ORIGINAL PAPER

Inequality and development: is the Kuznets curve in effect today?

Diego Martínez-Navarro[1] · Ignacio Amate-Fortes[1] ·
Almudena Guarnido-Rueda[1]

Abstract

The aim of this work is to study the Kuznets curve in order to examine whether the hypothesis on inequality and development that he posited in his 1955 article is verified or not when using the data at our disposal today; these data are more numerous, both for countries and periods available, than when Kuznets originally conducted his study. The approach that makes this research unique is that it will be performed by differentiating the sample in terms of underdeveloped and developed countries. In this regard, at present (with the data and methodologies of Cochrane–Orcutt and GMM System), the Kuznets hypothesis seems to be robustly verified because, when taking a variable other than the Log GDPpc (GDP per capita in logarithms) as a measure of development, such as the HDI or the proportional contribution of the agricultural sector on GDP, the relationship described by Kuznets still seems to be present; this is not a regularity when using the basic GDPpc variable. Moreover, it has been observed that, over the very long term, the Milanovic hypothesis seems to appear; namely, that inequality follows a sinusoid form rather than a concave curve. Finally, a section has been included in which we see how the 3 effects (scale, technique and composition) of world trade on inequality affect, as has been applied in recent years on CO_2 emissions in the Environmental Kuznets Curve.

✉ Diego Martínez-Navarro
   dmn011@ual.es

   Ignacio Amate-Fortes
   iamate@ual.es

   Almudena Guarnido-Rueda
   guarnido@ual.es

[1]   Faculty of Economics and Business, University of Almería, La Cañada de San Urbano, Ctra. Sacramento s/n, 04120 Almería, Spain

1 Introduction

Why does income distribution vary between countries? This has been one of the great issues of the last century. Why greater inequality exists in some countries than in others has been studied using several explanatory factors; yet even today, no single answer has emerged. Taking advantage of the increased availability of information, although not abundant, these studies have tried to find solid conclusions. In this paper, we attempt to contrast one of the theories that has generated most controversy regarding inequality; that proposed by Simon Kuznets in 1955 in his article "Economic growth and income inequality". Our novel approach uses the GDPpc in logarithms as the measure of development, expanding the model with variables that collate the country's level of democracy and education, as well as a classification that divides the sample into developed and underdeveloped countries. This theory needs revisiting because of the relevance of inequality seen in present-day economies, and for its capacity to undermine welfare and generate crises, as observed in Matos (2019).

Kuznets's proposal, based on his empirical observations from the U.S. in the 20th Century, advocates an inverted U-shaped relationship between economic development over time and inequality, arguing that economic development cannot be equitably distributed at every stage of a developing economy. Hence, as a traditionally agricultural economy grows, it will mainly benefit landowners, generating inequality; a situation that can then be alleviated by the tax decisions taken by the State, and other factors such as demographics. In other words, economic development will result in an income distribution that creates divergences in the early stages of development, reaching a point when it begins to generate more equitable distribution in the following stages of development.

The purpose of this work is to review the literature regarding Kuznets's inverted-U curve and then examine it using graphical and econometric analyses to provide empirical evidence in its favour from the spectrum of countries for which information is available. Contributing to the question by using an enlarged study sample (187 countries), the study method reduces the sample dispersion using the log GDP per capita, the econometric techniques employed and the approach that discriminates between countries based on whether they are developed or not (the criterion being whether they belong to the OECD or not) as well as studying the sample in its entirety.

Firstly, we will approach the Kuznets curve by detailing what it is, as formulated by its author Simon Kuznets in 1955, then we will present an overview of the literature written about it, both supporting and opposing the hypothesis. Following this, we will be in a position to enter the empirical realm: starting with a graphical analysis, a simple way for allowing us to intuit the results to be found in the subsequent econometric analysis, describing the model we will use and the data and methodology employed to understand the obtained results. These will be subjected to a robustness analysis by substituting the development measure (Log GDPpc) with other development-related measures (HDI and agriculture aggregated in proportion to GDP). This will be followed by an analysis of how scale, technical

and composition effects influence inequality. Finally, our own conclusions will be presented.

2 Theoretical framework

The Kuznets curve is the graphical representation of the relationship between two variables; one that reflects the level of development and the other that measures the inequality of a country. Kuznets argued that as soon as an economy began to develop from its subsistence level, economic inequality would increase due to two factors: first, because of the concentrated savings of the higher-income population–higher-income individuals save part of their income while those with lower incomes show savings near to zero. In the long run, this has a cumulative effect, which translates into a growing proportion of assets (saving performance) amongst the wealthy classes and their descendants. Secondly, industrialization—knowing that economic development progressively relegates primary activities for industrial ones, and that inequality amongst the rural population is lower than that for the urban population, the increase in the population dedicated to industry, moving from the countryside to the cities, will provoke increased inequality in the country. In addition, the difference between rural and urban populations per capita income does not necessarily diminish, even finding evidence that it is maintained or increased because per capita productivity in industry is higher than in agriculture. These are the two arguments that underlie the ascendant inequality proposed by Kuznets.

In the same way, once a country reached a certain level of development, the structure would change, favouring the greater influence of four factors over the two already explained, impeding rising inequality and even diminishing it, resulting in a long-term fall in inequality. The four factors behind the curve's descent are: first, the action of the State through direct measures such as taxation (inheritance taxes) or spending policies (social benefits); or indirect measures such as decreasing the value and/or performance of assets through monetary policy. Second, demographics—the birth rate amongst the wealthy classes is lower than that for the poorest classes, meaning there are ever fewer individuals from rich families and ever more from poor families, with its pertinent statistical effect (reducing inequality since there are more members within the low-middle income group). Third, technological change—the assets generated by industry lose out to the new technologies generated in the sector; thus, if the rich and their descendants do not sell their "old technology" assets to acquire rights over the " new technology ", their long-term gains will be reduced as the old technology loses value in the face of new developments. Fourth and finally, the income realised by the service sector population increases with respect to incomes from agriculture and industry; there are high-paying jobs in the service sector. Nonetheless, the wealthy classes have little chance of increasing their income by occupying these positions because they are already part of a very high income group, a totally opposite position to that of the middle and low-class members of the society, who, by accessing these positions, increase their income significantly, reducing the level of inequality in the economy.
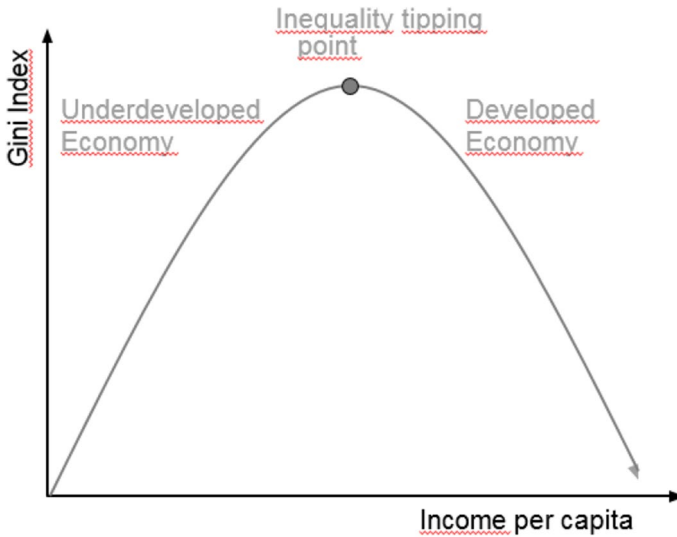
Fig. 1 Kuznets curve Source: compiled by the author

In short, Kuznets argued that the concentration of savings and migration from the countryside to the city would create inequality in the first instance. However, productivity differentials (which would lead to the reallocation of resources), the generation of new job opportunities, and the birth rates of the various segments of the population, would reduce the inequality created initially. That is why the form displayed by this hypothesis is inverted, as seen in Fig. 1. Interestingly, the empirical evidence has shown that if the abscissa axis is formulated in terms of time rather than expressed as per capita income, this hypothesis and curve also appears to be fulfilled (Lempert 1987).

Now that we know the Kuznets hypothesis, we must comprehend the measurements of economic development and economic inequality used for it. Firstly, the economy's production of goods and services is established; that is, the gross domestic product (GDP) for a period in per capita terms so that this production measure takes the country's population into account—the result being the GDP per capita (GDPpc), an indicator of a country's economic development.

The inequality can be measured based on different indices: The Theil Index (Theil 1979), the Palma Index (Palma 2016), or the best known of all and the one used in this work, the Gini Index (Gini and Edgeworth 1922), among many others. All these measures of inequality in some way try to quantify what the Lorenz curve shows.

Kuznets published his article in 1955, adding in the conclusions "this work is perhaps based on 5% empirical information and 95% speculation" (Kuznets 1955, p. 26). He invited successive generations of economists to continue investigating the issue he had launched in his article for the American Economic Review. The studies carried out by Fields (1989), Deininger and Squire (1996, 1998), Higgins and Williamson (1999), Barro (2008), Prados de la Escosura (2008) and Rattan

(2012) have provided evidence both to affirm the veracity of the Kuznets hypothesis and that no significant proof exists that said hypothesis is fulfilled.

To begin, we can cover the studies that support the Kuznets hypothesis by providing evidence of an inverted U-shaped curve among the countries studied.

In this regard, Barro (2008) confirmed the Kuznets curve in the period from 1960 to 2000 for a sample of heterogeneous countries; this study also added the effect that "openness to trade" had on economic inequality. The study found a relationship, a non-significant yet positive effect, suggesting that for a certain income, more trade would generate more economic inequality. However, in the same article, Barro (2008) states that the increase in trade could increase per capita income; therefore, even if it generated inequality, it would simultaneously have the compensatory effect of reducing poverty.

Higgins and Williamson (1999) also showed evidence of the Kuznets curve for a spread of countries around the world between 1969 and 1990. Their study had the peculiarity of being divided by age groups and openness to trade. In this case, they found that the adult cohorts had less economic inequality while the cohorts of young people showed greater inequality at the aggregate level. They observed that the developed nations, which tend to lengthen their life expectancy, having larger groups of elderly people than in developing nations, showed lower levels of inequality. As with Barro (2008), Higgins and Williamson (1999) found that the impact of globalization on inequality is small.

Curiously, as has already been mentioned, there are also numerous studies showing evidence that the Kuznets curve does not exist, and that its hypothesis was a fortuitous prediction of the changes in inequality that developing countries would experience.

Deininger and Squire (1996, 1998) developed a database looking at the countries for which they had information on economic inequality and inequality around the world in 1996. When analysing these countries, they found that the highest levels of inequality were in Latin America, the Caribbean and sub-Saharan Africa, with Gini indices of nearly 50%, whereas developed countries showed low Gini indices—the Kuznets hypothesis seemed to be fulfilled. However, Deininger and Squire (1996) did not find any systematic relationship between the aggregated income growth and changes in the Gini coefficient when comparing the changes in the inequality experienced in their sample countries during a decade that showed economic development. They found that for half the period, inequality increased, while for the other half, it decreased, arguing that the changes in the Gini coefficients are modest and not explained by changes in income; the only effect of these being to reduce poverty in periods of economic expansion. Furthermore, 2 years later, Deininger and Squire (1998) reaffirmed their hypothesis, showing that for low-income countries, the income coefficient relative to reduced inequality was only positive for two of the sample countries. An effect that disappeared by including in the study a fictitious variable for the countries that belonged to Latin America, leaving the coefficient negative. With these results showing that there is little empirical evidence of any Kuznets curve, and that cross-sectional studies can be misleading because Latin American countries as a whole have a middle-income level and generally show inequality levels belonging to high income countries.

Fields (1989) provided evidence showing that the Kuznets curve was not always fulfilled by looking at economic growth instead of development. According to this study, inequality increased at the same frequency in low-income countries during times of economic growth as in high-income countries when experiencing the same phenomenon. So, the only evidence for changes in inequality generated by economic growth came from reducing poverty by increasing national income. Moreover, Fields (1989) also studied the differences in inequality between Asia and Latin America, concluding that, in Latin America, there were higher rates of rising inequality than in Asia; however, statistically, the results did not diverge significantly.

The conclusions made by Deininger and Squire (1996, 1998), and by Fields (1989), formed the basis of what Rattan (2012) called "The Latin American Effect"; this being the distortion caused by countries in Latin America presenting middle-income levels and high inequality.

An interesting case that discards the "Latin American Effect" of Rattan (2012), is the extensive study by Prados de la Escosura (2008) regarding the individual case of Spain from 1850 to 2000. In this, Prados de la Escosura observed that inequality grew during periods of political instability whereas it decreased in periods of economic growth; thus, in terms of inequality, the Kuznets hypothesis was fulfilled according to the study. The study is relevant because, from colonial times to 1950, Spain followed the same pattern as Latin America. The pattern was broken in 1950, when it began converging with the most developed countries, especially those in Europe (decreasing its Gini index) as a result of the economic growth that Spain experienced during the 1950s. This has therefore been interpreted that Latin America, as with Spain then, has not yet reached its turning point on the Kuznets curve so it is still too early to consider whether the hypothesis is fulfilled or not for this economic bloc. In conclusion, according to this research, Latin America is still in the ascendant part of the Kuznets curve, just as Spain was before the 1950s; for this reason, the Kuznets hypothesis is not yet accepted when studying this group of countries.

More recently, in 2013, Piketty expanded the work done by Kuznets (1955), with the difference that: (1) economic inequality was calculated using taxpayers' declared income; (2) instead of using the GDPpc (a measure of development), it used the GDP growth rate (a measure of growth); and (3) it used more countries and periods than Kuznets. In particular, Piketty mainly analysed the pattern of inequality and income in France, Great Britain, Germany and Sweden, the United States and Japan. He noted that there has been a widespread tendency for increasing economic inequality in developed countries since the 1950s, he explains it as being because of income concentration, understanding income as the right of property over capital, land, stocks and bonds etc. If we assume that increases in GDP lead to increases in GDPpc, considering that developed countries have low birth rates, this would mean that the Kuznets prediction—that countries with high GDPpc levels are those with the least economic inequality—is not met.

Piketty (2013) proposed that the main source of economic inequality is: $r > g$; r being the average capital return rate (i.e. benefits, interests, income, dividends, etc.) and g being the GDP growth rate, which also represents population growth. Pointing out that the reduction in inequality in the United States during the first half of

the C20th that Kuznets noted, was not only due to natural market forces (i.e. to the USA's economic growth) but also the reduction in the capital return rate along with the simultaneous increase in the economic growth rate. Finally, Piketty concludes that, in order to reduce inequality and contain the force that generates this divergence, and thus fulfil Kuznets's prediction of a future with less economic inequality, then a tax on capital must be imposed to reduce income accumulation amongst the world's richest population cohort, and to transfer this income to the workers.

The Piketty book has also inspired controversy. His suggestion of using taxes to redistribute wealth has prompted defenders of free markets and capitalism to respond by trying to challenge his argument. Thus, authors such as Magness and Murphy (2015), McCloskey (2014) and Henderson (2014) have reviewed the Piketty database meticulously and claim that Piketty manipulated the data in order to support his hypothesis, questioning the overall validity of Piketty's thesis.

In addition, Piketty's model itself is subject to criticism; the most notable being the one made by Acemoglu and Robinson, authors of a relevant research study in 2012.[1] They noted that for Piketty's model, it is necessary to save 100% of the revenues earned by the owners of capital, a hypothesis that could never be fulfilled since they must allocate a minimum to consumption, and most likely invest owing to the pressure exerted by competitors. Another scholar of inequality, Lemieux (2016), criticized Piketty's research (2013) for focusing on the richest 1% in the population, noting that the remaining 99% can behave in a different way; thus limiting Piketty's research to the richest population cohort, while unable to explain the sources of the real inequality that affects the rest of the population. Lemieux (2016) criticizes Piketty's study (2013) as being specific for a certain segment of the population rather than a general study, as Piketty tries to present.

The fundamental difference between the conclusions made in the Piketty (2013) theory to that of Kuznets (1955) is the suggestion that, if tax is not imposed on capital in the future, economic inequality will tend to increase; Kuznets predicts the opposite, that it will tend to decrease, as has been seen before, and as is shown in the empirical evidence presented in the Sect. 3.2.3. Indeed, when the Kuznets curve is studied using the Gini variable for the whole sample, one observes a concave curve, like that described by Kuznets. However, if we differentiate between underdeveloped and developed countries, one can see that, in the very long term, the graph takes the form of a sinusoid, with an upward trend in its final stretch, as Piketty stated. In his book, he declared that if tax is not applied on capital to transfer income from those who earn more to those who earn less, inequality will indeed grow; this hypothesis is supported by Milanovic (2016), who also speaks of a sinusoidal form (as shown in Fig. 2) regarding inequality and economic development. Nonetheless, Milanovic, unlike Piketty, argues that the origin of the "second Kuznets curve" (this is how Milanovic defines the resurgent inequality) is due to the simultaneous action of technological progress and globalization, which come from an increased imperative to reduce fiscal pressure

---

[1] Acemoglu, D., and Robinson, J. TO. (2012). Why countries fail [_ 1]. Editorial Planeta Colombiana, Bogotá.
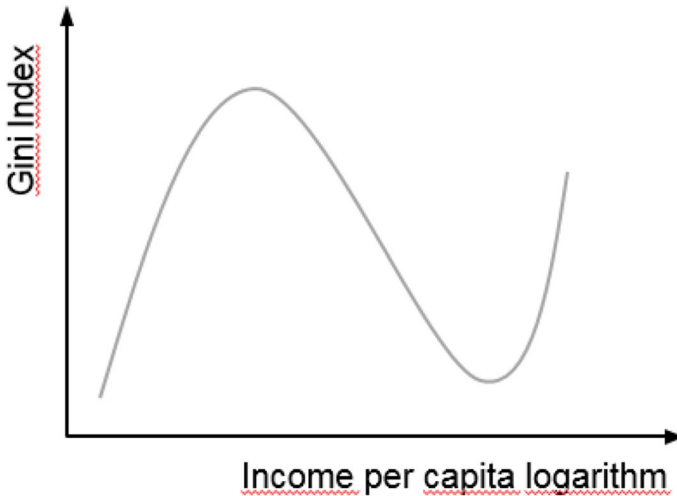
Fig. 2 Graphic approximation of the Milanovic sinusoid Source: compiled by the author

on mobile capital and skilled work, and the disappearance of the middle class in the West. Furthermore, Milanovic offers other explanations for the upturn in inequality, such as homogamy (people with high purchasing power marrying highly trained people) or the growing importance of money in politics, which enables the wealthy classes to impose rules that are favourable to them by financing political campaigns, thus reinforcing the dynamics of inequality. Consequently, we cannot reject Kuznets's or Piketty's proposals.

The Kuznets-curve hypothesis remains a hot issue even today, as indicated by the many articles written about it annually. You can find contributions that support it, such as those by Zhang (2014), Utari and Cristina (2015), Jauch and Watzka (2016), Nielsen (2017), VanHeuvelen (2018) and Comin (2019) or refute it, such as by Yusuf et al. (2014), Kiatrungwilaikun and Suriya (2015), Meneejuk and Yamada (2016), Kanbur (2017), Costantini and Paradiso (2018) and Baymul and Sen (2019). In addition, there are those who apply the Kuznets theory to other areas, as with Sulkowski and White (2016), who posited the Kuznets curve of happiness, or more recently, Auci and Trovato (2018), who successfully applied the Kuznets hypothesis to the environment by substituting the level of inequality with CO_2 emissions.

In summary, studies on the Kuznets curve have produced heterogeneous results. Nevertheless, it is possible to gather conclusions which seem to be met. On the one hand, all point to the existence of a relationship between economic development and economic inequality, although some, like Fields (1989), pointed out that, although this relationship exists, the Kuznets hypothesis was based on a group of countries, those of Latin America amongst them, which led to erroneous results. On the other hand, from the evidence provided, it also seems to be accepted that inequality itself has a detrimental effect on economic development, thereby fostering and justifying redistributive policies to improve economic growth and development.

Fig. 3 Kuznets curve (1970) with atypical value logarithmic scale Source: compiled by the author

Quadratic adjustment of inequality in the cross section of countries with a logarithmic scale (1970). With atypical value.
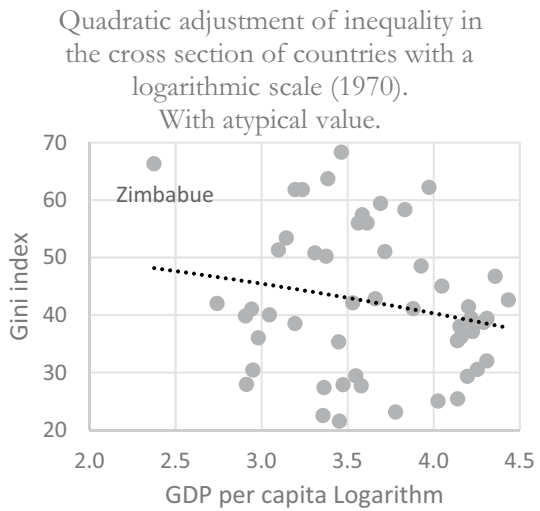
3 Empirical section

Initially an approximation of the data will be performed through graphical analysis. This will later be supplemented by econometric analysis.

3.1 Graphical analysis

Some economists, such as Barro (2000), Prados de la Escosura (2008) and Gallup (2012), often use econometric analysis to study the Kuznets hypothesis. So, this method has also been used to prove its existence. The simplicity of a graphical study can help form an intuition about it, which is later reinforced by the econometric study.

Initially, a graphical analysis was carried out for each decade from 1970 to 2010 to appreciate the evolution of the hypothesis raised over time, and finally for all the information as a whole. Moreover, it has been the medium used to detect atypical data that distort the analysis. Data are considered atypical if they clearly diverge from the rest of the point cloud. In order to clearly appreciate this discrimination, look at the following table (Fig. 3), which shows what Fig. 4 would look like if we had not omitted the atypical value of that graph; then, we see how the trend in Fig. 3 that includes the observation referring to Zimbabwe (considered atypical observation) disrupts the trend remarkably. Other atypical values include Zambia and Madagascar in 1980, Tanzania and Uganda in 1990, and Luxembourg and Singapore in 2010. In the 2000s, there were no indications of any atypical values. Likewise, it has been confirmed that these data are atypical when using the typical deviation; thus verifying that the typical deviation of the above-mentioned values exceeds the typical sample deviation by at least 3 times—this is analogous to the study performed by Vargas-Quesada et al. (2017) with the interquartile range used as the data's spatial proximity measurement. It is essential for the sake of the study to eliminate this data

Quadratic adjustment of inequality in the cross section of countries with a logarithmic scale (1970)

Quadratic adjustment of inequality in the cross section of countries without a logarithmic scale (1970)
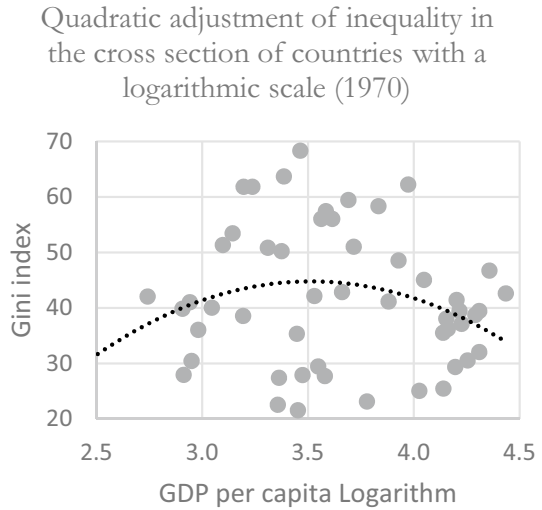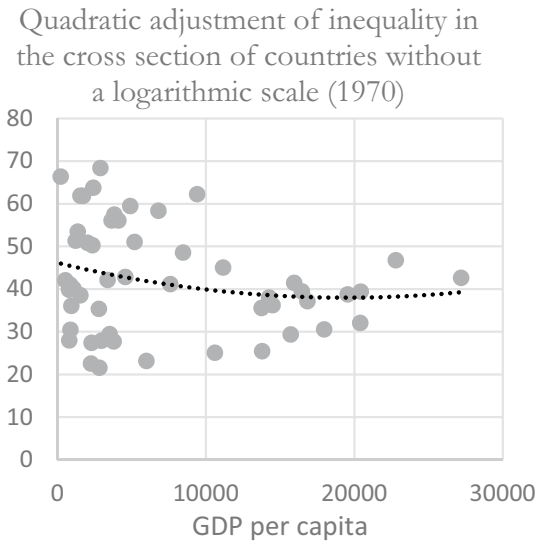
since only one atypical value (such as Zimbabwe in the case shown) can completely mask the relationship posed by Kuznets.

In Fig. 5, referring to the estimation of the Kuznets curve with a quadratic adjustment of inequality having level GDPpc (GDPpc refers to GDP per inhabitant), we observe that in the 1970's (as in the 1980's, 1990's and 2000's when represented in this way), there are similar decreasing tendencies of inequality with a small upturn at the end, making it possible to distinguish a gentle convexity.

However, the methodology carried out by Gallup (2012) reproduced Kuznets curves with quadratic adjustment of inequality, but this time taking the logarithm

Fig. 6 Kuznets curve with logarithmic scale (1980) Source: compiled by the author
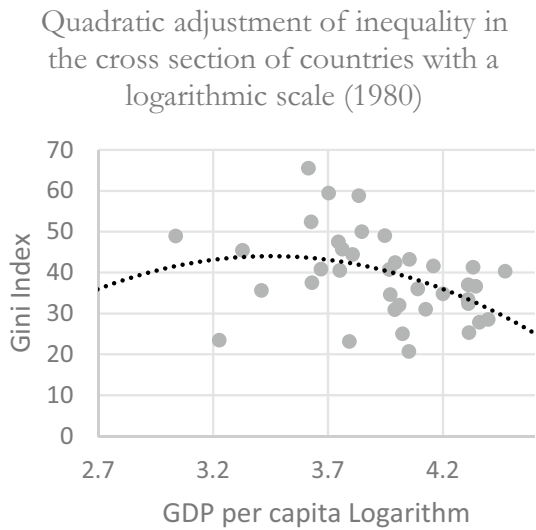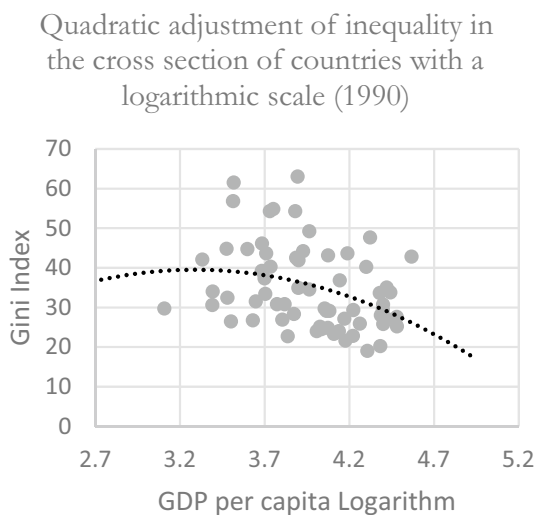

Quadratic adjustment of inequality in the cross section of countries with a logarithmic scale (1980)

Fig. 7 Kuznets curve (1990) with logarithmic scale Source: compiled by the author

Quadratic adjustment of inequality in the cross section of countries with a logarithmic scale (1990)

of GDPpc—"Log (GDPpc)" onwards—variable to soften the effect of the variance in level GDPpc values; thus, one can show the relationship that would otherwise be hidden. Accordingly, in Figs. 4, 6, 7 and 8, we can perfectly distinguish how the curve follows the inverted-U form described by Kuznets, seemingly fulfilling the relationship described in his hypothesis.

As it was done with the 1970s, for the decade of 2010 the Kuznets Curve will also be plotted using the GDPpc with and without logarithm on the abscissa axis. In Fig. 9 we can see the year 2010 without a logarithmic scale, in this one a similar relation is shown to that of Fig. 5; following clearly a descendent tendency without any upturn at the end. Likewise, Fig. 10 shows the GDPpc with logarithmic base,

Quadratic adjustment of inequality in the cross section of countries with a logarithmic scale (2000)

Quadratic adjustment of inequality in the cross section of countries without a logarithmic scale (2010)
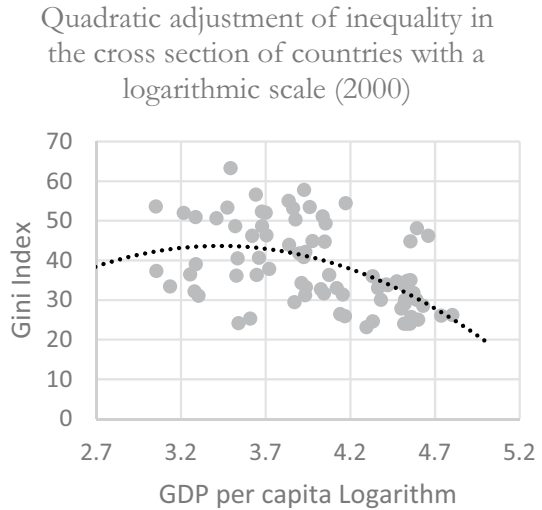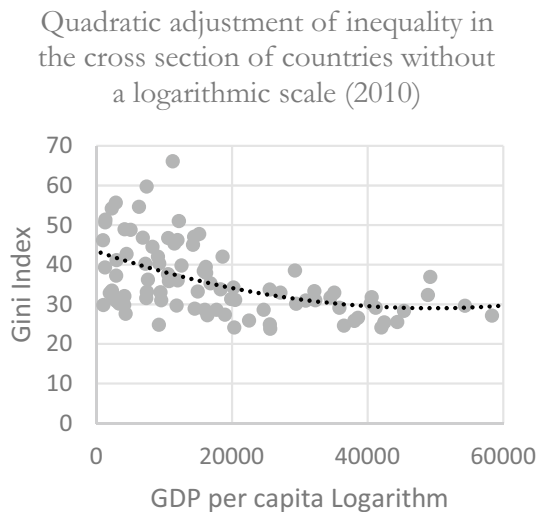
which shows similarity with the previous graphs when the income is expressed in logarithms (Figs. 4, 6, 7, 8). This contradicts our expectations, since we assumed a structural change in the pattern of the Kuznets Curve caused by the intensification of globalization, or by the Great Recession that began in 2008.

To conclude, we can affirm that the Kuznets hypothesis is discernible based on the graphical analysis as long as it is used as a measure of development with log GDPpc; otherwise, the variations of this magnitude from one country to another make observation difficult, and are not found if we take into account outliers, such as the case of Zimbabwe as shown in Fig. 3. Likewise, his observation is fulfilled (based on log GDPpc) in the other graphs shown above. However, given that

Fig. 10 Kuznets curve (2010) with logarithmic scale Source: compiled by the author

Quadratic adjustment of inequality in the cross section of countries with a logarithmic scale (2010)
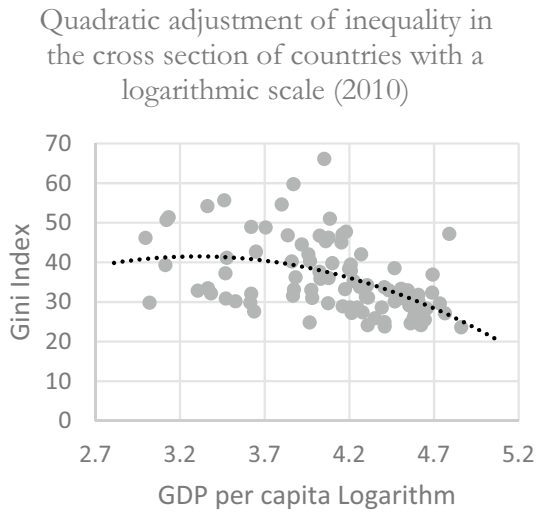
this observation is not fulfilled with the level GDPpc variable, it is a non-robust observation—this is because it depends on how the economic development variable is represented as to whether the curve is reproduced in the inverted-U form or not.

3.2 Model

In the exploratory graphical analysis, similar forms were found to the Kuznets curve for the countries taken as a whole although in a non-robust way. This section sets out the models for empirically contrasting the Kuznets hypothesis. The first model is estimated using the Log (GDPpc) variable; that is to say, using the same procedure as in the graphical analysis, after which the model will be expanded to include variables that can explain the inequality. Consequently, the basic model is as follows:

$$\text{Gini}_{i,t} = \alpha + \beta_1 \text{Log(GDPpc)}_{i,t-1} + \beta_2 \text{Log(GDPpc}^2)_{i,t-1} + u_{i,t}. \quad (1)$$

In addition, to better study the curve, the following expanded model will also be performed in which, with other variables available, we will attempt to summarize the Kuznets hypothesis more precisely if these suggest any consistent relationship in the different estimates. The variables that will be added are: a measure of education (the average total years of education received by the individual) following the recommendations of Benabou (1996); in his reverse causality study, he considered public education as a form of transfer from the wealthy classes to the more disadvantaged; this variable can also affect inequality. Similarly, given that Prados de la Escosura (2008) ruled out the "Latin American Effect" of Rattan (2012), we will examine the effect that a country's development level has using a variable that divides the sample into developed or underdeveloped nations; the criterion being whether or not they belong to the OECD. Lastly, we will employ a *proxy* variable on

the country's democratic level to see how the form of government affects inequality; a proposal made by Milanovic (2000).

$$
\begin{aligned}
\text{Gini}_{i,t} = \ &\alpha + \beta_1 \text{Log(GDPpc)}_{i,\,t-1} + \beta_2 \text{Log(GDPpc)}^2_{i,\,t-1} \\
&+ \beta_3 \text{Education}_{i,\,t-1} + \beta_4 \text{Polity IV}_{i,\,t} + \beta_5 \text{OECD}_{i,\,t} + u_{i,\,t},
\end{aligned}
\tag{2}
$$

where the Gini index is contained in the "Gini" variable; α is a constant; Log (GDPpc) is the logarithmic real per capita income of purchasing power parity expressed in international dollars for 2005; education is a variable that contains the average total years of education for a representative sample of individuals of each country; while Polity IV is a measure of the country's level of democracy; finally, OECD is a *dummy* variable that says whether or not the country belongs to the OECD, taking this as a sign of whether it is a developed country or not. The models also contain the subscripts *i* and *t*, which indicate the country and year, respectively, to which the observation refers. It is noteworthy that the income and education variables are delayed by a period (t − 1) because we think that the inequality of a year is determined by the income and education of the previous period, not of the same year, according to other authors such as Prete (2018). Finally, *u* is the model's random disturbance.

The models will also be estimated for the overall sample, for the segment of countries that belong to the OECD, and for the non-OECD segment of countries. If the Kuznets hypothesis is true, the coefficients obtained from the Log (GDPpc) regressions will be positive, and their respective squares negative. If not, one could not defend the Kuznets curve in the inverted-U form.

Conversely, the variables that extend the model are expected to have negative coefficients. So, for the countries that have the most developed human capital measured alongside the average total years that the population studies (the "Education" variable), the levels of inequality will be lower. With regard to the Polity IV and OECD variables—the first indicates how democratic the countries in the sample are and the second differentiates if the country is developed or not—one would expect that economies with more democracy in developed countries would have lower levels of economic inequality. In other words, these explanatory variables should more precisely specify the descending section of the Kuznets curve.

3.2.1 Data

Since inequality studies were initiated, from Kuznets (1955) to Milanovic (2016), all have encountered the same problems—the scarcity of data and the difficulty of collecting them, as well as the variables omitted from the models.

The data on inequality frequently lack overall comparability because these studies are not carried out by the same international body that obtains the measure of inequality (in this case, the Gini index), which consider equivalent standards for all countries. This information must therefore be compiled from different institutions. The Gini indices used come from the Standardized World Income Inequality Database (hereinafter and for the purposes of linguistic economy: SWIID) (Solt 2016), which was updated in September 2015; this database is a

compilation of various sources, containing information from the end of the 19th century to the beginning of the 21st, from all the available countries each year.

The rest of the world's regions do not yet have organizations that gather information about their level of inequality. However, in regions such as East and South Asia, and Africa, they can be inferred by means of family expenses or income surveys, used to calculate the Gini index by authors such as Leigh (2008), Saunders (2004), Brandolini (1998), Gusenleitner et al. (1996) and Guger (1989); in this way completing the information relating to this variable for countries absent from the initial databases, including African countries, and thus obtaining representation from all the continents.

The criteria for choosing the data for each country each year from the various estimates offered by SWIID were that:

- Households were the unit of account; only using individuals when there were no data for households; At the same time, trying to ensure that the calculation carried out on household income surveys covered all sources of income, or in its absence, estimated household consumption extracted from a representative sample in that country.
- The estimation sample should cover the whole population, not just the urban or rural segments, to obtain the greatest generality offered by the data.
- The quality of the time series should be as accurate as possible, using similar methods and definitions of inequality between the different countries. However, this criterion is the one most infringed upon by the countries lagging behind, which do not have efficient estimates.

The Penn World Tables database (Version 8.1, Feenstra et al. 2016) has been used, in which the incomes of all the countries in the world are recorded in terms of the Gross Domestic Product per capita, establishing them in purchasing power parity with respect to international dollars for 2005. Data on per capita income with delays were considered to avoid endogeneity, and because it is considered advisable that inequality in the present period be a consequence of income from the previous period. This variable has also been successfully utilised with logarithms, as proven by Gallup, facilitating the graphical visualization of the Kuznets curve, which could yield favourable statistical results.

In the expanded regressions, two numerical variables have been added: one is the estimation of the average years of overall education that individuals have received, prepared by United Nations Development Programme (UNDP) for the calculation of the HDI (2013) for the 1990–2017 period in 189 countries. The other variable was developed by a team led by Monty G. Marshall, working on the Polity IV Project (2013), which seeks to estimate the level of democracy in 167 countries, based on events that each country goes through: fractionalism, autocratic setbacks, revolutions, successful coups d'etats, transitions and political regimes established, etc. This variable is produced on a scale from − 10 to 10; however, it has been standardized between 0 (total autocracy) and 10 (perfect democracy) to facilitate its interpretation and inclusion into econometric models.

Table 1 Descriptive table of variables

Variable	Observations	Mean	Standard deviation	Minimum	Maximum	Countries
Gini	2.528	36.76	9.99	16.23	77.3	187
GDPpc	7.333	11,747.99	15,438.77	134	208.740	187
Log (GDPpc)	7.333	8.72	1.16	4.89	12.24	187
Education	4.843	6.59	3.01	0.09	13.18	181
Polity IV	6.259	5.92	3.63	0	10	167
HDI	4.737	0.64	0.16	0.21	0.951	187
Agriculture	5.798	15.38	12.74	0.03	79.04	187

All data provided are unweighted averages

Definition of the variables

GINI: The Gini index, a percentage expression of the Gini coefficient. It ranges from 0 (perfect equality) to 100 (perfect inequality)

GDPpc: The gross domestic product per capita calculated in purchasing power parity deflated to international dollars of 2005

Education: An estimation of the average total number of years that individuals from a country have received education

Polity IV: An estimation of a country's level of democracy, standardized between 0 (total autocracy) and 10 (perfect democracy)

HDI: The Human Development Index

Agriculture: The added value of agriculture as a percentage of GDP

No. countries: The number of countries for which there is data

It also contains a d*ummy* variable that divides the sample into two segments: developed countries and underdeveloped countries; classifying a country into one or other sub-group, whether they belong to the OECD (Organization for Cooperation and Economic Development).

Finally, the panel data includes time series for 187 in the Americas, Europe, Asia, Africa and Oceania; thus being representative of all the continents. These data cover the years 1970–2016. Despite the low annual variability of some of the variables that make up the database, the use of panel data allows us to analyze these small differences that occur from year to year and from country to country, making it possible for the analysis to consider more information and be more powerful than if these annual data were omitted, this perspective is supported by works such as that of Molina-Morales et al. (2013). Table 1 below provides basic information (the number of observations, mean, typical deviation, minimum value, maximum value and number of countries for which the observations were observed) for each relevant model. Regarding the atypical values, these have been eliminated so that the regression is true to the mass of data, which has also had them removed. Of these values, the most prominent is Zimbabwe, which completely distorted the tendency not to appreciate the Kuznets curve when it was present in the sample data.

We should add that, later on, two new variables will be introduced in the study so as to carry out a sensitivity analysis on the results; hence, it will be possible to verify if the conclusions observed in the main analysis are robust. These two new variables are the HDI and the aggregate value of agriculture in terms of GDP.

The Human Development Index (HDI) variable is an indication of human development in a country, produced and published by UNDP; in this case, collating data from the *Human Development Report 2006* (Watkins et al. 2006) and the *Human Development Report 2015* (Jahan et al. 2015); this index considers three variables in its calculation: health (approximated by the population's life expectancy), education (measured in terms of years of schooling, matriculation and literacy) and a dignified standard of living (measured using the GDPpc). Conversely, the agricultural variable, as a percentage of GDP, has been taken directly from the World Bank database.

3.2.2 Econometric methodology

In this work the Cochrane–Orcutt estimation technique has been used, discarding the fixed and random effects models for infringements of the basic hypotheses necessary for their correct use, as is explained shortly.

Thanks to the panel data technology, we can collect information from different countries at different periods, enriching the study with information that would not appear in the cross-sectional data within single-year observations. The disadvantage, perhaps, is that the qualities relevant to the model are not observable, so that individual errors would correlate with the observations. However, the Cochrane–Orcutt estimation (Cochrane and Orcutt 1949) contemplates this error in the model by taking charge of the AR (1) serial correlation in the linear models, the most suitable in this case.

Considering the model presented in Eq. (1), we recall that $u_{i,t}$ is the error term in the country *i* for the moment *t*. If we assume that the process of generating the residues follows a stationary first-order process with an autoregressive structure, we would have

$u_{i,t} = \rho \cdot u_{i,t-1} + \varepsilon_{i,t}$, $|\rho| < 1$, with the $(\varepsilon_{i,t})$ errors being white noise, the Cochrane–Orcutt procedure can be used to transform the model using quasi-difference:

$$\text{Gini}_{i,t} - \rho \cdot \text{Gini}_{i,t-1} = \alpha(1-\rho) + \beta_1 \left[\text{Log}(\text{GDPpc})_{i,t-1} - \rho \cdot \text{Log}(\text{GDPpc})_{i,t-2}\right]$$
$$+ \beta_2 \left[\text{Log}(\text{GPDpc}^2)_{i,t-1} - \rho \cdot \text{Log}(\text{GDPpc}^2)_{i,t-2}\right] + \varepsilon_{i,t}. \tag{3}$$

After this transformation, the error terms are white noise, so the statistical inference is now valid, since the sum of residual squares is reduced to the minimum with respect to the α and β parameters, conditioned to ρ.

The tests carried out on the regressions considering the whole sample were the Cumby-Huizinga autocorrelation test and the modified Wald heteroscedasticity test, the results of which are found in Table 2.

The Cumby-Huizinga autocorrelation test (Cumby and Huizinga 1992), that one of the main advantages of the C-H framework is that it can be used to test autocorrelation in delay orders $(q+1)\ldots(q+s)$ under the null hypothesis that the series being tested is MA(q), says that in our model, autocorrelation is present even up to the fourth delay. Also, we know that the GDPpc and the Gini index are variables that are very influenced by their trend and have little variation from 1 year to another, then

Table 2 Tests performed on the regressions for the total sample Source: compiled by the author

Autocorrelation test		Heteroscedasticity test		Breusch–Pagan Lagrangian multiplier		Im-Pesaran-Shin unit-root	
Chi^2	T	F	t	Chi^2	t	Statistic	t
45.430	0.000	0.0067	13.67	138.37	0.000	4.1369	0.004
45.186	0.000						
40.726	0.000						
40.170	0.000						

we were already expecting that there really was an autocorrelation, a fact confirmed by this test. Likewise, we performed the Lagrange multiplier heteroscedasticity test, modified by Wald, which is useful even when the normality assumption is violated. The null hypothesis of this test is that there are no heteroscedasticity problems. It indicates that our sample does have heteroscedasticity, so the regressions will be carried out under the assumption that the variance is not constant.

Additionally, Breusch and Pagan Lagrangian multiplier and the unitary root test of Im-Pesaran-Shin (2003) was also performed. From the first, Breusch and Pagan (1980) Lagrangian multiplier test, we can deduce by rejecting Ho that individual effects are relevant in this case, and for that reason it is better to use panel data. From the second, we can see that with a t value of 0.004 the Ho of unitary roots in all panels is rejected. Therefore we can say that the Gini variable is a stationary variable. Likewise, knowing that we do not have unitary roots, in this field it is worth highlighting the work of Baltagi et al. (2013) in which they show that for the regressions that have roots close to unity, without suffering unitary roots, they maintain certain asymptotic capacities that the regressions with non-unitary roots have and therefore they could be used to support conclusions. However, this is not our case since we can significantly rule out that our model suffers from unitary roots.

Conversely, the GMM System methodology has been used because there are indications that endogeneity exists between the level of inequality and the level of development; i.e., between the Gini and log (GDPpc) variables. In addition, this type of dynamic model analysis is recommended in case there are more individuals than periods (Roodman 2009b), having in this case four times as many individual variable (countries) than periods.

The results obtained through the Cochrane–Orcutt procedure will be contrasted with the GMM System model estimation (Arellano and Bond 1991). This is done because the inclusion of the delayed variable (especially in possible cases of endogeneity such as this) could lead to a "dynamic bias" in the panel, which would result in inconsistent estimates using any other estimator, other than that proposed by Arrellano and Bond. An econometric solution to the endogeneity of variables found in other works, such as Antonakakis and Collins (2018) or Windarti et al. (2019), is to estimate using the GMM System. For this reason, both techniques will be used to appreciate the estimation's robustness. Moreover, to minimize endogeneity, the model has the regressors out-of-phase more than one period; this is to say, *i* periods

(with i able to be 5 or 10 years). In this way, the estimators' consistency can be maintained despite possible skewness from endogeneity.

3.2.3 Discussion of the results

A peculiarity of the results obtained, in contrast to those obtained by Gallup (2012), is that the regressions have high determination coefficients, none of them lower than 0.69, meaning the goodness of fit is moderately good; i.e., the Gini variable's proportion of variation is explained by at least 69% by the independent variables. This determination coefficients cannot be higher because the model has a specification problem, known as omitting relevant variables, which generates this error that the variable effect forms part of the error term. However, we find that the coefficients are significant, even at 1% in most cases, when the whole sample is considered; thus, they can offer an intuition regarding the Gini behaviour. Also, as far as System GMM estimation is concerned, it is noteworthy that the m1 and m2 tests find a first-order serial correlation but not a second-order one, so there is no need to worry about autocorrelation. As well as the fact that Hansen's tests result between 0.5 and 0.25 in most ideally fitting to be able to say that the model is not over-identified according to Roodman (2009a). Finally, it should be noted that when the term constant is significant, it is always positive, from which we can extract that economies, tend to generate inequality by their very nature.

In the basic model, the estimates made using the Cochrane–Orcutt panel methodology (Table 3) we find that the Kuznets curve is confirmed, with Log (GDPpc) coefficients of 14.78 and Log (GDPpc)^2 coefficients equal to − 1.04, the representation of these being a curve that is similar to the graphical analysis view. These results contradict those obtained by Gallup (2012) In that article, the regression, which used Pooled OLS methodology, seems to indicate a U-shaped relationship; nonetheless, the results of Barro (2000) support our results. Furthermore, the GMM system regressions show favourable results.

The most interesting point about these regressions is when one jointly observes the regression results for the segmented sample (countries belonging to the OECD and those that do not). In the sub-sample of non-OECD countries, positive coefficients are shown for Log (GDPpc) and negative coefficients for Log (GDPpc)^2, whereas the opposite is observed for the estimation of those countries belonging to the OECD (i.e., negative coefficients for Log (GDPpc) and positive coefficients for Log (GDPpc)^2; the same occurs in the regressions made with System GMM estimation, making this a systematic result. Therefore, if we look at the generality of the regressions, we can take it as truer that underdeveloped economies comply with the Kuznets hypothesis, as can be seen in the work by Cheng and Wu (2017) for China, whereas in developed economies, this pattern does not remain in effect, as seen in the U.S. case study performed by Nasr et al. (2019).

Thus, if we assume that, in the past, OECD countries behaved like those that do not currently belong to the OECD, the inequality that followed the pattern described by the Kuznets curve would lead to an upturn in inequality as it reaches the developed economy level; i.e., we would be encountering sinusoidal behaviour, not just a concave

Table 3 Estimation of the models

Cochrane–Orcutt estimation

Gini$_t$

Regressors	Basic model						Expanded model					
	Complete		Non-OECD		OECD		Complete		Non-OECD		OECD	
	Estimates	t	Estimates	t	Estimates	t	Estimates	t	Estimates	t	Estimates	t
Log (GDPpc)$_{t-1}$	14.78***	3.71	6.41**	1.35	− 36.56**	− 2.01	11.57***	2.70	3.66	0.70	− 41.67*	− 1.62
Log (GDPpc)$^2_{t-1}$	− 1.04***	− 4.66	− 0.47**	− 1.71	1.63**	1.79	− 0.76***	− 3.15	− 0.26	− 0.90	1.76	1.41
Education$_{t-1}$							− 0.25	− 1.48	− 0.34*	− 1.56	0.35	1.74
Polity IV$_t$							0.39***	3.03	0.51***	4.05	− 2.22	− 2.70
OECD$_t$							− 5.33***	− 5.03	–		–	
Constant	8.46***	0.48	21.79***	1.08	234.35***	2.60	0.52*	− 0.03	30.73*	1.38	290**	2.27
R^2	0.75		0.74		0.69		0.78		0.73		0.77	

System GMM estimation

Regressors	Complete		Non-OECD		OECD		Complete		Non-OECD		OECD	
	Estimates	t	Estimates	t	Estimates	t	Estimates	t	Estimates	t	Estimates	t
Log (GDPpc)$_{t-1}$	6.06***	0.35	24.84*	0.83	− 62.91*	− 1.31	0.22**	0.02	13.87*	3.52	20.60*	0.67
Log (GDPpc)$^2_{t-1}$	− 0.11**	− 0.12	− 1.18*	− 0.69	3.01**	1.25	− 0.009*	− 0.02	− 0.90*	− 3.41	− 1.13*	− 0.77
Education$_{t-1}$							− 1.10***	− 2.58	− 1.85***	− 4.66	− 0.40*	1.73
Polity IV$_t$							0.71*	1.77	0.94***	1.20	6.07***	4.11
OECD$_t$							− 5.93**	− 2.48	–		–	
Constant	82.53***	1.06	165.50***	1.28	358.16***	1.51	49.36*	1.05	106	1.80	− 7.31	− 0.05
Num. Instr.	61		60		61		133		118		121	

Table 3 (continued)

Tests	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value	Estimates	p value
m1 test	− 4.30	0.000	− 3.57	0.000	− 2.56	0.011	− 4.27	0.000	− 3.28	0.001	− 3.10	0.002
m2 test	− 0.83	0.409	− 0.91	0.363	− 0.45	0.650	− 0.80	0.421	− 0.88	0.380	− 0.13	0.893
Hansen test	72.6	0.125	55.05	0.131	52.8	0.141	140.56	0.264	105.62	0.258	32.67	0.261

The null of the $m1$ and $m2$ test is the absence of first and second order serial correlation in the disturbances, respectively. The null of the Hansen test is the adequacy of moment conditions

*Significant at 10%

**Significant at 5%

***Significant at 1%

curve; this supports the hypothesis posited by Milanovic (2016) for the Kuznets curve over the very long term.

In the extended models, the Log (GDPpc) has a positive coefficient and the Log (GDPpc)^2 has a negative coefficient; consequently, they seem to respond correctly to the relationship between the explained variable (Gini) and the explanatory variable (Log GDPpc) according to Kuznets' hypothesis; although not in all cases the coefficients are so appropriate. The coefficients obtained for the estimation of the developed countries (members of the OECD) by the Cochrane–Orcutt method show the changed signs, that is, Log (GDPpc) is negative and Log (GDPpc)^2 is positive, although the latter is not significant. This lack of significance is also found in the sample for underdeveloped countries (non-OECD members) in the Cochrane–Orcutt regression and in the full-sample and sub-sample regressions for developed countries (OECD members) in the System GMM estimate. However, it can be restated that the Kuznets curve is observed in the regressions presented, using the Gini index as a measure of inequality and the Log (GDPpc) as a measure of development, as found in Kiatrungwilaikun and Suriya (2015).

With regard to the Education, Polity IV and OECD variables—the education and OECD variables; that is, the average total years of schooling of the population and whether the country is developed or not, have a negative influence on inequality with significant coefficients. This indicates that the countries with more education and development seem to be more equitable; totally expected results in accordance with paragraph 3.2., and similar to those obtained by Barro (2000), who also took into account the years of schooling, obtaining negative coefficients. On the other hand: Milanovic (2000) supports the results by also obtaining a negative coefficient in his education variable. This is also contrary to Piketty (2013), who argues that in the more developed countries, inequality is reemerging; in our case, we observe that developed countries have less tendency to inequality. On the other hand, as with Barro (2000), our Polity IV variable (representing the level of state democracy governing the country under study) has a positive coefficient although with practically no significance; thus, we cannot venture to interpret these coefficients lightly. Nonetheless, a possible explanation as to why the higher level of democracy in a country can positively affect its level of inequality could be the growing importance of money in politics; this argument was made by Milanovic (2016), who argues that the elites, who hold the money and capital, can influence policy to make legal reforms that benefit them in particular.

In short, Log (GDPpc) is a variable that collects information on income with less variance than GDPpc. With it, the relationship that Kuznets noted is observed over the period covered in 187 countries. In addition, empirical evidence is also found that, by separating the sample into developed and undeveloped countries, the Kuznets curve takes a sinusoidal form over the very long term, as Milanovic (2016) suggested.

3.3 Robustness analysis

We check if the results are robust by repeating the estimated models above, replacing the Log (GDPpc) variable with another measure of economic development. In this way, robustness analyses will be carried out using the HDI and the aggregated

Table 4 Estimation of the basic models of sensitivity—HDI_t (offsetting the GDPpc)

	Cochrane–Orcutt estimation					
	$Gini_t$					
	Basic model					
	Complete		Non-OECD		OECD	
Regressors	Estimates	t	Estimates	t	Estimates	t
HDI_{t-1}	28.16**	1.98	12.04**	0.71	− 19.85***	− 0.74
HDI^2_{t-1}	− 9.31***	− 1.61	− 10.05**	− 0.70	14.94**	0.76
Constant	34.82***	8.01	38.67***	7.85	39.31***	4.42
R^2	0.76		0.75		0.73	
	Basic model					
	Complete		Non-OECD		OECD	
Regressors	Estimates	t	Estimates	t	Estimates	t
HDI_{t-1}	51.26	0.59	108.10	0.92	− 64.90	− 0.68
HDI^2_{t-1}	− 26.88	− 0.42	− 67.85	− 0.75	44.38	0.98
Constant	58.81**	28.31	80.94**	2.21	53.71	1.59
Num. Instr.	77		75		77	
Tests	Estimates	p value	Estimates	p value	Estimates	p value
m1 test	− 4.17	0.000	− 3.09	0.002	− 3.22	0.001
m2 test	− 0.10	0.921	− 0.10	0.917	− 0.01	0.994
Hansen test	86.33	0.155	81.64	0.205	82.57	0.198

See footers on Table 3

value of agriculture with respect to GDP (%). This is done exclusively with the basic models because in these, we can clearly see the interrelationships between inequality and the various variables that measure development (in one way or another) in this study. As with the regressions using the Log (GDPpc) variable, the coefficient of determination obtained relatively high, around 0.75, for the GMM System methodology we obtain similar results to the previous ones in the tests carried out.

The analysis carried out using HDI as the independent variable over the medium term (Table 4); helps us to intuit that the form following the inequality (for the complete sample) has an inverted-U form; indeed, this is as the actual analysis points out. We can therefore expect that in the future, it will effectively decrease the inequality, just as Kuznets argued, as supported by Parente (2019). Again, the term constant is very significant and positive.

On the one hand, reviewing the segmented regressions together again, we note that in regressions with Cochrane–Orcutt methodology non-OECD countries fulfil the hypothesis and its complementary, whereas OECD members do not fulfil it over the long-term, again suggesting the Milanovic hypothesis, as this is posited over the (very) long term.

Table 5 Estimation of the basic models of sensitivity—Agriculture$_t$ (offsetting the GDPpc)

	Cochrane–Orcutt estimation					
	Gini$_t$					
	Basic model					
	Complete		Non-OECD		OECD	
Regressors	Estimates	t	Estimates	t	Estimates	t
Agriculture$_{t-1}$	0.087*	1.35	0.21*	2.84	− 0.08*	− 0.90
Agriculture$_{t-1}^2$	− 0.0009*	− 0.78	− 0.002*	− 1.80	0.0013*	0.60
Constant	41.15***	59.95	44.43***	55.12	32.59***	41.57
R^2	0.75		0.70		0.70	
	Basic model					
	Complete		Non-OECD		OECD	
Regressors	Estimates	t	Estimates	t	Estimates	t
Agriculture$_{t-1}$	0.048	0.18	0.25	0.76	− 0.55*	− 2.28
Agriculture$_{t-1}^2$	− 0.0004	− 0.10	− 0.002	− 0.46	− 0.015*	− 2.49
Constant	38.43***	15.90	43.69***	13.63	34.37***	14.67
Num. Instr.	87		84		87	
Tests	Estimates	p value	Estimates	p value	Estimates	p value
m1 test	− 4.33	0.000	− 3.09	0.002	− 3.22	0.001
m2 test	− 0.72	0.469	− 0.34	0.736	− 0.78	0.434
Hansen test	59.94	0.198	49.53	0.243	28.96	0.251

See footers on Table 3

On the other hand, in the regressions with System GMM methodology coefficients are insignificant. So it would be a mistake to enunciate any hypothesis. The only thing that we can observe with significance is that the term constant is positive, which means that in economies there is a prevalence of inequality.

The sensitivity analysis was repeated, but this time the percentage representing the agricultural sector in the GDP was employed as an independent variable, a measure of the development that might be more biased than the HDI but inevitably more related to the productive structure of a country.

The results of this second sensitivity test are collated in Table 5. These indicate that the Kuznets curve is verified in the same way as when considering the Log (GDPpc) and the HDI, as Baymul and Sen (2019) previously found, although in their case instead of using the percentage of agriculture over GDP they used the ratio of population employed in agriculture. Results are (for the complete sample) significant in the GMM System methodology; this was also found in the study by Zhou and Shi (2019), in which they observed that demographic transition generates a Kuznets curve. However, reproducing regressions for the developed and underdeveloped country subsamples, we see in the regressions

performed through the Cochrane–Orcutt method that the same occurs as in the previous cases. That is, it reinforces Milanovic's concept that in the long run inequality behaves in a sinusoidal pattern and not as a concave curve. With respect to estimates made using System GMM estimation, we again find that the coefficients are not significant, except in the case of OECD member countries that show little significance and their coefficients indicate a fall in inequality in both coefficients (Agriculture and Agriculture^2) which is the expected behaviour according to Kuznets' approach. Likewise, in this case it is also observed that the term constant is very significant and positive.

In summary, the Kuznets curve appears to be robust if the sensitivity analysis is carried out using the HDI and the aggregated agriculture variables for the complete sample. On the other hand, it is recognised that Milanovic's hypothesis is observable in the estimates made by the Cochrane–Orcutt method, being totally rejected by the estimates made by the system GMM method. Furthermore, it can be seen how the results of the agriculture variable show levels of significance that are very different from those obtained by Log (GDPpc) and the HDI; here we can find the cause of the controversy that the subject generates: that the different ways of measuring inequality and development strongly condition the results of the study.

4 World trade and inequality

The World Trade Organization (WTO) has observed an unprecedented expansion in the volume of world trade, which has multiplied by more than 27 from 1950 to the present day, and as a consequence this affects various areas of the economies.

For several years, these effects have been classified and studied, and used to measure how trade liberalization affects CO_2 emissions to investigate the Environmental Kuznets Curve (EKC) as authors such as Panayotou (2000), Cole and Elliott (2003), Dinda (2004) and Tsurumi and Managi (2010) have done. In our study we want to pose how these effects concern the original Kuznets curve, which focuses on inequality.

The scale effect refers to the impact on the country of an increase in production or economic activity as a result of trade liberalization. In principle, it is assumed that, if nothing else changes, trade liberalization will increase production and, as a consequence, more energy will be consumed, emitting more greenhouse gases. This effect is measured through GDP per km^2 Tsurumi and Managi (2010), which can be obtained through the World Bank database.

The composition effect, this encompasses the fact that in the face of trade liberalization, the country will focus its production on goods and services in which it has a comparative advantage. This will lead to a reallocation of resources that will lead the economy to emit more or less greenhouse gases depending on which sectors it has that comparative advantage. The end result of the composition effect is difficult to predict because each country has a different specialization. The composition effect can be measured in different ways, through the weight of industry in GDP (Cole and Elliott 2003) or through the capital-labor ratio (Tsurumi and Managi 2010), in our case we will use the second measure found in the Extended Penn World Tables.

Table 6 Descriptive table of data on scale, technique and composition effects

Variable	Observations	Mean	Standard deviation	Minimum	Maximum	Countries
Gini	2.528	36.76	9.99	16.23	77.3	187
GDP per capita	7.333	11747.99	15438.77	134	208.740	187
GDP per kilometre^2	7.333	3.53	20.09	0.00001	647.52	187
Capital-labor ratio	4.765	35395.35	39386.27	323.89	295076.7	187
Log (GDP per capita)	7.333	8.72	1.16	4.89	12.24	187
Log (GDP per km^2)	7.333	− 2.87	2.83	− 8.84	6.47	187
Log (capital-labor ratio)	4.765	9.64	1.49	5.78	12.59	187

All data provided are unweighted averages

Finally, the technical effect encompasses the improvement of production processes through two channels: First, as trade is liberalized, goods and technologies can be accessed at a lower cost. Another way could be that the increase in income as a result of trade could encourage society to demand greater environmental quality, thereby reducing greenhouse gas emissions. This effect is measured through GDP per capita as Tsurumi and Managi (2010).

A descriptive table of the variables to be used for this section can be seen in Table 6.

As we did before, we will begin the section through its graphical representations with respect to the Gini index. As in the previous case following Gallup (2012) and Tsurumi and Managi (2010), we will take the logarithmic scale effects variables to reduce their dispersion (Figs. 11, 12).

One thing that strikes us is that the scale and composition effects are diametrically opposed. We can find how countries with little "scale" show high levels of inequality, and these are reduced as it increases generating more equitable societies, to a certain point where inequality grows again, as well as greenhouse gas emissions, establishing this parallel with the EKC.

Likewise, the composition effect behaves as described by Kuznets, perhaps drawing the passage from economies based on the primary sector to those based on the secondary sector, resulting in an increase in inequality due to the disparity in wages between the two sectors. These economies finally show high capital-labor ratios, which could be a sign that they have reached the service sector, which according to Kuznets' approach would be one of the motors generating equality. It should be noted that in the approach of the EKC the composition effect is the only a priori unpredictable and in this case it has been found that depending on the bracket in which we situate the ratio can affect positively or negatively as seen graphically (Fig. 13).

Thirdly, the technical effect shows an increasing trend in the graph, although it is the one where the mass of points looks more concentrated and uniform. We could affirm that this one could impel the inequality although in the scope of the

Fig. 11 Representation of the
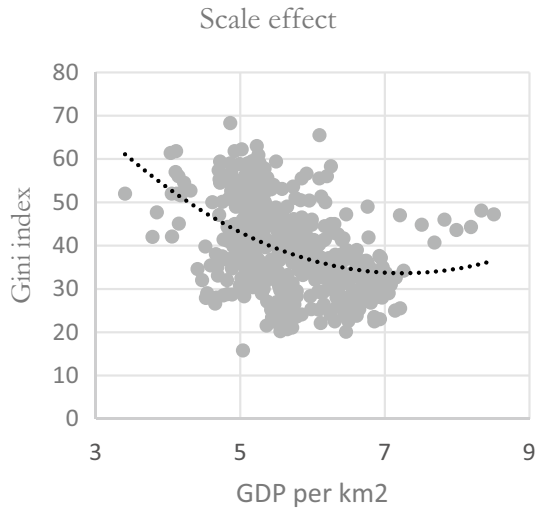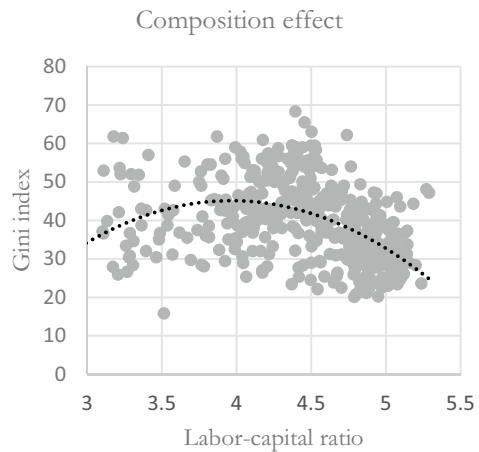scale effect Source: compiled by
the author

Fig. 12 Representation of the
composition effect Source:
compiled by the author

EKC diminishes the emissions of gases, being this effect different for the EKC and for the curve of original Kuznets.

In Table 7, we can see the result of estimating the Gini coefficient as an exogenous variable with respect to these three effects as endogenous variables, as shown in Eq. (4).

$$\text{Gini}_{i,t} = \alpha + \beta_1 \text{Log}\big(\text{GDP per km}^2\big)_{i,\,t} + \beta_2 \text{Log}(\text{GDP per capita})_{i,\,t}$$
$$+ \beta_3 \text{Log}(\text{capital} - \text{labor ratio})_{i,\,t} + u_{i,\,t}. \tag{4}$$

Insisting on what we have seen in the graph of the scale effect, when this coefficient is significant it is negative in all estimates. As a result, trade liberalization, in terms of the increase in national production generates equity, possibly due to the

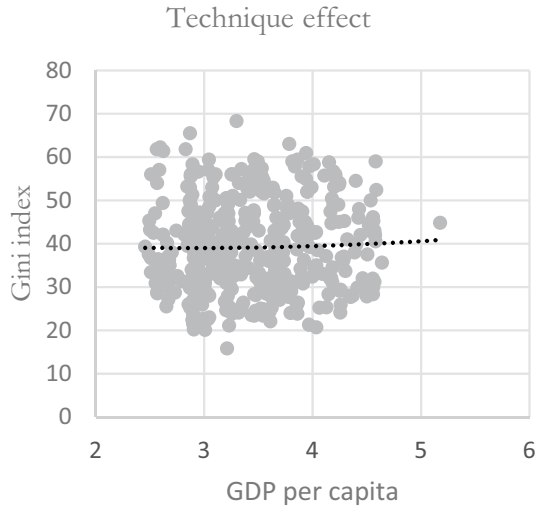**Fig. 13** Representation of the technique effect Source: compiled by the author

Table 7 Estimation of the scale, technique and composition effects

	Cochrane–Orcutt estimation					
	Gini$_t$					
	Basic model					
	Complete		Non-OECD		OECD	
Regressors	Estimates	t	Estimates	t	Estimates	t
Log (GDP per km^2)	− 2.54***	− 17.44	− 1.33***	− 8.63	− 2.27***	− 8.58
Log (GDP per capita)	1.79**	2.01	3.85***	4.39	1.78*	1.15
Log (capital-labor ratio)	1.61**	2.06	0.70**	0.93	4.09***	2.96
R^2	0.88		0.93		0.89	
	Basic model					
	Complete		Non-OECD		OECD	
Regressors	Estimates	t	Estimates	t	Estimates	t
Log (GDP per km^2)	− 2.59***	− 5.11	− 1.97***	− 3.88	− 1.84***	− 2.91
Log (GDP per capita)	3.83*	− 1.38	1.75*	0.64	4.35*	1.37
Log (capital-labor ratio)	6.25**	2.43	5.12**	2.10	6.38**	2.87
Num. Instr.	93		98		93	
Tests	Estimates	p value	Estimates	p value	Estimates	p value
m1 test	− 3.35	0.001	− 2.96	0.003	− 2.08	0.037
m2 test	− 0.03	0.972	0.24	0.807	− 0.15	0.877
Hansen test	73.68	0.198	56.77	0.243	66.41	0.223

See footers on Table 3

reduction in unemployment and the increase in income that derives from this growing production.

The technical effect, as we saw graphically, has significantly positive results, this could be interpreted by the second way that we saw in the approach of the technical effect; we said that as a consequence of the increase of the income by the commercial opening the society would demand greater environmental quality, nevertheless, as it is evident in the graph and in this table, this increase of the income is not equitable, reason why it generates inequality.

And finally, for the composition effect, as in the case of the technical effect, we obtain positive and significant coefficients, so we can deduce that the sectors in which the economies end up specializing, most possibly industrial and service sectors, generate inequality due to the different incomes obtained by the different workers of the productive system.

In conclusion, we can see how the scale effect, as with the EKC, increases both greenhouse gas emissions (Dinda 2004 and Tsurumi and Managi 2010) and inequality, although for the econometric analysis we have obtained the positive result, which is to be expected due to the fact that the majority of the section shown in the graph shows a marked trend with a negative slope in the data. The composition effect, unpredictable in the case of the EKC, has shown positive and negative results in the case of inequality, the determining factor being the capital-labor ratio itself, and this can be included in Kuznets' own statement (1955). And finally the technical effect, contrary to the results obtained by Dinda (2004) and Tsurumi and Managi (2010) for the EKC of decreases greenhouse gas emissions, in the original Kuznets curve shows that inequality increases.

5 Conclusions

Since the Kuznets curve was proposed in 1955, it has been studied in great detail, as we have already discussed. It is a very relevant hypothesis because it warns us that in any economy, there are mechanisms that encourage inequality as the economy develops. In the same way, it encourages us to fight against inequality, given that it is not a negative externality of totally safe development, it can be alleviated through public policies that promote the birth rate, redistribute wealth or encourage social mobility.

So, is the Kuznets curve still in effect? The response to this has caused controversy amongst economists since its inception. The most appropriate answer would be yes, with nuances. Yes, one can see that the Kuznets curve is fulfilled when we measure inequality using the Gini index and measure development using the logarithm of the GDP per capita. We are dealing with a presumably sensitive result because when we change the development measure to HDI or the agricultural level of a country, the expected relationship is verified if we rely on the Kuznets hypothesis, but the same does not occur when using the basic GDPpc variable (i.e., without applying logarithms) probably because it is a variable with greater dispersion than those mentioned above. It should be pointed out that in the sensitivity analyses of the subsamples, one could even observe that the pattern that follow the changes in

inequality are not those proposed by Kuznets, but paradoxically, the opposite. This fact occurs specifically in the subsample of developed countries (OECD members) and it is confirmed by the economic literature, which contains articles that support it and others that reject it.

In summary, we have found that the Kuznets Curve is found for the whole set of countries regardless of the econometric technique and inequality measurement variables used in general terms. For the sub-sample of non-OECD countries, the Kuznets hypothesis is still valid, as well as for its complementary, the OECD countries observe on some occasions that it follows the opposite pattern to that stated by Kuznets, more in line with Piketty (2013) and Milanovic (2016).

Perhaps even Kuznets himself would not be surprised to discover that his hypothesis is not totally certain; that is to say, not robust. In his own article he expressed the speculative nature that led him to postulate such a theory: "The paper is perhaps 5 percent empirical information and 95 percent speculation, some of it possibly tainted by wishful thinking" (Kuznets 1955, p. 26).

In the section in which the effects applied to the EKC have been analysed, it has been obtained that the scale effect act on greenhouse gas emissions as well as inequality. While for the technical effect it has been found that it exerts a totally opposite influence, it reduces gas emissions but increases inequality. And in the case of the composition effect, if in the field of the EKC it is unpredictable, in the case of inequality it has been shown that it follows the process described by Kuznets (1955) in his argumentation and example proposed in his paper published more than half a century ago.

Since the twentieth century, when studies on inequality first began, they have faced two fundamental problems: firstly: data—there is a scarcity of information available in many countries, and besides, it is of questionable quality and low comparability, given that there are no defined criteria in the international community; in calculating the Gini index, for example. Secondly, there are endogeneity issues and variables omitted, resulting in models with very little predictive capacity and doubtful analytical capability. Nevertheless, we have enough empirical evidence to say that there is a direct relationship between development and inequality, and that, so far, the Kuznets theory (1955) is the one which most approximates reality. Regardless, action by the state to combat inequality is believed necessary since, as found in Rodrik (2007), development alone is not sufficient to alleviate inequality.

Finally, the excellent question posed by Lyubimov (2017) for the case of Russia, but which here we cover in general, remains to be answered: what is the future of inequality in countries and what economic policy measures should be taken? Which hypothesis best represents the behavior of inequality, that of Piketty (2013) or that of kuznets (1955)? In our analysis we find that there is evidence that Kuznets' hypothesis is true, however, that would imply that inequality will fall and not increase as Piketty (2013) argues. How is this possible? First of all to say that, as Lemieux (2016) explained, Piketty's work focuses on a specific segment of the population, the richest, for a sample of countries that we can classify as highly developed countries, then his analysis focuses on the tail of the Kuznets hypothesis, on the side of very highly developed countries, which, if we look at Milanovic's analysis (2016) we can see that they are experiencing a second wave of inequality,

what this author calls "second Kuznets curve", thus generating in the long term a trend shaped like a sinusoid and not a U-inverted. This phenomenon, which Milanovic calls the "second Kuznets curve", is the form of U observed in studies on inequality such as that carried out by Blanco and Ram (2018).

Then, answering the question. For all countries in a cross-sectional analysis, the Kuznets curve is still valid; however, it seems that when these are highly developed, forces that increase inequality again come into play, as warned by Stiglitz (2012) and Piketty (2013) and Milanovic (2016), so it would be necessary for developed countries to carry out policies to contain and reduce inequality, although, citing Lyubimov (2017): "The specific form of such income redistribution is a topic for a separate discussion". The only thing we can say about it is that, according to Stiglitz (2012); the solution to inequality is not going to come from market forces because of the power exercised over it by the richest collectives, but from the state. However, concentrating market power in too few hands is just as bad as excessive regulation.

References

Antonakakis, N., & Collins, A. (2018). A suicidal Kuznets curve? *Economics Letters, 166,* 90–93.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58,* 277–297.

Auci, S., & Trovato, G. (2018). The environmental Kuznets curve within European countries and sectors: greenhouse emission, production function and technology. *Economia Politica, 35*(3), 895–915.

Baltagi, B. H., Kao, C., & Liu, L. (2013). The estimation and testing of a linear regression with near unit root in the spatial autoregressive error term. *Spatial Economic Analysis, 8*(3), 241–270.

Barro, R. (2000). Inequality and growth in a panel countries. *Journal of Economic Growth, 5*(1), 5–32.

Barro, R., (2008). *Inequality and growth revisited. Asian Development Bank: Working Paper Series on Regional Economic Integration*: 14. http://aric.adb.org/pdf/workingpaper/WP11_%20Inequality_and_Growth_Revisited.pdf. Accessed 3 Nov 2004

Baymul, Ç., & Sen, K. (2019). Kuznets revisited: What do we know about the relationship between structural transformation and inequality? *Asian Development Review, 36*(1), 136–167.

Benabou, R. (1996). Inequality and growth. *NBER Macroeconomics Annual, 11,* 11–92.

Blanco, G., & Ram, R. (2018). Level of development and income inequality in the United States: Kuznets hypothesis revisited once again. *Economic Modelling*. Available online from 1 Dec 2018.

Brandolini, A., (1998). Income distribution and growth in industrial countries (pp. 69–105). Cambridge Mass: The MIT Press.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies, 47,* 239–253.

Cheng, W., & Wu, Y. (2017). Understanding the Kuznets process—an empirical investigation of income inequality in china: 1978–2011. *Social Indicators Research, 134*(2), 631–650.

Cochrane, D., & Orcutt, G. H. (1949). Application of least Sauares regression to relationships containing autocorrelated error term. *Journal of the American Statistical Association, 44,* 32–61.

Cole, M. A., & Elliott, R. J. (2003). Determining the trade-environment composition effect: the role of capital, labor and environmental regulations. *Journal of Environmental Economics and Management, 46*(3), 363–383.

Comin, A. A. (2019). Economic development and inequalities in Brazil: 1960–2010. In *Paths of inequality in Brazil* (pp. 287–307). Springer, Cham.

Costantini, M., & Paradiso, A. (2018). What do panel data say on inequality and GDP? New evidence at US state-level. *Economics Letters, 168,* 115–117.

Cumby, R. E., & Huizinga, J. (1992). Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica, 60*(1), 185–195.

Deininger, K., & Squire, L. (1996). A new data set measuring income inequality. *The World Bank Economic Review, 10,* 26. https://doi.org/10.1093/wber/10.3.565.

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

Dinda, S. (2004). Environmental Kuznets curve hypothesis: A survey. *Ecological Economics, 49,* 431–455.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2016). The next generation of the Penn World Table. *American Economic Review, 105*(10), 3150–3182.

Fields, (1989). Changes in poverty and inequality in developing countries. *The World Bank Research Observer, 4*(2), 167–185. https://doi.org/10.1093/mbro/4.2.167.

Gallup, J. L. (2012). *Is there a Kuznets curve*?. Working Paper. Portland State University.

Gini, C., & Edgeworth, F. Y. (1922). *Indici di concentrazione e di dipendenza*. Unione tipografico-editrice torinese.

Guger, A. (1989). Einkommensverteilung und -verteilungspolitik in Österreich. Handbuch der Österreichischen Wirtschaftspolitik. Manz, Viena.

Gusenleitner, M., Winter-Ebmer, R., & Zweimüller, J. (1996). The distribution of earnings in Austria, 1972–1991. Johannes Kepler Universität Linz.

Henderson, D. (2014). An Unintended Case for More Capitalism. Cato Institute: Working Papers. pp. 58–61.

Higgins, M., & Williamson, J.G. (1999). *Explaining inequality the world round: Cohort size, Kuznets curves, and openness*. National Bureau of Economic Research Working Paper 7224:57. Accessed 3 2014. http://www.nber.org/papers/w7224.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics, 115,* 53–74.

Jahan, S., Jespersen, E., Mukherjee, S., Kovacevic, M., Bonini, A., Calderon, C., et al. (2015). *Human development report 2015: Work for human development*. New York: PNUD, EE.UU.

Jauch, S., & Watzka, S. (2016). Financial development and income inequality: A panel data approach. *Empirical Economics, 51*(1), 291–314.

Kanbur, R. (2017). *Structural transformation and income distribution: Kuznets and beyond*.

Kiatrungwilaikun, N., & Suriya, K. (2015). Rethinking inequality and growth: The Kuznets curve after the millennium. *International Journal of Intelligent Technologies and Applied Statistics, 8*(2), 159–169.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review, 45*(1), 1–28.

Leigh, A. (2008). Do Redistributive State Taxes Reduce Inequality? *National Tax Journal*, *61*(1), 81–104.

Lemieux, T. (2016). *Inequality and changes in task prices: within and between occupation effects. In inequality: causes and consequences* (pp. 195–226). UK: Emerald Group Publishing Limited.

Lempert, D. (1987). A demographic-economic explanation of political stability: Mauritius as a microcosm. *Eastern Africa Economic Review, 3*(1), 77.

Lyubimov, I. (2017). Income inequality revisited 60 years later: Piketty vs Kuznets. *Russian Journal of Economics, 3*(1), 42–53.

Magness, P. W., & Murphy, R. P. (2015). Challenging the empirical contribution of Thomas Piketty's capital in the 21st century. *Journal of Private Enterprise*, *30*, 1–43.

Matos, C. (2019). Inequality and crisis: conspicuous consumption as the missing link in the Portuguese case. *Journal of Economic Issues, 53*(1), 26–38.

McCloskey, D. N. (2014). Measured, Unmeasured, Mismeasured, and Unjustified pessimissim: A Review gjgjEssay of Thomas Piketty's Capital in the twenty-first century. *Erasmus Journal for Philosophy and Economics*, *7*(2), 73–115.

Meneejuk, P., & Yamada, W. (2016). *Analyzing the relationship between income inequality and economic growth: does the Kuznets curve exist in Thailand?*.

Milanovic, B. (2000). *Determinants of cross-country income inequality: An 'augmented' Kuznets hypothesis. Equality, participation, transition* (pp. 48–79). UK: Palgrave Macmillan.

Milanovic, B., (2016). *Introducing Kuznets waves: How income inequality waxes and wanes over the very long run*. Voxeu.

Molina-Morales, A., Amate-Fortes, I., & Guarnido-Rueda, A. (2013). Institutions and public expenditure on education in OECD countries. *Hacienda Pública Española, 204*(1–2013), 67–84.

Nasr, A. B., Balcilar, M., Saint Akadiri, S., & Gupta, R. (2019). Kuznets curve for the US: A reconsideration using cosummability. *Social Indicators Research, 142*(2), 827–843.

Nielsen, F. (2017). Inequality and inequity. *Social Science Research, 62,* 29–35.

Palma, J. G., (2016). *Do nations just get the inequality they deserve? The 'Palma Ratio're-examined* (no. 1627). Faculty of Economics, University of Cambridge.

Panayotou, T., (2000). *Economic growth and the environment, Center for International Development (CID) at Harvard University Working paper No. 56 and environment and development paper no. 4*.

Parente, F. (2019). Inequality and social capital in the EU regions: a multidimensional analysis. *Regional Studies, Regional Science, 6*(1), 1–24.

Piketty, T., (2013). *Capital in the twenty-first century*. The Belknap Press of Harvard University Press.

Polity IV Project (2013). *Political regime characteristics and transitions, 1800–2002*, disponible en http://www.systemicpeace.org/polity/polity4.htm.

Prados de la Escosura, L. (2008). Inequality, poverty and the Kuznets curve in Spain, 1850–2000. *European Review of Economic History, 12,* 38. https://doi.org/10.1017/S1361491608002311.

Prete, A. L. (2018). Inequality and the finance you know: Does economic literacy matter? *Economia Politica, 35*(1), 183–205.

Rattan, D. (2012). Exploring the Kuznets Curve and the 'Latin American Effect' in Latin America. In K. Kulkarni (Ed.), *International economic review: Post recession challenges & analyses* (pp. 277–287). Daryaganj: Matrix Publishers.

Rodrik, D. (2007). *The real exchange rate and economic growth: Theory and evidence*. USA: Harvard University.

Roodman, D. (2009a). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*(1), 135–158.

Roodman, D. (2009b). How to Do xtabond2: An introduction to "Difference" and "System" GMM in Stata. *Stata Journal, 9*(1), 86–136.

Saunders, P. (2004). Examining recent changes in income distribution in Australia. *The Economic and Labour Relations Review*, *15*(1), 51–73.

Solt, F., (2016). The standardized world income inequality database. *Social Science Quarterly, 97*. SWIID Version 5.1.

Stiglitz, J. E. (2012). *The price of inequality: How today's divided society endangers our future*. USA: WW Norton & Company.

Sulkowski, A., & White, D. S. (2016). A happiness Kuznets curve? Using model-based cluster analysis to group countries based on happiness, development, income, and carbon emissions. *Environment, Development and Sustainability, 18*(4), 1095–1111.

Theil, H. (1979). The measurement of inequality by components of income. *Economics Letters, 2*(2), 197–199.

Tsurumi, T., & Managi, S. (2010). Decomposition of the environmental Kuznets curve: Scale, technique, and composition effects. *Environmental Economics and Policy Studies, 11,* 19–36.

Utari, G. D., & Cristina, R. (2015). Growth and inequality in Indonesia: Does Kuznets curve hold? *Journal of Modern Accounting and Auditing, 11*(2), 93–111.

VanHeuvelen, T. (2018). Recovering the missing middle: A mesocomparative analysis of within-group inequality, 1970–2011. *American Journal of Sociology, 123*(4), 1064–1116.

Vargas-Quesada, B., Bustos-Gonzalez, A., & De, F., & Moya-Anegon, F. (2017). *Scimago institutions rankings: The most comprehensive ranking approach to the world of research institutions*.

Watkins, K., Carvajal, L., Coppard, D., Fuentes, R., Chosh, A., Giamberardini, C., et al. (2006). *Human development report 2006*. New York: PNUD. EEUU.

Windarti, N. A. N. I. K., Hlaing, S. W., & Kakinaka, M. (2019). Obesity Kuznets curve: international evidence. *Public Health, 169,* 26–35.

Yusuf, A. A., Sumner, A., & Rum, I. A. (2014). Twenty years of expenditure inequality in Indonesia, 1993–2013. *Bulletin of Indonesian Economic Studies, 50*(2), 243–254.

Zhang, W. (2014). *Has China crossed the threshold of the Kuznets curve?*. The University of Texas inequality project working paper. Retrieved from http://utip.gov.utexas.edu/papers.html.

Zhou, Q., & Shi, W. (2019). Socio-economic transition and inequality of energy consumption among urban and rural residents in China. *Energy and Buildings, 190,* 15–24.