CPB Discussion Paper

No 107
June 2008

Reassessing the relationship between inequality and development

Joseph F. Francois and Hugo Rojas-Romagosa

CPB Netherlands Bureau for Economic Policy Analysis
Van Stolkweg 14
P.O. Box 80510
2508 GM The Hague, the Netherlands

Telephone +31 70 338 33 80
Telefax +31 70 338 33 50
Internet www.cpb.nl

ISBN 978-90-5833-367-4

Abstract in English

We reassess the empirical relevance of the Kuznets Curve with a new inequality dataset. Using panel data estimations that account for the heterogeneity of inequality observations, we test for both the unconditional and the conditional hypothesis that includes alternative inequality determinants. We find that inequality and income levels are related in a cubic function or “tilde-pattern”. This novel finding does not contradict the traditional Kuznets hypothesis, but extends it. Increasing inequality in OECD countries during recent years suggests that inequality rises at high levels of economic development. This “tilde-pattern” is robust to different inequality indicators, estimation techniques and control variables.

Keywords: Income distribution, Kuznets curve, Gini coefficient, Atkinson index
JEL classification: D31, O15

Abstract in Dutch

Wij onderzoeken het bestaan van de Kuznets-curve met een nieuwe database en testen deze hypothese voorwaardelijk en onvoorwaardelijk gebruikmakend van verschillende indicatoren voor ongelijkheid. We gebruiken een paneldata-schattingsmethode die rekening houdt met de heterogeniteit van de data over ongelijkheid. De studie concludeert dat ongelijkheid en inkomen aan elkaar gerelateerd zijn volgens een derdemachtsvergelijking of ‘tilde-patroon’. Deze nieuwe uitkomst is niet in tegenspraak met the Kuznets-hypothese maar breidt deze hypothese juist uit. Toenemende ongelijkheid in de OESO-landen in de afgelopen jaren duidt er op dat ongelijkheid toeneemt met een hoog niveau van economische ontwikkeling. Dit ‘tilde-patroon’ is robuust voor verschillende indicatoren voor ongelijkheid, schattingstechnieken en control-variabelen.

Steekwoorden: Inkomensverdeling, Kuznets-curve, Gini-coefficient, Atkinson-index.

Contents

Summary

We reassess the empirical relevance of the Kuznets Curve with a new inequality dataset. An important limitation in the empirical assessment of the Kuznets curve has been the lack of long and consistent inequality time series. The appearance of the World Bank dataset (Deininger and Squire, 1996) eased some of these data constraints, but their compilation methodology has been strongly criticized (Székely and Hilgert, 1999; Atkinson and Brandolini, 2001; Pyatt, 2003; Galbraith and Kum, 2005). Our new inequality compilation methodology addresses some of these concerns and we account for the heterogeneity of inequality observations given by inequality source (i.e. gross or net income, or expenditure) and the reference units (i.e. person or household).

An improved feature of our econometric estimations is that we use weighted regressions to account for the fact that inequality statistics (and other cross-country statistics in general) are better collected in developed countries. When the error terms are weighted by GDP per capita levels, the Kuznets hypothesis is somehow modified. The quadratic function is no longer valid, but instead, a cubic function is highly significant and robust in all our empirical specifications.

This novel "tilde-pattern" does not contradict the traditional Kuznets hypothesis, but extends it. Increasing inequality in OECD countries during recent years suggests that inequality rises at high levels of economic development. This tilde-pattern is robust to several changes in the empirical specification. For instance, we use two different inequality indicators (e.g. the Gini coefficient and the Atkinson index) and two different panel data regressions techniques. We also test for both the unconditional Kuznets hypothesis, and the conditional hypothesis that includes alternative inequality determinants.

Among these inequality determinants, regional and ex-communist country dummies can account for much of the inequality variance. While the other inequality determinants: educational attainment, a demographic variable and a democracy index are significant, but not robust to different specifications. Moreover, these variables are highly correlated to the per capita income levels, which makes them problematic to use in the Kuznets regressions. We also use three different trade openness indicators, but we do not find any robust relation between them and inequality levels.

Therefore, we establish a strong empirical relation between inequality and income levels, which has been long debated due to the influence of the data previously used. The empirical regularity of the Kuznets curve provides a stylised fact regarding the evolution of inequality in the development path. It highlights that income levels can explain much of cross-country inequality differences by acting as a summary measure of other inequality determinants, such as education levels, redistribution policies and demographics. This empirical evidence contradicts the hypothesis that inequality does not change significantly over time and, in addition, it shows that inequality can still be an important element in poverty reduction strategies.

1 Introduction

In his seminal article of 1955, Kuznets introduced the concept that income inequality first increases and then decreases during the process of economic development. This hypothesis, now well-known as the Kuznets inverted-U curve, has been central to the subsequent literature on income inequality. However, due to inequality data limitations, the empirical testing of the hypothesis has produced mixed results.

Using a newly compiled income inequality dataset (Francois and Rojas-Romagosa, 2005) we find that the relationship between inequality and per capita income levels follows instead a cubic function or tilde-pattern. Thus, the path of inequality can be characterized by five distinctive stages. First, inequality initially increases at low per capita income levels. Then it flattens and reaches a peak in middle-income economies. In the third stage, inequality begins to diminish in high-income countries; and flattens in the fourth stage. Finally, inequality increases again at very high per capita income levels. The first three stages of this process follow the familiar inverted-U Kuznets hypothesis. However, recent inequality increases in developed countries indicates the empirical regularity of the last two inequality paths, which were not accounted for in the original Kuznets hypothesis.

An improved feature of our econometric estimations, is that we use weighted regressions to account for the fact that inequality statistics (and other cross-country statistics in general) are better collected in developed countries. When the error terms are weighted by GDP per capita levels, the Kuznets hypothesis is somehow modified. The quadratic function is no longer valid, but instead, a cubic function is highly significant.

These empirical results confirm the existence of the Kuznets curve in the relevant development levels. However, recent inequality phenomena in OECD countries introduces a new pattern of income dispersion that goes beyond the initial Kuznets explanations. This does not directly contradict the Kuznets hypothesis, but it does suggest that inequality does not monotonically decrease after a certain point on the development path.

These findings are robust to several changes in the empirical specification. For instance, we use two different inequality indicators (e.g. the Gini coefficient and the Atkinson index) and two different panel data regressions techniques. We also test for both the unconditional Kuznets hypothesis, and the conditional hypothesis that includes alternative inequality determinants.

Among these inequality determinants, regional and ex-communist country dummies can account for much of the inequality variance. While the other inequality determinants: educational attainment, a demographic variable and a democracy index are significant, but not robust to different specifications. Moreover, these variables are highly correlated to the per capita income levels, which makes them problematic to use in the Kuznets regressions. We also use three different trade openness indicators to assess the claim that increased trade liberalization has been responsible for the surge of inequality in some OECD countries. Even when these openness indicators have severe limitations of their own, we do not find any evidence that trade openness significantly affects income inequality.

The contribution of this paper is to establish a strong empirical relation between inequality and income levels, which has been long debated due to the influence of the data previously used. The empirical regularity of the Kuznets curve provides a stylised fact regarding the evolution of inequality in the development path. It highlights that income levels can explain much of cross-country inequality differences by acting as a summary measure of other inequality determinants, such as education levels, redistribution policies and demographics. This empirical evidence contradicts the hypothesis that inequality does not change significantly over time and, in addition, it shows that inequality can still be an important element in poverty reduction strategies. On the other hand, the theoretical underpinnings of the Kuznets hypothesis are scarce,[1] and the empirical evidence provided here can be used as a building block for future theoretical work.

[1] Galor and Tsiddon (1996) provide a theoretical mechanism in a general equilibrium context.

2 Literature overview

An important limitation in the empirical assessment of the Kuznets curve has been the lack of long and consistent inequality time series. No single country possess a long enough time series to test directly for the evolution of inequality during all its development stages. The hypothesis is meant to analyse the long-term relationship embodied in any development process and no single country has a long enough time-series that can gauge such a process. For instance, in countries that are already developed, the series do not go further back in time.[2] For developing countries, inequality data are even scarcer and since they by definition are not developed yet, one can only partially test the hypothesis. These limitations have given rise to the use of cross-country regressions, which assume that each country's income level can proxy for the development stages of a prototypical economy. The use of cross-country regressions has well known drawbacks and it is an imperfect substitute for time-series analysis. Nonetheless, given the strong inequality data limitations it is the most widely used methodology.[3]

Most of the literature before 1996 has been surveyed by Kanbur (2000). He claims that the consensus on the distributional effects of growth has cycled in the postwar period. In 1996, with the introduction of the World Bank inequality dataset (Deininger and Squire, 1996), many of the data limitations where overcome. This dataset compiled only inequality observations that fulfilled certain basic quality criteria. In addition, it was larger and more comprehensive than previously available compilations, which allowed the use of pooled estimation techniques. Afterwards, several empirical papers that tested for the presence of the Kuznets curve emerged (Deininger and Squire, 1998; Barro, 2000; Higgins and Williamson, 2002).

Following Higgins and Williamson (2002) we can distinguish two versions of the Kuznets hypothesis. The original (strong) version presented by Kuznets argues that labour demand drives income inequality during the development process. In the early stages of development, labour-saving technological change and structural change (urbanization and industrialization) widens inequality. Later on, these forces slow down and inequality is gradually reduced. In this unconditional version inequality is driven solely by labour demand.[4] Alternatively, the conditional (weak) version recognizes that other factors can also be involved. These factors can reinforce or offset the basic labour demand forces at play –e.g. demographic transitions, resource endowments, governmental and trade policies. For example, trade openness can increase the supply of labour-intensive goods in developed countries and thus decrease the

[2] The longest series is from the United States and starts in 1944.

[3] The alternative approach is to use case-studies, as in the original paper by Kuznets.

[4] The influential paper by Lewis (1954) had the same implications. He assumed two sectors, one traditional with labour surplus and low wages and a modern sector with high wages. Growth was achieved by moving labour from one sector to the other. Inequality initially increased and then decreased, as a bigger share of workers received the higher wages of the modern sector.

income of unskilled workers and increase inequality. Therefore, this version of the Kuznets hypothesis is conditional on alternative factors and provides a better theoretical basis to conduct empirical studies.

Among the authors that tested the unconditional version, Deininger and Squire (1998) and Li et al. (1998) did not find empirical support for the Kuznets curve. On the other hand, the conditional version tested by Barro (2000) and Higgins and Williamson (2002) does appear as en empirical regularity. However, all these studies used the dataset by Deininger and Squire, which has been strongly criticized recently (Székely and Hilgert, 1999; Atkinson and Brandolini, 2001; Pyatt, 2003; Galbraith and Kum, 2005). Thus, although the Deininger and Squire dataset was an important step forward, it did not solve all the measurement problems associated with inequality observations.

Finally, the inverse causal relation between inequality and growth has also been a debated topic. Using different theoretical backgrounds some papers find a negative relation between inequality and growth (e.g. Alesina and Rodrik, 1994; Bénabou, 1996), while other authors find a positive relation (see for example, Forbes, 2000). Banerjee and Duflo (2003) argue that a non-linear relation may also be possible and conclude that the existing empirical results are still premature.

3 Inequality data and empirical specification

To avoid some of the compilation and measurement problems involved with inequality observations, Francois and Rojas-Romagosa (2005) constructed an inequality dataset that consistently uses inequality sources (i.e. income and expenditure) and reference units. This dataset does not mix different inequality sources and for the empirical estimations we use the Gini coefficient based on gross income. In the last section we employ the observations based on net income for OECD countries.

In addition, this dataset includes estimates of Atkinson indexes from inequality grouped data. This alternative inequality indicator complements the widely used Gini coefficient. Formally, the Atkinson index is a family of indicators that diverge on the value of the relative inequality aversion parameter θ.[5] The Atkinson index is more sensitive to income changes at the extremes and the degree of sensitiveness is determined by θ. On the other hand, the Gini coefficient is more sensitive to income variations around the mean. The higher θ the more weight is given to the extremes of the income distribution. However, the index is more volatile to small changes when $\theta > 1$. Thus, the Atkinson indicator is usually estimated using values between 0.5 and 1, following the estimation of the associated relative risk aversion parameter from the macro literature. In this paper we use $\theta = 1$ throughout. [6]

Because our dataset extends into the recent past, we are also able to examine recent inequality trends in the higher-income range, where we are pushing beyond the development levels that underpinned earlier work. In this regard, we confirm recent findings that inequality has been rising in countries at the highest income levels (Gottschalk and Smeeding, 2000; Atkinson, 2003).

There are several ways in which to specify the panel-data regressions and to order the observations. A common feature of cross-country inequality observations is its unbalanced nature, i.e. there are many countries with few scattered observations and some others with complete or almost complete time-series. The later is common for OECD countries, while the former is representative of developing countries. To avoid assigning a higher weight to countries with more inequality observations, we follow Higgins and Williamson (2002) and Barro (2000), and organize the data by decades. However both studies diverge in the way they estimate the decadal observations. Higgins and Williamson use decadal averages, while Barro uses values centered on the years 1960, 1970, 1980 and 1990. In what follows, we name this last approach as “year-centred”, and we use both ways of organizing the data.

When using decadal averages our sample has 81 countries, 14 of which are OECD countries

[5] See Atkinson (1970).

[6] In a previous study, we found that higher values of theta are too volatile and $\theta = 0.5$ does not have much variance. On the contrary, with $\theta = 1$ we have a comparable index to the Gini with respect to levels and variance (Francois and Rojas-Romagosa, 2005).

and the number of countries per decade varies from 43 (in the 1960s) to 57 (in the 1990s). In the case of year-centered values, the sample consists of 82 countries, 15 of which are OECD countries. The number of countries per decade is 36, 40, 42 and 59 respectively.

Moreover, the basic specification has different regressions for each decade, instead of having a regression for each country varying in time, as is the common practice in panel data estimations. Thus, we estimate a panel of four regressions, with each separate regression having a number of observations equal to the number of countries:

$$\begin{aligned} I_{i\_60} &= \alpha + \beta_1 (\log GDP)_{i\_60} + \beta_2 (\log GDP)^2{}_{i\_60} + \gamma \mathbf{D}_{i\_60} + \varepsilon_{i\_60} \\ &\vdots \\ I_{i\_90} &= \alpha + \beta_1 (\log GDP)_{i\_90} + \beta_2 (\log GDP)^2{}_{i\_90} + \gamma \mathbf{D}_{i\_90} + \varepsilon_{i\_90} \end{aligned} \tag{3.1}$$

where I_{i_d} is the inequality measure for country i in decade d. GDP is gross domestic product per capita[7] and **D** is the vector of inequality determinants. Note that in this basic specification, the estimated coefficients are constant over decade and country.

Another divergence between both studies is the estimation technique employed. Higgins and Williamson use random-effects regressions, while Barro uses feasible generalized least squares (FGLS) that account for heteroskedasticity, and temporally and spatial correlated errors. Nevertheless, Beck and Katz (1995) have shown that for samples typically used in social sciences (many sections with few time observations) FGLS yield a small gain in the estimation efficiency, but produce overconfident standard errors. Therefore, we use also correlated panel corrected standard errors (PCSEs) to take account for the contemporaneous correlation of the standard errors between decades.[8] In addition, we later assume a specific form of heteroskedasticity and model it directly into the PCSEs estimations. Comparing the random-effects GLS and PCSEs results, both generate identical coefficients estimates but different standard errors.

To sum up our empirical specifications, we have 8 panel data combinations. Each one varies by the way data is organized (year centered or decadal average), the estimation method (random-effects GLS or PCSEs) and the inequality measure (Gini coefficient or Atkinson index).

[7] The income data is taken from the Penn World Tables (PWT), version 6.1.

[8] We employ a Breusch-Pagan LM test of the independence of the errors across panels and in all specifications we rejected the null hypothesis.

4 Inequality determinants

Besides the two income variables that assess the Kuznets hypothesis, we can include additional inequality determinants and check for the conditional Kuznets curve. Unfortunately, we lack a formal model to explain income inequality and to analyse the rich variety of interrelations between different economic variables and income inequality (Atkinson, 1997). Thus, we do not have a theoretical basis on which to choose inequality determinants, and we follow the literature and present some suggested indicators. To obtain a consistent panel set, we estimate these inequality determinants using decadal averages and year centered values.[9]

4.1 Educational attainment

Following Barro (2000) we include the average years of schooling in the total population over age 15 at three different levels: primary, secondary and higher education.[10] We expect that the primary and secondary education coefficients are negative and that higher education has a positive impact on inequality.

4.2 Country dummies

We add three different country-specific dummy variables. Socialist or ex-socialist countries, Latin American countries and Sub-Saharan African countries. It is expected that socialist countries have significant less inequality given the strong redistribution policies practiced in these societies. On the other hand, the other two types of countries have above average inequality, but the specific reasons for this pattern are not easy to explain. Barro tries to explain their significance by introducing colonial heritage and religious affiliation.[11]

4.3 Demographic factors

Higgins and Williamson argue that cohort size is a fundamental inequality determinant. The basic idea is that fat cohorts are associated with lower earnings and when these cohorts are located in the middle of the demographic distribution it smoothes the lifetime pattern of earnings. On the contrary, when fat cohorts are associated with the youngest population, then it tends to exacerbate the differences in earnings between different age groups and thus, increases inequality.

[9] In particular, we use only values that are not more than four year apart from the initial year of each decade.

[10] The data is taken from Barro and Lee (1994, 2001).

[11] Nevertheless, he does not present these results and uses both dummies throughout his estimations.

They proxy cohort size by the proportion of the adult population between 40 and 59 years and name this variable "mature".[12]

4.4 Democracy and institutional variables

We use a subjective measure of electoral rights (democracy index) from the Freedom House.[13] Barro also includes a subjective indicator measuring the maintenance of the rule of law, but the earliest value available is for 1982 and thus, we do not use it. The same limitation is present for the corruption index that is also compiled by Political Risk Services.

Thus, we only use the democracy index and in general, we expect that more electoral freedom decreases inequality, once we have accounted for the presence of socialist countries.

4.5 Trade openness indicators

There are a number of important analytical and practical problems involved when measuring trade openness (Edwards, 1998). As explained by Berg and Krueger (2003), the main concern is about policies that distort market allocations and there can be many instruments that can achieve this; among others, tariffs and non-tariffs barriers (NTBs), and discriminatory exchange rates.

Direct policy measures like average tariffs and NTBs are not available for the four decades we are analysing. In particular, UNCTAD and WTO data on tariffs, for a broad group of countries, is only available for the mid 1980s and 1990s.[14]

A strategy to overcome this has been to test the robustness of the results using different indicators.[15] Therefore, we use three different trade openness indicators: a) the Sachs and Warner index; b) a measure of trade volumes adjusted by country size and population; and c) the ratio of import duties to total imports. [16]

Finally, these openness indicators are included in the regressions both separately and combined with the level of income. In using this specification we follow the implications of standard trade theory, where the effects of openness are associated with the relative endowments of each country (see for example Francois and Rojas-Romagosa, 2004).

[12] We estimate this variable with data from the United Nations (Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, World Population Prospects: The 2002 Revision, http://esa.un.org/unpp).

[13] Where the index varies from 1 (highest) to 7 (lowest degree of freedom).

[14] In a related study, we use this tariff data to estimate the impact of trade openness on inequality for a cross-country sample centered around 1994 (Francois and Rojas-Romagosa, 2004).

[15] Among others, Levine and Renelt (1992); Edwards (1997, 1998); Greenaway et al. (1998).

[16] The construction and limitations of these indexes are explained in (Rojas-Romagosa, 2005, Chap. 6).

5 Empirical results using the new inequality dataset

First we present the results that are comparable to the previous empirical literature. Afterwards we deal with two econometric problems found in our specification: the high correlation between many of the inequality determinants with income and the presence of heteroskedasticity.

5.1 Initial results

Using our new inequality dataset we estimate the previously specified panel systems, with different combinations of the income variables with subsets of inequality determinants. These results are roughly comparable to those of the previous empirical literature. We obtain five main results:

1. The income variables have the sign and significance expected from the Kuznets hypothesis. The coefficients are robust to the different specifications, although the levels vary in some cases. In particular, we find evidence of both the unconditional and conditional hypothesis. However, the unconditional version has a low explanatory power, which is much increased by the inclusion of other inequality determinants. In Table 5.1 we report our results when using the Gini coefficient and year-centered data. Similar results are obtained with the Atkinson index and using decadal averages.
2. The trade openness variables are not significant in any specification. Nor in combination or not with other inequality determinants. This is true for both the cases were the variable appears alone or in conjunction with the log of GDP per capita. This result could be a consequence of the limitations intrinsic to the openness indicators used, but it can also suggest that trade liberalization plays only a minor role in the determination of income dispersion. Given these results, we do not employ the trade openness variables in the rest of our estimations.
3. Of the rest of inequality determinants, the dummy variables have the expected sign and are all significant throughout the different specifications. On the other hand, the educational attainment variables, the democracy index and the demographic variable (mature) also have the expected sign and in general are statistically significant, but this result is not robust to different specifications. Regressing the equation with the inequality determinants and without the dummy variables, we find that the explanatory power is similar, but lower than the specification where only dummy variables are used. Thus, these dummy variables can be acting as a composite variable that agglomerates different characteristics that affect inequality in different ways for each country.
4. We also tested if the estimated coefficients are stable over time. The Wald test of equal coefficients over decades is not rejected and this is the case for both the conditional and unconditional Kuznets specifications.

Table 5.1 Gini coefficient regressions using year-centered data

	Random effects GLS			PCSEs		
Log GDP	57.12	35.54	48.91	56.04	35.54	48.91
	[12.38]***	[9.83]***	[14.18]***	[14.92]***	[11.53]***	[15.35]***
Log GDP squared	− 3.57	− 2.26	− 2.88	− 3.53	− 2.26	− 2.88
	[0.73]***	[0.58]***	[0.85]***	[0.89]***	[0.68]***	[0.90]***
Dummy socialist		− 12.87	− 14.30		− 12.87	− 14.30
		[1.55]***	[2.91]***		[1.59]***	[2.69]***
Dummy LAC		9.14	5.58		9.14	5.58
		[1.29]***	[1.49]***		[1.66]***	[1.71]***
Dummy SSA		8.68	8.84		8.68	8.84
		[2.08]***	[2.94]***		[2.49]***	[3.17]***
Primary schooling			− 0.53			− 0.53
			[0.61]			[0.69]
Secondary schooling			− 1.47			− 1.47
			[0.84]*			[0.89]*
Higher schooling			6.13			6.13
			[3.69]*			[3.43]*
Democracy index			0.92			0.92
			[0.33]***			[0.36]***
Mature			− 33.82			− 33.82
			[17.29]*			[18.54]*
Number of observations	177	177	142	177	177	142
R-squared	0.19	0.62	0.68	0.19	0.62	0.68

Constant terms are not reported.
Standard errors in brackets, where * significant at 10%; ** significant at 5%; *** significant at 1%

5. To check for country-specific fixed-effects we use a different specification. To allow for unobserved country characteristics, we introduce a country specific constant in the panel system 3.1, so α is then a vector of country-specific terms (α_i), which are constant over time. We use only countries with at least two observations, which do not have to be adjacent.[17] As expected, the income variables are not significant in this specification, although they remain with the same signs. As explained before, four decades is not a long enough time span to measure changes in development levels for a specific country. Furthermore, the exclusion of between-country effects, which accounts for much of the inequality differences, leaves little variance in the sample to be exploited.

[17] This leaves a sample of 40 countries for the year centered series and 41 when decadal averages are used.

5.2 Multicollinearity of the alternative inequality determinants

A problem may arise if the alternative inequality determinants are correlated with the income variables. This is confirmed by the correlation estimates presented in Table 5.2, where the dummy variables are lowly correlated with income, but the rest of inequality determinants are highly correlated with the income variable.

Table 5.2 Correlations with income (Log GDP)

Data organization:	Year-centered	Decadal averages
Dummy socialist	0.054	0.053
Dummy LAC	− 0.154	− 0.148
Dummy SSA	− 0.368	− 0.387
Primary schooling	0.804	0.819
Secondary schooling	0.764	0.779
Higher schooling	0.635	0.651
Democracy index	− 0.616	− 0.683
Mature	0.680	0.725

From Table 5.2, it is clear that there is an multicollinearity problem in our initial specification. To avoid this problem, we only use the dummy variables as alternative inequality determinants, which already add most of the explanatory power that can be obtained when using schooling, democracy and mature together. In any case, our main results are not sensitive to the different uses and combinations of these other inequality determinants.

5.3 Heteroskedasticity corrected regressions

An important characteristic of cross-country data is that richer countries tend to produce better statistics than poorer ones. Hence, measurement error can be associated with the level of development. Using this insight, one can correct for heteroskedasticity by weighting the error term by GDP per capita levels.

We use both the Breusch-Pagan and the Szroeter to test for the presence of heteroskedasticity in our regressions. We run separate cross-section tests for each of the four decades. In each case, we cannot reject the presence of heteroskedasticity.

Although the intuition and heteroskedasticity tests seem very straightforward, this correction has not been widely applied in the earlier papers that tested the Kuznets curve.[18]

When we use weighted panel corrected standard errors (PCSE) the Kuznets curve disappears. This proves that the use of weighted regressions can be critical to the results. In Table 5.3 we

[18] Barro (2000) does not report any significant changes in his results when the measurement error is corrected to account for income levels.

report the regressions for the Gini coefficient the data is organized in decadal averages and year-centered. Similar results are obtained when using the Atkinson index.

Table 5.3 Gini coefficient weighted PCSEs regressions

	Year-centered		Decadal averages	
Log GDP	22.69	20.40	8.77	2.44
	[20.05]	[13.90]	[20.38]	[14.33]
Log GDP squared	− 1.62	− 1.37	− 0.80	− 0.35
	[1.13]	[0.79]*	[1.15]	[0.81]
Dummy socialist		− 13.20		− 11.20
		[1.49]***		[1.74]***
Dummy LAC		9.98		10.34
		[1.56]***		[1.65]***
Dummy SSA		11.75		16.87
		[2.89]***		[3.95]***
Number of observations	177	177	193	193

Constant terms are not reported.
Standard errors in brackets, where * significant at 10%; ** significant at 5%; *** significant at 1%

6 The OECD challenge to the Kuznets hypothesis

In general, we found strong empirical support for the Kuznets curve. The only specification in which the Kuznets hypothesis is rejected is when we weighted the regressions by income, reflecting the likely association of measurement errors and development levels. However, this apparent contradiction can be explained by the recent rise of inequality in OECD countries and the introduction of an income cubic function to account for this event.

The inequality pattern of OECD countries in the last two decades has changed. As observed in other studies (Gottschalk and Smeeding, 2000; Atkinson, 2003; Francois and Rojas-Romagosa, 2005), there is an measurable increase in the inequality levels in rich countries that contrasts with the previous decreasing trend for the 1960s and 1970s. This has produced an U-pattern of inequality in OECD countries for the last four decades. When we run the regressions using only OECD countries[19] the signs of the income variables reverse –indicating an U-pattern instead of the Kuznets curve. In Tables 6.1 we present these results.

Table 6.1 PCSEs regressions for OECD countries, decadal averages data

	Gross income		Net income	
	Gini	Atkinson	Gini	Atkinson
Log GDP	− 94.53	− 133.55	− 371.26	− 327.49
	[126.74]	[181.53]	[86.40]***	[95.94]***
Log GDP squared	4.84	6.77	19.09	16.79
	[6.60]	[9.44]	[4.53]***	[5.02]***
Number of observations	54	51	55	49
R-squared	0.03	0.06	0.35	0.27

Constant terms are not reported.
Standard errors in brackets, where * significant at 10%; ** significant at 5%; *** significant at 1%

Although the coefficients are not significant when the inequality indexes are defined for gross income, they are significant when net income is used for decadal averages. Given the relevant redistribution policies that characterize most of the OECD countries, inequality based on net income is a better concept to measure income dispersion. Thus, we see that there has been a reversal of the inequality trend in our OECD sample.

Furthermore, this reversal of the effects of the income variables can be influencing our results, since the weighted regressions assign more importance to inequality observations from OECD countries. To assess this possibility we introduce a cubic income function and we present the results in Tables 6.2 and 6.3.

[19] Our sample of OECD countries is: Australia, Belgium, Canada, Germany, Denmark, Spain, Finland, France, Great Britain, Greece (only for the year-centered data), Japan, Netherlands, Norway, New Zealand, Sweden and United States.

Table 6.2 Gini coefficient weighted PCSEs regressions using a cubic income function

	Year-centered		Decadal averages	
Log GDP	758.46	409.33	1005.91	674.65
	[214.46]***	[155.00]***	[221.48]***	[155.68]***
Log GDP squared	− 87.19	− 46.52	− 116.61	− 78.26
	[24.97]***	[17.88]***	[25.74]***	[17.99]***
Log GDP cubic	3.29	1.74	4.45	2.99
	[0.96]***	[0.68]**	[0.99]***	[0.69]***
Dummy socialist		− 12.85		− 10.91
		[1.42]***		[1.61]***
Dummy LAC		9.44		9.25
		[1.50]***		[1.53]***
Dummy SSA		12.45		17.68
		[2.86]***		[3.75]***
Number of observations	177	177	193	193
Medium income peak	2,380	2,508	2,368	2,320
High income hollow	19,377	23,087	16,243	16,516

Constant terms are not reported.
Standard errors in brackets, where * significant at 10%; ** significant at 5%; *** significant at 1%

Table 6.3 Atkinson index weighted PCSEs regressions using a cubic income function

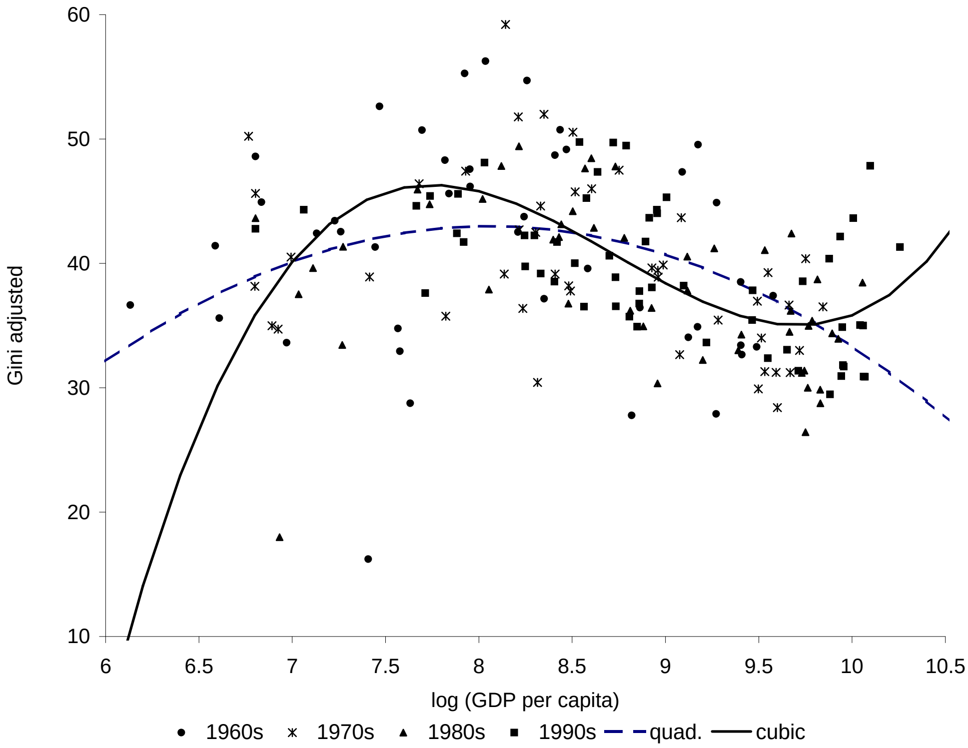
	Year-centered		Decadal averages	
Log GDP	944.62	587.55	1036.12	670.97
	[289.68]***	[238.21]**	[316.80]***	[241.35]***
Log GDP squared	− 106.92	− 66.09	− 119.13	− 77.81
	[33.43]***	[27.26]**	[36.46]***	[27.68]***
Log GDP cubic	3.98	2.45	4.51	2.97
	[1.28]***	[1.04]**	[1.39]***	[1.05]***
Dummy socialist		− 13.05		− 9.81
		[2.35]***		[2.51]***
Dummy LAC		11.47		11.51
		[1.92]***		[2.03]***
Dummy SSA		19.09		27.89
		[4.74]***		[5.44]***
Number of observations	136	136	135	135
Medium income peak	2,747	2,993	2,506	2,362
High income hollow	21,835	21,925	17,801	16,190

Constant terms are not reported.
Standard errors in brackets, where * significant at 10%; ** significant at 5%; *** significant at 1%

Our results clearly indicate the presence of an income cubic function in our data. This specification is robust for all the different combinations of the conditional and unconditional curve using either data organized by decadal averages or centered on the first year of the decade, as well as for both inequality indexes. In addition, we report the two income level turning points for which the inequality trends change direction.

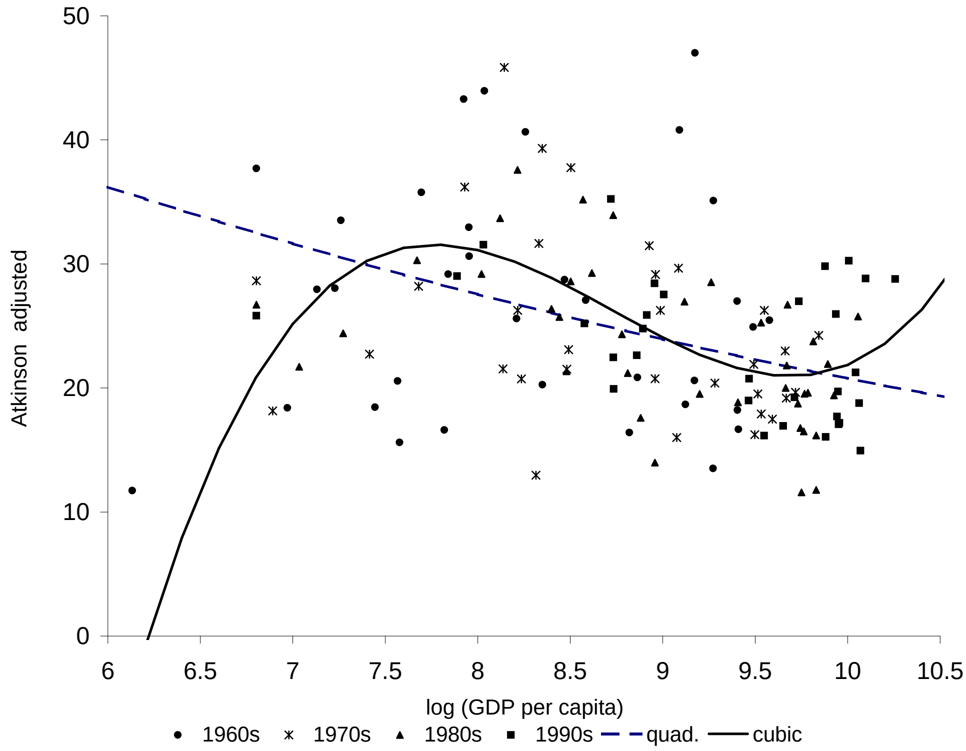
In Figure 6.1 we plot the estimated quadratic and cubic functions with respect to the logs of GDP per capita and the unexplained part of the Gini coefficient when it is adjusted to account for the effect of the dummy variables. The cubic function itself does not contradict the Kuznets hypothesis, but adds an increasing trend after a certain high income threshold level is reached. Moreover, we organized the data by decade, and it can be observed that this threshold level has only been surpassed by some countries in the 1980s and specially in the 1990s. In fact, when we estimated the quadratic functions without the last decade, the Kuznets hypothesis is again valid.

Figure 6.1 Adjusted Gini coefficient quadratic and cubic trends, decadal averages data



Finally, in Figure 6.2 we present the scatter plot when the Atkinson index is presented. For this case, we can draw the same conclusions as for the Gini coefficient. Note that both quadratic specifications have a bad fit into the data and for the Atkinson, we even have a slight U-pattern.

Figure 6.2 Adjusted Atkinson index quadratic and cubic trends, decadal averages data



7 Conclusions

We used an improved inequality dataset that consistently uses the concept measured and thus, reduces the measurement error implicit in the widely used Deininger and Squire dataset. Following previous empirical specifications to test for inequality determinants (Higgins and Williamson, 2002; Barro, 2000) we use panel estimations for four decades and a representative sample of countries. We find clear support for the Kuznets hypothesis, for both the conditional and the unconditional version. Furthermore, this result is consistent over time, econometric techniques and two different ways to organize the data. Throughout the different empirical specifications, we employ the Atkinson index to assess the robustness of the results obtained from the Gini coefficient as an inequality indicator.

Additional inequality determinants also produce similar outcomes to previous studies. In particular, cohort size, democracy, schooling and dummy variables for socialist, Latin American and Sub-Saharan countries are generally significant and exhibit the expected effects. However, these determinants are not robust and they are correlated with the income variables, which creates possible multicollinearity problems when they are included in the regressions alone or together. A simple solution is to include only the dummy variables, which account for much of the increase of explanatory power characteristic of the conditional Kuznets version.

Moreover, we do not find evidence that trade openness influences inequality in a significant way. We use three different openness indicators, combine them with income levels to account for cross-country endowment differences. A possible reason of this lack of influence, is that openness measurement is not adequately captured by the Sachs and Warner index, trade adjusted variables and the share of collected duties with respect to imports. Direct tariff indicators are only available for the 1990s and in a related study Francois and Rojas-Romagosa (2004) we used a cross-country estimation centered around 1994 and found that average tariffs significantly affect inequality as predicted by the Stolper-Samuelson theorem.

When we use weighted regressions to account for data quality heterogeneity among countries, we find that the cubic function is highly significant and robust to all specifications. This behaviour can be explained by the recent inequality trend in OECD countries, where income dispersion has been increasing. Since the observations from these rich countries have more weight, it can account for a new inequality phenomena, that of a tilde-pattern where income inequality increases after a certain high income level is attained.

These results do not contradict the original Kuznets effect, since for a range of incomes the curve is clearly discernible and statistically significant, but it may be a sign that highly-industrialized information-driven economies may be experimenting new inequality consequences from changes in labour demand. This process can be related to the surge of skilled-labour demand widely reported in the literature. In particular, there is a general agreement that the wage inequality increase experienced by most developed countries after the

1980s was caused by a shift in the relative demand for skilled labour. However the sources of this shift are controversial. Most studies argue that skilled-biased technological change (SBTC) had increased the demand and the returns to skilled labour Katz and Autor (1999). On the other hand, some authors claimed that trade liberalization and increased North-South trade had been responsible (Feenstra and Hanson, 2003). While the decline of the welfare state in some OECD countries can also be contributing to this inequality outcome (Gottschalk and Smeeding, 2000; Atkinson, 2003).

Although our empirical results are compelling, our theoretical underpinnings are not. Explaining most of the variation on inequality through dummy variables and per capita income levels is distressing. These explanatory variables are only indirectly capturing the influence of other inequality determinants, such as labour demand and technological changes, governmental redistribution policies and other country-specific characteristics that we are not able to identify. Thus, our findings are stylised facts that must be explained within a comprehensive theoretical setting.

References

Alesina, A. and D. Rodrik, 1994, Distributive politics and economic growth, *Quarterly Journal of Economics*, vol. 109, no. 2, pp. 465–490.

Atkinson, A.B., 1970, On the measurement of inequality, *Journal of Economic Theory*, vol. 2, pp. 244–263.

Atkinson, A.B., 1997, Bringing income distribution in from the cold, *Economic Journal*, vol. 107, pp. 297–321.

Atkinson, A.B., 2003, Income inequality in OECD countries: Data and explanations, *CESifo Economic Studies*, vol. 49, no. 4, pp. 479–513.

Atkinson, A.B. and A. Brandolini, 2001, Promise and pitfalls in the use of ’secondary’ data-sets: Income inequality in OECD countries, *Journal of Economic Literature*, vol. 39, pp. 771–799.

Banerjee, A.V. and E. Duflo, 2003, Inequality and growth: What can the data say?, *Journal of Economic Growth*, vol. 8, no. 3, pp. 267–299.

Barro, R., 2000, Inequality and growth in a panel of countries, *Journal of Economic Growth*, vol. 5, no. 1, pp. 5–32.

Barro, R. and J.W. Lee, 1994, Data set for a panel of 138 countries, Mimeo.

Barro, R. and J.W. Lee, 2001, International data on educational attainment: Updates and implications, *Oxford Economic Papers*, vol. 3, pp. 541–563.

Beck, N. and J.N. Katz, 1995, What to do (and not to do) with time-series cross-section data, *American Political Science Review*, vol. 89, no. 3, pp. 634–647.

Bénabou, R., 1996, Inequality and growth, *NBER Macroeconomics Annual*, vol. 11, pp. 11–74.

Berg, A. and A. Krueger, 2003, Trade, growth, and poverty: A selective survey, IMF Working Paper WP/03/30, Washington D.C.

Deininger, K. and L. Squire, 1996, A new data set measuring income inequality, *World Bank Economic Review*, vol. 10, no. 2, pp. 565–591.

Deininger, K. and L. Squire, 1998, New ways of looking at old issues: Inequality and growth, *Journal of Development Economics*, vol. 57, pp. 259–287.

Edwards, S., 1997, Trade policy, growth, and income distribution, *American Economic Review*, vol. 87, no. 2, pp. 205–210.

Edwards, S., 1998, Openness, productivity and growth: What do we really know?, *Economic Journal*, vol. 108, pp. 383–398.

Feenstra, R.C. and G.H. Hanson, 2003, Global production sharing and rising inequality: A survey of trade and wages, in E.K. Choi and J. Harrigan, eds., *Handbook of International Trade*, pp. 146–185, Blackwell Publishing, Malden, MA.

Forbes, K.J., 2000, A reassessment of the relationship between inequality and growth, *American Economic Review*, vol. 90, no. 4, pp. 869–887.

Francois, J.F. and H. Rojas-Romagosa, 2004, Trade policy and the household distribution of income, CEPR Discussion Paper 4436.

Francois, J.F. and H. Rojas-Romagosa, 2005, The construction and interpretation of combined cross-section and time-series income inequality datasets, World Bank Policy Research Working Paper 3748.

Galbraith, J.K. and H. Kum, 2005, Estimating the inequality of household incomes: A statistical approach to the creation of a dense and consistent global data set, *Review of Income and Wealth*, vol. 51, no. 1, pp. 115–143.

Galor, O. and D. Tsiddon, 1996, Income distribution and growth: The Kuznets hypothesis revisited, *Economica*, vol. 63, no. 250, pp. S103–S117.

Gottschalk, P. and T.M. Smeeding, 2000, Empirical evidence on income inequality in industrialized countries, in A.B. Atkinson and F. Bourguignon, eds., *Handbook of Income Distribution*, Elsevier, Amsterdam.

Greenaway, D., W. Morgan and P. Wright, 1998, Trade reform, adjustment and growth: What does the evidence tell us?, *Economic Journal*, vol. 108, no. 450, pp. 1547–1561.

Higgins, M. and J.G. Williamson, 2002, Explaining inequality the world round: Cohort size, Kuznets curves, and openness, *Southeast Asian Studies*, vol. 40, no. 3, pp. 268–302.

Kanbur, R., 2000, Income distribution and development, in A.B. Atkinson and F. Bourguignon, eds., *Handbook of Income Distribution*, Elsevier, Amsterdam.

Katz, L.F. and D. Autor, 1999, Changes in the wage structure and earnings inequality, in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, vol. 3A, pp. 1463–1555, Elsevier.

Kuznets, S., 1955, Economic growth and income inequality, *American Economic Review*, vol. 45, no. 1, pp. 1–28.

Levine, R. and D. Renelt, 1992, A sensitivity analysis of cross-country growth regressions, *American Economic Review*, vol. 82, no. 4, pp. 942–963.

Lewis, W., 1954, Economic development with unlimited supplies of labor, *Manchester School of Economic and Social Studies*, vol. 22, pp. 139–181.

Li, H., L. Squire and H.F. Zou, 1998, Explaining international and intertemporal variations in income inequality, *Economic Journal*, vol. 108, pp. 26–43.

Pyatt, G., 2003, Development and the distribution of living standards: A critique of the evolving data base, *Review of Income and Wealth*, vol. 49, no. 3, pp. 333–358.

Rojas-Romagosa, H., 2005, *Essays on Trade and Equity*, Tinbergen Institute Research Series, Book No. 349, Thela Thesis, Doctoral dissertation, Erasmus University Rotterdam.

Székely, M. and M. Hilgert, 1999, What's behind the inequality we measure: An investigation using Latin American data, Research Department Working Paper 409, Inter-American Development Bank.