KUZNETS INVERTED U-CURVE HYPOTHESIS EXAMINED ON UP-TO DATE OBSERVATIONS FOR 145 COUNTRIES

Oksana Melikhova, Jakub Čížek*

Abstract:

The Kuznets hypothesis of inverted U-curve dependence of the income inequality on the absolute value of the average income is still an unresolved issue despite the growing number of theoretical and empirical research on this topic. This paper analyzes the historical data on the average income and the income inequality for the period 1979-2009 collected for 145 countries. We found that the income inequality is influenced predominantly by governmental policy on subsidies and social transfers. Different amount of subsidies and social transfers across various countries makes the data biased. The inverted U-curve was found in countries with low amount of social contribution. However, increasing amount of social contributions makes the U-curve flat and shifts its maximum to higher values of the average income. Based on the experimental data a model describing the influence of both governmental policy and the level of economic development was developed.

Keywords: Kuznets curve, inverted U-curve, income inequality, comparative theory.

JEL Classification: O15, C23

1. Introduction

Simon Smith Kuznets (1955) was the first who suggested existence of a general relationship between the income inequality and the income *per capita*. His hypothesis states that the income inequality initially rises with economic development but after reaching its maximum it subsequently falls in advanced stages of economic development. Hence, the relationship between the income inequality and the average income expressed as GDP *per capita* has shape of inverted U-curve.

The Kuznets inverted U-curve hypothesis can be undoubtedly considered as one of the most influential statements ever made on inequality and development (Moran, 2005). Since Kuznets pioneering work in 1955 a number of theories have been proposed to explain the inverted U-curve relationship between the income inequality and the level of economic development. Anand and Kanbur (1993a) followed the original Kuznets idea and developed an inter-sectoral migration model. Williamson (1991) proposed that change in technology initially increases the income inequality, but it declines when skills

---

*   Institute of Economic Studies, Faculty of Social Sciences, and Faculty of Mathematics and Physics, Charles University in Prague, V Holešovičkách 2, Praha 8, Czech Republic (oksana.melikhova@ mff.cuni.cz; jakub.cizek@mff.cuni.cz).

related to the new technology become widespread. Barro (2000) suggested that each implementation of new technology is accompanied by dynamic Kuznets-type effect on the distribution of income. Aghion and Bolton (1997) attributed the Kuznets U-curve to credit market imperfections which cause different behaviour among the poor and the rich: in early stages of economical development the rich get richer, while the poor remain poor, at later stages of economical development the accumulation of wealth by the rich people pushes down their interest rates and allow the poor to become richer.

Although many theoretical models can predict the Kuznets U-curve, empirical evidence for the validity of the Kuznets hypothesis is still a matter of controversy. Empirical research on the validity of Kuznets hypothesis was performed by many authors during last 30 years, but obtained results are controversial and not conclusive. Ahluwalia (1976) in his early work found support for the Kuznets hypothesis. However, Anand and Kanbur (1993a) re-analyzed later the same data as Ahluwalia (1976) and did not find any evidence supporting the Kuznets hypothesis. In addition, Anand and Kanbur (1993b) analyzed also additional data than those used by Ahluwalia and again did not find any evidence for the Kuznets hypothesis. Deininger and Squire (1998) performed both cross-country analysis and examination of country specific time series. They did not found any support for the Kuznets inverted U-curve neither in the cross-country analysis nor in the country specific inter-temporal data. On the other hand, Jha (1996) analyzed observations for 76 countries for the period 1960-92 and found that Kuznets hypothesis holds. Similarly Milanovic (2000) reported that Kuznets hypothesis was supported by data for 80 countries during the 1980s. Bulir (2001) came to similar conclusion from analysis of cross-sectional data for 75 countries. Parametric and semi-parametric testing of Kuznets hypothesis performed by Lin *et al.* (2006) on the same data as Bulir (2001) gave also support for the Kuznets hypothesis. But recently Tam (2008) analyzed 'high quality' data for 84 countries and did not find any support for the Kuznets hypothesis.

The income inequality is usually measured by the GINI coefficient while the GDP *per capita* characterizes the level of economic development. Most studies followed the approach of Ahluwalia (1976) and performed linear regression of GINI index on the logarithm of GDP *per capita* and its square term, *i.e.* the relationship between GINI index and GDP *per capita* is assumed in the form

$$\text{GINI}_i = \beta_0 + \beta_1 \log(\text{GDP})_i + \beta_2 \log(\text{GDP})_i^2 + \varepsilon_i \, . \tag{1}$$

Positive coefficient β_1 and negative β_2 obtained from regression are viewed as a support of the validness of Kuznets hypothesis. The maximum of the inverted U-curve described by Equation (1) occurs at GDP *per capita*

$$\text{GDP}_{\text{TP}} = 10^{-\frac{\beta_1}{2\beta_2}}, \tag{2}$$

which according to the Kuznets hypothesis represents the 'turning point', *i.e.* the level of development from which inequality should decrease with further economical development. The maximum value of GINI coefficient predicted by the Kuznets U-curve is

$$\text{GINI}_{max} = -\frac{\beta_1^2}{4\beta_2} + \beta_0.$$ (3)

The both quantities GDP_{TP} and GINI_{max} calculated from results of analysis performed previously by various authors are listed in Table 1.

Table 1

Summarized Results of Previous Data Analysis Testing Validity of the Kuznets Hypothesis

Period	Number of countries	Number of pairs	Fraction of complete data (%)	GINI_{max} (%)	GDP_{TP} (constant 2000 US$)	Reference
1965-1971	60	60	0.6	57.6	642	Ahluwalia (1976)
1970-1990	75	75	0.8	62.7	2221	Bulir (2001)
1990-2000	44	45	0.5	45.0	2575	Hayami (2005)
1965-2003	82	82	0.8	46.0	2570	Iradian (2005)
1970-1990	75	75	0.8	45.9	912	Lin et al. (2006)
1979-2008	145	630	6.3	43.8	1528	This work

Note: The columns of the table show the time period covered by analyzed data, number of countries included in the analysis, number of samples (*i.e.* matching pairs of GINI index and GDP *per capita*), the fraction of the complete data set (approximately 10,000 pairs) represented by number of samples used in analysis, the maximum value GINI_{max} of the fitted model function and the position of the turning point GDP_{TP} of the fitted model function. Results of linear regression of data in Figure 1 by Equation (1) are shown in the table for comparison.

One can see in Table 1 that GDP_{TP} and GINI_{max} exhibit a big scatter which hinders any meaningful prediction of the turning point position. Since in the Kuznets hypothesis GDP_{TP} and GINI_{max} are principal quantities describing a fundamental economical law, the failure of various analyses to agree at GDP_{TP} and GINI_{max} arises some doubts about the validity of the Kuznets hypothesis.

The empirical analyses of Kuznets hypothesis suffer from a small number of data available. The income inequality is being systematically traced from 1960, *i.e.* for about of 50 years. Taking into account the number of countries in the world, a complete data set containing GINI index and GDP *per capita* pairs measured annually since 1960 would contain approximately 10,000 pairs. Unfortunately only a small fraction of these data is available mainly due to lack of information about GINI index. Table 1 summarizes the time period, the number of countries involved in investigation and the number of GINI index - GDP *per capita* pairs used in the previous studies published in literature. One can see in Table 1 that the number of GINI index - GDP *per capita* pairs analyzed so far represents less than 1% of the complete data set.

In this work we performed a detailed analysis of the relationship between the income inequality and the level of economic development on a large sample representing 6.3% of the complete dataset and we developed a model which describes the empirical observations.
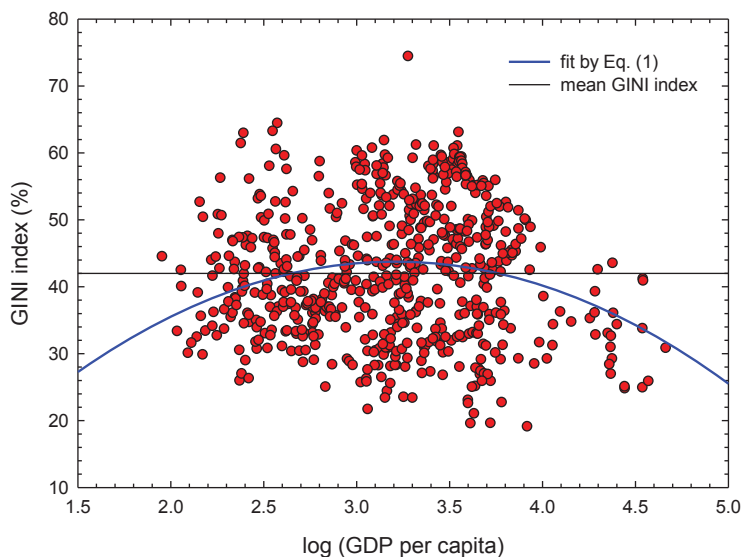
2. Results

Analyzed sample consists of data available in the online World Bank database World Development Indicators (WDI) (2011) for 145 countries for which GINI index was measured at least once in the period 1979-2009. In total WDI contains 6,228 records about GDP *per capita* and 644 records about GINI index for various countries in the period 1979-2009. There are 630 matching pairs of GINI index and GDP *per capita* for the same country and the same year. Hence, our sample represents 6.3 % of the complete data set, which is almost 8 times larger size of the sample compared to previous works. Figure 1 shows all data as a scatter plot of GINI index against the common logarithm of GDP *per capita* (expressed in constant US$ 2,000) for the same country in the same year.

Basic statistical description of all available observations and data for which matching pairs can be found are given in Table 2. Because of the limited number of GINI and GDP *per capita* pairs, it is important to check whether data in Figure 1 can be considered as a representative sample from the complete data set. Figure 2(a) and (b) show marginal probability density function (pdf) for GINI index and GDP *per capita*, respectively. The marginal pdf constructed from all available data irrespective if they form matching pairs or not, *i.e.* 644 records of GINI index and 6,228 values of GDP *per capita* are plotted by dashed lines, while solid lines show marginal pdf constructed only from the data forming matching pairs of GINI index and GDP *per capita*, *i.e.* 630 records. Since there are only a few values of GINI index for which the corresponding GDP *per capita* value is not available, shape of both marginal pdf's in Figure 2(a) is very similar. More interesting is comparison of marginal pdf's for logarithm of GDP *per capita*. One can see in Figure 2(b) that there is a reasonable agreement between the marginal pdf constructed from all 6,228 GDP records and that created only from GDP values for which the corresponding GINI index value is available. In comparison with the marginal pdf constructed from all available GDP records the pdf constructed from the matching pairs only exhibit slightly enhanced contribution from middle-income countries (log GDP *per capita* falling in the range from 3 to 3.8) at the expense of the contribution of high-income countries (log GDP *per capita* > 4). The student's t-test was employed to check if there is a statistically significant difference between the mean of the marginal pdf constructed from all GDP *per capita* records and that created from the matching pairs only. The p-value from the two sided student's t-test (assuming as a null hypothesis that both distributions have the same mean) is P = 0.30. Hence, we can conclude that there is no statistically significant difference in the mean of both pdf's in Figure 2(b) and the data in Figure 1 can be considered as a representative sample.

Visual inspection of data in Figure 1 did not reveal any regular pattern which would suggest the inverted U-curve dependence of the inequality on the logarithm of GDP *per capita*. Correlation tests were employed to examine possible relationship between the GINI index and the logarithm of GDP *per capita*. Results of these tests are summarized in Table 3.

Figure 1
Dependence of the GINI Index on the Common Logarithm of GDP per Capita for 145 Countries

Note: All historical observations in period 1979-2008 are plotted. Curved line shows result of linear regression using the model function described by Equation (1). Straight line shows the mean GINI index calculated as arithmetic average of all entries.

Source: WDI (2011), authors' calculations.

Table 2
Basic Statistic Description of GINI Index, Log GDP per Capita, Social Contributions (% GDP) and Subsidies and Other Transfers (% of GDP)

	samples	mean	median	std	min	max
GINI	644	42.0	41.5	10.2	19.4	74.3
log GDP *per capita*	6228	3.35	3.31	0.70	1.79	5.07
social contributions (% of GDP)	1220	6.26	5.29	5.32	0.00	20.25
subsidies and transfers (% of GDP)	1630	11.14	9.23	8.41	0.027	47.1
GINI with matching GDP per capita pairs	630	42.2	41.7	10.0	19.5	74.3
GINI with matching social contributions (% of GDP) pairs	242	41.2	38.86	10.6	21.6	65.8
GINI with matching subsidies and transfers (% of GDP) pairs	232	39.1	36.8	9.91	21.6	61.0
log GDP per capita with matching GINI pairs	630	3.19	3.23	0.52	1.96	4.67
social contributions (% of GDP) with matching GINI pairs	242	5.87	4.99	4.64	0.00	19.4
subsidies and transfers (% of GDP) with matching GINI pairs	232	11.15	9.84	7.43	0.169	31.9

Figure 2

Marginal pdf of (a) the GINI Index and (b) the Logarithm of GDP per Capita
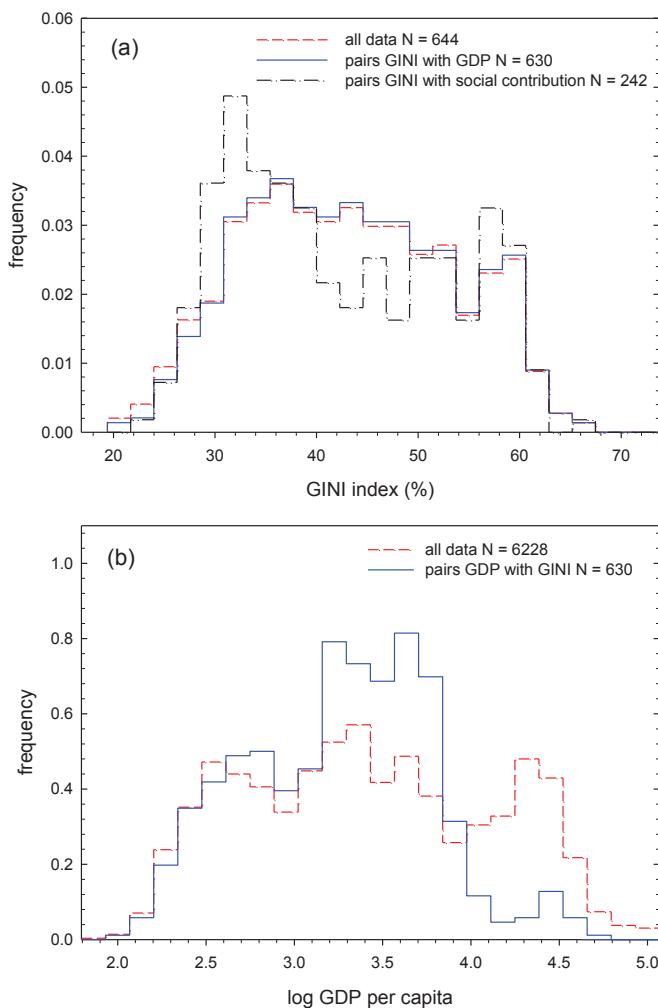
Note: Dashed lines show marginal pdf constructed from all available data of the GINI index (N = 644) and the GDP *per capita* (N = 6228) irrespective if they form matching pairs or not. The solid lines show marginal pdf constructed only from data forming matching pairs of the GINI index and the GDP *per capita* for the same country and the same year (N = 630). Dash-dotted line in the upper panel shows marginal pdf of the GINI index observations for which corresponding social contributions data forming matching pair for the same country and the same year are available (N = 242).

Source: WDI (2011), authors' calculations.

The Pearson sample correlation coefficient $\rho = 0.008$ was calculated for the data plotted in Figure 1. The P-value of 0.47 determined for the Pearson correlation coefficient by a two sided permutation test (assuming as a null hypothesis zero correlation between GINI and GDP *per capita*) testifies that there is no statistically significant Pearson correlation. This result should be taken with caution since the Pearson coefficient is a measure of a linear

dependence between random variables, while the Kuznets hypothesis predicts quadratic relationship between the GINI index and the logarithm of GDP *per capita*. Therefore, we analysed the data in Figure 1 also using the Spearman rank order correlation test which examines whether there is a monotonic relationship between both quantities but it is independent on shape of this dependence. Results of this analysis shown in Table 3 do not show any statistically significant Spearman correlation.

Table 3

Results of Statistical Correlation Tests and the χ² Statistical Test of Independence

Method	Test statistics: correlation coefficient / χ² per degree of freedom	95 % confidence interval	P-value
(i) GINI index versus logarithm of GDP per capita			
Pearson correlation test	0.008	(-0.086, 0.086)	0.47
Spearman correlation test	0.061	(-0.081, 0.081)	0.14
χ² test of independence	0.94	(0.89, 1.11)	0.32
(ii) GINI index versus social contributions (expressed in % of GDP)			
Pearson correlation test	-0.620	(-0.12, 0.12)	< 0.001
Spearman correlation test	-0.621	(-0.10, 0.10)	< 0.001
χ² test of independence	3.22	(0.81, 1.19)	< 0.001
(iii) GINI index versus subsidies and other transfers (expressed in % of GDP)			
Pearson correlation test	-0.497	(-0.12, 0.12)	< 0.001
Spearman correlation test	-0.504	(-0.10, 0.10)	< 0.001
χ² test of independence	2.94	(0.81, 1.19)	< 0.001
(iv) GINI – $\beta_{s,1}$ s_i versus logarithm of GDP per capita			
Pearson correlation test	0.329	(-0.12, 0.12)	< 0.001
Spearman correlation test	0.371	(-0.10, 0.10)	< 0.001
χ² test of independence	1.85	(0.89, 1.11)	< 0.001

Note: The tests are applied on (i) pairs of the GINI index and the logarithm of GDP *per capita*; (ii) pairs of the GINI index and the subsidies and other transfers (expressed in % of GDP); (iii) pairs of the GINI index and the social contributions (expressed in % of GDP); (iv) pairs of the part of GINI index unexplained by social contributions ($GINI_i - \beta_{s,1} s_i$) and the logarithm of GDP *per capita*. The 95% confidence interval for the correlation coefficients was determined using permutation test. The P-value in the last column of table was calculated assuming zero correlation or independence of investigated quantities as the null hypothesis in correlation and χ² test, respectively.

The Kuznets hypothesis predicts inverted U-curve relation between the GINI index and the logarithm of GDP *per capita*. If by chance the position of maximum of this dependence is located close to the median of the logarithm of GDP *per capita*, then the resulting correlation coefficients could be close to zero even if the Kuznets hypothesis holds. For this reason we analysed data in Figure 1 also by the χ² test of independence which can be considered as an ultimate test for existence of *any relationship* between the GINI index and the logarithm of GDP *per capita*. Assuming as a null hypothesis that the GINI index and the logarithm of GDP *per capita* are independent variables, the χ² test yields

the value of χ^2 *per* degree of freedom $\chi^2/\nu = 0.94$ and the p-value of 0.32 (see Table 3). Hence, the χ^2 test testifies that there is no reason to reject the null hypothesis and no support for the inverted U-curve dependence of the GINI index on the GDP *per capita* was found. This result indicates that the income inequality is influenced predominantly by *other factors* than the level of economic development.

Table 4

Results of Linear Regressions of Empirical Data

| Quantity	Coefficient	Standard error	P-value
(i) Kuznets curve, samples $N = 630$			
$GINI_i = \beta_0 + \beta_{GDP,1}(\log GDP)_i + \beta_{GDP,2}(\log GDP)_i^2 + \varepsilon_i$			
β_0	-14	11	0.20
$\beta_{GDP,1}$	36.3	7.2	< 0.001
$\beta_{GDP,2}$	-5.7	1.1	< 0.001
R^2 (R^2 adj)	0.039 (0.036)		
F	16.1	$P(F\|H_0)$	2.8×10^{-7}
(ii) Linear regression on transfer and subsidies (% of GDP), samples N = 232			
$GINI_i = \beta_0 + \beta_{t,1} t_i + \varepsilon_i$			
β_0	46.51	0.99	< 0.001
$\beta_{t,1}$	-0.637	0.074	< 0.001
R^2 (R^2 adj)	0.247 (0.244)		
F	61.6	$P(F\|H_0)$	1.6×10^{-13}
(iii) Linear regression on social contribution (% of GPD), samples $N = 242$			
$GINI_i = \beta_0 + \beta_{s,1} s_i + \varepsilon_i$			
β_0	48.26	0.82	< 0.001
$\beta_{s,1}$	-1.233	0.079	< 0.001
R^2 (R^2 adj)	0.505 (0.503)		
F	246.1	$P(F\|H_0)$	1.1×10^{-38}
(iv) Linear regression on social contribution (% of GPD) and Kuznets curve, samples N = 242			
$GINIi = \beta_0 + \beta_{s,1} s_i + \beta_{GDP,1}(\log GDP)_i + \beta_{GDP,2}(\log GDP)_i^2 + \varepsilon_i$			
β_0	-22	12	0.056
$\beta_{GDP,1}$	38.7	7.2	< 0.001
$\beta_{GDP,2}$	-4.9	1.1	< 0.001
$\beta_{s,1}$	-1.64	0.10	< 0.001
R^2 (R^2 adj)	0.560 (0.555)		
F	103.1	$P(F\|H_0)$	8.3×10^{-43}

Note: The linear regressions of empirical data were made using: (i) Equation (1), *i.e.* dependence of the GINI index on the logarithm of GDP *per capita* and its square, see Figure 1; (ii) Equation (5) *i.e.* dependence of the GINI index on the volume of subsidies and other transfers t_i (expressed in % of GDP), see Figure 6; (iii) Equation (4), *i.e.* dependence of the GINI index on the amount of social contributions s_i (expressed in % of GDP), see Figure 5; (iv) Equation (6) containing a contribution of the social transfers and also square and linear term of the logarithm of GDP *per capita*. Goodness-of-fit is characterized by the R^2 coefficient and the R^2 adjusted to the number of degrees of freedom (R^2 adj) and also by F-test: F-value and the probability $P(F\|H_0)$ that the null hypothesis (H_0 : all regression coefficients except of β_0 are zero) is true.

Table 4 shows result of linear regression of data in Figure 1 by Equation (1). The parameters GDP$_{TP}$ and GINI$_{max}$ calculated from fitted parameters by equations (2) and (3) are listed in Table 1 and the model curve is plotted in Figure 1 by solid line. From fitting we obtained a negative value of β_2 and a positive value of β_1 coefficient which is in accordance with the inverted U-curve hypothesis. Both coefficients β_2 and β_1 are statistically significant (at the 1% significance level), which was confirmed also by the Fisher F-test, while uncertainty of the absolute coefficient β_0 is very large making this parameter statistically insignificant. The R^2-value of 0.039 achieved in fitting by Equation (1) is rather low which testifies to a poor agreement of the inverted U-shape curve with empirical observations. Thus, GDP *per capita* can explain only a small portion of the income inequality data. Similar conclusion has been drawn by Barro (2000).

Figure 1 presents aggregated data for the whole period 1979-2009. However, both the income inequality and the GDP *per capita* are quantities undergoing certain development with time. To examine this effect the mean value of GINI index and the logarithm of GDP *per capita* were calculated for each year and are plotted in Figure 3 (a) and (b), respectively. The number of records available in each year is shown in the figure as well.

Figure 3
Time Development of the Mean Value of (a) the GINI Index, (b) the Logarithm of GDP per Capita
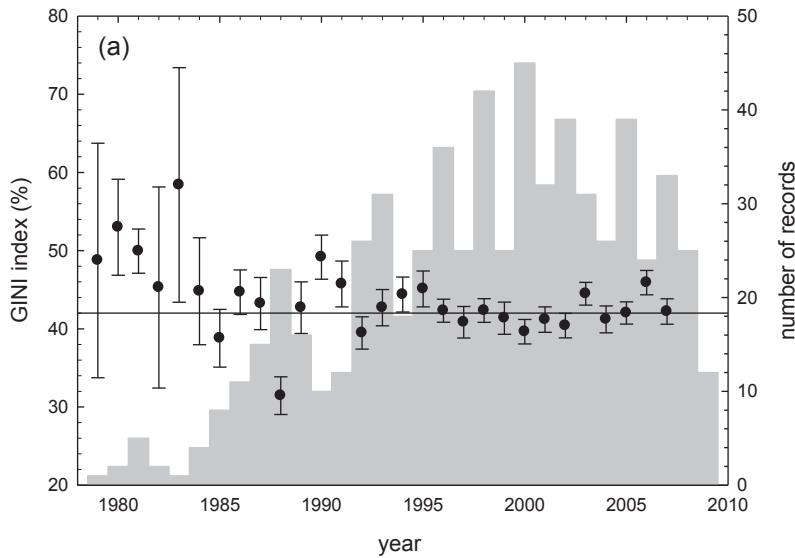

Note: Error bars indicate uncertainty of the mean values expressed as one standard deviation of the arithmetic average estimated from the spread of data in each year. Gray columns show the number of records for the GINI index and the GDP *per capita* available in the WDI for each year. The solid line in the upper panel shows GINI Index long-term average.

Source: WDI (2011), authors' calculations.

From inspection of Figure 3 it becomes clear that the mean value of the GINI index remains approximately constant and exhibits only statistical fluctuations around a long-term average indicated in Figure 3 (a) by a solid line. On the other hand, the mean value of the logarithm of GDP *per capita* strongly increases in the period 1995-2007, see Figure 3 (b). This marked increase in the GDP *per capita* remains, however, unnoticed by the GINI index, which supports the picture drawn above that the GINI index is predominantly influenced by other factors than the GDP *per capita*.

The mean value of the national GINI indexes calculated as an arithmetic average of all data in Figure 1 is $GINI_{mean} = (42 \pm 1)$ %, see the solid line in Figure 1 and Figure 3 (a). Banerjee and Yakovenko (2010) showed that when limited amount of money is distributed randomly between many agents then the principle of maximum entropy tells us that under constraints of constant total number of agents and total amount of money the equilibrium distribution of money between the agents is given by exponentially decaying function. Drăgulescu (2001) calculated that the GINI index for such exponential distribution equals 50 %. The mean value of the national GINI indexes is lower than 50 % corresponding to a random distribution of resources among inhabitants in each country. This is most probably due to active policy of governments in majority of countries with purpose to decrease the inequality of the income distribution of their citizens. Attempts to decrease the inequality of the income distribution are realized most frequently by social transfers of money from the rich members of society towards the poor citizens. Hence, if this policy works one should expect that the GINI index is influenced by the volume of the governmental social transfers. WDI contains data about social contributions (expressed in % of revenue). The social contributions multiplied by the revenue (expressed in % of GDP) give a measure of resources used by government to decrease the income inequality. From WDI we obtained 1,220 records for social contributions and these observations form 242 matching pairs with GINI for the same country and the same year (see Table 2). Hence, this sample represents 2.4% of the complete dataset. Again it was checked whether the matching pairs of GINI index and social contributions can be considered as a representative sample. Dash-dotted line in Figure 2(a) shows the marginal pdf of the GINI index records for which corresponding data about the social contributions are available. The shape of this distribution is in very reasonable agreement with pdf constructed from all available values of GINI index. Application of the student's t-test did not reveal any statistically significant difference in the mean of both distributions. Comparison of the marginal pdf of the percentage of social contributions constructed from all available data (N = 1220, dashed line) and that constructed only from data forming matching pairs with GINI index (N = 242, solid line) is shown in Figure 4.

Figure 4
Marginal pdf of the Social Contributions
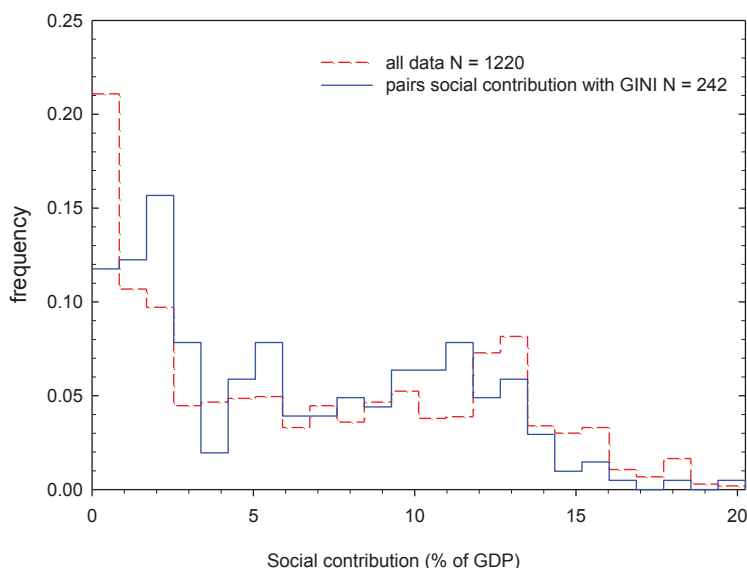
Note: Dashed line shows marginal pdf constructed from all available data (N = 1220) while solid line shows marginal pdf constructed only from data forming matching pairs with the GINI index for the same country and the same year (N = 242).

Source: WDI (2011), authors' calculations.

Shape of both pdf's is very similar and the student's t-test confirmed that there is no significant difference between both distributions. Hence 242 matching pairs of the GINI index and the social contribution can be considered as a representative sample.

Figure 5 shows a scatter plot of the GINI index plotted against the social contributions (expressed in % of GDP). One can see in the figure that there is indeed a decreasing trend of the income inequality with increasing fraction of social contributions. The relationship between the GINI index and the social contributions was examined again using the Pearson and Spearman correlation coefficients. Results of these tests listed in Table 3 testify that there is statistically *highly significant* (P-value < 0.001) *negative correlation* between the GINI index and the social contributions. The χ^2 test of independence employed to pairs of GINI index and the social contribution revealed that these quantities are almost certainly (p-value < 0.001) *not independent*. The relation between the GINI index and the social contributions s_i can be well fitted by a linear dependence

$$\text{GINI}_i = \beta_0 + \beta_1 s_i + \varepsilon_i \qquad (4)$$

The coefficients β_0, β_1 obtained from heteroscedasticity corrected weighted least square linear regression of data in Figure 5 are listed in Table 4. The linear coefficient β_1 is

negative and highly statistically significant. The R^2 value of 0.505 obtained in fitting by Equation (4) suggests that social contributions can explain a large portion of the cross-country differences in the income inequality.

If the social contributions are zero then linear regression by Equation (4) predicts GINI index (48.3 ± 0.8) % (see Table 4). This value is comparable with abovementioned GINI index of 50 % corresponding to a random distribution of limited resources among inhabitants. From the results in Table 4 one can easily calculate that fully equal income distribution characterized by GINI = 0 is achieved when the volume of social contributions becomes $s = (39.1 ± 0.1)$ expressed in % of GDP. Note that such volume of social contributions is roughly two times higher than the maximum s_i observation in our data set $s_{i,max} = 20.25$ % corresponding to France in 1996.

Figure 5
Dependence of the GINI Index on the Social Contributions
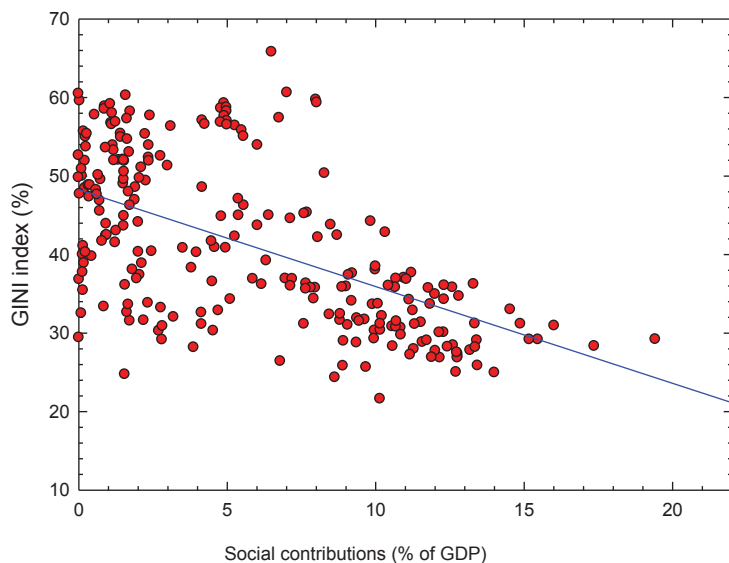
Note: The solid line shows linear regression by Equation (4). This plot was constructed from 242 matching pairs of the GINI index and the social contributions (expressed in % of GDP).
Source: WDI (2011), authors' calculations.

The volume of governmental subsidies and transfers (expressed in % GDP) can be considered as another indicator of the government attempts to reduce the inequality of the income distribution. This indicator was used together with the logarithm of GDP *per capita* and its square term as one of the explanatory variables in linear regression of the GINI index by Milanovic (2000). The WDI (2011) contains 1,630 records for volume of subsidies and other transfers (expressed in % of GDP) for the period 1979-2009 (see Table 2). From these data we can make 232 matching pairs of the GINI index and the percentage of subsidies and other transfers representing 2.3% of the complete dataset.

Again using the t-test it was confirmed that the matching pairs of the GINI index and the percentage of subsidies and other transfers can be considered as a representative sample. Figure 6 shows dependence of the GINI index on the volume of subsidies and other transfers (expressed in % of GDP).

Figure 6
Dependence of the GINI Index on the Volume of Subsidies and Other Transfers
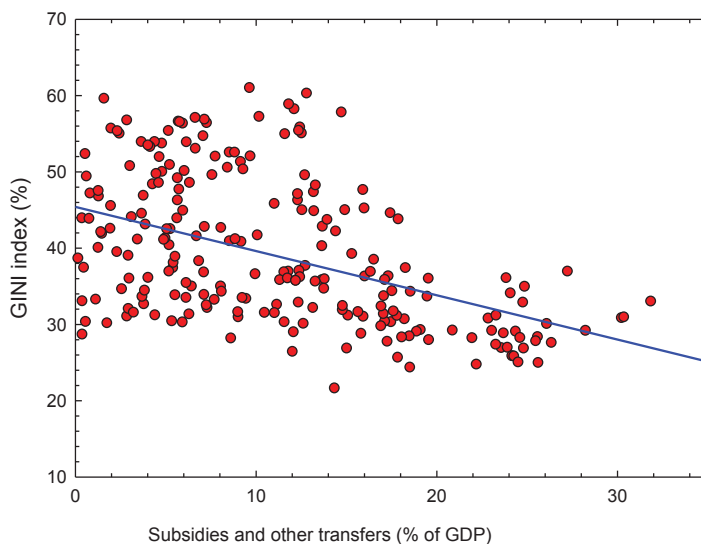
Note: The volume of subsidies and other transfers is expressed in % of GDP. The solid line shows linear regression by Equation (5). This plot was created from 232 matching pairs of the GINI index and the volume of subsidies and other transfers.

Source: WDI (2011), authors' calculations.

The GINI index decreases with increasing volume of subsidies and other transfers. Statistically significant negative correlation between the GINI index and the volume of subsidies and other transfers was found using both the Pearson and the Spearman correlation test (see Table 3). However, from inspection of Table 3 it is clear that GINI index is less correlated with the volume of subsidies and other transfers than with the social contributions. The dependence of GINI index on the volume of subsidies and other transfers t_i was fitted by a linear dependence

$$\text{GINI}_i = \beta_0 + \beta_1 t_i + \varepsilon_i \tag{5}$$

and results are shown in Table 4. A weighted least square method was used to account for heteroscedasticity of data. Similarly to dependence of the GINI index on the social contribution the linear coefficient β_1 obtained from regression is negative and highly statistically significant. However, the R^2 value for the linear fit of the GINI index on the volume of subsidies and other transfers by Equation (5) is significantly lower than the R^2 value for the linear fit of the GINI index on the social contributions by

Equation (4). Taking into account that number of samples in fitting by Equation (4) and (5) was comparable one can conclude that the volume of subsidies and other transfers is less explanatory variable than the amount of social contributions.

Figure 7 shows the volume of subsidies and other transfers plotted *versus* the social contributions. As expected these two variables are strongly correlated, the Pearson correlation coefficient is $\rho = 0.873$. However, detailed examination revealed that subsidies and other transfers contain not only transfers intended for decreasing the income inequality but also transfers made for other purposes. This can be well seen in Figure 7 which contains some records with relatively large volume of subsidies and other transfers, while the social contribution is very low or even zero. For this reason social contributions represent better measure of government attempts to decrease inequality of income distribution than the volume of subsidies and other transfers.

Figure 7

The Volume of Subsidies and Other Transfers Plotted versus the Social Contributions
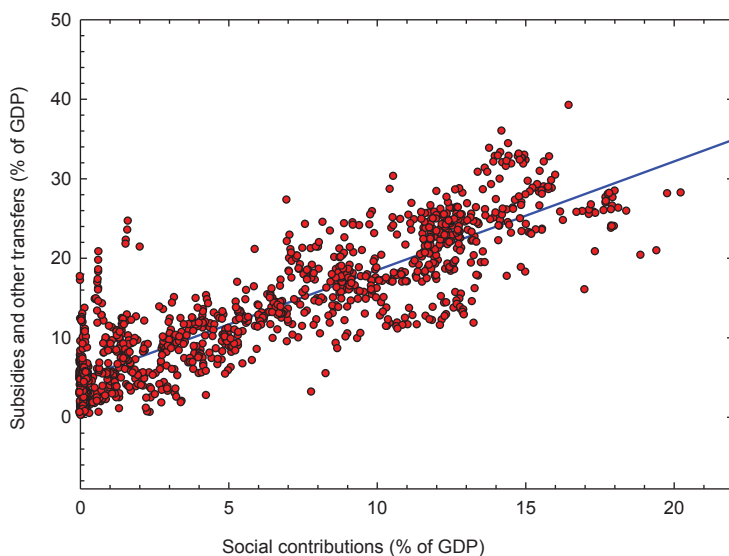
Note: The solid line shows linear regression. This plot was created from 1,131 matching pairs of the volume of subsidies and other transfers and the social contributions.
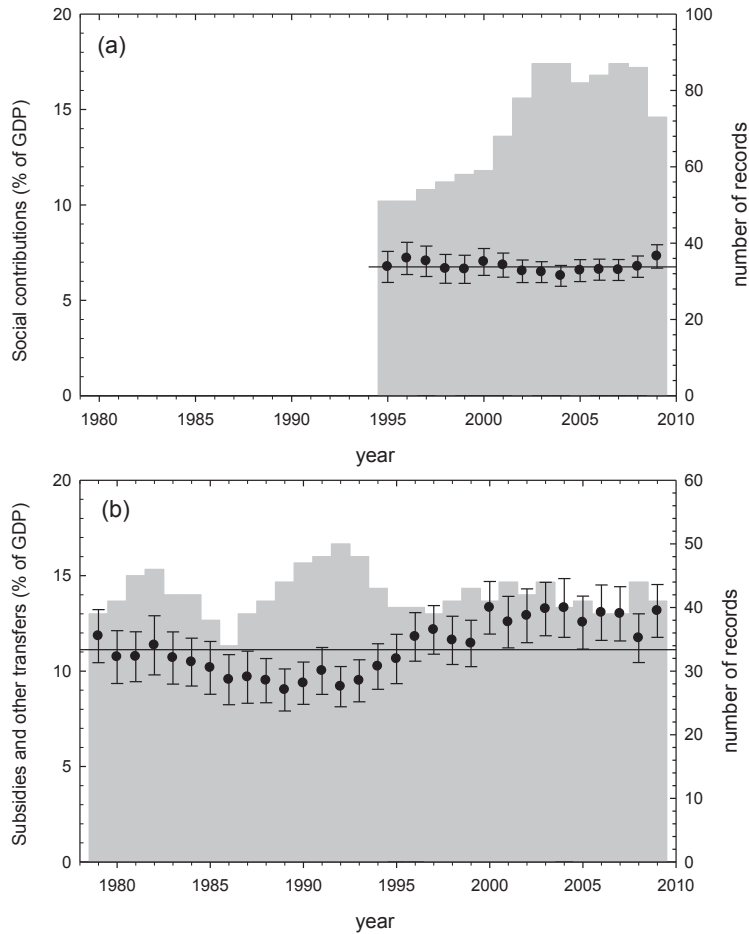
Source: WDI (2011), authors' calculations.

Figure 8 (a) shows time development of the mean value of social contributions calculated as an arithmetic average for each year for the time period 1995-2009 for which the data about the social contributions are available in WDI (2011). For comparison the time dependence of the mean value of the volume of subsidies and other transfers is plotted in Figure 8 (b). In contrast to the GDP *per capita* plotted in Figure 3(b) which increases in the period 1995-2007, the social contributions remain approximately constant and also the subsidies and other transfers exhibit only very slight variation with time. This is

consistent with behaviour of the GINI index which remains essentially constant and only fluctuates around a long-term average, see Figure 3(a).

Thus, our analysis confirmed that social transfers applied by governments represent the main factor which strongly influences the GINI coefficient. Various amounts of social contributions in various countries make the data in Figure 1 biased and causes large scatter of the data points in the figure. Thus, it is clear that the amount of social contributions have to be included into regression of data in Figure 1 leading to equation

$$\text{GINI}_i = \beta_0 + \beta_{s,1} s_i + \beta_{GDP,1} \log(\text{GDP})_i + \beta_{GDP,2} \log(\text{GDP})_i^2 + \varepsilon_i. \tag{6}$$

Figure 8

Time Development of the Mean Value of (a) the Social Contributions and (b) the Subsidies and Other Transfers

Note: The long-term averages are shown by the solid lines. Gray columns show the number of records for the social contributions and the subsidies and other transfers available in the WDI (2011) for each year.
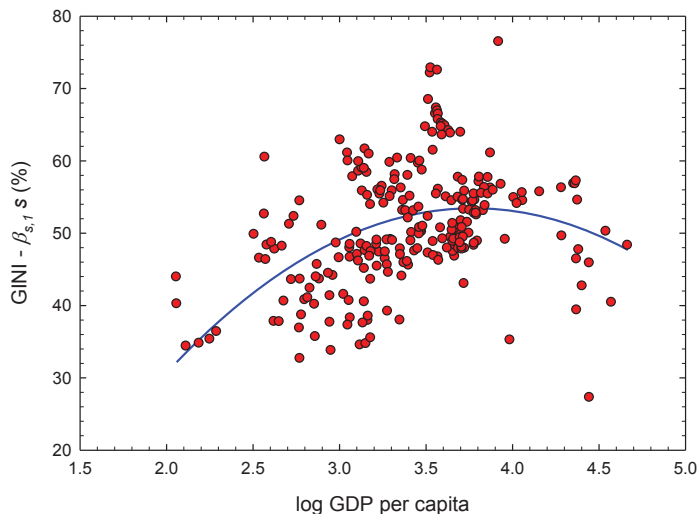
Source: WDI (2011), authors' calculations.

Results of heteroscedasticity corrected linear regression by Equation (6) are listed in Table 4. Negative coefficient $\beta_{GDP,2}$ and positive coefficient β_{GDP1} obtained from fitting (both statistically significant) are in agreement with the Kuznets hypothesis. The coefficient $\beta_{s,1}$ is negative and not far from that obtained from fitting by Equation (5). The R^2 value obtained in fitting by Equation (6) is only slightly improved compared to that achieved in fitting by Equation (5) assuming only the influence of the amount of social contributions. This testifies that the income inequality is influenced predominantly by the amount of social contributions.

However, results of fitting by Equation (6) indicate that *some remaining part* of the income inequality could be explained by the inverted U-curve dependence of the inequality on the level of economic development. Figure 9 shows a scatter plot of $GINI_i$ - $\beta_{s,i} s_i$, *i.e.* the portion of the income inequality *unexplained* by the amount of social contributes, plotted verses the logarithm of GDP *per capita*. Contrary to the original data in Figure 1 the unexplained part of the income inequality in Figure 9 shows clearly the inverted U-curve dependence. The data in Figure 9 were examined by the correlation tests and results are listed in Table 3. Both Pearson and Spearman correlation coefficients were found to be positive because the number of points on the increasing part of the inverted U-curve is much higher than in the decreasing part. The χ^2 test of independence led to rejection of the null hypothesis which indicates that data in Figure 9 cannot be considered as independent variables. The solid line in Figure 9 shows the best fit of data points by a parabolic curve.

Figure 9
Scatter Plot of GINI – $\beta_{s,1}s$ versus the Common Logarithm of GDP per Capita

Note: The GINI – $\beta_{s,1}s$ represents the part of the income inequality unexplained by the social contributions. The solid line shows best fit by a parabola.

Source: WDI (2011), authors' calculations.

It is a question whether the shape of the inverted U-curve is influenced by the amount of social contribution, *i.e.* whether shape of the Kuznets curve in countries with high and low amount of social contributions differs. In a hypothetical country with extremely high amount of social contributions *where all income is equally distributed among all citizens the Kuznets curve cannot exist* because the GINI index must be always zero independently on the level of economic development. This suggests that increasing amount of social contributions could make that Kuznets curve weaker. To examine this problem data in Figure 1 were divided into four groups according to the amount of social contributions. In Figure 10 the dependence of the GINI index on the logarithm of GDP *per capita* is plotted separately for countries with various amount of the social contributions: low amount 0-5%, Figure 10(a), intermediate amount 5-10%, Figure10(b), high amount 10-15%, Figure 10(c) and very high amount > 15%, Figure 10(d).

Figure 10

Scatter Plots of the GINI Index versus the Logarithm of GDP per Capita for Countries with Various Amount of Social Contributions
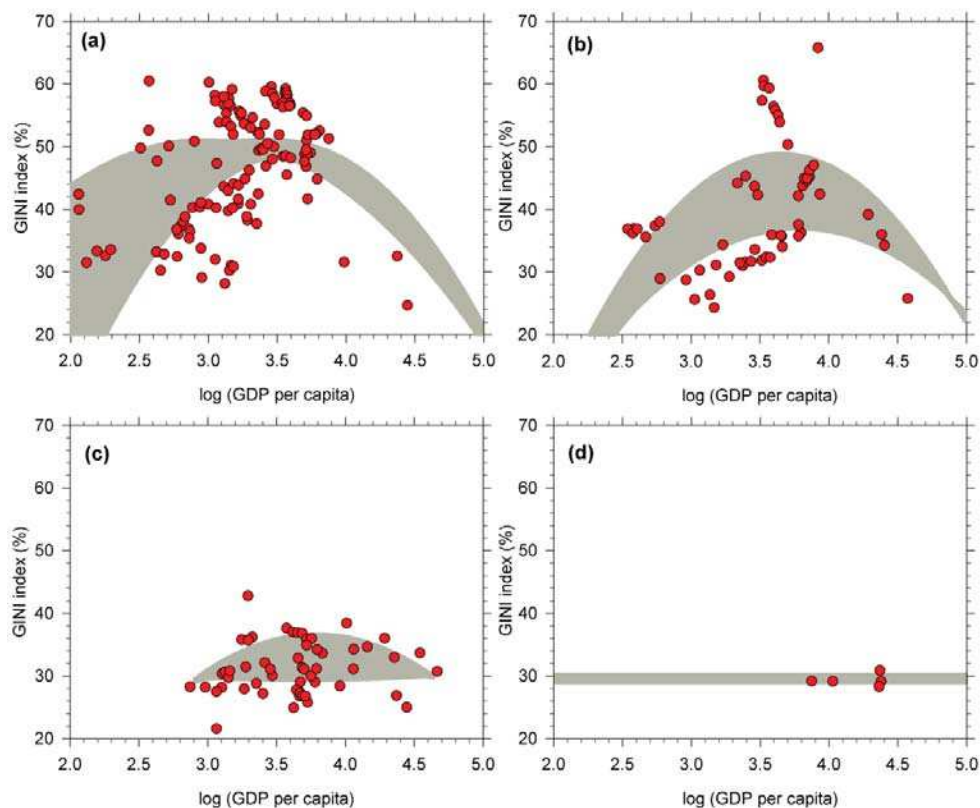
Note: The amount of social contributions is expressed in % of GDP: (a) low amount (< 5 %), (b) intermediate amount (5-10 %), (c) high amount (10-15 %), (d) very high amount (> 15 %).The shadowed areas show region predicted for countries with corresponding amount of social contributions by the model described by Equation (10).

Source: WDI (2011), authors' calculations.

The dependence of the GINI index on the logarithm of GDP *per capita* for countries with low and intermediate amount of social contributions exhibits the inverted U-curve predicted by the Kuznets hypothesis, see Figure 10(a, b). With increasing amount of social contributions the dependence of the GINI index of the logarithm of GDP *per capita* becomes more and more flat and is pushed down to the lower values of the GINI index. The dependence of the GINI index on the logarithm of GPD *per capita* for countries with high amount of social contributions shown in Figure 10(c) becomes almost flat and compared to the curves in Figures 10(a, b) is shifted to significantly lower values of the GINI index. Finally one can see in Figure 10(d) that the dependence of the GINI index on the logarithm of GDP *per capita* for countries with very high amount of social contributions (>15%) is completely flat although only few points are available here because the number of such countries is quite low.

For further analysis of the empirical data it is more appropriate to rewrite Equation (1) in the form

$$\text{GINI}_i = \text{GINI}_{\max}\left(1 - \frac{1}{w}\left(\log \text{GDP}_i - \log \text{GDP}_{\text{TP}}\right)^2\right) + \varepsilon_i, \tag{7}$$

where GDP_{TP} is the position of the turning point, $\text{GINI}_{\max}$ is the maximum of the Kuznets curve and w is a measure of the width of the inverted U-curve since full width at half maximum (FWHM) of the Kuznets parabola is $\text{FWHM} = (2w)^{1/2}$. From inspection of panels in Figure 10 it becomes clear that with increasing amount of social contributions the parabolic U-curve becomes wider and its maximum is pushed down while the position of its maximum is shifted towards higher values of GDP *per capita*. Thus, the parameters GDP_{TP}, $\text{GINI}_{\max}$ and w depend on the amount of social contributions and the Kuznets curve can be expressed as

$$\text{GINI}_i = \text{GINI}_{\max}\left(1 - f_{\max}\left(s_i\right)\right)\left(1 - \frac{1 - f_w\left(s_i\right)}{w}\left[\log \text{GDP}_i - \log \text{GDP}_{\text{TP}}\left(1 - f_{TP}\left(s_i\right)\right)\right]^2\right) + \varepsilon_i, \tag{8}$$

where the functions $f_{\max}(s)$, $f_w(s)$, and $f_{TP}(s)$ describe the dependence of the maximum, width and the turning point position on the amount of social transfers s, *i.e.* a decrease of maximum, an increase of width and a shift of the turning point position of the inverted U-curve. The functions $f_{\max}(s)$, $f_w(s)$, and $f_{TP}(s)$ are assumed in polynomial form

$$f_k\left(s\right) = \sum_{l=1}^{n_k} a_{k,l} s^l, \tag{9}$$

where $a_{k,l}$ are constants and the identifier k stands for max, w or TP. The Equation (8) can be then expressed in the form

$$\text{GINI}_i = \sum_{k=0}^{2}\sum_{l=0}^{n} c_{k,l}\left(\log \text{GDP}\right)_i^k s_i^l + \varepsilon_i, \tag{10}$$

where $c_{k,l}$ are parameters of the model which are fitted to the empirical data. Our analysis showed that the model function described by Equation (10) can be restricted only to terms with $n \leq 3$. Inclusion of higher-order terms does not lead to further improvement

in agreement of the model function with empirical data and the higher-order terms are statistically insignificant. Moreover, the coefficients $c_{1,3}$ and $c_{2,3}$ were found to be statistically insignificant and were, therefore, fixed at zero, *i.e.* $c_{1,3} = c_{2,3} = 0$. Results of heteroscedacity corrected linear regression by Equation (10) are shown in Table 5.

Table 5

Results of Linear Regression of Empirical Data in Figure 1 by Model Function Described by Equation (10)

| Quantity	Coefficient	Standard error	P-value
Kuznets curve with parameters the amount of social contributions, samples $N = 242$			
$$\mathrm{GINI}_i = \sum_{k=0}^{2}\sum_{l=0}^{n} c_{k,l}\left(\log \mathrm{GDP}\right)_i^k s_i^l + \varepsilon_i, \quad (n = 3,\ c_{1,3} = c_{2,3} = 0)$$			
$c_{0,0}$	-18.0	8.0	0.009
$c_{1,0}$	47	14	< 0.001
$c_{2,0}$	-7.9	2.4	< 0.001
$c_{0,1}$	-45	10	< 0.001
$c_{0,2}$	3.49	0.81	< 0.001
$c_{0,3}$	0.0203	0.0036	< 0.001
$c_{1,1}$	22.8	5.5	< 0.001
$c_{2,1}$	-2.72	0.75	< 0.001
$c_{1,2}$	-2.06	0.43	< 0.001
$c_{2,2}$	0.259	0.057	< 0.001
R^2 (R^2 adj)	0.745 (0.735)		
F	75.5	$P(F|H_0)$	6.8×10^{-64}
F_k	62.5	$P(F_k|H_{0,k})$	1.2×10^{-53}

Note: The model function represents the Kuznets curve with maximum, width and turning point position modified by the amount of social contributions. Goodness-of-fit is characterized by the R^2 coefficient and the R^2 adjusted to the number of degrees of freedom (R^2 adj) and also by F-test: F-value and the probability $P(F|H_0)$ that the null hypothesis (H_0 : all regression coefficients except of a constant term $c_{0,0}$ are zero) is true. In order to check the significance of the terms going beyond the classical Kuznets curve an additional F-test was performed using null hypothesis $H_{0,k}$ that all regression coefficients expect of $c_{0,0}$, $c_{1,0}$ and $c_{2,0}$, *i.e.* the terms included in the classical Kuznets curve, are zero. The F_k value obtained from the F-test and the probability $P(F_k|H_{0,k})$ that null hypothesis is true are given in the last row of the table.

All parameters $c_{k,l}$ used in the model are statistically significant at the 1% level of significance. The $R^2 = 0.745$ obtained from the linear regression testifies that the model described by that Equation (10) explains large portion of empirical data. Moreover, F-test was employed to check the statistical significance of our new model described by Equation (10) with respect to the traditional Kuznets curve, *i.e.* the inverted U-curve relationship between GINI index and logarithm of GDP *per capita* given by Equation (1). The results of the F-test listed in the last row of Table 5 clearly showed that our model describes empirical data significantly better than the traditional Kuznets curve.

Figure 11 shows a two-dimensional plot of the model curve given by Equation (10) with the parameters $c_{k,l}$ obtained from linear regression of experimental data. Note that the model function is plotted only in the region below a 'boarder line' $s_i = 10 \, (\log \text{GDP})_i - 18.5$ which defines the upper level of the volume of social contributions affordable at given level of economic development. There are no empirical data above this boarder line. One can clearly see in Figure 11 that dependence of the GINI index on the logarithm of GDP *per capita* has indeed shape of inverted U-curve, but with increasing amount of social contributions the maximum of the inverted U-curve is shifted to higher values of GDP *per capita* and the U-curves become more and more flat.

Figure 11
Two Dimensional Plot of the Model Described by Equation (10)
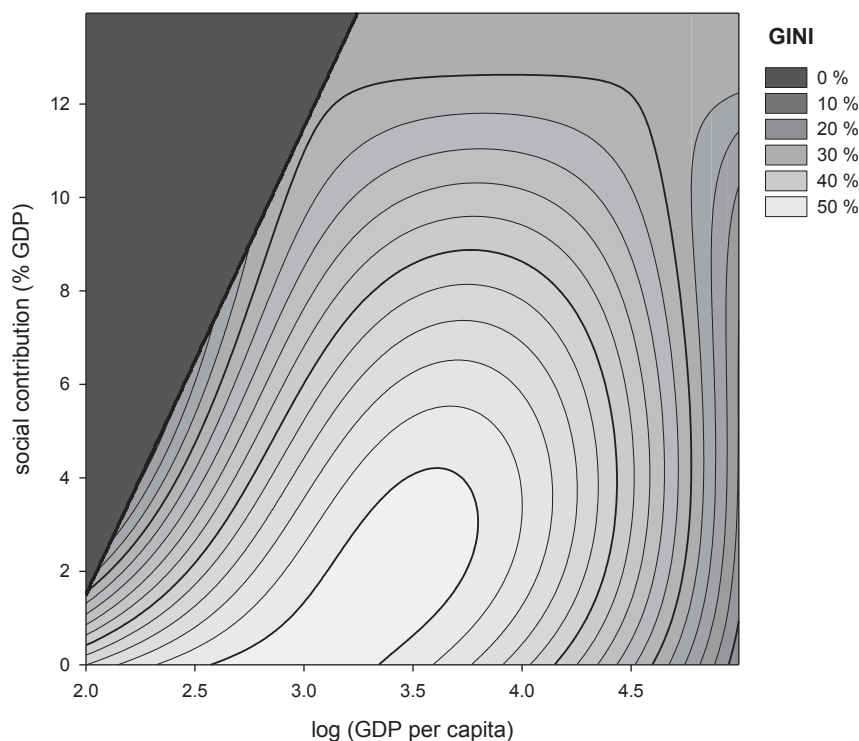
Note: Gray intensity coded GINI index is plotted as a function of the logarithm of GDP *per capita* and the amount of social contributions (expressed in % GDP).

Source: WDI (2011), authors' calculations.

The dependence of GDP_{TP}, *i.e.* the position of maximum of the inverted U-curve representing the turning point, is plotted in Figure 12. Increasing amount of social contributions causes a shift of the turning point to higher GDP *per capita* and trajectory of this shift exhibits S-shape when plotted versus logarithm of GDP *per capita*. In countries with very high amount of social contributions the position of turning point

approaches infinity and the Kuznets U-curve becomes so flat that it practically vanishes. The empirical data for various countries are shown in Figure 12 in the form of a bubble plot where size of each bubble is proportional to the GINI index value. Bubbles located close to the position of the turning point are on average larger compared to those located far away which is in agreement with the inverted U-curve.

Figure 12
**Comparison of Turning Point Position with Empirical Data**
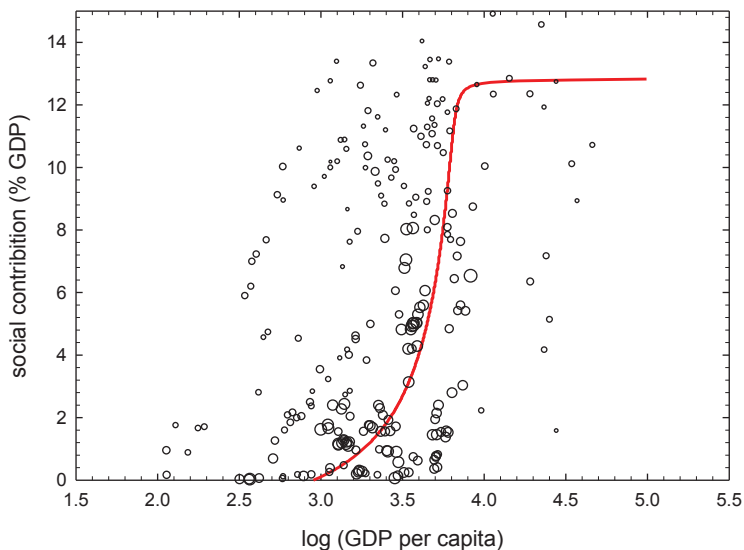
Note: Solid line shows the position of the turning point calculated from the model described by Equation (10). Empirical data are included in the figure in the form of a bubble plot. Size of each bubble is proportional to the income inequality measured by the GINI index.

Source: WDI (2011), authors' calculations.

Figure 13

The Maximum of the Inverted U-curve (GINI$_{max}$) Calculated by Equation (10) Plotted As a Function of the Amount of Social Contributions (expressed in % GDP)
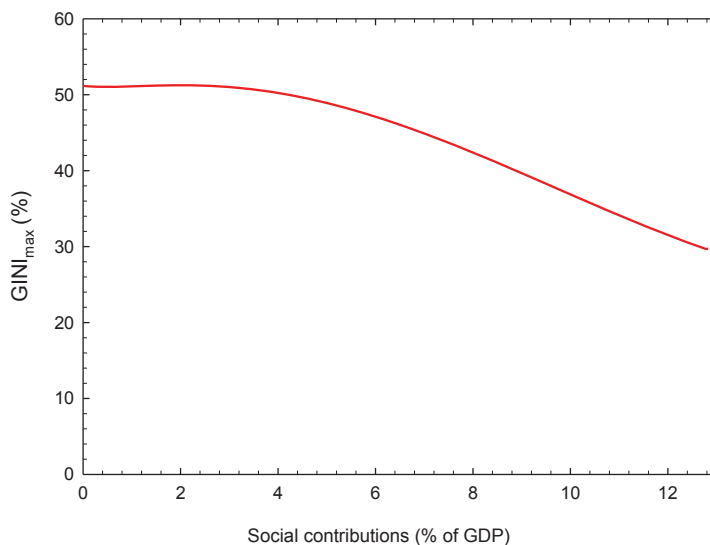
Source: WDI (2011), authors' calculations.

Figure 13 shows dependence of the maximum of the inverted U-curve, GINI$_{max}$, on the amount of social contributions. GINI$_{max}$ monotonically decreases with increasing amount of social contributions and this together with shift of GDP$_{TP}$ make the inverted U-curve flat. The shadowed areas in Figure 10 show a band defined by merged model curves calculated by Equation (10) for countries with low (< 5%), intermediate (5-10%), high (10-15%) and very high (> 15%) amount of social contributions. The area defined by Equation (10) is in reasonable agreement with empirical data.

3. Conclusions

The empirical test of the Kuznets hypothesis was performed on data for 145 countries in the period 1979-2009. It was found that the income inequality is strongly influenced by the amount of social contributions which makes the panel data for various countries biased. The inverted U-curve was found in the countries with low amount of social contributions. With increasing amount of social contribution the inverted U-curve flattens, its maximum decreases and position of the turning point is shifted to higher GDP *per capita*. Hence, our analysis suggests that long standing controversies regarding the existence or non-existence of the Kuznets inverted U-curve were likely caused by the fact that the U-curve was blurred due to various amount of social contributions which influence the income inequality much more than the level of economic development.

References:

Aghion, P., Bolton, P. (1997), "A Trickle-Down Theory of Growth and Development." *Review of Economic Studies*, Vol. 64, No. 2, pp.151–172.

Ahluwalia, M. S. (1976), "Inequality, Poverty and Development." *Journal of Development Economics*., Vol. 3, No. 4, pp. 307–342.

Anand, S., Kanbur, R.S.M. (1993a), "The Kuznets Process and the Inequality-Development Relationship." *Journal of Development Economics*, Vol. 40, No. 1, pp. 25–52.

Anand, S., Kanbur, R.S.M. (1993b), "Inequality and Development: A Critique." *Journal of Development Economics*, Vol. 41, No. 1, pp.19–43.

Banerjee, A., Yakovenko, V. M. (2010), "Universal Patterns of Inequality." *New Journal of Physics,* Vol. 12, No. 7.

Barro, R. J. (2000), "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth*, Vol. 5, No. 1, pp. 5–32.

Bulir, A. (2001), "Income Inequality: Does Inflation Matter?" *IMF Staff Papers*, Vol. 48, No. 1, pp. 139–159.

Deininger, K., Squire, L. (1998), "New Ways of Looking at Old Issues: Inequality and Growth." *Journal of Development Economics,* Vol. 57, No. 2, pp. 259–287.

Drăgulescu, A. A., Yakovenko, V. M. (2001) "Evidence for the Exponential Distribution of Income in the USA." *The European Physical Journal B,* Vol. 20, No. 20, pp. 585–589.

Hayami, Y., Godo, Y. (2005), *Development Economics: From the Poverty to the Wealth of Nations.* Oxford: Oxford University Press, pp. 200–204.

Iradian, G. (2005) "Inequality, Poverty, and Growth: Cross-Country Evidence." International Monetary Fund Washington Working Paper No. 05/28.

Jha, S. (1996) "The Kuznets Curve: A Reassessments." *World Development,* Vol. 24, No. 4, pp. 773–780.

Kuznets, S. (1955) "Economic Growth and Income Inequality." *American Economic Review*, Vol. 45, No. 1, pp.1–28.

Lin S.-C., Huang H.-C., Weng H.-W. (2006) "A Semi-Parametric Partially Linear Investigation of the Kuznets' Hypothesis." *Journal of Comparative Economics,* Vol. 34, No. 3, pp. 634–647.

Milanovic, B. (2000) "Determinants of Cross-Country Income Inequality: An Augmented Kuznets' Hypothesis," in Franicevic, V., Uvalic, M., eds., *Equality, Participation, Transition: Essays in Honour of Branko Horvat*. London: Macmillan & Co, pp. 48–79.

Moran, T. P. (2005) "Kuznets's Inverted U-Curve Hypothesis: The Rise, Demise, and Continued Relevance of a Socioeconomic Law." *Sociological Forum,* Vol. 20, No. 2, pp. 209–244.

Tam, H. (2008) "An Economic or Political Kuznets Curve?" *Public Choice,* Vol. 134, NO. 3–4, pp. 367-389.

Williamson, J.G. (1991) "British Inequality during the Industrial Revolution: Accounting for the Kuznets Curve" in Brenner, Y.S., Kaelble H., Thomas, M., eds., *Income Distribution in Historical Perspective*. Cambridge: Cambridge University Press.

WDI (2011), *World Development Indicators*. Washington, The World Bank, available at: http://databank.worldbank.org.