Review of Income and Wealth
Series 51, Number 1, March 2005

ESTIMATING THE INEQUALITY OF HOUSEHOLD INCOMES: A STATISTICAL APPROACH TO THE CREATION OF A DENSE AND CONSISTENT GLOBAL DATA SET

by James K. Galbraith* and Hyunsub Kum

LBJ School of Public Affairs, The University of Texas at Austin

The deficiencies of the Deininger and Squire data set on household income inequality are well known to include sparse coverage, problematic measurements, and the combination of diverse data types into a single data set. Yet many studies have relied on this data due to the lack of available alternatives. In this paper we show how the UTIP-UNIDO measures of manufacturing pay inequality can be used, with other information, to estimate measures of household income inequality. We take advantage of the systematic relationship between the UTIP-UNIDO estimates and those of Deininger and Squire. The residuals from this exercise provide a map to problematic observations in the Deininger and Squire data, and the estimated coefficients provide a way to construct a new panel data set of estimated household income inequality. This new data set provides comparable and consistent measurements across space and through time.

1. Introduction

In recent years the master compilation of statistics on household income inequality by Klaus Deininger and Lyn Squire of the World Bank (Deininger and Squire, 1996, hereafter DS) has become a staple of development economics research.[1] It is the raw material underlying Sala-i-Martin's highly publicized claim that global inequality has declined since 1975 (Sala-i-Martin, 2002a).[2] Others have used it to reassess the relationship between inequality income and economic growth. For example, Forbes deploys it to find that higher levels of inequality can produce higher subsequent growth *rates* (Forbes, 2000), a finding that controverts both the traditional Kuznets hypothesis (Kuznets, 1955),[3] and also more recent arguments that egalitarianism might be good for growth (Birdsall *et al.*, 1995; Perotti, 1996).

Yet many scholars remain uneasy about the quality of the information contained in the DS data set. To begin with, the coverage is sparse and unbalanced. With fewer than 700 country/year observations in the most widely used versions,[4]

*Correspondence to: James Galbraith, Lyndon B. Johnson School of Public Affairs, P.O. Box Y, University of Texas at Austin, Austin, TX 78713-8925, USA (Galbraith@mail.utexas.edu).

[1]There are several alternative income inequality data sets available particularly the Luxembourg Income Study (LIS) and World Income Inequality Data set (WIID). But the former is restricted to wealthy western countries and the latter is an expanded compilation of which the DS data are the core part.

[2]Dollar and Kraay (2001) make a similar argument using the DS data.

[3]Kuznets postulated an inverted "U" relationship between the level of income and the level of inequality. Interestingly Ram (1997) finds both inverted and upright "U" relationships between inequality and economic development in the DS data, depending on whether ordinary least squares or a fixed-effects specification is used.

[4]This data is available at http://www.worldbank.org/research/growth/dddeisqu.htm. We restrict our attention to the figures denoted as "high quality" and as giving nation-wide coverage.

TABLE 1

Different Types of Inequality in the DS 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

Notes:
*Indicates whether the measure of income is gross or net of taxes.
**Indicates whether the survey measure is of expenditure or income.

DS offer only infrequent measures of inequality for much of Africa, Latin America and Asia. The United States, Great Britain, Bulgaria, India and Taiwan are among the few countries for which DS provide annual or nearly annual observations over long periods of time. This means that studies attempting to assess the time trend of inequality worldwide must not only allow themselves to be affected by the bias that may be associated with a history of regular surveys of income inequality, but also either restrict their attention to a subset of the data in order to achieve a better semblance of balance, or else attempt to fill in the gaps by interpolation. The first approach is taken in Forbes (2000) who uses five-year intervals, and in Aldersen and Nielson (2002) who deal with only 16 OECD countries. Sala-i-Martin (2002b) takes the second approach: among other things, where only a single observation is available, Sala-i-Martin assumes that no change occurred over the whole time period under study.[5]

Atkinson and Brandolini (2001) present a critique of DS that focuses, in part, on the many different types of data that are mixed up in the data set, even after the "high-quality" filters suggested by DS have been applied.[6] As shown in Table 1, these include measures of expenditure inequality and of income inequality, measures of inequality of gross and of net income, and measures of inequality of both personal and household income.[7] The comparability of these various measures is questionable, but what can one do? Expenditure surveys are prevalent in some parts of the world, and income surveys in others; there is no way to go back and convert one into the other.

DS (1996 and 1998) 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. But Atkinson and Brandolini are skeptical: "we doubt whether a simple additional or multiplicative adjustment is a satisfactory solution to the

[5]Obviously, this procedure will be without bias only if it happens that there is no systematic pattern in the global evolution of inequality. See Milanovic (2002b) for a detailed critique of Sala-i-Martin's interpolations.

[6]According to DS (1996), a data point is deemed "high-quality" if the underlying survey meets three criteria: (a) coverage of all types of income, including in-kind income, (b) coverage of urban and rural households, and (c) focus on households rather than individuals.

[7]There are four types of household size adjustments applied in the DS data: household, household equivalent (weighted by the number of persons), person, and person equivalent (wherein the effective number of household members is assumed to be the square root of the actual number).

heterogeneity of the available statistics. Our preference is for the alternative approach of using a data-set where the observations are as fully consistent as possible." All in all, Atkinson and Brandolini reject the use of "macro" data sets collated from disparate studies, and urge reliance instead only on studies from which the underlying micro information can be recovered. This is the approach taken by Milanovic (2002a) in his efforts to measure the "true" evolution of household income inequality. However, this approach is limited by its own cost, complexity and by the limited availability of surveys. To date, Milanovic has produced global household inequality measures for only three years.

Even within individual countries, the range of fluctuation in the DS 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, 1989–90. We detect nine cases in which changes of over 5 Gini points happened over a single year. We think changes of such speed and magnitudes are unlikely, except when they coincide with moments of major social upheaval. And unfortunately, at such moments and places household income surveys are rarely undertaken.

The University of Texas Inequality Project (UTIP) has produced an alternative global inequality data set, based on the Industrial Statistics database published annually by the United Nations Industrial Development Organization (UNIDO). This data set has approximately 3,200 observations over 36 years (1963–99). It is also based on source data that are much more likely to be accurate and consistent, both through time and across countries.[8] However, the data do not measure household income inequality. UTIP-UNIDO is a set of measures of the dispersion of pay, using the between-groups component of a Theil index (Theil, 1972), measured across industrial categories in the manufacturing sector.[9] While there is evidence that the UTIP-UNIDO measures provide a sensitive index of changes in distribution generally, the exact nature of the correlation between an establishment-based measure of manufacturing pay inequality and a survey-based measure of household income inequality is not clear, particularly in comparisons across countries.

In this paper, we offer an approach that combines the information in the DS data with the information in the UTIP-UNIDO data, along with a certain amount of additional information, in order to accomplish two objectives. The first is to separate the useful from the doubtful information in the DS data set itself. The second is to permit a more informed filling-in of missing information about household income inequality. In effect, we replicate the coverage of the UTIP-UNIDO data set with *estimated* measures of household income inequality, based on the relationship between inequality of household incomes, inequality of industrial pay, and other variables. The result is a data set for estimated household income inequality that is much more comprehensive than DS and that is consistently adjusted to reflect a household income inequality basis.

[8]Rodrik (1999) and Berman (2000) have recently endorsed the comparability and accuracy of the UNIDO compilation of employees and payments measures on which the calculation of the UTIP-UNIDO inequality measure is based.

[9]We also calculate elasticities based on a cross-industries Gini coefficient computed following Pyatt (1976), which is reported in Appendices B and C.

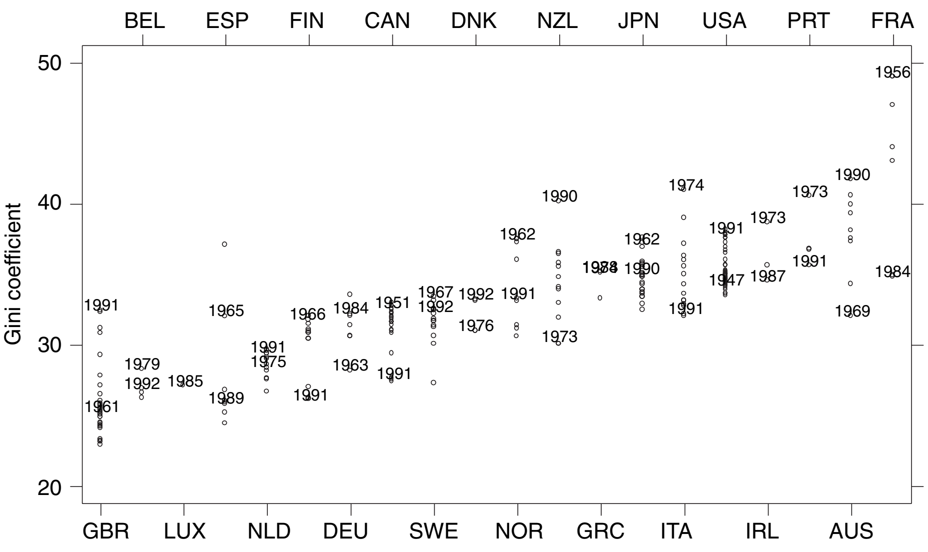


Figure 1. Rank and Distribution of DS Gini for 20 OECD Countries
*Years represent the first and last observations for each country.

2. The Comparability Problem in Deininger and Squire

The first issue is how to diagnose the comparability problem in the DS data. We take two approaches. First, we try to assess each value in DS using information in the data set itself. Here a principal concern is the different types of source data: expenditure and income, net and gross income, household and per capita surveys. Bias from this source may well be systematic, not random, since certain countries tend systematically to conduct one type of survey and not the other.

Our second approach is to find other variables that are reliably and systematically related to the DS inequality measures. If such relationships can be found, they can be used to assess and also to expand the DS data set. In the next section, we relate the DS data to four economic variables for which data are available on a global scale: the UTIP-UNIDO measures of pay inequality, the share of manufacturing employment in total population, the degree of urbanization, and the rate of population growth. We will discuss the theoretical justification for these variables below.

As mentioned above, the DS data is a compilation of fragmented information across countries and through time. It is easy to find apparently anomalous measurements in this data set, as a simple graphical illustration will show.

Figure 1 presents a summary of DS Gini coefficients for 20 OECD countries, ranked in order from lowest to highest, and showing also the reported direction of movement of inequality over time. The first and last observed years for each country are also denoted. Conceição and Galbraith (2001) and Galbraith and Kum (2003) have already remarked that some of the readings—such as lower inequality for Spain (ESP) than for Sweden (SWE), such as France (FRA) as the most unequal country in Europe but with a huge leveling of incomes over time, such as steadily falling inequality in Italy—would raise the eyebrows of any informed observer.

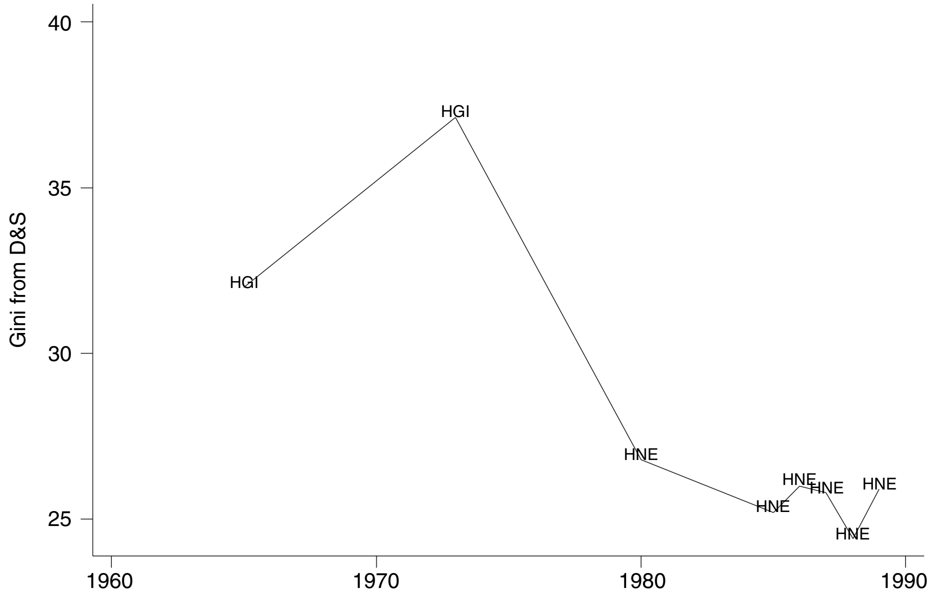


Figure 2. Inequality in Spain, as Reported by DS
HGI: Household Gross Income; HNE: Household Net Expenditure.

To see this point closely, we review the source characteristics of the DS data as shown in Table 2. The "high quality" DS data includes inequality measures based on three distinguishable sources. Some are expenditure-based and some are income-based. Some are per capita and some relate to households. Among the income measures, some are gross and others are net of tax. If household gross income (HGI) is assumed to be the reference category, only 39 percent of DS observations worldwide literally fit into this category. If household net income (HNI) is added, the combined share increases to 52 percent.[10] In other words, at least 48 percent of the DS data cannot be classified as measures of household income.

Table 2 shows a clear divergence of inequality measures by source. The simple mean differences between expenditure-based and income-based inequality, and between household and per capita inequality, are significant and substantial. The distribution of sources across regions is also notably unbalanced. Most South Asian, African and Middle East countries use expenditure surveys, most Eastern European countries use per capita income, and only half of inequality measures from Latin American countries are household income. Even among OECD members only half (52 percent) of observations are based on household gross income.

Furthermore, sources of inequality sometimes vary even in the same country. For instance, inequality measures for Spain are based on two different sources: household gross income (HGI) and household net expenditure (HNE). The shift from one to the other no doubt partly explains both the decline in measured inequality (Figure 2) and why the average level of inequality appears low in the DS data as seen in Figure 1.

[10]This table is based on the 652 observations whose categorical information is available in the DS data.

TABLE 2

DIFFERENT SOURCES OF GINI IN THE DS DATA (n = 652)

		Non-OECD						OECD					
Region		HGI	HNI	HNE	PGI	PNI	PNE	HGI	HNI	HNE	PGI	PNI	PNE
East Asia & Pacific	N	36			14	26	8	44					
	Mean	42.53			34.73	29.62	34.47	35.32					
East Europe & Cent Asia	N	5	5		61	19							
	Mean	41.4	27.48		25.76	22.91							
Latin America	N	57		2	32		12						
	Mean	50.07		49.93	51.48		42.43						
Middle East & N. Africa	N			3			16						
	Mean			40			41.33						
North America	N							68					
	Mean							33.92					
	N	22		8		1	33						
South Asia	Mean	39.73		31.55		30.06	32.44						
Sub Saharan	N	5	3	1			36						
Africa	Mean	50.7	57.82	54.21			43.86						
	N							17	76	9		33	
WE	Mean							36.77	32.06	28.63		26.19	
Total	N	125	8	14	107	46	105	129	76	9		33	
	Mean	45.75	38.86	37.6	34.63	26.86	39	34.78	32.06	28.63		26.19	

Notes:
HGI = Household Gross Income.
HNI = Household Net Income.
HNE = Household Net Expenditure.
PGI = Per capita Gross Income.
PNI = Per capita Net Income.
PNE = Per capita Net Expenditure.

We find similar situations from 30 out of 104 countries (4 from the OECD[11] and 26 from outside the OECD[12] including 14 Latin American countries), where the information is available (n = 652). Figure 3 presents DS measures for Brazil, Columbia, Jamaica and Peru. These examples show that generally-perceived wild fluctuations of inequality in Latin American countries are partly due to differences among the various sources that comprise the DS data.

Table 3 shows the results when the DS inequality measures are regressed on dummies indicating the different sources and additional regional dummies. Only dummies for source characteristics are included in the first row; these estimates indicate that, on average, net income and per capita-based measures of inequality are lower than household and gross income-based measures.[13] Of course, it is possible that these differences reflect real differences in inequality, independent of data type. But this conjecture appears less compelling after we control for regional differences as shown in the next row. On average, Eastern Europe shows the lowest level of inequality, while Latin America, Africa and the Middle East show much higher levels of inequality than Western Europe. Controlling for regions, the type of data remains a significant determinant of the measure, with one exception: the mean difference between income and expenditure measures of inequality disappears. It appears that income-expenditure differences are highly correlated with regional differences that are now controlled explicitly. Of course, this finding does not tell us whether the observed differences in inequality measures are "true" differences across regions, or an artifact of the systematic practice of some regions to use one type of measure or the other.

3. Estimating the Relationship Between Inequalities of Pay and Income

Pay inequality and income inequality are different economic concepts. But they are not unrelated. In most countries, manufacturing pay[14] is a significant component of all pay. And pay is everywhere the largest single element in income. Moreover, the manufacturing sector is not sealed off from the economy at large. Largely unskilled (and low-wage) workers in manufacturing are substitutes for unskilled (and similarly low-wage) workers in services and agriculture, and vice

[11]This includes Spain, Germany, Denmark and Finland.

[12]This includes Brazil, Chile, Columbia, Costa Rica, Guatemala, Guyana, Honduras, Jamaica, Mexico, Panama, Peru, Venezuela, Sri Lanka, Pakistan, Mauritius, Zambia, Seychelles, Malaysia, Philippines, Bulgaria, Czech Republic, Hungary, Poland, Romania, Russia and Yugoslavia.

[13]Grun and Klasen (2003) and Dollar and Kraay (2002) also found that household-based and gross income-based measures are typically higher than expenditure measures. These various findings are all derived from differing data and model specifications. Dollar and Kraay use the expanded DS data (n = 814) with fixed-effects; Grun and Klasen use WIID (2000) data (n = 2,033) with more detailed reference units; we use a subset of the original DS data (n = 652) with the three dummies specified in the Table 3. We experimented with a fixed-effects model with similar results. One inconsistency between our analysis and other studies is that in a simple OLS specification expenditure-based measures are higher than income-based measures in our study. But as discussed later in the main text, this estimate loses its significance when other control variables enter into the model. We use the estimates from Table 4 in our EHII estimation.

[14]This refers to what is reported as payroll in manufacturing surveys, including wages, salaries and fringe benefits.

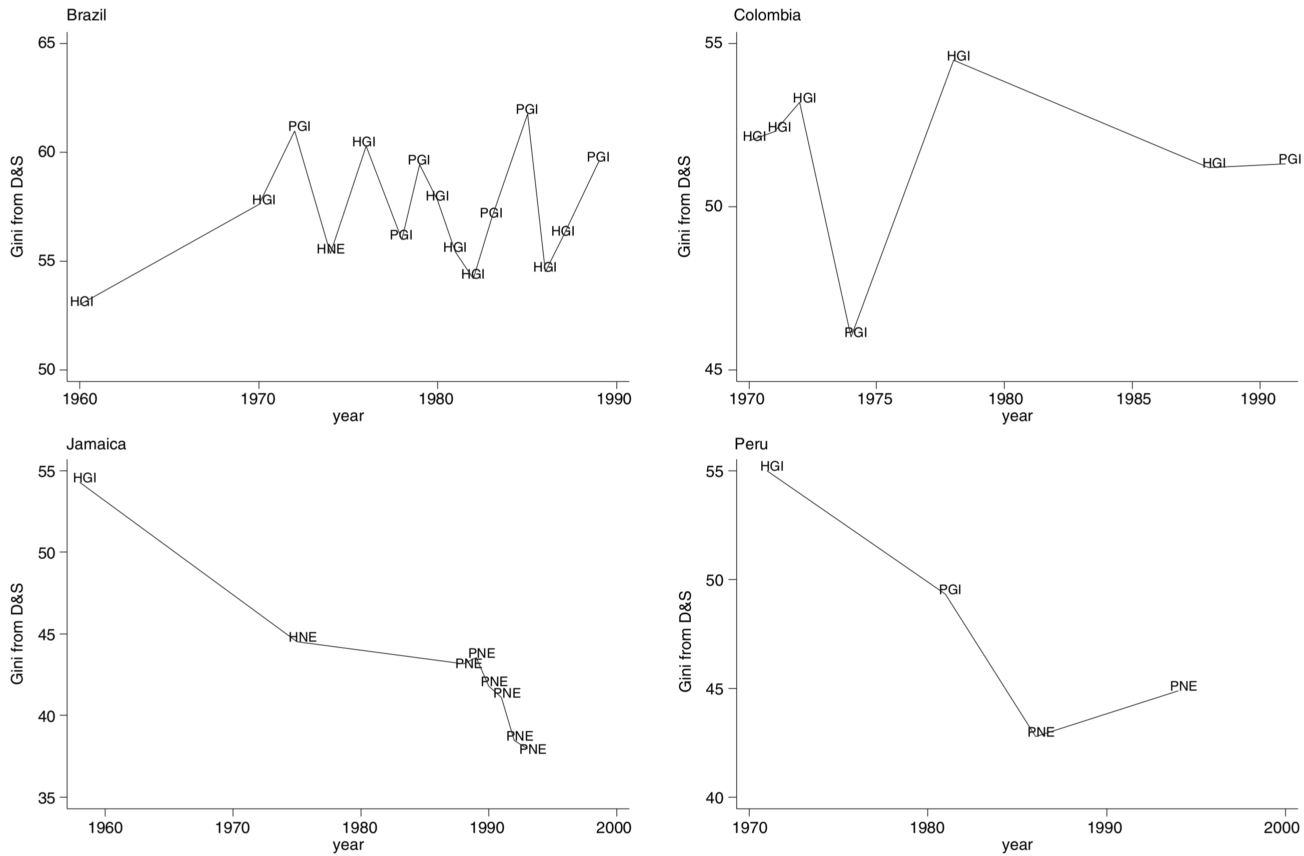


Figure 3. Inequality in Selected Latin American Countries, as Reported by DS Data

TABLE 3

EFFECT OF DATA TYPE AND REGION IN THE DS DATA (n = 652)

	Expenditure	Person	Net	Constant	EAP	ECA	LAC	MENA	NA	SAS	SSA
Coefficient	0.296	–0.147	–0.214	3.661							
	(10.97)**	(7.72)**	(10.09)**	(282.14)**							
Coefficient	0.010	–0.109	–0.119	3.551	0.092	–0.182	0.407	0.365	–0.030	0.120	0.439
	(0.42)	(7.72)**	(6.52)**	(191.68)**	(4.52)**	(7.45)**	(17.47)**	(9.33)**	(1.20)	(4.62)**	(14.91)**

Notes:
Income = 0, Expenditure = 1.
Household = 0, Person = 1.
Gross = 0, Net = 1.
EAP: East Asia and Pacific.
ECA: Eastern Europe and Central Asia.
LAC: Latin America.
MENA: Middle East and North Africa.
NA: North America.
SAS: South Asia.
SSA: Sub Saharan Africa.
Western Europe (base dummy, omitted).
*Significant at 5%, **significant at 1%.

versa. For this reason, it is likely (though *not* certain) that changes in inequality inside manufacturing will tend to mirror changes in inequality in the structure of pay overall.[15]

Figure 4, adapted from Galbraith and Kum (2003), gives weight to this argument. It portrays the trends of UTIP-UNIDO pay and DS income inequality for Great Britain (left) and the U.S. (right) in matching time frames. This simple graphical comparison indicates that it is likely (though not certain) that changes in inequality inside manufacturing will tend to mirror changes in inequality beyond formal industry pay.

Moreover, as noted by Atkinson (1997),[16] overall wage inequality has been widely used as an alternative to income inequality in the literature. For example, Williamson (1982) argues that the "wage differential and its development seems to parallel broader trends in income distribution"; he regards wage inequality as a "simplified phenomenon of the evolution of overall inequality." Acemoglu (1997) identifies increased earnings and wage inequality as the main components of rising income inequality in the U.S. In Brenner *et al*. (1991), a number of studies test the Kuznets hypothesis using measures of wage inequality. Kuznets would have approved: in his seminal 1955 piece he calls for the exclusion of the incomes of the economically inactive, "to avoid complicating the picture" (Kuznets, 1955).

Suppose, then, that we have two data sets. One of them, DS, attempts to measure household income inequality, but does so imperfectly, owing to inconsistencies in the underlying measurements and other problems. The other, UTIP-UNIDO, measures the dispersion of manufacturing pay across industrial sectors, a much narrower economic concept, but does so with precision.[17] Let us assume that measurement errors in DS are—apart from that related to type of data—random for practical purposes. While patterns may exist, we have no reason to suspect that they were designed into the construction of the data set.

In that case, we propose the following model:

$$I = \alpha + \beta * T + \gamma * X + \varepsilon$$

[15]Wade (2002) concurs with this conclusion.

[16]Atkinson (1997) also finds close similarity in the movements of household income inequality and individual earning inequality over 1970s and 1980s in the U.K., even though he is cautious for direct relationship between them since there are other income sources like capital income and transfers.

[17]The UNIDO Industrial Statistics from which the UTIP measures are calculated report just two measurements for each industrial category: total employment and payroll in nominal domestic currency units. Calculating this inequality measure requires no adjustment for inflation, or purchasing power parity, and poses no other issues of method. The major difficulty in extracting comparable Theil coefficients from the data set lies in the occasional discontinuities in the number of industrial categories UNIDO reports for different countries and years. In most cases, we have overcome this difficulty by reconstructing the original categories from the published data. On rare occasions, missing measurements of payroll or employment were filled in by interpolation. A fuller discussion of the issues involved in measuring dispersions of manufacturing pay by these means appears in Galbraith and Kum (2003).

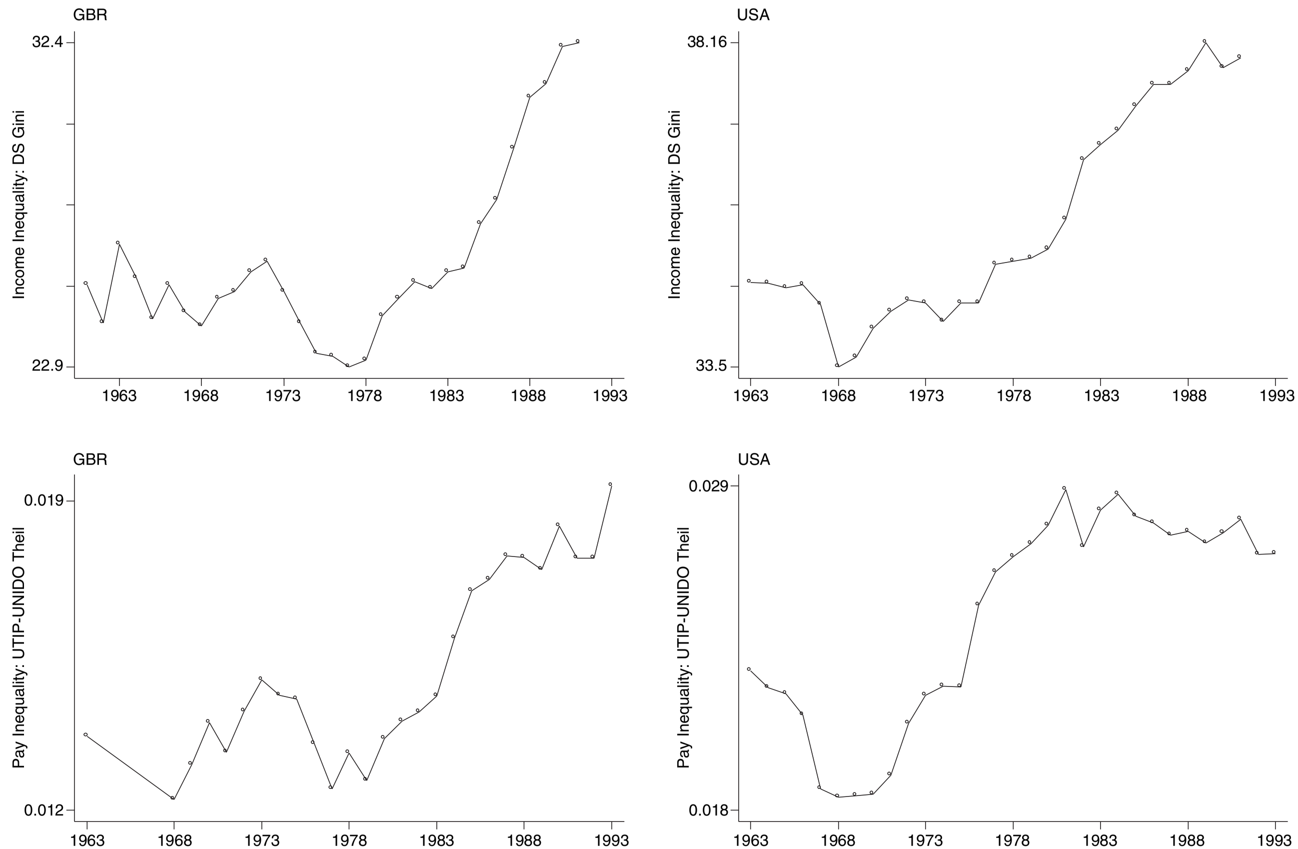


Figure 4. Inequality in Income and in Manufacturing Pay for the U.S. and Great Britain

where I represents the DS measure of inequality (in Gini coefficients), T represents the measured dispersion of manufacturing pay,[18] and X is a matrix of conditioning variables including dummies for the three types of data source (G, H, I),[19] and other relevant economic variables.

We are able to assemble three economic variables for which coverage was sufficient and for which a good theoretical rationale exists for considering them as determinants of income inequality. These are (a) the ratio of manufacturing employment to population (MFGPOP), (b) the share of urban population (urban),[20] and (c) population growth rate (POPGROWTH),[21] and we are able to match these independent variables to just under 500 observations in the DS "high-quality" data set.[22]

A word of theoretical justification is appropriate in each case. First, it is obvious that the importance of the manufacturing sector in total economic activity varies widely from place to place (and in some places also over time). The ratio of manufacturing employment to population provides a crude-but-effective measure of the relative size and importance of manufacturing, and conversely of the relative size and importance of services, agriculture, natural resource extraction, and government taken together. In general, since manufacturing tends to be more heavily unionized than the other sectors, and since industrialization is associated historically with the development of the middle class, we expect higher shares of manufacturing employment in population to be associated with lower inequality.

To justify the inclusion of urbanization, we look to Kuznets (1955), who noted that urban centers tend to encompass more diverse and complex forms of economic activity than rural areas—which are, virtually by construction, the domain of agriculture.[23] Wealthy people live in cities. Thus urbanization should be associated with greater inequalities, other things equal, at least so long as there remains a significant rural population against which the wealth of the cities can be compared. Note that as incomes rise two phenomena occur together: urbanization (associated with rising inequality of incomes) and industrial deepening (associated with declining inequality in manufacturing pay). The effect of urbanization on inequality thus offsets, to a degree, that of the industrialization per se

[18]To improve the efficiency of the estimates, particularly since the UTIP-UNIDO measures are strongly log-normal in their distribution, we take the log of both inequality measures. Thus the coefficient will be a measure of the elasticity of income inequality with respect to a Theil measure of manufacturing pay dispersion.

[19]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 DS data.

[20]This is derived from World Bank Macro Table.

[21]Population variable is derived from WDI 2002, World Bank Macro Table (2003) and Penn World Table 6.1.

[22]We have often been advised to include a measure of government transfer payments in this exercise, but there are two problems. First, paucity of data cuts down the degrees of freedom drastically. Second, when we ran the regression on the reduced data set, the coefficient on transfers as a share of GDP was not significant. An evident explanation is that the equality of the pay structure is a good predictor of the generosity of social security systems.

[23]Kuznets (1955) noted that "other conditions being equal, the increasing weight of urban population means an increasing share for the more unequal of the two component (*rural and urban*) distributions." We thank Branko Milanovic for calling this remark to our attention.

TABLE 4

LINEAR REGRESSION RESULTS

	Model 1	Model 2	Model 3	Model 4	Model 5
Expenditure	0.272	–0.015	–0.139	–0.124	–0.146
	(3.89)***	(0.19)	(1.64)	(1.45)	(1.96)*
Person	–0.145	–0.121	–0.081	–0.072	–0.081
	(1.92)*	(2.49)**	(1.88)*	(1.71)*	(2.16)**
Net	–0.179	–0.086	–0.042	–0.048	–0.025
	(2.84)***	(1.60)	(0.83)	(0.95)	(0.58)
Ln(Theil)		0.165	0.118	0.117	0.106
		(5.47)***	(4.99)***	(5.02)***	(4.82)***
mfgpop			–0.002	–0.002	–0.002
			(3.88)***	(3.80)***	(3.31)***
urban				0.001	0.001
				(0.89)	(1.23)
popgrowth					5.687
					(2.98)***
Constant	3.611	4.249	4.205	4.156	3.984
	(98.47)***	(37.40)***	(46.91)***	(39.56)***	(35.44)***
Observations	484	484	484	481	481
R–squared	0.24	0.49	0.59	0.59	0.63

Notes:
Dependent variable is natural logarithm of Gini from DS.
Income = 0, Expenditure = 1.
Household = 0, Per Capita = 1.
Gross = 0, Net = 1.
*Significant at 10%; **significant at 5%; ***significant at 1%.

on household incomes, and it is appropriate to include it in a regression relating pay inequalities to income inequality.

Population growth is, for us, merely an available proxy for the age structure of the underlying population. A population which is growing rapidly will include a larger number of children and young people, necessarily. Households will accordingly be larger on average and of greater variability in size, and it is likely that households with lower income have more children than their wealthier counterparts. This may work to increase per capita income inequality and it could have an effect on inequality measured across households.

Table 4 presents the results of an OLS regression with robust standard errors, introducing the conditioning variables seriatim. Since the current data structure is in panel format, estimate the model assuming that the observations are independent across countries, but not necessarily within country.

We begin with a model including only three dummies for the types of source (Income/expenditure, Household/per capita, Gross/net) in the DS data (Model 1). The result indicates that inequality measures based on income and expenditure are significantly different. Whether income inequality is measured on gross or net basis also makes a considerable difference. However, these patterns are not very robust when other conditioning variables are added; the gross/net variable loses significance while the income/expenditure remains significant only at the ten percent level in models three and five. On the other hand, the household/per capita difference is significant at the 10 percent level through all the models, and at the 5 percent level in two of them.

We find that the UTIP-UNIDO pay inequality measure (T) is, strongly associated with the DS income inequality (I) measure. T alone accounts for almost 25 percent of variation in I; adding in dummies for the types of source raises the R^2 to around 49 percent (Model 2). Running the model in log-log form generates elasticity estimates, which are between 0.106 (Model 5) and 0.165 (Model 2). Thus a rise in the Theil measure of manufacturing pay dispersion between 6.06 and 9.43 percent is estimated to correspond to a 1 percent increase in a Gini coefficient for household income inequality. Given the much greater volatility of the Theil measure,[24] and also the greater volatility of manufacturing pay compared with household income,[25] this is a reasonable value in our view.[26]

The ratio of manufacturing employment to population (MFGPOP) has the expected negative sign with significance at the 1 percent level consistently. This indicates that an economy with a larger manufacturing sector shows lower income inequality, other things being equal. By adding this to manufacturing pay inequality and the types of data (Model 3), almost 60 percent of all the variation in the DS data set is accounted for.[27]

Adding the variables of urbanization and population growth (Model 5) raises the proportion of variation explained by another 3 percentage points together. Population growth enters positively at the 1 percent significance level. Consistent with Kuznets' expectation, the urbanization ratio is estimated as a positive factor, but the coefficient is not significant.

We offer in Table 5 the results of fixed-effects and random-effects estimations, in which we control separately for the particular characteristics of each country in the data set. It is well known that the variation of income inequality is much larger across country rather than through time. Thus, an explicit control for country may better capture the evolutionary relationship among variables.[28]

The model is following:

$$I_{it} = \alpha + \beta * T_{it} + \gamma * X_{it} + \upsilon_i + \varepsilon_{it}$$

As the table shows, pay inequality continues to have a very significant relationship with income inequality in all cases. The estimated coefficients are between 0.079 and 0.119 in both random and fixed effects models, and they are reasonably consistent with the previous results from OLS. The fact that the elasticities are lower than in the pooled regression does suggest a difference between the within-country

[24]The number of ISIC categories used in the calculation of Theil measures is not identical for each year and country.

[25]Household income includes incomes from other sources such as non-labor wage, land and capital.

[26]In the appendix we report a table of elasticities calculated after recomputing our Theil measures as a between-industries Gini coefficient. The elasticity of this pay Gini to the DS Ginis is about 0.25.

[27]We check the robustness of estimated coefficients for Theil and MFGPOP by separating the data into groups by type of source: income, expenditure, gross, net, household, per capita. Estimates of Theil are all significant at the 1 percent level and those of MFGPOP are also significant except in one case (expenditure only). Signs of estimates are all expected and not much change in the magnitude of estimates is found. These results are available from the authors on request.

[28]The properties of Fixed and Random effects models are discussed in Greene (2000) and Baltagi. (1995). In our analysis, the fixed-effects model is preferred to the random-effects model in all cases. Hausman-test statistics are all significant at a less than 1% level.

TABLE 5

FIXED AND RANDOM EFFECTS MODEL ESTIMATION RESULTS

	Model 1F	Model 1R	Model 2F	Model 2R	Model 3F	Model 3R
Expenditure	–0.151	–0.011	–0.160	–0.059	–0.175	–0.059
	(3.09)***	(0.29)	(3.36)***	(1.57)	(3.62)***	(1.54)
Person	–0.049	–0.061	–0.045	–0.052	–0.048	–0.051
	(2.86)***	(3.64)***	(2.66)***	(3.20)***	(2.81)***	(3.15)***
Net	–0.034	–0.084	–0.021	–0.057	–0.016	–0.057
	(1.19)	(3.26)***	(0.74)	(2.26)**	(0.59)	(2.24)**
Ln(Theil)	0.099	0.119	0.084	0.094	0.079	0.094
	(8.63)***	(11.47)***	(7.18)***	(8.75)***	(6.60)***	(8.73)***
mfgpop			–0.001	–0.002	–0.001	–0.002
			(4.29)***	(6.72)***	(4.50)***	(6.50)***
urban					0.001	0.000
					(1.57)	(0.30)
popgrowth					–0.578	0.491
					(0.81)	(0.74)
Constant	3.961	4.136	3.985	4.129	3.893	4.112
	(84.61)***	(92.58)***	(86.32)***	(97.79)***	(51.38)***	(71.76)***
N	484	484	484	484	481	481
Country	81	81	81	81	81	81

Notes:
F and R represent fixed effects model and random effects model respectively.
Dependent variable is natural logarithm of Gini from the DS.
Income = 0, Expenditure = 1.
Household = 0, Per Capita = 1.
Gross = 0, Net = 1.
*Significant at 10%; **significant at 5%; ***significant at 1%.

and the between-country relationship between manufacturing pay inequality and overall income inequality; the unobserved country fixed effects appear to account for part—but by no means all—of the relationship.

The share of manufacturing employment to total population (MFGPOP) retains its separate significance at the 1 percent level and the coefficients in all cases are positive and stable as expected. Interestingly the magnitudes of both coefficients (T and MFGPOP) do not change much in different specifications, which means their effects are relatively independent from those of the additional variables. On the other hand, the addition of controls for country obliterates the significance of the latter two conditioning variables, urbanization and population growth, showing that these variables influence inequality only to the extent that they differ across countries. Accordingly, while this exercise does not discredit the use of urbanization and population growth in the regression, it inclines us to regard pay inequality and manufacturing employment share as very robust independent determinants of income inequality.

4. FINDING THE PROBLEMS IN DS: A STUDY OF RESIDUALS

The residuals from the ordinary least squares regressions can, we believe, usefully indicate those countries in the DS data set where Gini coefficients may be either too high or too low. Note that we implicitly assume that there is no *systematic* bias in the DS data. There is no reason to suspect any such bias, and no

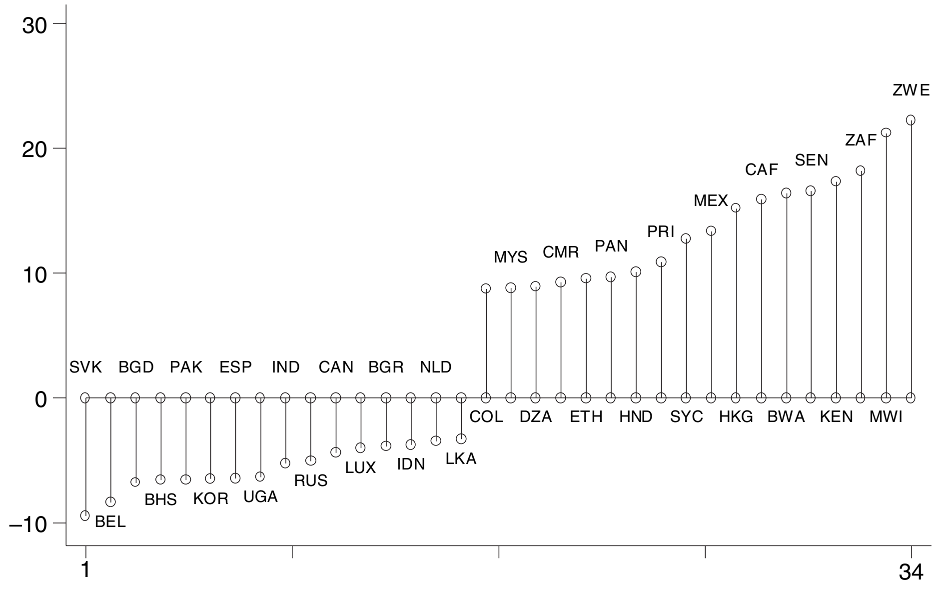


Figure 5. Selected Residuals: DS Gini Compared with Predicted Values

way, on the basis of our exercise, to correct for it. The deeper concern here is with cases where the DS household income inequality measures have yielded results that are simply out of character with pay dispersions and related factors after controlling for the differences in data sources (H, G, I): either implausibly low (undervalued), or implausibly high (overvalued). Figure 5 presents selected countries whose average Gini values are out of line with the predicted Gini values of our Model 3, using OLS estimation with 484 observations.[29] The y-axis in this figure indicates the difference in Gini values between the predicted and the observed.

This figure includes some very important cases. Five major South Asian countries—India (IND), Sri Lanka (LKA), Indonesia (IDN), Pakistan (PAK) and Bangladesh (BGD)—all exhibit reported Gini coefficients considerably lower than their manufacturing employment shares and pay dispersions would appear to justify. The prevalence of expenditure-based surveys is known to play an important role in this region, but even beyond this the estimates appear too low. The same (not surprisingly) appears true for Spain (ESP), which remains an incongruous choice, in our view, to be Europe's most egalitarian country.

Among positive outliers, South Africa (ZAF) stands out, with a Gini measure 18.2 points higher than would be justified by manufacturing pay differentials and manufacturing share. Some—perhaps most—of this may be quite real, owing to South Africa's unique history of racial repression, perhaps to distinctive features of its current population structure, and perhaps also to markedly different abilities to survey income in white and non-white communities. The problem is that we

[29]Residuals from Model 5 produce similar results. These are available from the authors on request.

have no evidence on any of these points. Moreover, South Africa is an industrial country. Since part of the South African manufacturing labor force is comprised of non-whites, and they are no doubt more heavily represented in low-wage industrial sectors, it seems to us that some of the effects of apartheid on pay should be captured in the observed manufacturing pay dispersion. The final resolution of this discrepancy remains open to further research.

The DS values for six other sub-Saharan African countries are also quite high. In Sub-Saharan Africa generally it is possible—here we speculate—that the combination of open agricultural and herding country (hence, an absence of money incomes or direct taxes) with a large proportion of mining income (hence, a concentration of high incomes) generates income inequalities that are out of proportion to observed pay inequalities. However, given that we can match only 19 inequality observations for 15 sub-Saharan countries—in most cases, only one observation per country—these comparisons should be treated with great caution.

Other high measures in the DS data set include Latin American countries: Mexico (MEX), Puerto Rico (PRI), Honduras (HND), Panama (PAN), and Colombia (COL). Mexico is an interesting case, as it is notable that Mexico's *manufacturing* pay dispersion across industries is not very different from that found in the United States. For most of the period under study, moreover, Mexico maintained effective protection for staple agriculture, which surely worked to reduce urban-rural differentials below what one often observes in the Third World. Yet, surveys report Mexican income inequality on a par with that in Brazil, where racial and agricultural patterns are very different. Finally, we note the case of Hong Kong, where we estimate DS Gini coefficients to be over 15.2 Gini points higher than our model would predict. This is telling case, in our view, since Hong Kong is a city-state with no agriculture to speak of and therefore no urban-rural differential.

Table 6 assesses regional patterns in the residuals, by averaging them across the major regions. Several major regions have roughly offsetting high and low estimates, but others have a systematic tendency to come in high or low. The largest consistent apparent underestimates of inequality are in South Asia, as we suspect already, where DS characteristically report Gini values comparable to those given

TABLE 6

REGIONAL PATTERN IN MEAN RESIDUALS (n = 484)

Region	DS Gini	Estimates	Residuals	N
South Asia	34.04	38.59	–4.55	45
North America	33.76	35.85	–2.09	49
East Europe & Central Asia	25.13	26.95	–1.82	71
Western Europe	30.52	31.03	–0.51	121
East Asia & Pacific	35.83	34.79	1.05	109
Latin America	47.87	42.26	5.61	55
Middle East & North Africa	41.13	34.89	6.24	16
Sub Saharan Africa	47.68	37.38	10.30	18

for Northern Europe and Scandinavia. Parts of East Asia and the Pacific region are also apparently strongly underestimated. But very high values for Malaysia (a heavily industrialized country with a 30 percent manufacturing share) and Hong Kong bring the average up. On the other hand, the largest apparent overestimates of income inequality are in Latin America and Sub-Saharan Africa—one of the most urbanized developing regions, and one of the most rural.

Absent compelling evidence to the contrary, which has so far not been presented in the literature that we know of, we believe the more likely explanation for discrepancies between our estimates and the DS values lies in the data. Quite apart from the problem of quality, household surveys are necessarily creatures of the cultures in which they are taken. Systematic differences in income measurement across regions with different cultural and political characteristics, in the way surveys are administered, and in the way they are responded to, should not be surprising.

5. Building a Deep and Balanced Income Inequality Data Set

As noted, the "high-quality" subset of the DS data set has fewer than 700 observations. The UTIP-UNIDO data set has just fewer than 3200 observations. On the assumption that the relationship between the UTIP-UNIDO Theil and the DS household income inequality has been estimated accurately, it is thus possible to calculate an estimated household income inequality measure to match each of UTIP-UNIDO pay dispersion measures. We present this in a data set denoted EHII. It is based on just two exogenous variables: pay inequality and manufacturing share, plus dummies for data type;[30] the variables for urbanization and population growth are dropped, as they add little to the explanatory power of the regression while imposing some restrictions on the coverage. EHII is calculated from OLS estimates with conditioning variables in Model 3 as described above.[31]

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

[30]A version based on all four exogenous variables is also posted on the site http://utip.gov.utexas.edu; the differences are minor. A particular problem occurs in a handful of countries, notably in the Persian Gulf, where the combination of extra-high values for urbanization, population growth and wage inequalities generates estimated Gini coefficients that are above the boundary values for that coefficient.

[31]The choice whether to include or exclude the coefficient estimates on the dummy variables is a judgment call, which we make in favor of inclusion for a number of reasons, including our priors, the evidence from selected countries such as Spain where both types of survey are available, and the fact that these variables are significant in the fixed and random effects models. On the other hand, after reflection we decided against including regional dummies or calculating the EHII data from models that included country fixed effects. Such an approach in our judgment would have amounted to assuming the correctness of the DS data, when one purpose of the exercise is to identify those countries and regions where discrepancies exist and further study is needed.

population (MFGPOP). The intercept (α) and coefficients (β and γ) are deterministic parts extracted from OLS estimation of Model 3 in Table 4.[32]

This data set has, we believe, three distinct advantages over that of DS. 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 DS. 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 as a reference (denoted as α),[33] and unexplained variations in the DS income inequality measures (previously ε) are treated for what they probably are: as inexplicable. They are therefore disregarded in the calculations of the EHII Gini coefficients.[34]

We call attention particularly to those cases where the EHII estimates are much lower than the DS Gini coefficients. In fact, 11.1 percent of the DS data are higher than 50 Gini points, whereas EHII data suggest that that pay inequality and manufacturing employment share could produce such values in only a few cases. If the DS values are accurately measured, they must be reflecting phenomena occurring in other parts of the economy.[35]

Figure 6 provides estimates for income inequality in the OECD countries, corresponding to Figure 1's compilation of measures from the DS data. It is worth noting that the estimated Gini coefficients are more narrowly spaced over time than those reported by DS, which indicates the changes of inequality in the OECD countries are much smaller or stabilized than those of DS. They are more consistent in increasing from the start to the finish of the data set: in most cases, later inequality is higher. Also the rank order places the Scandinavian countries at the low end of OECD countries, with the Mediterranean countries ranking consistently high. No surprising phenomena like Spain and France in Figure 1 turn up.

[32]It is possible there are some instances of selection bias. For instance, inequality will be understated where the unemployment rate is high since industrial job losses affect mainly low-income workers. Also in very rich countries, trends in capital income can lead to large differences between the trends of pay inequality and of income inequality.

[33]It would be a small matter to recompute the estimates to any basis desired: expenditure, gross income, net income, household or per capita.

[34]We invite researchers to download and examine these data sets, to use them in their research into the evolution of global inequality, and to send us reactions and suggestions for improvement. We remain open especially to persuasive reasons to transfer additional information from the DS data set to the estimation of our own measures (for instance by finding additional statistically valid predictors of the measured inequality in the DS data). But our philosophical position is to approach this issue conservatively. We will add new information to the underpinnings of our estimates when there is strong reason to believe that the resulting estimates would be markedly improved, and only when the sacrifice in terms of coverage is not great.

[35]EHII has a higher sample mean than DS: 41.4 Gini points compared to 34.7; this reflects the larger proportion of values for non-OECD countries. The standard deviation is smaller for EHII: 7.5 against 8.7 Gini points. The minimum EHII value is 19.7 compared with 17.8 for DS, the maximum is 64.7 compared to 62.3. We are skeptical of the higher values, insofar as the assumption of linearity is less likely to hold for extreme values.

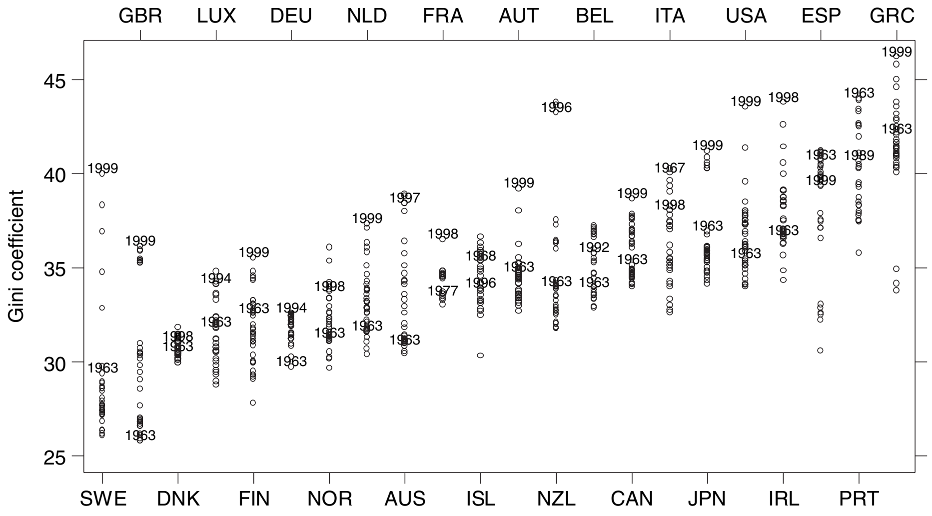


Figure 6. Estimated Household Income Inequality for OECD Countries

*Years represent the first and last observations for each country.

Figure 7 presents mean differences between the EHII estimates of income inequality and those of DS by regions, alongside 95 percent confidence intervals. The figure illustrates the discrepancies between the two data sets especially for South Asia (SAS), Latin America (LAC) and Middle East and North Africa (MENA), and the fact that for other regions discrepancies are far less. 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.

It is possible, of course, that some of these differences are rooted in reality, in systematic regional differences captured by surveys but not reflected in the EHII estimates. However, as noted above, no one has yet provided a persuasive account—based on statistical evidence as opposed to conjecture—of what that reality might consist. We suspect that the most likely reason for the large inter-regional differences in measures of income inequality—after controlling for the effects of observed patterns of pay and manufacturing share—may lie in different cultural views of the nature of income, and in different characteristic responses to efforts to inquire into this topic. But this of course is only conjecture on our part.

We next turn to the question of perhaps greatest interest and controversy in this field. Is household income inequality rising or not? Figure 8a presents unweighted average values of income inequality for each year from DS, grouped into two large categories: OECD and non-OECD member countries. For each group and year, a bar indicates the standard error of the observations for that year.

The answer given by the DS data is somewhat confusing. Overall there is no trend in the data for OECD member countries. There does appear to be a rising trend outside the OECD after 1982, but the average values do not rise above their

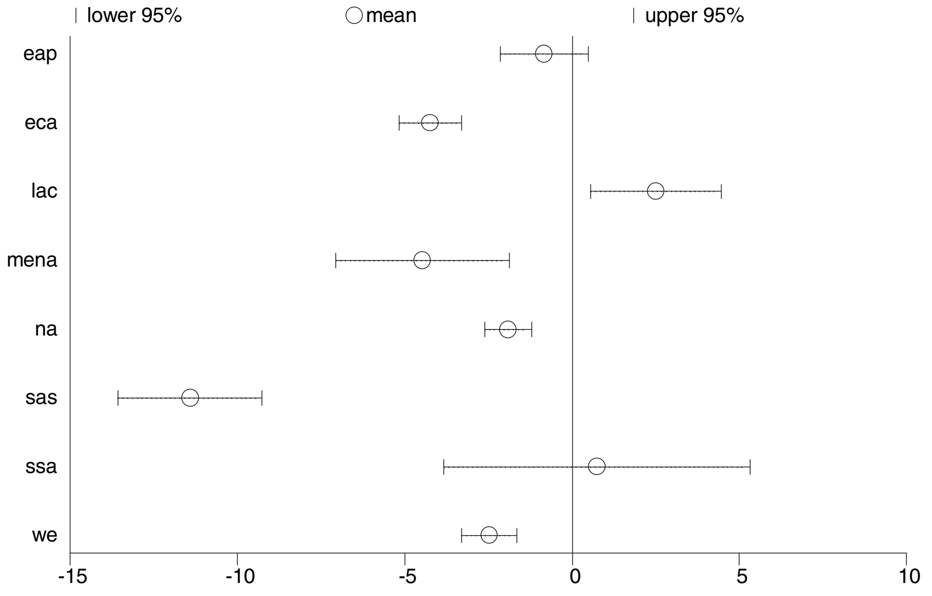


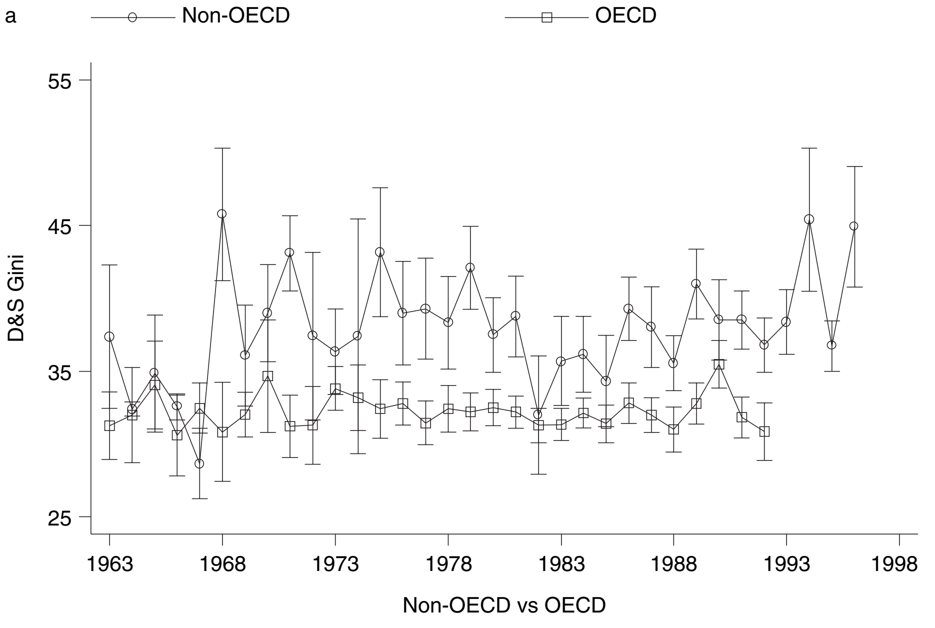
Figure 7. Mean Value and Confidence Interval (95%) for the Difference of DS and EHII

eap: East Asia and Pacific; eca: Eastern Europe and Central Asia; lac: Latin and Central America; mena: Middle East and North Africa; na: North America; sas: South Asia; ssa: Sub Saharan Africa; we: Western Europe.

values in the mid-1960s. And the extent of the upward trend depends very much on the degree to which one credits that a sharp downward trend in average inequality in the developing world from 1979 to 1982—over 10 Gini points in only three years—actually did occur. Of course, it is easier to believe this, than that inequality in the entire developing world jumped nearly 20 Gini points in 1968 alone, or that it bounced down some eight Gini points in 1995, only to bounce back the same amount in 1996.

The main reason for the instability is simply the very sparse and unbalanced character of the DS data set. The sample selection changes so radically from one year to the next, that no very meaningful generalizations can be drawn from movements in the mean or the standard deviation.

Figure 8b gives the answer that would be presented by the EHII data set, were the observations restricted to the same countries and years included in DS. The EHII data set has some clear advantages. The big bump of 1968 is now merely the rebound from a (still-implausible) down-blip in 1967. And it does appear that outside the OECD inequality has reached new highs lately—no doubt partly (as Squire 2002 has recently emphasized) due to the rise of inequality in the post-communist states. Still the implausible downdraft of 1982 remains visible in this data. The reason turns out to be simple: the DS data set for 1982 reports observations only for a handful of non-OECD countries, and all of them (Bulgaria, China, Korea, Hungary, Poland and Taiwan) happen to be low inequality countries in everybody's measures. Similar changes in sample also account for much of the



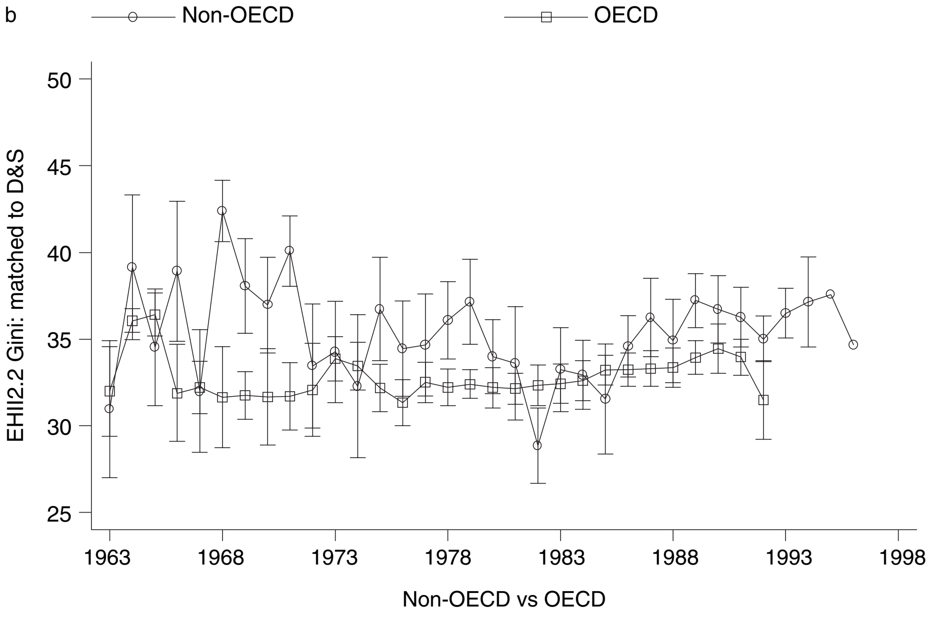


Figure 8. Trends of Inequality in the DS (Top) and Subset of EHII Matched to DS (Bottom)

other year-to-year volatility, especially in 1994–96. And this points out a key pitfall of the DS data set. No matter how accurate the individual data points may be, if coverage is so sparse, variable and erratic, then observations about averages are inevitably at risk for a high degree of selection bias.

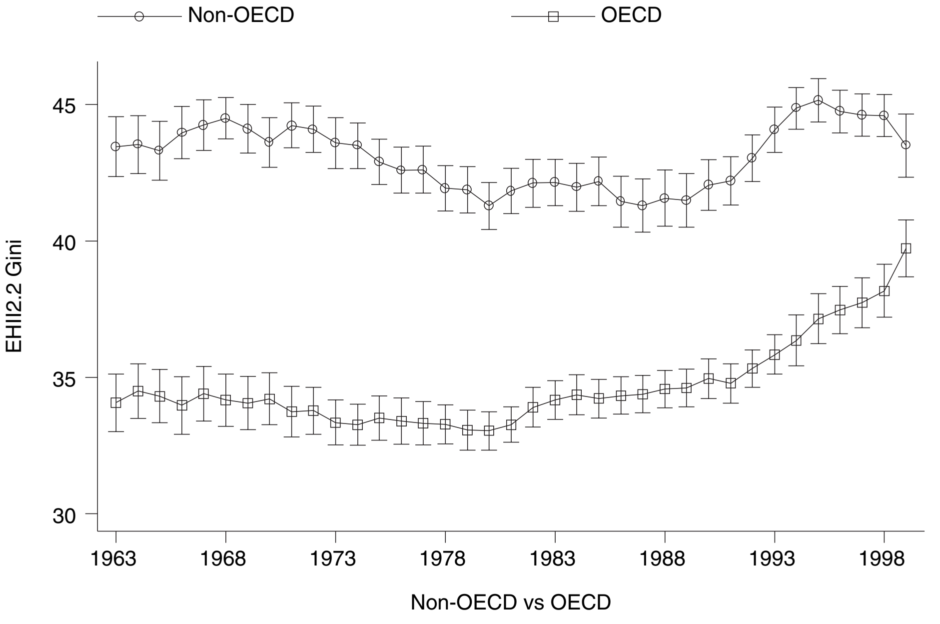


Figure 9. Trends of Inequality in EHII (N = 3,179)

The advantage of the EHII data set, on the other hand, is highly extensive coverage. We illustrate this in Figure 9, which is based on all of our observations. What is instantly visible is the fact that average values stabilize, and standard errors narrow dramatically, when compared to the particular sample of countries and years used by DS. The EHII data set gives fairly unambiguous testimony as to the direction of movement of inequality in the global economy. It is strongly and continually upward for the OECD countries beginning in 1979, which coincides with the advent of Thatcherism and monetarism, and eventually of Reaganism and supply-side economics. This is the period of high real interest rates and enforced liberalization, of steady attack on the welfare state—and it shows. Among non-OECD countries, the relationship between the UTIP-UNIDO and the DS data is likely to be somewhat weaker, since pay (and especially manufacturing sector pay) is a smaller part of a complex structure of formal and informal incomes. It is interesting that a secular downward trend ends in 1982 but a sharp rising pattern, in these measures, only begins around 1987. This finding is in some contrast to findings based on measures of pay dispersion alone (see Galbraith and Kum, 2003), which find the clear upturn in those measures beginning in 1982 for both OECD and non-OECD countries. The period of rising inequality after 1989 appears to peak around 1995 though we suspect that the lower average for 1999 is spurious, owing to lags and missing observations[36] in the reporting of underlying data to UNIDO and other agencies.

[36]The number of countries for the year 1999 is reduced from over 50 to 17.

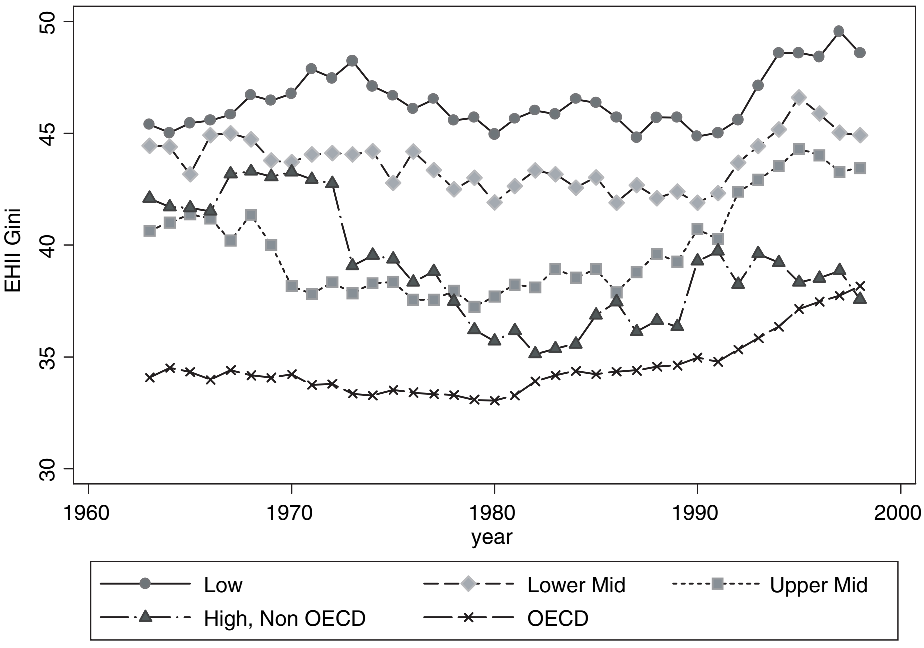


Figure 10. Trends of Inequality in the EHII by Income Level

However, we can report now that rising inequality outside the OECD after 1987 or 1989 is not mainly a phenomenon of the transition countries, as Squire (2002) conjectures. Rather, as shown in Figure 10, it occurs in all income categories,[37] except that of high-income–non-OECD countries—a mixed lot including the small oil sheikhdoms. We find that there is a general pattern of rising inequality in the non-OECD world in the age of globalization, consistent with Galbraith and Kum (2003), but starting somewhat later than they find for manufacturing pay. Second, the long downtrend through 1989 in non-OECD countries is more striking, given (once again) that the EHII data are constructed in part from manufacturing pay inequality data which are clearly rising dramatically after 1982. There may be selection effects here as the composition of the sample changes. However, the most plausible conjecture not involving bias is that increasing manufacturing activity outside the OECD worked to offset the effect of rising inequalities in the pay structure on household income. This would certainly be an interesting twist to the globalization debate. However these data do show—assuming they are estimated with tolerable accuracy—that rising household income inequality did become a general worldwide phenomenon in the late 1980s and thereafter.

[37]This categorization is based on national income level adopted from the World Development Indicators (WDI 2002).

6. Conclusions

The evidence of manufacturing pay dispersions, alongside other broad demographic and developmental indicators, can be brought to bear on the issue of global household income inequality. This approach draws upon the systematic information contained in the World Bank's income inequality data set, while excluding information that cannot be accounted for by statistical means. In so doing, it permits us to extract the more useful measures from the DS data set, while pinpointing and calling attention to the wide range of measures that are, we believe, deeply problematic.

The results suggest several conclusions. First, there is good reason to believe that household income inequality is much more consistently distributed across space than the DS data set would have one believe. Countries similarly situated and economically open to each other (in North Europe, for instance) usually do not display widely differing income dispersions. Second, income inequality measures do not, in real life, change over time with the high speed and amplitude found in the DS numbers, either within countries or cross-country averages. Third, while Gini coefficients above 55 may exist on the planet, outside the Middle East they would have to be accounted for by factors entirely separate from and unrelated to manufacturing pay dispersions, urbanization and population growth. We believe that the literature on high inequality in Africa and Latin America needs to take account of this finding. While inequality on those continents is undoubtedly high, it may not be as high as many have believed. Fourth, we believe there is evidence that inequality in the major countries of South Asia (and also in Indonesia) is much higher than a casual reading of the DS data would suggest. Some of this is clearly due to the reliance on expenditure surveys, and we present here what we believe is a reasonable way to correct for the differences in measurement so introduced.

There is good reason to believe that inequality did in fact rise, through most of the world (but not everywhere) in the age of globalization. These increases are consistently visible in our measures for OECD countries beginning in the early 1980s. The strong correspondence of this trend to previously observed trends in manufacturing pay may reflect the importance of manufacturing pay to income shifts in industrial countries. Outside the OECD, where manufacturing is a smaller and more variable component of economic activity, it appears that the large increases in household income inequality started later. It may be that the large forces of development, including the general processes of industrialization, worked to offset the rise in pay disparities imposed by globalization—until the late 1980s.

Finally, and perhaps most important, we have used the statistical estimates of the effect of manufacturing pay inequality and the other conditioning variables to generate estimates of household income inequality for up to 3,179 country-year observations. We present this data set to the research community for evaluation and comment, in the hope that our approach will help to expand the information available to researchers on the important topics in this area.

Appendix A

Mean Values of Gini for DS and EHII

	Gini Coefficients		Number of Observations	
Country	DS	EHII	DS	EHII
Algeria	38.7	38.5	1	28
Australia	37.9	33.1	9	35
Bahamas	45.8	50.0	11	3
Bangladesh	34.5	42.9	10	26
Barbados	47.2	44.0	2	28
Belgium	27.0	35.1	4	30
Bolivia	42.0	47.4	1	30
Botswana	54.2	46.5	1	15
Brazil	57.3	41.7	15	25
Bulgaria	23.3	30.8	28	36
Burkina Faso	39.0	45.1	1	10
Cameroon	49.0	51.0	1	24
Canada	31.3	35.7	23	37
Central African Rep	55.0	48.0	1	19
Chile	51.8	45.3	5	37
China	32.7	31.0	12	10
Colombia	51.5	44.0	7	37
Costa Rica	46.0	41.4	9	18
Cote d'Ivoire	38.9	47.8	5	22
Czech Republic	22.3	21.2	12	29
Denmark	32.1	30.6	4	36
Dominican Rep.	46.9	46.7	4	23
Ecuador	43.0	45.3	1	37
Egypt	38.0	42.2	4	36
El Salvador	48.4	45.5	1	29
Ethiopia	44.2	44.1	1	9
Fiji	42.5	43.2	1	27
Finland	29.9	32.0	12	37
France	43.1	34.0	7	17
Gabon	61.2	49.4	2	8
Gambia	39.0	44.9	1	8
Germany, West	31.2	31.7	7	32
Ghana	35.1	50.8	4	28
Greece	34.5	42.0	3	37
Guatemala	55.7	48.8	3	26
Honduras	54.5	45.9	7	26
Hong Kong	41.6	29.4	7	27
Hungary	24.6	30.5	9	37
India	32.6	48.4	31	37
Indonesia	33.5	48.7	11	29
Iran	43.2	43.1	5	30
Ireland	36.3	37.8	3	36
Italy	34.9	36.9	15	32
Jamaica	42.9	49.9	9	27
Japan	34.8	36.2	23	37
Jordan	39.2	48.0	3	32
Kenya	54.4	49.3	1	36
Korea	34.2	39.5	14	37
Kyrgyz Rep	35.3	44.9	1	6
Latvia	27.0	28.6	1	6
Lesotho	56.0	50.0	1	7
Lithuania	33.6	39.8	1	5
Luxembourg	27.1	31.3	1	32
Madagascar	43.4	45.0	1	22
Malawi	62.0	49.4	1	32

Appendix A (*continued*)

Country	Gini Coefficients		Number of Observations	
	DS	EHII	DS	EHII
Malaysia	50.4	41.2	6	32
Mauritania	40.2	54.8	2	2
Mauritius	40.7	42.2	3	32
Mexico	53.9	40.2	9	23
Moldova	34.4	36.2	1	9
Morocco	39.2	48.4	2	26
Nepal	30.1	47.5	1	9
Netherlands	28.6	33.5	12	37
New Zealand	34.4	34.7	12	34
Nicaragua	50.3	41.8	1	21
Nigeria	38.5	45.3	3	26
Norway	34.2	32.3	9	36
Pakistan	31.5	45.8	9	30
Panama	52.4	46.7	4	35
Peru	48.0	48.2	4	12
Philippines	47.6	46.6	7	35
Poland	25.7	31.3	17	30
Portugal	37.4	40.0	4	27
Puerto Rico	51.1	45.1	3	15
Romania	25.8	30.2	3	12
Rwanda	28.9	48.7	1	12
Senegal	54.1	44.1	1	24
Seychelles	46.5	36.2	2	11
Sierra Leone	60.8	54.0	1	2
Singapore	40.1	39.0	6	37
Slovakia	20.5	33.6	2	6
Slovenia	27.1	29.0	2	12
South Africa	62.3	43.3	1	33
Spain	27.9	39.5	8	37
Sri Lanka	41.7	45.8	9	17
Sudan	38.7	46.7	1	1
Sweden	31.6	29.2	15	37
Taiwan	29.6	31.6	26	25
Tanzania	40.4	48.9	3	23
Thailand	45.5	48.4	8	19
Trinidad & Tobago	46.2	49.1	4	23
Tunisia	42.5	46.7	5	25
Turkey	50.4	44.0	3	36
U.S.S.R./Russia	26.9	40.0	5	6
Uganda	36.9	50.2	2	14
Ukraine	25.7	36.8	1	9
United Kingdom	26.0	32.5	31	33
United States	35.3	36.6	45	37
Venezuela	44.4	44.4	9	32
Yugoslavia	32.6	42.1	10	5
Zambia	49.6	47.2	4	18
Zimbabwe	56.8	45.3	1	36

Appendix B

Linear Regression Results with UTIP-UNIDO Gini

	Model 1	Model 2	Model 3	Model 4	Model 5
Income	0.245	–0.035	–0.168	–0.15	–0.177
	(8.06)**	(1.14)	(5.60)**	(4.89)**	(6.06)**
Household	–0.152	–0.125	–0.085	–0.075	–0.086
	(7.30)**	(7.28)**	(5.44)**	(4.53)**	(5.49)**
Gross	–0.171	–0.079	–0.034	–0.04	–0.016
	(7.60)**	(4.08)**	(1.93)	(2.27)*	(0.95)
Ln(UGini)		0.336	0.243	0.242	0.213
		(15.15)**	(11.31)**	(11.29)**	(10.38)**
mfgpop			–0.002	–0.002	–0.002
			(11.11)**	(11.20)**	(8.60)**
urban				0.001	0.001
				(2.34)*	(2.94)**
popgrowth					6.002
					(7.56)**
Constant	3.609	2.811	3.173	3.123	3.06
	(246.99)**	(52.04)**	(54.64)**	(49.99)**	(51.41)**
Observations	468	468	468	465	465
R–squared	0.23	0.49	0.59	0.6	0.64

*Significant at 5%; **significant at 1%.

Appendix C

Fixed and Random Effects Model Results with UTIP-UNIDO Gini

	Model 1F	Model 1R	Model 2F	Model 2R	Model 3F	Model 3R
Income	–0.185	–0.029	–0.195	–0.08	–0.211	–0.079
	(3.62)**	(0.74)	(3.89)**	(2.08)*	(4.16)**	(2.05)*
Household	–0.049	–0.06	–0.045	–0.051	–0.047	–0.05
	(2.77)**	(3.45)**	(2.57)*	(3.05)**	(2.70)**	(3.02)**
Gross	–0.037	–0.089	–0.023	–0.061	–0.019	–0.061
	(1.31)	(3.46)**	(0.84)	(2.43)*	(0.68)	(2.42)*
Ln(UGini)	0.192	0.241	0.162	0.19	0.152	0.192
	(8.42)**	(11.57)**	(6.92)**	(8.80)**	(6.36)**	(8.82)**
mfgpop			–0.001	–0.002	–0.001	–0.002
			(4.21)**	(6.66)**	(4.41)**	(6.43)**
urban					0.001	0
					(1.40)	(0.29)
popgrowth					–0.74	0.466
					(1.03)	(0.70)
Constant	2.851	2.921	3.012	3.123	2.959	3.102
	(47.70)**	(46.39)**	(43.12)**	(46.08)**	(33.92)**	(40.38)**
Observations	468	468	468	468	465	465
Number	76	76	76	76	76	76

*Significant at 5%; ** significant at 1%.

References

Acemoglu, Daron, "Matching, Heterogeneity, and the Evolution of Income Distribution," *Journal of Economic Growth*, 2, 61–92, 1997.

Alderson, Arthur S. and Francois Nielsen, "Globalization and the Great U-Turn: Income Inequality Trend in 16 OECD countries," *American Journal of Sociology*, 107, 1244–99, 2002.

Atkinson, Anthony, "Bringing Income Distribution in from the Cold," *Economic Journal*, 107, 297–321, 1997.

Atkinson, Anthony and Andrea Brandolini, "Promise and Pitfalls in the Use of Secondary Data-Sets: Income Inequality in OECD Countries as a Case Study," *Journal of Economic Literature*, 34, 771–99, 2001.

Baltagi, Badi H., *Econometric Analysis of Panel Data*, John Wiley & Sons, New York, 1995.

Berman, Eli, "Does Factor-Biased Technological Change Stifle International Convergence? Evidence from Manufacturing," *National Bureau of Economic Research*, working paper 7964, October 2000.

Birdsall, Nancy, David Ross, and Richard Sabot, "Inequality and Growth Reconsidered: Lessons from East Asia," *World Bank Economic Review*, 9(3), 477–508, September 1995.

Brenner. Y. S., H. Kaelble, and M. Thomas, *Income Distribution in Historical Perspective*, Cambridge University Press, Cambridge, 1991.

Conceição, Pedro and James Galbraith, "Towards a New Kuznets Hypothesis: Theory and Evidence on Growth and Inequality," in Galbraith and Berner (eds), *Inequality and Industrial Change: A Global View*, Cambridge University Press, Cambridge, 139–60, 2001.

Deininger, Klaus and Lyn Squire, "A New Data Set Measuring Income Inequality," *World Bank Economic Review*, 10, 565–591, http://www.worldbank.org/research/growth/dddeisqu.htm, 1996.

———, "New Ways of Looking at Old Issues: Inequality and Growth," *Journal of Development Economics*, 57, 259–87, 1998.

Dollar, David and Aart Kraay, "Growth Is Good for the Poor," *Journal of Economic Growth*, 7(3), 195–225, 2002.

Forbes, Kristin, "A Reassessment of the Relationship Between Inequality and Growth," *American Economic Review*, 90, 869–87, 2000.

Galbraith, James and Hyunsub Kum, "Inequality and Economic Growth: A Global View based on Measures of Pay," *CESifo Economic Studies*, 49, 527–56, 2003.

Greene, William, *Econometric Analysis*, 4th edition, Prentice Hall, 2000.

Grun, Carola and Stephan Klasen, "Growth, Inequality and Well-being: Intertemporal and Global Comparisons," *CESifo Economic Studies*, 49, 617–59, 2003.

Kuznets, Simon, "Economic Growth and Income Inequality," *American Economic Review*, 45, 1–28, 1955.

Milanovic, Branko, "True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone," *Economic Journal*, 112, 51–92, January 2002a.

———, "The Ricardian Vice: Why Sala-i-Martin's Calculation of World Income Inequality Cannot Be Right," Mimeo, September 2002b.

Perotti, Roberto, "Growth, Income Distribution, and Democracy: What the Data Say," *Journal of Economic Growth*, 1, 149–87, 1996.

Pyatt, Graham, "On the Interpretation and Disaggregation of Gini Coefficients," *Economic Journal*, 86, 243–55, 1976.

Ram, Rati, "Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World," *Southern Economic Journal*, 64, 576–8, 1997.

Rodrik. Dani, "Democracies Pay Higher Wages," *Quarterly Journal of Economics*, 114, 707–38, 1999.

Sala-i-Martin, Xavier, "The Disturbing Rise of Global Income Inequality," *National Bureau of Economic Research,* working paper 8904, http://www.nber.org/papers/w8904, August 2002a.

———, "The World Distribution of Income: Estimated from Individual Country Distributions," *National Bureau of Economic Research,* working paper 8933, http://www.nber.org/papers/w8933, May 2002b.

Theil, Henry, *Statistical Decomposition Analysis: With Application to the Social and Administrative Science*, North Holland, Amsterdam and London, 1972.

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

Wade, Robert H., "Globalization, Poverty and Income Distribution: Does the Liberal Argument hold?" *DESTIN* Working Paper, 33, 2002.

UNU/WIDER-UNDP World Income Inequality Database, http://www.wider.unu.edu/wiid/wiid.htm, 2000.

Williamson, Jefferey, "The Structure of Pay in Britain, 1710–1911," *Research in Economic History*, 7, 1–54, 1982.