Inequality and Economic Growth: Data Comparisons and Econometric Tests

James K. Galbraith and Hyunsub Kum

Galbraith@mail.utexas.edu   Kumh@mail.utexas.edu

The University of Texas Inequality Project
LBJ School of Public Affairs
The University of Texas at Austin
Austin, Texas 78713

Abstract

This paper discusses two issues in the relationship between inequality and economic growth: the data and the econometrics. We first review the inequality data set of Deininger and Squire, which, we argue, fails to provide adequate or accurate longitudinal and cross-country coverage. We then introduce our own measures of the inequality of manufacturing pay, based on the UNIDO Industrial Statistics. In our view, these provide indicators of inequality that are more stable, more reliable, and more comparable across countries than those of Deininger and Squire. Turning to the relationship between inequality and development, we diagnose several common econometric problems in the literature, including measurement error, omitted variable bias, serial correlation in longitudinal data, and the possible persistence of lagged dependent variables. By taking steps to account for these problems, we seek more reliable inferences concerning the relationship between inequality, national income and economic growth. We find evidence that generally supports Kuznets' specification for industrializing countries: inequality tends to decline as per capita income increases. However, after 1981 two problems emerge. First, per capita GDP growth slows dramatically in most countries. Second, there is a worldwide trend toward rising inequality in our data, independent of GDP or its changes. The timing and geographic pattern of these increases suggest a link to the high real interest rates and global debt crisis of the period beginning in 1982.

Date of Present Version: April 5, 2002

I. Data Problems in Research on the Kuznets Hypothesis

For students of the relationship between inequality and economic growth, the quality of inequality measures has been a problem ever since Kuznets (1955) formulated his hypothetical inverted "U" relationship between inequality and national income. While some studies have treated this relationship as "stylized fact" (Ahluwalia, 1976) or "economic law" (Robinson, 1976), longitudinal data required to test it rigorously have always been scarce, and measurement controversies persist in available data to this day.

In this respect, Deininger and Squire's effort (hereafter D&S, 1996) is monumental. D&S collected many disparate surveys of income inequality, and compiled those meeting certain criteria of process[1] into a single "high-quality" panel, offering 693 country/year observations since 1950. This is now a standard reference, on which dozens of papers have been based.

However, the D&S data do not generate a consistent relationship between income inequality and either income levels or rates of growth. In different papers many different, even contradictory, forms of this relationship have been specified, including both inverted (Barro, 2000) and upright (Ram, 1997) "U" relationships between inequality and income levels, and both downward (Deininger and Squire, 1998) and upward–sloping (Forbes, 2000) relationships between inequality and subsequent economic growth.

The differences stem partly from data issues. Despite the apparently large number of observations, the coverage of the D&S data set remains limited and unbalanced. And 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, recipient units and processing procedures that cannot be wholly reconciled to each other, even with "high-quality" filtering.[2] <Table 1> shows the various sources of the D&S Gini coefficients by income definition and recipient units.

Table 1. The Distribution of Inequality Measures by Different Definitions in D&S Data

	Recipient 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 Gini coefficient is calculated based on income or expenditure
** indicates whether the recipient unit is the person, the household, the household equivalent (weighted by the number of persons), 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 based on gross or net of taxes.

[1] 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.
[2] Simple ANOVA indicate this fact more clearly. The most variance in the D&S inequality measures stems from differences across countries and through time. But even so, whether inequality is measured on gross or net income and or on income or expenditure generates significant mean differences at the one-percent level.

In order to remedy these inconsistencies in part, D&S (1998) suggest adding 6.6 points (on a scale of 100) to Gini coefficients measured from expenditure data.[3] Li, Squire and Zou (1998) and Forbes (2000) follow this suggestion when they use the D&S data. Barro (2000) uses dummy variables for different recipient units to account for this in his estimations. However, no remedy has fully satisfied the critics. Atkinson and Brandolini conclude, "that differences in definitions may be quantitatively important, but we doubt whether a simple additional or multiplicative adjustment is a satisfactory solution to the heterogeneity of the available statistics. Our preference is for the alternative approach of using a data-set where the observations are as fully consistent as possible."[4]

This consideration partly motivates the present study. In empirical work, we argue, it is important to use comparable and consistently measured data. Statistical transformations to overcome inconsistencies must be doubted. If they are inadequate, they could generate biased regression coefficients in associated econometric work.[5]

A better way is to look for better data. However, such an effort faces severe difficulties. Comparable and consistent measures of income inequality, whether on a household or per head basis, are difficult, almost implausible, to collect in practice.[6] The Luxembourg Income Studies (LIS), now undertaking a meticulous examination of micro-level data, provide the best source for such comparisons. But the LIS coverage is restricted mainly to a few of wealthiest countries and recent years, making it inadequate for a study of the Kuznets relation.

II. Toward a Consistent and Comparable Inequality Data Set

Our strategy is to narrow our focus to the measures of inequality of pay, more specifically to measures of inequality in manufacturing pay. While this may seem an extreme concession, it is motivated by several considerations.

First, pay is a major source of total income. Thus, changes in pay inequality are reflected in income inequality. Indeed, pay inequality has been widely used as an alternative to income inequality in many studies. For example, Williamson (1982) argues that the "wage differential and its development seems to parallel broader trends in income distribution;" Williamson regards pay inequality as a *simplified phenomenon of the evolution of overall inequality* (emphasis added)*."* Acemoglu (1997) identifies increased

---

[3] This adjustment could be valid only if there is little deviation around the suggested mean difference of 6.6. However, our tests show that before adding 6.6, the mean difference between income and expenditures measures of inequality is 1.89 and a t-test against the mean difference being zero is rejected at the five percent level. After adding 6.6, the mean difference is 8.49 and a t-test shows this difference is significant at one percent level.

[4] Deininger and Squire (1996) also agree to be prudent in using their own data set by saying that "the most justifiable way to ensure cross-country comparability of inequality measures is to use only measures that are defined consistently."

[5] This can be regarded as systematic measurement error problem.

[6] In particular, since the only consistent formal definition of "income" comes from tax law, which is nationally specific, a comparative concept of income is not easily achieved even where detailed data on different national income-measures exist.

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 that test the Kuznets hypothesis from measures of pay inequality are collected and reported.

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. If this is the case, as we believe, the behavior of industrial wage inequality becomes an appropriate focus of research into the evolution of income inequality as a whole.

Third, as Barro (2000) points out, recent studies on inequality and development go beyond the shift of persons from agriculture to industry as a source of the evolution of inequality. One new focus is the role of technological change. In Galor and Tsiddon (1997) and Aghion and Howitt (1997), technological change raises the concentration of skilled workers in the advanced sectors against unskilled worker in backward sectors. Of course, manufacturing is the sector most affected by modern technological change. Therefore, if this proposition holds, income inequality would certainly have an inter-industrial feature that would show up in changing pay differentials between advanced and backward manufacturing industries.

Fourth, 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.

Our 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

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

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 denotes group identity.

We capture T^B as our 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.[7]

---

[7] Our measure of inequality is based on variation across industrial categories, necessarily a partial measure. Galbraith et al. (2001), however, provides the theoretical and empirical evidence that within and between

Galbraith (1998) and the papers in Galbraith and Berner (2001) explore the properties of this component of Theil's T in detail.[8]

The UNIDO source permits calculation of inequality measures for nearly 3200 country/year observations, covering over 150 countries during the period 1963 to 1999. We have computed these measures for the University of Texas Inequality Project and refer to them hereafter as the UTIP-UNIDO data set.[9] We then match this data to real gross domestic product (GDP) per capita, as a measure of economic development, from the Penn World Tables version 5.6. Including only countries with 4 or more observations on both variables, this matching reduces our data to 2834 country/year observations. The coverage of observations in region and time is tabulated in <Table 2>. Observations are annual for virtually all of the Americas, Europe, and Asia; only in Africa and for small island countries do we face significant gaps in coverage.

Table 2. UTIP-UNIDO Inequality Measures: Distribution of Observations in Region/ Time

Continent	Before 1965	1966 – 1970	1971 - 1975	1976 – 1980	1981 – 1985	1986 – 1990	1991 – 1995	1996 – 1999
Africa	28	91	111	127	123	87	96	40
Central & North America	24	48	62	58	67	55	49	20
Asia	33	73	87	99	104	100	86	33
Europe	52	99	105	110	115	122	106	48
Oceania	9	17	20	20	24	24	16	5
South America	11	21	27	35	41	46	43	17

The summary statistics of our inequality measures and GDP per capita are seen in <Table 3>. Due to their skewed distribution, both variables are subjected to log transformations for parametric estimation. Distributions of the UTIP-UNIDO Theil measures before and after log transformation are presented in <Figure 1>.

Table 3. Summary Descriptive Statistics

Variable	N	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
Log (Theil)	2834	-3.36	0.97	-6.93	0.03	-0.35	3.17
Log (GDPPC)	2834	8.10	0.98	5.61	10.65	-0.08	2.10

* Theil and GDPPC indicate UTIP-UNIDO inequality measure and GDP per capita respectively.

---

groups inequalities will show similar patterns of dynamics. In Acemoglu (1997), he also notes that within and between measures of inequality show similar patterns of evolution over time.

[8] The most recent version of this data set may be downloaded from the UTIP web-site at http://utip.gov.utexas.edu

[9] To clarify, responsibility for the inequality calculations rests entirely with UTIP; UNIDO is the supplier of the underlying data set.

Figure 1. Distribution of the UTIP-UNIDO Theil Inequality Measures

Several preliminary findings may be noted. First, as shown in <Figure 2>, the United States experienced rising inequality in industrial pay from the early1970s, as did Britain though with more fluctuations.[10] This finding is matched well in many other studies, for instance, Levy and Murnane (1981), Juhn et al. (1983) and Acemoglou (1997). Second, our data provide many single-source observations for developing countries. For example, Singapore and Korea show a downward pattern of inequality in pay since early 1970s, while in South America inequality rises in most countries after 1980. The difference between East Asia and Latin America in the evolution of inequality has been discussed in many studies but no study has shown this point on a more wholly comparable basis.

Figure 2. Inequality in Manufacturing Pay for Selective Countries

---

[10] There is a sharp increase in measured inter-industrial pay inequality in the United States after the 1997,due almost entirely to rising earnings in the computer sector.

<Figure 3a> presents a simple series of unweighted means of log (Theil) UTIP-UNIDO pay inequality measures, annually for developed (OECD) and less developed (non-OECD) countries, together with bands indicating the standard error of the series. From this, we can see that (a) in general, within-group inequality measures are higher for developing countries; (b) both OECD and non-OECD countries experienced increasing pay inequality since the early 1980s, and (c) the gap in pay inequality between developed and developing countries remains nearly steady over four decades.

When the same procedure is applied to the D&S data, <Figure 3b>, great fluctuations both within and between groups are found, from year to year. In 1964, 1966 and 1982, but not in other years, non-OECD countries appear actually to enjoy less income inequality on average than OECD countries. And since the early 1980s, while non-OECD countries appear to have experienced increased income inequality, OECD countries appear to have not.

Figure 3. Time-series of means of inequality measures, OECD and non-OECD countries, with standard errors.

Figure 3a. UTIP-UNIDO Data   Figure 3b. D& S Data

III. Inequality and National Income: An Empirical Analysis

(1) Model and Expectation

The main purpose of this study is to seek for *any* systematic relationships of inequality with economic development. For this purpose, we estimate pay inequality as a function of economic development, as measured by per capita national income. While many additional control variables are discussed in the literature, we confine our focus to a simple, unconditional relationship. We choose two equations.

(1) $Ln(I_{it}) = \beta_1 * LnY_{it} + \alpha_i + \varepsilon_{it}$

(2) $Ln(I_{it}) = \beta_2 LnY_{it} + \beta_3 * (LnY_{it})^2 + \alpha_i + \varepsilon_{it}$

Here, 'Y' indicates GDP per capita measured in 1985 international dollar (GDPPC), and 'I' represents and inequality measure, in this case, the UTIP-UNIDO Theil index. Both variables are in logs.[11] The error term ε_{it} is assumed to satisfy white noise assumptions; subscripts $_i$ and $_t$ indicate country and year respectively. α_i refer to country-specific effects in panel estimation; these effects will capture country-specific differences in excluded control variables.

We expect the relationship between wage inequality and real income to be negative, insofar as (a) poor countries are more unequal than richer countries, as a rule, and (b) increases in average income are associated with declining inequality, especially in industrial pay. Thus in equation (1) we expect $\beta_1 < 0$. Our approach differs from many previous studies that have tried to test the original inverted-U shape of the Kuznets hypothesis. We believe that even though Kuznets' inverted-U curve may be regarded as a general depiction of inequality with respect to income, there is no reason why the (complete or symmetric) inverted-U curve should be found in data regardless of its source, coverage in time and region, and underlying factors. Our data comes from manufacturing payroll, and mostly from after the 1960s. With this restriction in time and character, it is not reasonable to seek evidence for Kuznets' original hypothesis, which was based in part on 19th century experience. Williamson and Lindert (1980) also emphasize this point. They argue that the upward portion of the Kuznets curve is hard to detect; the goodness of fit of an upward portion, if any exists, is not enough to identify an inverted-U in data from the industrial era.[12]

Equation (2) provides another way to test our reasoning. In this equation, $\beta_2 > 0$ and $\beta_3 < 0$ ($|\beta_2| > |\beta_3|$) is usually expected in testing for the Kuznets inverted-U curve. In this case, the expected turning point would be in the middle of observations as shown in (A) below. However, if our data are collected mostly from the downward portion of an inverted-U shaped curve, then $\beta_2 < 0$ and $\beta_3 < 0$ ($|\beta_2| > |\beta_3|$) are possible. In this case, the inverted U curve is asymmetric, with an elongated right tail. Thus, the expected turning point would be on the left of the income scale, as sketched in (B). A third possibility is based on recent findings of rising inequality in several developed countries (Conceicao, 2001). If these observations are accurate, then a new upward turn could be added to the original Kuznets inverted-U curve. In this case, a downward slope could be assumed over most of the range, which means $\beta_2 < 0$ and $\beta_3 > 0$ ($|\beta_2| > |\beta_3|$). The turning point would then be found on the right of the income range as depicted in (C).

---

[11] We employ a log transformation of GDPPC for two reasons: (1) its distribution is much more like the Normal than that of GDPPC, and (2) it is superior in a J-test for a non-nested model (Davidson & MacKinnon, 1981). Additional support comes from the test result for linearity and log-linearity, also proposed by Davidson & MacKinnon (1983) and Greene (2000).

[12] Kuznets also faced this limitation. As Lindert (1991) observes, despite "his fairly certain argument on decreasing inequality with economic growth, he was much less certain about earlier trends, voicing only the hunch that there may have been a slight movement toward wider gaps between rich and poor in the earlier phases of modern economic growth."

(A) (B) (C)

(2) Estimation and Econometric Issues

To begin a process of estimation with prevalent and traditional methods, we apply standard OLS and Huber/White robust estimators to pooled cross-section data. In this case, $\alpha_i = \alpha$ and the time subscripts are ignored in equation (1) and (2). As can be seen in <Table 4>, the estimate of β_1 in linear equation (1) is negative and significant as expected, and the estimates of β_2 and β_3 appear to support an inverted U curve. A non-parametric approach provides similar evidence from another angle. When least-absolute-value (LAV) regression, the running mean smoother and Kernel regression are applied to pooled data, a downward quadratic curve emerges. <Figure 4> is a graphical presentation of these non-parametric regressions.

Table 4. Pooled Cross-section Regression Estimates

Estimator	OLS	White Robust	LAV regression[13]	OLS	White Robust	LAV regression
Log(GDPPC)	-0.426	-0.426	-0.487	0.476	0.476	2.005
	(25.34)**	(5.85)**	(29.15)**	(1.82)	(0.38)	(6.47)**
Square of Log(GDPPC)[14]				-0.056	-0.056	-0.155
				(3.44)**	(0.69)	(8.06)**
Constant	0.091	0.091	0.638	-3.488	-3.488	-9.254
	(0.67)	(0.16)	(4.68)**	(3.33)**	(0.71)	(7.47)**
Observations	2834	2834	2834	2834	2834	2834
R-squared	0.18	0.18		0.19	0.19	

Dependent variable is Log (Theil); relevant t-statistics in parentheses
* Significant at 5%; ** significant at 1%

Several considerations render this cross-sectional approach undesirable. First, we are concerned mainly with the within-country evolution of inequality during the course of economic development, whereas a cross-sectional approach relies mainly on between-country variations. Second, country-specific factors that may be unobservable or excluded from the model are not controlled for in this framework. Third, a cross-section approach is valid only if the relationship of inequality to economic development is homogeneous across countries – i.e., if countries tend to follow identical development

---

[13] This is known as Quantile regression, using the median.
[14] This square term is subject to a severe multicollinearity problem (VIF = 245.05). Koopmans' method (1987) of transformation to avoid this problem was carried out, but the estimated results are not much different from those presented here.

paths separated only by differences in time. Since this is an implausible assumption, it is not safe to rely on the estimates from cross-sectional analysis exclusively.

Figure 4. Nonparametric Regression Curve

Cleveland's running mean smoother

Nadaraya-Watson Kernel Regression

We next turn to panel estimation, usually referred to as fixed-effects and random-effects models. The aforementioned problems are fairly easily handled in this framework. Panel estimation allows us to control for unobservable country-specific effects, which could result in omitted variable bias in cross-sectional regressions. In the fixed-effects model, these effects are handled by adding country-specific dummy variables to the equation.[15] The same logic can be applied to control for equally unobservable time-related omitted variables; this is done by adding a time-specific dummy variable (υ_t) to the equations (1) and (2).

In the random-effects model, country-specific effects (α_i) are assumed to be normally distributed and uncorrelated to any other explanatory variable in the equation.[16] Since these two assumptions appear too strong for our data, our tentative preference is for a fixed-effects model. Particularly, since only log (GDPPC) and its squared term are included as explanatory variables, the likelihood of correlation between country-specific effects and log (GDPPC) would be high. Thus, a fixed-effects model that does not require these assumptions seems more reasonable, despite some loss of efficiency.[17] Table 5 presents the estimates from fixed and random effects models using equation (1) and (2).

Before assaying the interpretation of estimates, we examine several issues relating to the overall fit of the model. All fixed-effects models suffer from heteroscedasticity. In our case, modified Wald test statistics are all significant at any conventional level. However, this is not surprising considering our data and specification: more than 75 percent of variations in inequality stem from cross-country differences rather than from variation through time within country.[18]

---

[15] Greene (2000) also notes that this can alleviate potential heteroscedasticity across countries.

[16] Violations of these assumptions would result in inconsistent estimates.

[17] Whereas random-effects model take into account of both within and between variations for efficiency, fixed-effects model makes use of the variation through time within country.

[18] Only 0.8% of total variation is explained by time variable in D&S data and it is significant at 5% level. In our data, this magnitude is 3%, which is still small, but statistically significant at any conventional level.

Table 5. Panel Regression Estimates

Estimator	Fixed effects [19]	Random Effects[20]	Fixed Effects[21]	Random effects[22]
Log(GDPPC)	-0.011	-0.066	-1.110	-1.194
	(0.31)	(2.06)*	(3.07)**	(3.48)**
Square of Log(GDPPC)			0.066	0.068
			(3.05)**	(3.30)**
Constant	-3.267	-2.764	1.228	1.811
	(11.51)**	(10.31)**	(0.82)	(1.28)
Observations	2834	2834	2834	2834
R-squared	0.19	0.18	0.77	
Countries	118	118	118	118

Dependent variable is Log(Theil); relevant t-statistics are in parentheses.
* Significant at 5%; ** significant at 1%.
F test shows fixed effects are significantly different from zero.

We perform two formal specification tests. 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 regressors 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. Our test results show that a random-effects model provides inconsistent estimates in equation (1) and (2). Based on these test results, the estimates from fixed-effects model appear more robust in present circumstances.

(3) Coefficient Estimates and the Augmented Kuznets Relation

The estimates of β_1 in equation (1) by fixed and random effects models are consistently negative and significant, which corresponds to our expectation. Equation (2), however, suggests another aspect of the evolution of inequality. When country-specific effects are controlled, an ordinary-U shape, instead of the inverted-U, is captured: ($\beta_2 < 0$ and $\beta_3 > 0$) as Fields and Jakubson (1994) and Ram (1997) suggest.[23] The fixed-effects model suggests $4,488 in real GDP per capita as the predicted turning point, while the random-effects model suggests $6,499. Both of predicted turning points are located in the middle-right area of the income scale (50 percentile income is $3,262 and 95 percentile income is $14,128). Can this be evidence to refute our expectation and moreover the Kuznets hypothesis?

---

[19] Modified Wald test for groupwise heteroscedasticity: chi2 (118)=43217.93, significant at any level.
[20] Breusch and Pagan LM test for random effects model: chi2(1)= 12240.71, Hausman specification test against random effects model: chi2(1) = 16.35. Both test statistics are significant at any level.
[21] Modified Wald test for groupwise heteroscedasticity: chi2 (118)=45359.41, significant at any level.
[22] Breusch and Pagan LM test for random effects model: chi2(1) = 11942.24, Hausman specification test against random effects model: chi2(2) = 15.43. Both test statistics are significant at any level.
[23] Especially Ram (1997) uses D&S data and fixed-effects model.

A closer look suggests otherwise. In equation (1) and (2) the error term (ε_{it}) is naively supposed to be white noise, satisfying the standard I.I.D.~$(0,\sigma^2)$ assumption. However, this is not so reasonable in longitudinal data. At least two issues emerge. First, 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 and random-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

(3) $Ln(I_{it}) = \beta_1 * LnY_{it} + \alpha_i + \varepsilon_{it}$
         where $\varepsilon_{it} = \rho\varepsilon_{it-1} + \eta_{it}$

and equation (2) is modified as

(4) $Ln(I_{it}) = \beta_2 LnY_{it} + \beta_3 * (LnY_{it})^2 + \alpha_i + \varepsilon_{it}$
         where $\varepsilon_{it} = \rho\varepsilon_{it-1} + \eta_{it}$

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

The estimation of equations (3) and (4) is presented in <Table 6>. As can be seen, the estimates of β_1 exactly correspond to our expectation. Compared with <Table 5>, the magnitude of the coefficient estimate and its significance level both increase surprisingly. As the autocorrelation coefficient ($\rho = 0.79$) indicates, the serial correlation problem in the error term is serious enough to hamper reliable inference in the earlier specification.

Table 6. Panel Regression Estimates with Autoregressive Error

Estimator	Fixed effects	Random effects	First order Autocorrelation (AR1) Fixed effects	Random effects
Log (GDPPC)	-0.331	-0.190	-0.587	-1.249
	(15.16)**	(3.96)**	(7.34)**	(2.41)*
Square of Log (GDPPC)			0.031	0.066
			(3.32)**	(2.06)*
Constant	-0.596	-1.759	-0.598	2.441
	(16.01)**	(4.55)**	(16.07)**	(1.17)
Observations	2716	2834	2716	2834
Countries	118	118	118	118

Dependent variable is Log (Theil) and relevant t-statistics are in parentheses
* Significant at 5%; ** significant at 1%

The estimates from equation (4) consistently indicate an ordinary-U curve with high significance. However, if this result is examined carefully, it is apparent that most of observations are placed on the downward part of this U-curve, which is surprisingly different from the finding in <Table 5>. The predicted turning points are $12,936 from the fixed-effects model and $12,863 from the random-effects model when AR (1) is

allowed. These are on the far right side of income scale, and thus we have evidence in support of Conceicao's (2001) conjecture, which he calls the "augmented Kuznets hypothesis."[24] This conjecture relates rising inequality in rich countries to a dualism of advanced technology and services, and also takes account of the highly unequal character of certain wealthy monoculture economies, notably the oil principalities of the Persian Gulf region.

A second way to look at the serial correlation problem is more complicated. If serial correlation in residuals (ε_{it}) comes from another source, that is, from some influence of omitted lagged dependent variables, then not only standard errors of the estimates but also coefficient estimates could 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

(5) $Ln(I_{it}) = \gamma_1 * Ln(I_{it-1}) + \beta_1 * LnY_{it} + \alpha_i + \varepsilon_{it}$

However, this model is also under severe restrictions. To get unbiased and consistent estimates, the lagged dependent variable $[Ln(I_{it-1})]$ should not be correlated with current error term: $E(Ln(I_{it-1}), \varepsilon_{it}) = 0$ and the time dimension (t) should be expanded to infinity, which is particularly not feasible in this study. To deal with this problem we adopt the popular method suggested by Arellano and Bond (1991), 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 (5) can be rewritten as

(6) $[Ln(I_{it}) - Ln(I_{it-1})] = \gamma_1 * [Ln(I_{it-1}) - Ln(I_{it-2})] + \beta_1 * [LnY_{it} - LnY_{it-1}] + [\varepsilon_{it} - \varepsilon_{it-1}]$

Two assumptions are critical to get consistent estimates from this estimator. First, the independent variable (LnY_{it}) should be predetermined (weakly exogenous): $E(LnY_{it}, \varepsilon_{is}) = 0$ for $s \geq t$. Second, there should be no presence of second-order autocorrelation in the first-differenced residuals, whereas first-order autocorrelation is allowed. The estimates from this estimator are presented in <Table 7>.

We estimate the equation (6) with and without yearly dummy variables, and so does quadratic equation. We first notice that the model still suffers from heteroscedasticity, invalidating the Sargan test to check the existence of over-identifying restrictions. However, the test for second-order serial correlation is satisfied for each equation. Thus, tentatively we can treat the results as valid. The coefficients from equation (6) are again negative and significant. Again an ordinary U curve emerges from the quadratic equation. In this case, the predicted turning point is $14,445, which is again on the far right side of the income range.

---

[24] See also Galbraith (1998) and Conceicao and Galbraith (2001).

Table 7. Panel Regression Estimates with Autoregressive Error

Estimator	Arellano and Bond			
	w/o year	w/ year	W/o year	w/ year
Log(Theil)	0.600	0.544	0.583	0.372
	(109.84)**	(19.67)**	(66.34)**	(6.39)**
Log(GDPPC)	-0.391	-0.155	-2.452	-2.406
	(20.75)**	(2.99)**	(5.58)**	(2.06)*
Square of Log(GDPPC)			0.128	0.142
			(4.74)**	(1.95)
Constant	0.014	0.014	0.012	-0.002
	(25.17)**	(8.16)**	(18.24)**	(0.11)
Observations	2316	2316	2246	2246
Countries	117	117	99	99

Dependent variable is Log (Theil); relevant t-statistics in parentheses
* Significant at 5%; ** significant at 1%

Our conclusion is that the relationship between pay inequalities and income, for most countries in the period since 1963, remains essentially downward sloping. Increases in income are associated with declining inequality, and poor countries have higher inequality, in general, than rich ones. There are some exceptions to this rule recently, but in our industrial age these mainly lie at the top, rather than at the bottom, of the income scale. Our estimated upward-turn makes the pattern of inequality look like an ordinary U, but with a very short right tail. This pattern is consistent and significant in various estimators, especially when various econometric problems are corrected, all of which suggests that there is something more than by chance. However, the number of cases is too small to generalize from with this data at this stage.

IV. Rising Inequality as a Global Pattern

Our data permit us to go beyond the patterns of recent research in another respect. Virtually all recent work assumes – or concludes (e.g., Dollar & Kraay 2002) -- that national characteristics govern the evolution of inequality and economic growth. From this it follows that national policy choices are the key to higher growth rates.

Our panel permits us to estimate not only country effects, showing the importance of persisting national institutions and industrial structures to (industrial earnings) inequality, but also a full set of yearly time effects. These show the changes in inequality that are common to the world economy. As <Figure 5> illustrates, we find striking evidence of a *global* trend toward higher industrial earnings inequality after 1980, independent of changes in levels of per capita GDP (Galbraith, 2002). Thus, any approach to reducing inequality must address changes in global economic conditions since the early 1980s, one cannot rely on national policy measures alone. We suggest that such global factors as rising real interest rates and the debt crisis that began in 1982 played a strong role in driving up inequality in many countries.

Figure 5. Industrial Pay Inequality: Global Time-effects from a Fixed-effects Model

V. Conclusions

This paper adds to an already-substantial burden of doubts concerning the reliability of the standard source of inequality measures in development economics research. The D&S data set is one from which scholars have been unable to come to a consensus view about a matter of importance, namely the relationship between inequality, national income, and economic growth. We suggest that the limitations of this data set cannot be overcome, as many have hoped, by the application of increasingly sophisticated technique to the raw material.

We have introduced a new source of information about cross-country differences and annual trends in inequality, based on measures of the dispersion of pay across industrial categories in a standard international data set. The advantage of this approach is consistent, comparable and reliable annual measurement for many countries of a variable which, while not representing the whole of income inequality, nevertheless has an undoubted influence on income inequality and is also interesting in its own right for theoretical and practical reasons.

Our variable is particularly important for an assessment of the Kuznets hypothesis relating inequality and economic development, especially insofar as that hypothesis is formulated as a relationship mainly relating national income to inequalities of pay.

We show that there is a clear downward-sloping relationship between inequality and income in this data, vindicating a core premise of the Kuznets hypothesis that inequality would tend to decline in the process of successful industrialization. Most of our observations lie clearly on this downward-sloping surface. However, there is some evidence that for the richest countries the relationship may reverse, yielding rising inequality as incomes increase, and an upright, rather than inverted, Kuznets curve with a turning point at a high income level.

On a discouraging note, we find strong evidence that this (mainly) downward-sloping Kuznets relationship has been shifting relentlessly outward in the years since 1982, both for developing and developed countries. This evidence points to changes in the global economic environment, independent of national policies and income gains or losses, generating a general climate of higher inequalities almost everywhere. Seeking the sources of such a global shift, which we suspect lie in changing global macroeconomic conditions, is a project for continuing research.

References

Acemoglu, D. (1997): Matching, Heterogeneity, and the Evolution of Income Distribution, *Journal of Economic Growth*, 2, 61-92

Aghion, P. and Howitt, P. (1998): Endogenous Growth Theory. Cambridge, MA. MIT Press

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

Allison, P.D. (1978): Measures of Inequality, *American Sociological Review*, 43, 865-880

Anand, S. and Kanbur, R.S.M. (1993): Inequality and Development: A Critique, *Journal of Development Economics*, 41, 19-43

Arellano, M. and Bond, S. R. (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297

Atkinson, A. 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*, 34, 771-799

Baltagi, B.H. (1995): Econometric Analysis of Panel Data, New York, John Wiley & Sons

Baltagi, B.H. and Wu, P.X. (1999): Unequally Spaced Panel Data Regression with AR (1) Disturbances, *Econometric Theory*, 15, 814-823

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

Barro, R.J. and Lee, J.W. (1994): Data set for a Panel of 138 Countries, Available at **http://www.worldbank.org/research/growth/ddbarle2.htm**

Berman, E. (2000): Does Factor-Biased Technological Change Stifle International Convergence? Evidence from Manufacturing, *Boston University and National Bureau of Economic Research*, W7964, October

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

Conceição, P. and Galbraith, J.K. (2001): Towards a New Kuznets Hypothesis: Theory and Evidence on Growth and Inequality, in Galbraith and Berner, 139-160

Davidson, R. and MacKinnon, J. (1981): Several Tests for Model Specification in the Presence of Alternative Hypotheses, *Econometrica*, 49, 781-793

Deininger, K and Squire, L. (1996): A New Data Set Measuring Income Inequality, *World Bank Economic Review*, 10, 565-591, Available at **http://www.worldbank.org/research/growth/dddeisqu.htm**

------------------------------------ (1998): New Ways of Looking at Old Issues: Inequality and Growth, *Journal of Development Economics*, 57, 259-287

Dollar, D. and Kraay, A. (2001): Growth Is Good for the Poor, World Bank Development Research Group, Manuscript

Easterly, W. and Sewadeh, M. (2001): Global Development Network Growth Database, World Bank, Available at **http://www.worldbank.org/research/growth/GDNdata.htm**

Fields, G. (1980): Poverty, Inequality, and Development, New York: Cambridge University Press.

Fields, G. and Jakubson, G.A. (1994): New Evidence on the Kuznets Curve, Manuscript, Cornell University

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

Galbraith, J.K. (1998): Created Unequal: The Crisis in American Pay*,* New York: Free Press

Galbraith, J.K. and Berner, M. eds. (2001): *Inequality and Industrial Change: A Global View,* Cambridge: Cambridge University Press.

Galbraith, J.K. (2002), A Perfect Crime: Inequality in the Age of Globalization, *Daedalus,* Winter, 11-25.

Galor, O. and Tsiddon, D. (1997): The Distribution of Human Capital and Economic Growth, *Journal of Economic Growth*, 2, 93-124

Greene, W.H. (2000): Econometric Analysis. Fourth Edition. Prentice Hall.

Juhn, C., Murphy, K.M. and Pierce, B. (1993): Wage Inequality and the Rise in Returns to Skill, *The Journal of Political Economy*, 101, 410-442

Koopmans, L.H. (1987): Introduction to Contemporary Statistical Methods, 2nd ed. Boston, MA: Duxbury Press

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

Levy, F. and Murnane, R. (1992): U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations, *Journal of Economic Literature*, 30, 1333-1381

Li, H., Squire, L. and Zou, H. (1998): Explaining International and Intertemporal Variations in Income Inequality, *The Economic Journal*, 108, 26-43

Lindert, P.H. and Williamson, J.G. (1985): Growth, Equality, and History, *Exploration in Economic History*, 22, 341-377

Lindert, P.H. (1991): Toward a Comparative History of Income and Wealth Inequality, in Brenner et al., Cambridge University Press

MacKinnon, J., White, H. and Davidson, R. (1983): Tests for Model Specification in the Presence of Alternative Hypotheses: Some Further Results, *Journal of Econometrics*, 21, 53-70

Nadaraya, E.A. (1989): Nonparametric Estimation of Probability Densities and Regression Curves, Kluwer Academic Publishers

Ram, R. (1997): Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World, *Southern Economic Journal*, 64, 576-578

Randolph, S. and Lott, W. (1993): Can the Kuznets Effect Be Relied on to Induce Equalizing Growth?, *World Development*, 21, 829-840

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

Theil, H. (1972): Statistical Decomposition Analysis: with Application to the Social and Administrative Science, Amsterdam and London, North Holland

Ullah, A. (1988): Nonparametric Estimation of Econometric Functionals, *Canadian Journal of Economics*, 21, 625-658

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

UTIP (University of Texas Inequality Project, 2001), available at **http://utip.gov.utexas.edu**

Williamson, J.G. (1982): The Structure of Pay in Britain, 1710-1911, *Research in Economic History*, 7, 1-54

Williamson, J.G. and Lindert, P.H. (1980): American Inequality: A Macroeconomic History, New York, Academic Press