Income Inequality in the EU – Dynamic Panel Model[1]

Joanna Muszyńska, Jarosław Oczki, Ewa Wędrowska
Nicolaus Copernicus University, Toruń, Poland

ABSTRACT: The article aims at examination of the shape of relationship between income inequality and the level of economic development measured by GDP per capita in 27 European Union countries in the period of 2004-2014. It also aims at identification of determinants of income inequality. Specifically, we test for the existence of an inverted U-shaped curve, as it is predicted by the standard Kuznets hypothesis, and J-shaped curve following the approach adopted by Deutsch and Silber (2004) and Anand and Kanbur (1993). The data come from Eurostat EU-SILC database (European Union Statistics on Income and Living Conditions), World Bank and International Monetary Fund.
In the EU-27 group of countries we contradict the Kuznets hypothesis – our results provide evidence for a U-shaped, rather than the inverted U relationship. It also follows from our analysis that our data cover only the descending part of the U, that is a shape of inverted J.

KEYWORDS: income inequality, Kuznets curve, Gini index, panel data model, EU-SILC

JEL classification: D63, C23, O15, E24

Introduction

The phenomenon of growing dispersion of incomes has attracted growing attention of researchers and policy makers. Numerous empirical studies find that since the 1980s income inequalities in the developed countries have been rising (OECD, 2011; Salverda *et al*., 2014; Franzini *et al.*, 2016). Some researchers argue that this trend is a result of the rise of computerization and increasing prevalence of information technologies. Autor, Katz and Kearney (2006) describe a new pattern in income inequality in the US as "polarization" of the labour market into high-wage and low-wage jobs at the expense of middle-wage work. The authors find that computerization strongly complements the non-routine, abstract, cognitive tasks of high-wage employees, while directly substituting for the routine tasks characteristic for traditional middle-wage jobs. Kiatrungwilaikun and Suriya (2015) find that the growth of information technology positively influences human capital of employees, their productivity and wages in industrial sector, while it has little impact on incomes of agricultural workers. Other authors argue that the increase in income inequalities in the developed countries can be attributed to growing foreign trade and international fragmentation of production – particularly imports from low labour cost countries and moving labour-intensive production processes to low-wage countries. Findings from an empirical study based on data on United States, UK and Japan (Galbraith *et al.*, 2001) confirm that inequalities in these countries have increased in response to rising internationalization. Raitano (2016) reports an increase in inequality after the outbreak of the global financial crisis in 2008.

One of the most debated theoretical frameworks for analysing income inequality is Kuznets hypothesis. Kuznets first published his research results on the relationship between

[1] This paper is prepared within the research project: *Income Inequalities in the European Union Countries* (Eurostat, No.: 164/2016-EU-SILC) and is based on data from Eurostat, EU Statistics on Income and Living Conditions – EU-SILC CROSS-SECTIONAL UDB 2015 – version 2 of March 2017. The responsibility for all conclusions drawn from the data lies entirely with the authors.

income inequality and the level of economic development in 1955 (Kuznets, 1955). The hypothesis based on empirical evidence from time-series data on England, Germany and United States, predicts that the relationship between Gini index and GDP per capita can be described by an inverted U-shaped curve (Kuznets curve). A country at the early stage of its economic development, when a share of employment in agriculture is relatively high, experiences low income inequality. During a transition from a pre-industrial to an industrial stage of economic development inequality increases, because the productivity, as well as wages, of workers employed in emerging industrial sector are higher than in agriculture. The gap between wages in both sector widens in response to further growth of industry and shrinking of agriculture. As economic development proceeds, a change of the trend occurs – wage dispersion stops rising and starts to fall as a result of labour force shift from the agricultural sector towards the industry. Higher supply of labour in industry limits the wage growth in this sector and lower labour supply in agriculture creates favourable conditions for an increase of incomes of agricultural workers. These trends are reinforced by stronger growth of agricultural wages which are an effect modernization in farming and improvements in productivity. In effect, wage disparity between both sectors further narrows.

Early empirical studies on Kuznets hypothesis confirmed the statistically significant inverted U-shaped relationship between income inequality and GDP per capita (Paukert, 1973; Ahluwalia, 1976). However, later studies based on longer sample periods including 1980s and 1990s, and utilising better quality cross-sectional and panel data on income inequalities, provide no empirical evidence of the existence of the Kuznets curve (Deininger *et al.*, 1998; Bruno *et al.*, 1996; Ram, 1997; Barro, 2000; Thornton 2001; Phahan *et al.*, 2010). The latest empirical evidence on the subject has been mixed, but tends to contradict Kuznets hypothesis. Gallup (2012) using panel data of 87 countries found the existence of anti-Kuznets curve – a statistically significant U-shaped relationship between income inequality and the level of economic development. Similar results were obtained by Kiatrungwilaikun and Suriya (2015), Castells-Quintana, Ramos and Royuela (2015) and Muszynska, Oczki and Wedrowska (2017). In 1993 Anand and Kanbur proposed a modified functional form of the traditional Kuznets relationship – a model of income inequality regressed on GDP per capita and its inverse (Anand *et al.*, 1993). This specification was also applied by Ravallion (1995), Savvides and Stengos (2000), Checchi (2001), Gregorio and Lee (2002) and Frazer (2006). Deutsch and Silber (2004) using cross-section data on 23 developing and developed countries and decomposing the Gini index by income sources found that the Kuznets type relationship has a shape of inverted J.

The aim of this paper is to examine the relationship between income inequality and GDP per capita in a panel of European Union countries in the period of 2004-2014. Specifically, we aim at testing Kuznets hypothesis of the inverted U-shaped relationship as well as identifying determinants of income inequality. The plan of the paper is as follows. In section 2 we describe the model, the data set, and discuss estimation methods. Section 3 presents estimation results for static and dynamic specifications of Kuznets U-curve for EU countries. Section 4 provides estimation results of modified Kuznets functional form – the J-shaped relationship between Gini and GDP per capita. Section 5 concludes the paper.

1. Methodology and data

In order to empirically verify the thesis of the paper static and dynamic panel data models are applied. The use of panel data allows for greater number of degrees of freedom than mere time-series or cross-sectional data and improve the accuracy of parameter estimates (Hsiao, 2003, p.3).

In the most general form panel data model can be written as follows:

$$y_{it} = \alpha + X'_{it}\beta + u_{it},\ (i = 1,...,N),(t = 1,...T), \tag{1}$$

$$u_{it} = \mu_i + \varepsilon_{it} \tag{2}$$

where:

y_{it} – dependent variable,

$\mathbf{X}_{it}$ – vector of independent variables,

α – constant term,

β – vector of parameters,

u_{it} – error term,

μ_i – unobservable, individual–specific effects,

ε_{it} – unobservable, random term.

The static models can be estimated as fixed and random effects models. The former is based on the assumption that the individual effects are constant parameters, while in the latter it is assumed that the individual effects are a random variable (independent and identically distributed $\mu_i \sim IID\left(0, \sigma_\mu^2\right)$).

The parameters of the fixed and the random effects models are calculated with Least Squares Dummy Variables (LSDV) and Generalised Least Squares (GLS) estimators, respectively. In order to confirm the relevance of the decomposition of the constant term and/or the error term the Wald and the Breusch and Pagan Lagrange Multiplier tests are used. The appropriate specification of the models is checked with Hausman[2] specification test. The significance of variables and the degree of statistical fit of the estimated equations are also verified as well as the residuals of the models. The results of the statistical and economic verification enable us to choose the best models.

The dynamic panel data models are characterised by the presence of a lagged dependent variable among the regressors. In this case the model described by the equations (1) and (2) can be specified as follows:

$$y_{it} = \alpha + \delta y_{it-1} + X'_{it}\beta + u_{it},\ (i = 1,...,N),(t = 1,...T), \tag{3}$$

$$u_{it} = \mu_i + \varepsilon_{it}$$

The use of the lagged dependent variable as a right-hand regressor renders the Ordinary Least Squares (OLS) estimator to be biased and inconsistent. The GLS estimator is also biased in a dynamic panel data model.

The basic problem arising from the inclusion of the lagged dependent variable is its correlation with the error term. Since dependent variable is by definition a random variable, then it can be expected that its lagged values are correlated with the error term. A number of methods have been proposed for the estimation of dynamic panel models, which account for the endogeneity of the explanatory variables. Most of the estimators proposed base on the Generalised Method of Moments (GMM). As Bond (2002, p.160) points out, GMM is "particularly useful when the model of interest contains endogenous or predetermined explanatory variables, but the process generating these series are not completely specified".

The idea behind the GMM is that the population moment conditions can be replaced by the sample moment conditions. The properties of the regressors instruments imply that the moment conditions for the errors with the instruments are equal to zero. Based on that, GMM

[2] Both methods and the tests used to assess the specification of panel data models are widely discussed in the literature, inter alia, by Baltagi (2005) and Wooldridge (2010).

estimator establishes population moment conditions and then uses the sample moment conditions to compute parameter estimates.

The form and the number of moment conditions that can be used in the GMM estimation process depend on the assumptions concerning the correlation between the variables X_{it} and the components: μ_i and ε_{it}. Assuming that ε_{it} is not correlated over time $E(\varepsilon_{it}\varepsilon_{it-s})=0$ and X_{it} are correlated with μ_i, it is possible to adopt three alternative assumptions about the correlation between X_{it} and ε_{it}. The explanatory variables can be:

– strictly exogenous, if $E(X_{it}\varepsilon_{is})=0$ for all *t* and *s*,
– predetermined, if $E(X_{it}\varepsilon_{is})\neq 0$ for $s<t$, but $E(X_{it}\varepsilon_{is})=0$ for all $s\geq t$,
– endogenous, if $E(X_{it}\varepsilon_{is})\neq 0$ for $s\leq t$, but $E(X_{it}\varepsilon_{is})=0$ for all $s>t$.

Adopting additional assumptions about the lack of correlation between variables X_{it} and u_{it} allows to indicate additional instrumental variables, and thus additional moment conditions.

In our study we apply two methods based on GMM: first-differenced GMM (FDGMM), proposed by Arellano and Bond (1991), and the system GMM (SYSGMM), developed by Blundell and Bond (1998). The FDGMM transforms all regressors, usually by differencing, in order to remove individual-specific effects μ_i and then uses instruments to form moment conditions. Lagged levels of the dependent variable, the predetermined variables, and the endogenous variables are used to form GMM-type instruments. First differences of the strictly exogenous variables are used as standard instruments.

Since the lagged-level instruments in the FDGMM estimator become weak if the autoregressive process is too persistent or the ratio of the variance of the panel-level effects to the variance of the idiosyncratic error is too large, the SYSGMM method uses additional assumption that allows for the introduction of more instruments and improves efficiency.

In addition to the moment conditions of lagged levels as instruments for the differenced equation, the SYSGMM estimator applies moment conditions in which lagged differences are used as instruments for the level equation. The additional moment conditions are valid only if first differences of instrument variables are uncorrelated with the fixed effects (Blundell *et al.*, 1998a; 1998b; 2000).

We verified the quality of the estimated dynamic models on the basis of statistical tests: the Arellano-Bond test for autocorrelation and the Sargan test of over-identifying restrictions (Arellano *et al.*, 1991, p.282). The latter examines if over-identifying restrictions omitted from the estimation process were correct. More precisely, the test evaluates correctness of the selection of instrumental variables in estimation process in the sense of them being uncorrelated with the error terms of the first difference model. The null hypothesis of the test states that the moment conditions are valid. The applied instruments are correct if the test provides no grounds for rejecting the null hypothesis.

The Arellano-Bond test verifies the assumption regarding first- and second-order autocorrelation in the first-differenced errors. The model is properly specified, i.e. the GMM method provides consistent estimator, if the test fails to reject the null hypothesis about the absence of the second-order autocorrelation of the first difference model error term. Presence of the first-order autocorrelation is expected, resulting from the model construction. When the idiosyncratic errors are independently and identically distributed (i.i.d.), the first differenced errors are first-order serially correlated. We also applied Sargan difference test (Blundell *et al.*, 2000, p.21) to verify the validity of the additional instruments resulting from the SYSGMM estimator. Thus, we examined whether the use of SYSGMM method leads to more precise results than FDGMM estimator.

Numerous studies on income inequality, especially early publications on the subject, were criticized for poor quality of data (see Atkinson *et al.*, 2003). In this article we use highly reliable, internationally comparable Eurostat EU-SILC (European Union Statistics on Income and Living Conditions) data on Gini coefficients based on equivalised disposable household income before social transfers (pensions are included in social transfers). Data for our models come from Eurostat, World Bank and World Economic Outlook Database from International Monetary Fund. The set of countries in our sample include: EU-15 states: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, United Kingdom, and 12 new member states of the European Union: Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Slovenia. The EU-SILC data on all sample countries is only available for the time period of 2004-2014. In the countries of Eastern Europe which joined the European Union in 2004 and 2007 the EU-SILC survey started from 2005. The survey carried out in 2005 provides data on 2004. The latest available EU-SILC data for all countries come from 2015 survey thus limiting our sample period to 2014.

2. Empirical results and discussion

In the first step of our investigation we tested the Kuznets hypothesis of the inverted U-shaped relationship between income inequality and GDP per capita. Thus, we estimated the static panel model:

$$ln\,GINI_{it} = \alpha + \beta_1\, ln\,GDP_{it} + \beta_2 (ln\,GDP_{it})^2 + Z_{it}^{'}\gamma + \mu_i + \varepsilon_{it} \qquad (1)$$

In our analysis of Gini-GDP per capita relationship we also consider the influence of a set of control variables on income inequality. The list of dependent variables and their descriptions are presented in table 1.

Table 1. Description of variables

variable	definition	source of data
Gini	Gini coefficient of equalised disposable income before social transfers (pensions included in social transfers)	Own calculations based on Eurostat / EU-SILC
GDP	GDP per capita: real gross domestic product per capita based on purchasing-power-parity (PPP) measured in 2010 international dollars	Own calculations based on IMF database
Depend	Old-age dependency ratio: the ratio of the number of persons aged 65 and over to the number of persons aged between 15 and 64	Eurostat
Selfemp	Share of self-employed: the number of self-employed as a share of total number of employed	World Bank
Unemp	Unemployment rate: the number of unemployed persons as a percentage of economically active population	Eurostat
School	Tertiary education attainment: the number of persons holding tertiary education diploma as a share of population aged 15 and over	Eurostat
Open	Trade openness: (export+import)/GDP	World Bank
Agr	The number of people working in agriculture as a share of total number of employed	Eurostat

Source: own elaboration.

At this stage of our analysis we chose between fixed and random effects panel model specifications. In random-effects models, it is assumed that country-specific effects are normally distributed and that there is no correlation between these effects and other explanatory variables. In models based on samples including a small number of objects, such as ours, these assumptions may not hold. Our sample contains a relatively small number of objects (12, 15 or 27 EU countries) and in random-effect specification the probability of correlation between the country-specific effects and lnGDP and sq_lnGDP could be high. Thus, a fixed-effects model that does not require these assumptions should be preferred, despite some loss of efficiency (Galbraith and Kum, 2002, p. 10). By performing the Hausman test, we confirm that country-specific effects are correlated with a vector of explanatory variables. Thus, we continue our analysis with estimating the fixed-effects static model. The results of estimation of static panel models are presented in table 2.

Table 2. The estimates of the fixed-effects models – Kuznets specification

	EU-27		**EU-15**		**EU-12**	
	model 1		**model 2**		**model 3**	
lnGDP	-3.6040	***	-7.2906	**	-2.1555	
sq_lnGDP	0.1629	***	0.3352	**	0.0915	
selfemp	-0.0072	***	-0.0061		-0.0067	**
depend	0.0108	***	0.0114	***	0.0062	
unemp	0.0045	***	0.0040	**	0.0042	***
school	0.0023	*	0.0006		0.0053	**
open	0.0005	**	0.0007	**	0.0003	
agr	-0.0013		-0.0118		0.0017	
constant	23.4150	***	43.2253	**	16.0237	**
R2 within	0.5737		0.5527		0.6375	
R2 between	0.0557		0.0013		0.5160	
R2 overall	0.1068		0.0490		0.5185	
F (μ_i)	33.46	***	22.32	***	20.73	***
AR(1)	0.42		0.43		0.40	

*Not*e: ***,**,*: 1%, 5% and 10% statistical significance respectively.
Source: own calculations.

The models 1-3 have been estimated with LSDV (Least Squares Dummy Variables) method. In case of two country groups: EU-27 and EU-15, we obtain statistically significant parameters on both GDP variables: lnGDP and sq_lnGDP. The signs of the parameters $\beta_1 < 0$, β_2>0, ($|\beta_1|>|\beta_2|$) mean that the relationship between income inequality and the level of economic development has a shape of U rather than inverted U, as it is predicted by Kuznets hypothesis. Our results are consistent with the findings by Ram (1997), Gallup (2012) and Muszyńska, Oczki and Wędrowska (2017). Also Kiatrungwilaikun and Suriya (2015) provide the evidence on statistically significant U-shaped relationship between inequality and GDP per capita in a group of 91 countries in the period of 2000-2012, and Castells-Quintana, Ramos and Royuela, (2015) find this type of relationship in a panel of EU regions at NUTS 1 level.

In a case of EU-12 country group we find no evidence of Kuznets hypothesis. Alternatively, following the approach by Rodríguez-Pose and Tselios (2008) we tested simpler specification assuming linear relationship between lnGini and lnGDP (table 3). We find statistically significant negative relationship between these variables. Hausman test indicates that the country effects are random. As Rodríguez-Pose and Tselios claim, these results can be due to two reasons. Firstly, because of relatively short sample period the data cover only one part of the expected U-shaped curve. Secondly, the sample include a group of relatively homogeneous countries while numerous studies on Kuznets hypothesis are based on much wider pool of countries including the less developed as well as highly developed ones.

Table 3. The estimates of the models for EU-12 – fixed and random effects

UE12	**fixed effects (FE)**		**random effects (RE)**	
	model 4		**model 5**	
lnGDP	-0.3491	***	-0.2673	***
selfemp	-0.0060	**	-0.0041	*
depend	0.0066		0.0050	
unemp	0.0038	**	0.0051	***
school	0.0055	***	0.0050	***
open	0.0003		0.0003	
agr	0.0020		0.0042	
constant	7.0934	***	6.2678	***
R2 within	0.6324		0.6256	
R2 between	0.5788		0.6701	
R2 overall	0.5677		0.6499	
F (μ_i)	22.75	***	-	
AR(1)	0.42		-	
Hausman	8.118	p-value	0.3223	

*Not*e: ***,**,*: 1%, 5% and 10% statistical significance respectively.
Source: own calculations.

Since the error terms in our fixed-effects models 1 and 2 do not satisfy no autocorrelation assumption (what results in biased standard errors of the estimates and biased test statistics) we next apply an autoregressive specification AR(1). The equation (1) is now modified and can be specified as:

$$ln\,GINI_{it} = \alpha + \beta_1\, ln\,GDP_{it} + \beta_2 (ln\,GDP_{it})^2 + Z_{it}'\gamma + \mu_i + \varepsilon_{it}$$

where: $\varepsilon_{it} = \rho\varepsilon_{it-1} + \eta_{it}$ (2)

AR(1) approach did not bring the expected results – residuals in our models are still serially correlated. Autocorrelation of residual terms can result from a number of reasons. If it is a result of omitted lags of the dependent variable then not only standard errors but also estimated parameters could be biased. In such case a dynamic panel model can be used (Galbraith *et al.*, 2002, p. 13). Therefore in the next step of our investigation we add a lagged dependent variable lnGINI to the set of regressors in equation 3 similarly to the approach adopted by Chong *at al*. (2009, p. 16). Parameters are estimated with FDGMM and SYSGMM methods. Since these methods require a relatively large number of instrumental variables, only the model for EU-27 countries, based on a sample of 270 observations, has been estimated. In EU-15 and EU-12 country sets the number of objects would be too small and as a result the parameter estimates would be biased. The dynamic model is described by equation 3.

$$ln\,GINI_{it} = \alpha + \delta ln(\,GINI_{it-1}\,) + \beta_1\, ln(\,GDP_{it}\,) + \beta_2 (ln\,GDP_{it}\,)^2 + Z'_{it}\gamma + \mu_i + \varepsilon_{it}\,, \qquad (3)$$

The GMM estimator is consistent when the lagged values of the explanatory variables in equation 3 are valid instrumental variables (Chong *et al.*, 2009, p. 17). We investigate this issue in our model by using two specification tests: Sargan test of over-identifying restrictions suggested by Arellano and Bond (1991) and error term serial correlation test by Arellano and Bover (1995). Failure to reject the null hypothesis of the Sargan test leads to the conclusion that the instrumental variables used in the model are valid. With the latter test we examine the null hypothesis of absence of second-order serial correlation of the differenced error term. If the test fails to reject it, then there is no serial correlation of the original error term in levels.

Table 4. Dynamic panel models – EU-27

	strictly exogenous		predetermined		endogenous			
	FDGMM	**SYSGMM**	**FDGMM**	**SYSGMM**	**FDGMM**	**SYSGMM**	**WG**	**OLS**
	model 6	**model 7**	**model 8**	**model 9**	**model 10**	**model 11**	**model 12**	**model 13**
lagged Gini	-0.13394	0.56456 ***	0.43419 ***	0.57887 ***	0.42091 ***	0.57203 ***	0.41590 ***	0.85200 ***
lnGDP	-2.58683 *	-4.61717 ***	-4.88895 ***	-2.03663 ***	-4.93711 ***	-1.20271 *	-2.02570 *	-0.01340
sq_lnGDP	0.10993 *	0.21950 ***	0.23486 ***	0.09639 ***	0.23732 ***	0.05796 *	0.08970 *	0.00140
selfemp	-0.00133	0.00380 *	0.00188	0.00371 **	0.00178	0.00252	-0.00430 *	0.00020
depend	0.00927 **	0.00472 *	0.00183	0.00626 ***	0.00199	0.00710 ***	0.00340	0.00220 ***
unemp	0.00346 *	0.00085	0.00323 **	0.00152	0.00326 **	0.00225 **	0.00320 ***	0.00180 ***
school	0.00338 **	0.00263 *	0.00229	0.00186 *	0.00229	0.00111	0.00220 *	0.00050
open	0.00065 **	-0.00014	0.00031	-0.00002	0.00032	-0.00002	0.00030	0.00002
agr	-0.00959	-0.01392 ***	-0.01459 **	-0.00437	-0.01438 **	-0.00121	-0.00760 *	0.00070
constant		25.78476 ***		12.13429 ***		7.64408 **	13.53470 **	0.47580
no of inst.	17	27	33	63	33	61		
Sargan	13.8022	23.5095	25.8874	70.5998	26.9493	58.6163		
p-value	0.0870	0.1330	0.3590	0.0530	0.3070	0.2160		
diff-Sargan		9.7073		44.7124		31.667		
p-value		0.3747		0.0314		0.2446		
AR1 (p-value)	0.7345		0.0000		0.0000			
AR2 (p-value)	0.8964		0.6079		0.6168			

Note: ***,**,*: 1%, 5% and 10% statistical significance, respectively.
Source: own calculations.

Even though we estimate the models on a basis of the sample of all 27 countries, the number of objects is relatively small. Because of this limitation we introduce only the first lag of the dependent variable serving as a regressor. Each additional lag of dependent variable added to the model requires further expansion of the matrix of instrumental variables and can cause a bias in FDGMM and SYSGMM estimates (Kiviet, 1995). The bias can become significant and outweigh the gains in efficiency from the use of the GMM (Ziliak, 1997, p. 419; Dańska-Borsiak, 2011, p. 120). In order to limit the number of instruments in the model only certain lags, instead of all available lags, for instrumental variables can be used (Roodman, 2006, p. 16).

Since in case of small samples lagged variables can be weak instruments what can result in biased and inefficient FDGMM estimators we use also SYSGMM. The instruments are weak when the dependent and explanatory variables exhibit strong persistence and/or the relative variation of the fixed effects increases (Araujo and Cabral, 2015). In autoregressive model, such as ours, this is the case when the value of coefficient on the lagged variable is close to 1 and when variance of group effects rises along with variance of the error term (Dańska-Borsiak, 2011, p. 94). FDGMM and SYSGMM can be one step or two step estimators. We use one-step estimators since in small samples the asymptotic standard errors for the two-step estimators can be biased downward (Arellano *et al.*, 1991; Blundell *et al.*, 1998a; Pagano, 2004).

Up to this point we have assumed that GDP per capita variable is exogenous. However, in reality, the reverse relationship can exist, that is income inequality can influence GDP per capita (see Forbes, 2000; Naguib, 2015 and Pagano, 2004). In the next step of our analysis we estimate models for three cases. Firstly, it is assumed that all explanatory variables are strictly exogenous, secondly, that lnGDP and sq_lnGDP are predetermined, and thirdly, that both these variables are endogenous. Table 4 presents the results of FDGMM and SYSGMM estimation of equation 3. The values of Sargan test indicate that the instrumental variables used in all three types of models (strictly exogenous, predetermined and endogenous) are valid. Arellano-Bond test confirms first-order correlation and no second-order autocorrelation of error terms in first differenced models based on assumptions of predetermination and endogeneity of explanatory variables. Serial correlation of order higher than one implies that moment conditions are not valid, that is the set of instruments used in the GMM estimation method is not proper. In the first-differenced model with exogenous dependent variables, the lagged dependent variable is insignificant. Also, there is no error term serial correlation.

The choice of additional instruments in SYSGMM estimation is examined with a use of difference-Sargan test. We conclude that the SYSGMM outperforms FDGMM estimator in the case of exogenous and endogenous independent variables – models 7 and 11, respectively. Test value for model 8 (predetermined explanatory variables) shows that the optimal estimator is FDGMM. Difference-Sargan test is also used in order to compare models 7, 8 and 11. The most accurate model describing the relationship between income inequality and economic development in EU-27 countries is model 11, estimated with SYSGMM and assuming endogeneity of lnGDP and sq_lnGDP.

Additionally, we examine whether the estimates of coefficients on the lagged dependent variable are unbiased by confirming that their values are between those obtained from application of OLS and WG (within-group) methods as in Tam (2008), Arellano and Bond (1991), Soto (2009), and Arnone and Presbitero (2010). We expect the following relationship to be true:

$$\delta_{WG} < \delta_{GMM} < \delta_{OLS} \tag{4}$$

OLS estimates tend to be biased upward whilst within-group method gives a downward biased estimates of this coefficient (Blundell *et al.*, 1998b, p.9). In our analysis the estimate of coefficient δ on the lagged dependent variable takes the value of δ= 0,57203, that is between 0,4158 and 0,8520 obtained with WG method (model 12) and with OLS method (model 13), respectively.

Economic growth is not the only factor influencing the dispersion of income. Parameters on control variables in model 11 indicate that unemployment rate and old-age dependency ratio statistically significantly determine income inequality in EU countries. The higher unemployment rate and the higher the share of persons aged 65 and over in the number of persons aged between 15 and 64, the greater the inequality. These results are not controversial. High unemployment rate leads to a more dispersed income since unemployment is one of the main factors influencing standard of living and poverty, especially in our data on income before social transfers. The old age dependency ratio proved significant because our data on income do not include pensions, thus demographic structure of the society significantly influences income inequality. The share of self-employed, and persons employed in agriculture in total employment, share of persons holding tertiary education in total population and trade openness all proved insignificant in model 11, although some of these variables significantly influenced Gini index in models assuming strictly exogenous and predetermined dependent variables.

Static model 2 describing the U-shaped relationship in EU-15 country group indicates that there is a significant influence of the same two control variables – unemployment rate and old-age dependency ratio. Also, income inequality is determined by trade openness - the variable describing the degree of internationalization of the economies. The higher the share of the sum of imports and exports in GDP the greater income inequality. The positive sign of the parameter has been expected because globalization and openness of the economies is pointed out as one of the main causes of rising income inequalities in developed countries (Galbraith *et al.*, 2001). A slightly different set of control variables proved significant in static panel model 4, based on the sample of EU-12 countries. In the new member states of the EU unemployment rate and the share of population holding tertiary education degree in total population influenced the Gini index in period 2004-2014. Our results on the significant and positive influence of higher education attainment support the findings by Barro (2000).

3. Alternative specification - J-curve

Our empirical results confirm the U-shaped curve describing income inequality and economic development measured by GDP per capita. In the next step we examine an alternative specification of the model and check whether the resulting U curve is symmetrical. In the empirical studies two approaches to this issue are used. Galbraith and Kum (2002) identify the turning point on the U curve. Another approach is estimation of the hyperbolic Anand-Kanbur specification that regresses an inequality index on per capita income and its inverse, or on logarithm of per capita income and the inverse of the logarithm of income (Anand *et al.*, 1993). Hyperbolic model assumes that the underlying relationship between Gini coefficient and GDP per capita has the shape of J or inverted J depending on the signs of parameters, and can be written as:

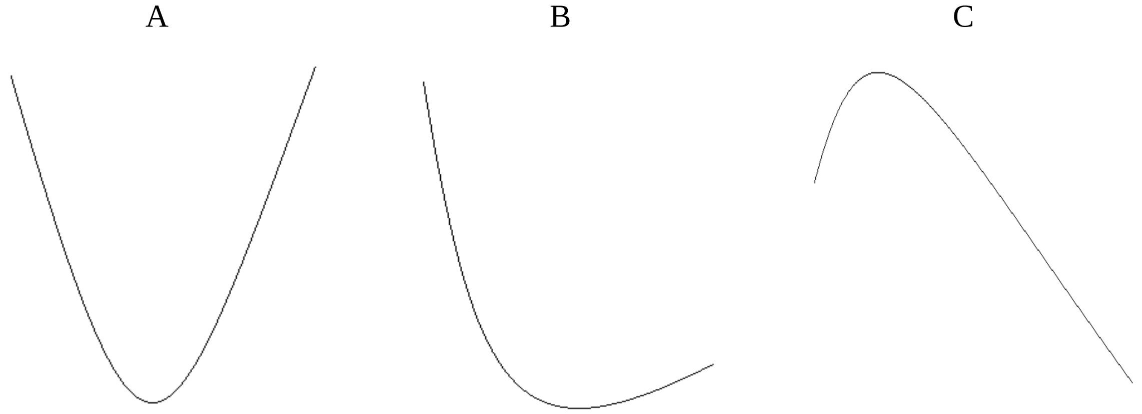
$$\ln(GINI) = a + b\ln(GDP) + c\frac{1}{\ln(GDP)} \tag{5}$$

If a turning point exists, and it can be calculated as $\sqrt{c/b}$, then it is a local extreme point (maximum or minimum value depending on the sign of *b* and *c*).

First derivative of function (5) is $\frac{b(\ln GDP)^2 - c}{(\ln GDP)^2}$ and it is equal zero if $\ln(GDP) = \pm\sqrt{c/b}$.

If $b > 0$ and $c > 0$ function (5) is convex and achieves a minimum value at $\ln(GDP) = \sqrt{c/b}$, because the second derivative $\frac{2c}{(\ln GDP)^3} > 0$ (curve B). If $b < 0$ and $c < 0$ function (5) is concave and achieves a maximum at $\ln(GDP) = \sqrt{c/b}$, because the second derivative $\frac{2c}{(\ln GDP)^3} < 0$, as sketched in C.

Figure 1. The shapes of considered relationships



Source: Author’s illustration.

Coefficients b and c describing curves B and C are both either positive or negative. If they have the opposite signs then the function 5 does not have minimum nor maximum. When $b < 0$ and $c > 0$ the function is convex and strictly decreasing, and when $b > 0$ and $c < 0$ it is concave and strictly increasing.

Similarly to the case of U-shaped relationship model, we have also included control variables (Z.) and the equation (5) can be written as:

$$ln\,GINI_{it} = \alpha + \beta_1\, ln\,GDP_{it} + \beta_2 \frac{1}{ln\,GDP_{it}} + Z'_{it}\beta_3 + \mu_i + \varepsilon_{it}, \qquad (6)$$

where Z represents a vector of control variables.

In the first step of the analysis of the J-shaped relationship we estimated static models for all three country groups. The results are presented in table 5.

Table 5. The estimates of the fixed effects models – J curve

	EU-27 model 14		EU-15 model 15		EU-12 model 16	
lnGDP	1.3690	**	3.8005	**	0.6060	
inv_lnGDP	170.8489	***	447.0019	**	92.8375	
selfemp	-0.0073	***	-0.0059		-0.0067	**
depend	0.0108	***	0.0111	***	0.0063	
unemp	0.0045	***	0.0038	*	0.0042	***
school	0.0022	*	0.0005		0.0053	**
open	0.0005	**	0.0007	**	0.0003	
agr	-0.0011		-0.0134		0.0017	
constant	-27.1136	**	-78.8381	**	-11.7241	
R2 within	0.5737		0.5555		0.6379	
R2 between	0.0580		0.0024		0.5146	
R2 overall	0.1096		0.0417		0.5175	
F (μ_i)	33.36		22.59		20.82	
AR(1)	0.42		0.43		0.40	

Note: ***,**,*: 1%, 5% and 10% statistical significance respectively.
Source: own calculations.

We find statistically significant influence of GDP variables – lnGDP and inv_lnGDP, on inequality in fixed-effects models for EU-27 and EU-15. There is no evidence of significant J-curved relationship in a group of countries EU-12. Thus we conclude that the relationship between logarithm of inequality and logarithm of GDP per capita in EU-12 country set is linear, as it follows from model 4 presented in table 3. Error terms in models 14 and 15 are serially correlated thus we estimate parameters again with AR(1) procedure similarly to the case of U-curve analysis. The resulting equation can be written as follows:

$$ln\, GINI_{it} = \alpha + \beta_1\, ln\, GDP_{it} + \beta_2 \frac{1}{ln\, GDP_{it}} + Z_{it}^{'}\gamma + \mu_i + \varepsilon_{it}$$

where: $\varepsilon_{it} = \rho\varepsilon_{it-1} + \eta_{it}$ (7)

As it was in the case of U-curve modelling this approach does not solve the problem of autocorrelation so we add lagged lnGINI variable on the right hand side of the equation:

$$ln\, GINI_{it} = \alpha + \delta ln(\, GINI_{it-1}\,) + \beta_1\, ln(\, GDP_{it}\,) + \beta_2 \frac{1}{ln\, GDP_{it}} + Z_{it}^{'}\gamma + \mu_i + \varepsilon_{it}, \tag{8}$$

Table 6 presents the equation (8) estimated with FDGMM and SYSGMM assuming: strict exogeneity, predetermination and endogeneity of lnGDP and the inverse of the logarithm of Gini index for EU-27 countries. The choice of instruments used in the estimators are confirmed by the Sargan test. The values of Arellano-Bond test for first-difference models indicate that there is first-order correlation and no serial correlation of second order in error terms in models assuming predetermination and endogeneity. In the model based on the assumption of strict exogeneity the lagged dependent variable proved statistically insignificant. Also there is no serial correlation of residuals.

The use of difference-Sargan test leads to a conclusion that SYSGMM estimator is superior over FDGMM in models 18 and 22, that is models based on assumptions of strict exogeneity and endogeneity of GDP per capita, respectively. FDGMM is a better estimator in model assuming predetermined GDP per capita variables. Specification 22, estimated with SYSGMM and assuming endogeneity of lnGDP and inv_lnGDP is indicated by difference-Sargan test as superior over all other J-curve specifications. We obtain the statistically significant coefficients on lnGDP and inv_lnGDP variables ($\beta_1 > 0$ and $\beta_2 > 0$) thus we conclude the resulting Gini-GDP per capita relationship has a shape of inverted J, as it is depicted in figure B.

Model 22 indicates to the same set of statistically significant control variables as model 11 describing the U-shaped relationship. Unemployment rate and old-age dependency ratio both have a positive impact on income inequality in the EU countries. The rest of control variables: the share of self-employed, and persons employed in agriculture in total employment, the share of persons holding tertiary education in total population and trade openness do not influence income inequality. In the EU-15 country group (model 15) apart from the statistically significant J-shaped relationship between lnGini and lnGDP, income inequality is influenced by unemployment rate, old-age dependency ratio and trade openness, that is the same set of determinants as in the model 2.

Table 6. The estimates of the dynamic panel models for EU-27 – J curve

	strictly exogenous		predetermined		endogenous			
	FDGMM	**SYSGMM**	**FDGMM**	**SYSGMM**	**FDGMM**	**SYSGMM**	**WG**	**OLS**
	model 17	**model 18**	**model 19**	**model 20**	**model 21**	**model 22**	**model 23**	**model 24**
lagged Gini	-0.14802	0.56592 ***	0.43106 ***	0.57251 ***	0.42075 ***	0.56963 ***	0.41510 ***	0.85180 ***
lnGDP	0.80365	2.15527 ***	2.43543 ***	1.04829 ***	2.47058 ***	0.66707 **	0.71860	0.03720
inv_lnGDP	118.86010 *	237.62550 ***	261.96240 ***	116.14190 ***	264.95620 ***	71.61200 **	94.72750 *	2.34010
selfemp	-0.00149	0.00396 *	0.00164	0.00376 ***	0.00159	0.00254	-0.00440 *	0.00020
depend	0.00942 **	0.00477 *	0.00182	0.00653 ***	0.00189	0.00726 ***	0.00340	0.00220 ***
unemp	0.00358 *	0.00098	0.00342 **	0.00156	0.00349 **	0.00226 **	0.00320 ***	0.00180 ***
school	0.00337 **	0.00243 *	0.00220	0.00180 *	0.00220	0.00108	0.00220 *	0.00050
open	0.00066 **	-0.00010	0.00032	-0.00001	0.00033	-0.00002	0.00030	0.00002
agr	-0.00936	-0.01393	-0.01469 **	-0.00446	-0.01446 **	-0.00137	-0.00750 *	0.00070
constant		-43.76436 ***		-20.67356 ***		-12.41235 *	-14.41120	-0.12340
no of instr.	17	27	33	63	33	61		
Sargan	13.877	23.6182	24.7135	69.1786	25.5332	57.3528		
p-value	0.085	0.130	0.422	0.067	0.377	0.251		
diff-Sargan		9.7412		44.4651		31.8196		
p-value		0.3718		0.0331		0.2388		
AR1 (p-value)	0.7662		0.0000		0.0000			
AR2 (p-value)	0.8776		0.6025		0.6135			

Note: ***,**,*: 1%, 5% and 10% statistical significance respectively.
Source: own calculations.

Conclusions

Empirical evidence on the relationship between income inequality and the level of economic development measured by GDP per capita has been mixed. Recent studies based on data from periods including the end of the twentieth century and the beginning of the present century seem to contradict the traditional theory of Kuznets which predicts the inverted U-shaped relationship. In case of many developed countries income inequality has not been declining and has not followed the trend predicted by the inverted U-curve.

Our results do not support Kuznets hypothesis. We confirm the U-shaped rather than inverted U-shaped relationship between Gini index and GDP per capita in EU-27. It also follows from our analysis that our data cover the descending part of the U, that is a shape of inverted J. Static model estimate results show that the above statistically significant relationships are also present in EU-15 countries. In case of country group EU-12 we provide no evidence of Kuznets-type relation, however we find a significant negative relationship between lnGini and lnGDP. We conclude that dynamic panel models are a better tool for describing Kuznets type relationship than static ones – in all specifications the lagged dependent variable proved statistically significant. We also find that the model assuming endogeneity of GDP per capita variable best describes the relationship between income inequality and income per capita. We conclude that SYSGMM estimator is superior over FDGMM in models based on assumptions of strict exogeneity and endogeneity of independent variables. Summing up, J-shaped specification estimated with System Generalized Method of Moments and assuming endogeneity of lnGDP and inv_lnGDP is indicated as superior over all other models we examine in the article. The statistically significant coefficients on lnGDP and inv_lnGDP variables and their signs show that the resulting Gini-GDP per capita relationship has a shape of inverted J.

The level of economic development is not the only factor influencing the dispersion of income. In both models: U- and J-shaped the same set of control variables proves statistically significant. Unemployment rate and old-age dependency ratio both have a positive impact on income inequality in the EU-27 countries. Other variables: share of self-employed, employed in agriculture in total employment, the share of population holding a university degree and trade openness do not have statistically significant impact on Gini index. Positive relation between unemployment rate and income inequality is not surprising taking into consideration that the equivalent disposable income before social transfers, we use in our analysis, is mostly employment related income. The higher share of persons receiving no employment related income in economically active population the greater income inequalities. Also, the significant positive impact of the percentage of older population in total population on income inequality could have been expected. High share of pensioners in a society means that proportionally large number of individuals receive relatively small pensions as compared to wages of the economically active majority.

Analysis of the differences in country groups, based on static panel models, leads to the conclusion that unemployment rate is the only control variable which significantly influences income inequality in both, EU-15 and EU-12. In "old" member states income inequality is also positively related to old-age dependency ratio and trade openness, while in EU-12 it is negatively influenced by the share of self-employed in total employment and positively related to the share of population holding university degree. The finding that the share of exports and imports in GDP which served as a proxy of the degree of internationalization of the economies positively influences income inequalities in EU-15 countries, not in EU-12 group, confirms our expectations that international specialization and foreign trade has more significant impact on incomes in high-wage countries. Greater imports from low labour cost countries can put downward pressure on incomes of unskilled employees in high-wage countries. Moreover, in developed countries exports can be seen as beneficial for well earning highly skilled employees rather than the low-skilled. It is more difficult to explain the negative relationship between the share of self-employed and income inequalities in the "new" member states. One could expect, however, that the higher incidence of forced self-employment in Eastern Europe, where this form of employment frequently substitutes a regular employment contract, but also is an alternative for being unemployed, leads to smaller income inequalities, as it improves incomes of individuals who are less likely to find and maintain employment. Positive impact of the share of individuals with tertiary education on income inequality in EU-12 can result from relatively small percentage of highly educated individuals in these countries or it can follow from particularly high returns to education. Further research on the determinants of income inequality should address these issues.

References

Anand, S., Kanbur, S.M.R. (1993), "The Kuznets Process and the Inequality–Development Relationship", *Journal of Development Economics*, Vol. 40, pp. 25–42.

Anyanwu, J.C. (2016), "Empirical Analysis of the Main Drivers of Income Inequality in Southern Africa", *Annals of Economics and Finance*, Vol. 17, No. 2, pp. 337–364.

Araujo, J.A., Cabral, J. (2014), "The Relationship Between Income Inequality and Economic Growth in Brazil: 1995-2012", *Problemas del desarrollo*, Vol. 46, No. 180, available at, https://probdes.iiec.unam.mx/en/revistas/v45n180/body/v45n180a6_1.php, referred on 10/04/2017.

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

Arellano, M., Bover, O. (1995), "Another look at the instrumental variable estimation of error-components models", *Journal of Econometrics*, Vol. 68, pp. 29-52.

Arnone, M., Presbitero F.P. (2010), *Debt Relief Initiatives: Policy Design and Outcomes*. Ashgate Publishing, Ltd.

Atkinson, A. B., Brandolini, A. (2001), "Promise and Pitfalls in the Use of .Secondary. Data-Sets: Income Inequality in OECD Countries as a Case Study", *Journal of Economic Literature*, 39(3), pp. 771-799.

Autor, D.H., Katz, L.F., Kearney, M. S. (2006), "The Polarization of the U.S. Labor Market", *American Economic Review*, 96 (2), pp. 189–94.

Baltagi, B.H. (2005), *Econometric Analysis of Panel Data*, 3rd Edition, Chichester: Wiley.

Barro, R.J. (2000), "Inequality and Growth in a Panel of Countries", *Journal of Economic Growth*, 5(1), pp. 5–32.

Blundell, R., Bond, S.R. (1998a), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, Vol. 87, No. 1, pp. 115-143.

Blundell, R., Bond, S.R. (1998b), "GMM Estimation with Persistent Panel Data: An Application to Production Functions", IFS, Working Paper Series, No. W99/4.

Blundell, R.W., Bond, S.R., Windmeijer, F. (2000). "Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator", IFS, Working Paper Series, No. W00/12.

Bond, S.R. (2002), "Dynamic panel data models: a guide to micro data methods and practice", *Portuguese Economic Journal*, Vol. 1, pp. 141-162.

Castells-Quintana, D., Ramos, R., Royuela V. (2015), "Income inequality in European Regions: Recent trends and determinants", *Review of Regional Research*, 35(2), pp. 123–146.

Checchi, D. (2001), “Education, Inequality and Income Inequality”. Distributional Analysis Research Programme, Discussion Paper, No. DARP 52.

Chong, A., Gradstein, M., Calderon, C. (2009), “Can Foreign Aid Reduce Income Inequality and Poverty?”, *Public Choice*, Vol. 140 (1/2), pp. 59-84.

Dańska-Borsiak B. (2011). *Dynamiczne modele panelowe w badaniach ekonomicznych,* Łódź: University of Łódź, [Dynamic panel models in economic research, *in Polish*].

Deutsch, J., Silber, J. (2004), “Measuring the Impact of Various Income Sources on the Link between Inequality and Development: Implications for the Kuznets Curve”, *Review of Development Economics*, Vol. 8, No. 1, pp. 110–127.

Forbes, K.J. (2000), “A Reassessment of the Relationship between Inequality and Growth”, *American Economic Review*, Vol. 90, pp. 869–887.

Franzini, M., Pianta, M. (2016), “The Engines of Inequality”, *Intereconomics,* 51(2), pp. 49-55.

Frazer, G. (2006). “Inequality and Development Across and Within Countries”, *World Development*, Vol. 34, No. 9, pp. 1459–1481.

Galbraith, J.K., Conceição, P. (2001), “Towards a New Kuznets Hypothesis: Theory and Evidence on Growth and Inequality”, in: Galbraith, J.K., Berner, M. (eds.): *Inequality and Industrial Change: A Global View*, New York: Cambridge University Press, pp. 139-160.

Galbraith, J.K., Kum, H. (2002), Inequality and Economic Growth: Data Comparisons and Econometric Tests, available at, http://utip.lbj.utexas.edu/papers/utip_21rv.pdf, referred on 10/04/2017.

Gallup, J.L. (2012), Is there a Kuznets curve?, available at, https://www.pdx.edu/econ/sites/www.pdx.edu.econ/files/kuznets_complete.pdf, referred on 28/03/2017.

Gregorio, J., Lee, J-W. (2002), “Education and Inequality: New Evidence from Cross-country Data”, *Review of Income and Wealth*, Vol. 48, No. 3, pp. 395-416.

Hsiao, C. (2003), *Analysis of Panel Data*, 2nd Edition, Cambridge: Cambridge University Press.

Kiatrungwilaikun, N., Suriya, K. (2015), “Rethinking Inequality and Growth: The Kuznets Curve after the Millennium”, *International Journal of Intelligent Technologies and Applied Statistics*, Vol. 8, No. 2, pp. 159-169.

Lachenmaier, S., Rottmann, H. (2007), “Effects of Innovation on Employment: A Dynamic Panel Analysis”, CESifo Working Paper, No. 2015.

Muszyńska, J., Oczki, J., Wędrowska, E. (2017), “Kuznets Hypothesis of Income Inequality: Empirical Evidence from EU”, University of Hradec Králové, Hradec Economic Days Conference Proceedings, pp. 643-651.

Naguib, C. (2015), “The Relationship Between Inequality and GDP Growth: An Empirical Approach”, LIS Working Paper, No. 631.

OECD (2011), Divided we stand. Why Inequality Keeps Rising, available at, https://www.oecd.org/els/soc/dividedwestandwhyinequalitykeepsrising.htm, referred on 15/03/2017.

Pagano, P. (2004), “An Empirical Investigation of the Relationship between Inequality and Growth”, Bank of Italy, Economic Research Department, Economic Working Papers No. 536.

Raitano, M. (2016), “Income Inequality in Europe Since the Crisis”, *Intereconomics*, 51(2), pp. 67-72.

Ram, R. (1997), “Level of Economic Development and Income Inequality: Evidence from the Postwar Developed World”, *Southern Economic Journal*, Vol. 64, No. 2, pp. 576-583.

Ravallion, M. (1995), “Growth and poverty: Evidence for developing countries in the 1980s”, *Economics Letters*, Vol. 48, pp. 411-417.
Roodman, A. (2007), “Note on the Theme of Too Many Instruments”, Center for Global Development, Working Paper, No. 125.
Rodríguez-Pose, A., Tselios, V. (2008), “Education and Income Inequality in the Regions of the European Union”, SERC Discussion Paper, Vol. 11, pp. 2-38.
Salverda, W., Nolan, B., Checchi, D., Marx, I., McKnight, A., Tóth, I. G., van de Werfhorst, H. (2014), *Changing Inequalities in Rich Countries: Analytical and Comparative Perspectives*. Oxford: Oxford University Press.
Savvides, A., Stengos, T. (2000), “Income inequality and economic development: evidence from the threshold regression model”, *Economics Letters*, Vol. 69, pp. 207-212.
Soto, M. (2009), “System GMM estimation with a small sample”, UFAE and IAE Working Papers, No. 395, available at, http://www.iae.csic.es/investigatorsMaterial/a111411124703archivoPdf29632.pdf, referred on 10/04/2017.
Tam, H. (2008), “An economic or political Kuznets curve?” *Public Choice*, Vol. 134, pp. 367–389.
Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data.* MIT Press.
Ziliak, J.P. (1997). “Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators”, *Journal of Business and Economic Statistics*, Vol. 15, No. 4, pp. 419-431.