44.88333	3
Cameroon	52.1	2
Canada	31.16512	29
Central African Republic	60.43333	1
Chile	51.66795	31
China	27.83222	30
Colombia	53.19947	25
Costa Rica	46.10402	19
Cote d`Ivoire	44.75644	9
Croatia	26.4628	9
Cuba	27.65	2
Cyprus	27.3	2
Czech Republic	22.6775	13
Denmark	33.50015	27
Djibouti	43.3	1
Dominican Republic	47.91688	10
Ecuador	53.14095	9
Egypt	35.72347	7
El Salvador	49.88745	12
Estonia	33.16107	13
Ethiopia	39.10819	3

Ecuador	45.31475	37
Egypt	42.23195	36
El Salvador	45.54604	29
Equatorial Guinea	50.34188	2
Eritrea	45.73096	25
Ethiopia	44.08959	9
Fiji	43.22512	27
Finland	32.04009	37
France	34.01536	17
Gabon	49.42605	8
Gambia, The	44.94524	8
Ghana	50.77894	28
Greece	41.95503	37
Guatemala	48.83312	26
Haiti	46.80135	21
Honduras	45.9015	26
Hong Kong	29.40661	27
Hungary	30.48976	37
Iceland	34.13809	29
India	48.40397	37
Indonesia	48.66917	29
Iran	43.08644	30
Iraq	43.18431	27
Ireland	37.84561	36
Israel	39.19804	34
Italy	36.9141	32
Jamaica	49.92812	27
Japan	36.1572	37
Jordan	48.00379	32
Kenya	49.25794	36
Korea, Republic of	39.4934	37
Kuwait	52.198	31
Kyrgyzstan	44.85311	6
Latvia	28.58869	6
Lesotho	50.0021	7
Liberia	50.03639	3
Libya	44.19023	17
Lithuania	39.7679	5
Luxembourg	31.32481	32
Macao	26.19321	20
Macedonia	37.66941	10
Madagascar	45.00645	22
Malawi	49.36376	32
Malaysia	41.22763	32
Malta	35.03087	34
Mauritania	54.84586	2
Mauritius	42.16082	32
Mexico	42.90316	30
Moldova	36.15182	9
Mongolia	55.96124	6

Fiji	44.97641	4
Finland	28.78795	19
France	33.30054	14
Gabon	50.21583	4
Gambia, The	57.65484	4
Georgia	34.9162	9
Germany	30.05679	28
Ghana	41.11208	7
Greece	39.95333	18
Guatemala	48.099	5
Guinea	50.7	2
Guinea-Bissau	50	2
Guyana	48.475	2
Haiti	51.5	1
Honduras	54.30067	13
Hong Kong	45.00917	10
Hungary	24.40304	26
India	32.89601	25
Indonesia	35.04109	15
Iran	45.11867	9
Ireland	34.87556	9
Israel	39.18357	9
Italy	36.17863	26
Jamaica	51.3377	15
Japan	33.44111	30
Jordan	37.85508	7
Kazakhstan	30.75	7
Kenya	58.77325	12
Korea, Republic of	34.24489	20
Kyrgyzstan	36.8704	13
Laos	33.2	2
Latvia	28.95312	14
Lesotho	60.56	5
Liberia	43	1
Lithuania	30.96275	13
Luxembourg	26.33741	9
Macedonia	29.87575	11
Madagascar	46.25	4
Malawi	54.54167	6
Malaysia	49.11755	12
Mali	48.43333	2
Mauritania	51.74167	6
Mauritius	39.66	5
Mexico	52.79245	12
Moldova	34.59103	13
Mongolia	35.25583	3
Morocco	47.92602	9
Mozambique	39.4	1
Namibia	73.9	1
Nepal	45.44167	4

Morocco	48.42512	26
Mozambique	52.25397	14
Namibia	43.28016	1
Nepal	47.45246	9
Netherlands	33.51756	37
Netherlands Antilles	45.60092	1
New Zealand	34.65019	34
Nicaragua	41.81116	21
Nigeria	45.29105	26
Norway	32.27953	36
Oman	50.37454	6
Pakistan	45.76312	30
Panama	46.67994	35
Papua New Guinea	49.78689	27
Paraguay	40.10873	1
Peru	48.1579	12
Philippines	46.6467	35
Poland	31.32388	30
Portugal	40.04113	27
Puerto Rico	45.11338	15
Qatar	54.52974	8
Romania	30.20028	12
Russia	40.01353	6
Rwanda	48.67767	12
Samoa	48.68798	2
Saudi Arabia	53.67166	1
Senegal	44.10613	24
Seychelles	36.16103	11
Sierra Leone	53.9525	2
Singapore	38.99526	37
Slovak Republic	33.56722	6
Slovenia	28.9798	12
Somalia	46.51789	14
South Africa	43.34604	33
Spain	39.47535	37
Sri Lanka	45.82809	17
St.Vincent & Grenadines	53.50308	2
Sudan	46.66663	1
Suriname	45.80422	20
Swaziland	49.19136	26
Sweden	29.19287	37
Syria	45.30579	36
Taiwan	31.60108	25
Tanzania	48.91225	23
Thailand	48.44694	19
Togo	49.31594	14
Tonga	46.46229	15
Trinidad &Tobago	49.07033	23
Tunisia	46.69935	25
Turkey	43.97338	36

Netherlands	29.61886	19
New Zealand	42.25083	30
Nicaragua	53.97835	2
Niger	46.2	3
Nigeria	47.83621	11
Norway	31.18193	22
Pakistan	32.85504	21
Panama	51.10439	13
Papua New Guinea	50.4	1
Paraguay	50.62319	6
Peru	51.12627	11
Philippines	46.0856	8
Poland	26.0567	27
Portugal	36.76111	9
Puerto Rico	47.64325	6
Romania	27.68101	11
Russia	35.71255	13
Rwanda	28.9	1
Senegal	48.02202	4
Serbia and Montenegro	26.99054	3
Seychelles	46	1
Sierra Leone	56.26111	4
Singapore	44.39236	29
Slovak Republic	21.74879	11
Slovenia	25.33433	12
South Africa	53.715	10
Spain	30.9196	15
Sri Lanka	39.32225	10
Sudan	42.61	3
Suriname	52.80917	1
Swaziland	61.43653	2
Sweden	32.86602	37
Switzerland	34.33463	5
Taiwan	29.92744	30
Tajikistan	31.09722	6
Tanzania	49.9037	9
Thailand	45.58806	15
Trinidad &Tobago	46.27861	6
Tunisia	43.82778	6
Turkey	48.34415	9
Turkmenistan	28.64815	9
Uganda	41.11833	3
Ukraine	29.75945	18
United Kingdom	30.0841	37
United States	41.35095	37
Uruguay	39.86698	17
Uzbekistan	29.10417	6
Venezuela	44.43475	25
Vietnam	36.13333	3
Yemen	30.55	2

Uganda	50.15773	14
Ukraine	36.80093	9
United Arab Emirates	45.70167	4
United Kingdom	32.46824	33
United States	36.55882	37
Uruguay	41.70581	24
Venezuela	44.37909	32
Yemen	48.72914	19
Zambia	47.18901	18
Zimbabwe	45.27099	36

Zambia	58.06875	8
Zimbabwe	64.125	4

Descriptive Statistics for D & S
Categorized by values of COUNTRY
Sample (adjusted): 1963 1996
Included observations: 617 after adjustments

COUNTRY	Mean	Obs.
Algeria	38.73	1
Armenia	39.39	1
Australia	37.88444	9
Bahamas	45.77273	11
Bangladesh	34.514	10
Barbados	48.86	1
Belarus	28.526	1
Belgium	27.00583	4
Bolivia	42.04	1
Botswana	54.21	1
Brazil	57.58071	14
Bulgaria	23.30464	28
Burkina Faso	39	1
Cameroon	49	1
Canada	31.1935	20
Central African Republic	55	1
Chile	51.844	5
China	32.68333	12
Colombia	51.50714	7
Costa Rica	45.495	8
Cote d`Ivoire	38.946	5
Czech Republic	27.428	2
Denmark	32.08435	4
Djibouti	38.1	1
Dominican Republic	46.9375	4
Ecuador	43	1
Egypt	36.66667	3
El Salvador	48.4	1
Estonia	34.6561	3
Ethiopia	44.2	1
Fiji	42.5	1
Finland	29.93333	12
France	40.75298	5
Gabon	61.225	2
Gambia, The	39	1
Germany	31.21896	7
Ghana	35.13	4
Greece	34.53	3
Guatemala	55.68	3
Guinea	40.4	1
Guinea-Bissau	56.12	1
Guyana	40.22	1

Honduras	54.49286	7
Hong Kong	41.58286	7
Hungary	24.48569	8
India	31.19895	19
Indonesia	33.49273	11
Iran	43.228	5
Ireland	36.31333	3
Italy	34.934	15
Jamaica	41.47275	8
Japan	34.71364	22
Jordan	39.18667	3
Kazakhstan	32.67	1
Kenya	54.39	1
Korea, Republic of	34.39083	12
Kyrgyzstan	35.32	1
Laos	30.4	1
Latvia	26.98	1
Lesotho	56.02	1
Lithuania	33.64	1
Luxembourg	27.1277	1
Madagascar	43.44	1
Malawi	62	1
Malaysia	50.35833	6
Mali	54	1
Mauritania	40.165	2
Mauritius	40.67333	3
Mexico	53.85286	7
Moldova	34.43	1
Morocco	39.195	2
Nepal	30.06	1
Netherlands	28.59455	12
New Zealand	34.3625	12
Nicaragua	50.32	1
Niger	36.1	1
Nigeria	38.54667	3
Norway	33.79375	8
Pakistan	31.50444	9
Panama	52.425	4
Peru	47.99	4
Philippines	47.504	5
Poland	25.68818	17
Portugal	37.4425	4
Puerto Rico	51.11	3
Romania	25.83333	3
Russia	26.938	5
Rwanda	28.9	1
Senegal	54.12	1
Seychelles	46.5	2
Sierra Leone	60.79	1
Singapore	40.115	6

Slovak Republic	19.49	1
Slovenia	27.072	2
South Africa	62.3	1
Spain	27.9	8
Sri Lanka	40.95125	8
Sudan	38.72	1
Sweden	31.63255	15
Taiwan	29.61692	26
Tanzania	40.36667	3
Thailand	46.08143	7
Trinidad &Tobago	46.27	3
Tunisia	42.508	5
Turkey	50.36333	3
Uganda	36.89	2
Ukraine	25.71	1
United Kingdom	26.06897	29
United States	35.53517	29
Venezuela	44.41556	9
Vietnam	35.71	1
Zambia	49.5775	4
Zimbabwe	56.83	1

X. Appendix 2: Fixed effect AR1 on unconditional Kuznets curve

Dependent Variable: DS96- Global Dataset
Method: Panel Least Squares
Date: 08/24/09 Time: 08:00
Sample (adjusted): 1964 1994
Cross-sections included: 41
Total panel (unbalanced) observations: 255

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.22549	24.75271	0.493905	0.6219
LNGDPC	4.147293	5.606783	0.739692	0.4603
LNGDPC2	-0.192547	0.317635	-0.606189	0.5450
AR1	0.654697	0.063435	10.32082	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.952224	Mean dependent var	33.62499
Adjusted R-squared	0.942488	S.D. dependent var	7.141513
Log likelihood	-474.8817	F-statistic	97.80131
Durbin-Watson stat	1.789663	Prob(F-statistic)	0.000000

Dependent Variable: WIID2 - Global Dataset
Method: Panel Least Squares
Date: 08/24/09 Time: 08:03
Sample (adjusted): 1964 1999
Cross-sections included: 96
Total panel (unbalanced) observations: 860

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-33.27065	22.05387	-1.508608	0.1318
LNGDPC	17.31444	5.009350	3.456426	0.0006
LNGDPC2	-1.039360	0.284751	-3.650067	0.0003
AR1	0.532491	0.029745	17.90159	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.887227	Mean dependent var	37.46310
Adjusted R-squared	0.872704	S.D. dependent var	9.938891
Log likelihood	-2256.317	F-statistic	61.09243
Durbin-Watson stat	2.130290	Prob(F-statistic)	0.000000

Dependent Variable: EHII - Global Dataset
Method: Panel Least Squares
Date: 08/24/09 Time: 08:04
Sample (adjusted): 1964 1999
Cross-sections included: 134
Total panel (unbalanced) observations: 2661

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	95.53095	5.593567	17.07872	0.0000
LNGDPC	-12.58250	1.272092	-9.891193	0.0000
LNGDPC2	0.725874	0.071955	10.08787	0.0000
AR1	0.817338	0.011992	68.15831	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.954130	Mean dependent var	41.70234
Adjusted R-squared	0.951658	S.D. dependent var	6.983294
Log likelihood	-4846.484	F-statistic	386.0354
Durbin-Watson stat	1.996493	Prob(F-statistic)	0.000000

Dependent Variable: UTIP - Global Dataset
Method: Panel Least Squares
Date: 08/24/09 Time: 08:05
Sample (adjusted): 1964 1999
Cross-sections included: 135
Total panel (unbalanced) observations: 2668

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.128879	0.107941	-1.193973	0.2326
LNGDPC	0.043129	0.024547	1.757017	0.0790
LNGDPC2	-0.002531	0.001388	-1.823162	0.0684
AR1	0.554307	0.017216	32.19769	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.744927	Mean dependent var	0.051956
Adjusted R-squared	0.731115	S.D. dependent var	0.057144
Log likelihood	5673.588	F-statistic	53.93241
Durbin-Watson stat	1.761264	Prob(F-statistic)	0.000000

Dependent Variable: DS96 - OECD
Method: Panel Least Squares
Date: 08/24/09 Time: 08:18
Sample (adjusted): 1964 1994
Cross-sections included: 16
Total panel (unbalanced) observations: 143

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.87310	64.40240	0.355159	0.7231
LNGDPC	0.258903	13.59202	0.019048	0.9848
LNGDPC2	0.071882	0.717015	0.100252	0.9203
AR1	0.766603	0.064851	11.82097	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.930378	Mean dependent var	31.94903
Adjusted R-squared	0.920272	S.D. dependent var	4.263735
Log likelihood	-219.2526	F-statistic	92.05866
Durbin-Watson stat	1.639461	Prob(F-statistic)	0.000000

Dependent Variable: WIID2 - OECD
Method: Panel Least Squares
Date: 08/24/09 Time: 08:19
Sample (adjusted): 1964 1999
Cross-sections included: 26
Total panel (unbalanced) observations: 390

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-66.92445	55.49646	-1.205923	0.2286
LNGDPC	27.02984	11.96505	2.259067	0.0245
LNGDPC2	-1.721175	0.645215	-2.667598	0.0080
AR1	0.682531	0.039372	17.33558	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.777441	Mean dependent var	32.94115
Adjusted R-squared	0.760178	S.D. dependent var	7.064825
Log likelihood	-1022.386	F-statistic	45.03713
Durbin-Watson stat	2.331112	Prob(F-statistic)	0.000000

Dependent Variable: EHII - OECD
Method: Panel Least Squares
Date: 08/24/09 Time: 08:21
Sample (adjusted): 1964 1999
Cross-sections included: 25
Total panel (unbalanced) observations: 756

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	332.3512	7.631065	43.55240	0.0000
LNGDPC	-66.19066	1.640478	-40.34840	0.0000
LNGDPC2	3.660186	0.088235	41.48228	0.0000
AR1	0.957264	0.015643	61.19351	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.965958	Mean dependent var	34.87936
Adjusted R-squared	0.964696	S.D. dependent var	3.932554
Log likelihood	-829.6961	F-statistic	765.0934
Durbin-Watson stat	2.010739	Prob(F-statistic)	0.000000

Dependent Variable: UTIP - OECD
Method: Panel Least Squares
Date: 08/24/09 Time: 08:23
Sample (adjusted): 1964 1999
Cross-sections included: 26
Total panel (unbalanced) observations: 763

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.786021	0.053383	14.72415	0.0000
LNGDPC	-0.172096	0.011476	-14.99551	0.0000
LNGDPC2	0.009606	0.000617	15.56082	0.0000
AR1	0.934911	0.021123	44.26075	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.874841	Mean dependent var	0.020704
Adjusted R-squared	0.870067	S.D. dependent var	0.014336
Log likelihood	2949.588	F-statistic	183.2344
Durbin-Watson stat	1.942810	Prob(F-statistic)	0.000000

XI. Appendix 3: STATA output file for annualized global dataset

```
  ___  ____  ____  ____  ____ tm

  Notes:

1.  (/m# option or -set memory-) 10.00 MB allocated to data
2.  (/v# option or -set maxvar-) 5000 maximum variables

 1 . insheet using "F:\research paper\results\090622_final _dataset_standard.csv", comma
   (33 vars, 6956 obs)

 2 . xtset cid year
        panel variable:   cid (strongly balanced)
         time variable:   year, 1963 to 1999
                 delta:   1 unit

 3 . xtabond ds96 lngdpc lngdpc2 openk labor_agri  lnedu male4059 female4059 , lags(1)
 artests(
   > 2)

  Arellano-Bond dynamic panel-data estimation  Number of obs        =          169
  Group variable:  cid                          Number of groups     =           23
  Time variable:  year
                                               Obs per group:    min =            1
                                                                 avg =     7.347826
                                                                 max =           27

  Number of instruments =     166               Wald chi2( 8)         =      225.14
                                               Prob > chi2          =      0.0000
  One-step results
```

ds96	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ds96						
L1.	.5861216	.0654786	8.95	0.000	.4577859	.7144573
lngdpc	15.3279	7.651911	2.00	0.045	.3304266	30.32537
lngdpc2	-.7136282	.4157659	-1.72	0.086	-1.528514	.101258
openk	.0329749	.0299013	1.10	0.270	-.0256306	.0915805
labor_agri	-.1051728	.16202	-0.65	0.516	-.4227261	.2123806
lnedu	.6931384	.846382	0.82	0.413	-.9657398	2.352017
male4059	-1.6373	.8541432	-1.92	0.055	-3.31139	.0367895
female4059	.8017151	.4465972	1.80	0.073	-.0735993	1.67703
_cons	-61.67916	35.2944	-1.75	0.081	-130.8549	7.496591

```
  Instruments for differenced equation
          GMM-type:  L(2/.).ds96
          Standard:  D.lngdpc D.lngdpc2 D.openk D.labor_agri D.lnedu D.male4059
                     D.female4059
  Instruments for level equation
          Standard:  _cons

    Sunday August 2 15:32:36 2009   Page 2

 4 .
 5 . xtabond wiid2 lngdpc lngdpc2 openk labor_agri  lnedu male4059 female4059 , lags(1)
 arte
   > sts(2)
```

```
Arellano-Bond dynamic panel-data estimation  Number of obs         =        529
Group variable:  cid                          Number of groups      =         49
Time variable:  year
                                              Obs per group:    min =          1
                                                                avg =   10.79592
                                                                max =         35

Number of instruments =      465              Wald chi2( 8)         =     120.63
                                              Prob > chi2           =     0.0000
One-step results
```

wiid2	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
wiid2						
L1.	.1405927	.0427165	3.29	0.001	.05687	.2243154
lngdpc	36.15331	7.355838	4.91	0.000	21.73613	50.57048
lngdpc2	-2.258695	.4567337	-4.95	0.000	-3.153877	-1.363514
openk	.0021054	.0127175	0.17	0.869	-.0228204	.0270312
labor_agri	-.1511069	.1164693	-1.30	0.194	-.3793825	.0771688
lnedu	-.692441	.9541223	-0.73	0.468	-2.562486	1.177604
male4059	-2.547316	.9523872	-2.67	0.007	-4.413961	-.6806713
female4059	1.798439	.4883471	3.68	0.000	.8412957	2.755581
_cons	-116.0031	30.69637	-3.78	0.000	-176.1669	-55.83929

```
Instruments for differenced equation
        GMM-type:  L(2/.).wiid2
        Standard:  D.lngdpc D.lngdpc2 D.openk D.labor_agri D.lnedu D.male4059
                   D.female4059
Instruments for level equation
        Standard:  _cons

6 .
7 . xtabond ehii lngdpc lngdpc2 openk u_pop lnedu male4059 female4059 , lags(1)
artests(2)

Arellano-Bond dynamic panel-data estimation  Number of obs         =       2038

Time variable:  year
```

```
Number of instruments =      629                Wald chi2( 8)          =    1124.73
                                              Prob > chi2          =      0.0000
One-step results
```

ehii	Coef.	Std. Err.	z	P>|z|	[95% Conf.	interval]
ehii						
L1.	.6652992	.0241174	27.59	0.000	.6180299	.7125685
lngdpc	-5.183618	3.873696	-1.34	0.181	-12.77592	2.408686
lngdpc2	.237235	.2226065	1.07	0.287	-.1990656	.6735357
openk	-.0025363	.0024294	-1.04	0.296	-.0072977	.0022252
u_pop	.0571881	.020772	2.75	0.006	.0164758	.0979005
lnedu	.3553396	.1879268	1.89	0.059	-.0129901	.7236692
male4059	-.1445232	.2755202	-0.52	0.600	-.6845329	.3954865
female4059	.1962164	.1454644	1.35	0.177	-.0888885	.4813213
_cons	33.7439	16.91576	1.99	0.046	.5896316	66.89818

```
Instruments for differenced equation
        GMM-type:  L(2/.).ehii
        Standard:  D.lngdpc D.lngdpc2 D.openk D.u_pop D.lnedu D.male4059
        D.female4059
Instruments for level equation
        Standard:  _cons
```

sunday August 2 15:32:36 2009   Page 3

```
8 .
9 . xtabond utip lngdpc lngdpc2 openk labor_agri  lnedu male4059 female4059 , lags(1)
artes
  > ts(2)

  Arellano-Bond dynamic panel-data estimation  Number of obs         =          2015
  Group variable:  cid                          Number of groups      =            88
  Time variable:  year
                                               Obs per group:    min =             1
                                                                 avg =      22.89773
                                                                 max =            35

  Number of instruments =      627                Wald chi2( 8)          =     599.67
                                                Prob > chi2          =      0.0000
  One-step results
```

utip	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
utip						
L1.	.5469018	.0242607	22.54	0.000	.4993518	.5944519
lngdpc	.1064748	.0567912	1.87	0.061	-.004834	.2177836
lngdpc2	-.0076705	.0032148	-2.39	0.017	-.0139713	-.0013697
openk	.0000181	.0000376	0.48	0.630	-.0000557	.0000919
labor_agri	-.0007455	.0002919	-2.55	0.011	-.0013175	-.0001734
lnedu	.0055451	.0030612	1.81	0.070	-.0004548	.0115449
male4059	.0034561	.0045067	0.77	0.443	-.0053768	.012289
female4059	-.0005964	.0024349	-0.24	0.806	-.0053687	.0041758
_cons	-.3421769	.2545129	-1.34	0.179	-.8410131	.1566593

```
Instruments for differenced equation
        GMM-type:  L(2/.).utip
        Standard:  D.lngdpc D.lngdpc2 D.openk D.labor_agri D.lnedu D.male4059
                   D.female4059
Instruments for level equation
        Standard:  _cons
```