

Income inequality and economic development: evidence from the threshold regression model

Andreas Savvides^a, Thanasis Stengos^{b,*}

^a*Department of Economics, Oklahoma State University, Stillwater, OK 74078, USA*

^b*Department of Economics, University of Guelph, Guelph, Ontario N1G 2W1, Canada*

Received 27 August 1999; accepted 25 April 2000

Abstract

Using panel data we investigate the relationship between income inequality and economic development. We employ a threshold regression model that allows for endogenous income thresholds and find strong evidence of a two-regime split in our sample. © 2000 Elsevier Science S.A. All rights reserved.

Keywords: Income inequality; Kuznets hypothesis; Threshold regression

JEL classification: O15

1. Introduction

According to the well-known Kuznets (1955) inverted-U hypothesis, income inequality increases during the early stages of economic development and after reaching a turning point declines. The relationship has been subject to extensive empirical testing [see Bruno et al. (1998) for a recent survey].

Most studies divide countries into two groups (developed and less developed) as a prerequisite to testing the Kuznets hypothesis. The implicit assumption is that the nature of the relationship differs according to a country's level of economic development. Therefore, the sample should be split into groups reflecting the level of economic development. It is important to note, however, that the criterion according to which countries are divided into two groups depends on a researcher's subjective notion of the category (LDC or otherwise) to which each country belongs or on a notional level of per capita income that divides the two groups.

The purpose of this paper is to reexamine the inequality–development relationship using recently

*Corresponding author. Tel.: +1-519-824-4120; fax: +1-519-763-8497.

E-mail address: thanasis@css.uoguelph.ca (T. Stengos).

developed econometric techniques that test endogenously the existence of a threshold level and, thus, allow for the possibility of endogenous sample separation. The threshold regression (*TR*) model allows the level of economic development (per capita income) to determine the existence and significance of a threshold level in the Kuznets relationship rather than imposing a priori an arbitrary classification scheme. If indeed there exists a well-defined relationship between income inequality and per capita income that depends on the level of economic development, the *TR* model can identify the threshold level and test simultaneously for such a relationship above and below the threshold.

The following section describes the *TR* model as it applies to the Kuznets hypothesis. Section 3 presents the empirical results.

2. The threshold regression (*TR*) model and the Kuznets hypothesis

In this paper we follow Hansen (1999) who suggests a bootstrap procedure to test the null hypothesis of a linear formulation against a *TR* alternative. The *TR* model assumes the data is given by $\{y_t, x_t, q_t\}_{t=1}^n$, where y_t and q_t are observations on the dependent variable and a threshold variable, respectively, and x_t is a $p \times 1$ vector of independent variables. The threshold variable q_t splits the sample into different groups and may be part of x_t . The *TR* model is given by

$$y_t = x_t^T \beta_1 + \varepsilon_t, \quad q_t \leq \gamma \quad (1)$$

$$y_t = x_t^T \beta_2 + \varepsilon_t, \quad q_t > \gamma \quad (2)$$

The model can be written in a single equation form by defining a dummy variable $d_t(\gamma) = I(q_t \leq \gamma)$, where $I(\cdot)$ denotes the indicator function. If we write the variable $x_t(\gamma) = x_t d_t(\gamma)$, Eqs. (1) and (2) can be expressed as

$$y_t = x_t^T \beta + x_t^T(\gamma) \theta + \varepsilon_t \quad (3)$$

where $\beta = \beta_2$ and $\theta = \beta_1 - \beta_2$. Eq. (3) allows all the regression coefficients to differ between sample groups. Hansen (1999) provides an algorithm that searches over values of γ using conditional *OLS* regressions based on a sequential search over all $\gamma = q_i$, for $i = 1, \dots, n$. The procedure also provides estimates of β and θ .

An important hypothesis for our test of the Kuznets hypothesis is whether the *TR* model is statistically significant relative to a simple linear specification. The null hypothesis in this case describes the simple linear specification and can be expressed as:

$$H_0: \beta_1 = \beta_2 \quad (4)$$

Devising a test for the null hypothesis in (4) runs into a serious obstacle, namely the threshold parameter γ is not identified under H_0 . We follow Hansen (1999) who suggests a heteroskedasticity-consistent Lagrange Multiplier (*LM*) bootstrap procedure to test the null hypothesis of a linear formulation against a *TR* alternative. Since γ is not identified under the null hypothesis of the no-threshold effect, the p values are computed by a fixed bootstrap method. In that case the x_t values are used as regressors and the bootstrap-dependent variable is generated from $N(0, \hat{\varepsilon}_t^2)$, where $\hat{\varepsilon}_t$ is the *OLS* residual from the estimated threshold model. Hansen (1999) shows that this procedure yields

asymptotically correct p values. It is important to note that if the hypothesis in (4) is rejected and a threshold level is identified, we should test again the TR model against a linear specification after dividing the original sample according to the threshold thus identified. This procedure is carried out until the null in (4) can no longer be rejected.

In testing the inverted-U hypothesis we consider two alternative specifications of the TR model in (3) as follows:

$$INEQ_{it} = a_0 + a_1 INC_{it} + a_2 (1/INC_{it}) + \varepsilon_{it} \quad (5)$$

$$INEQ_{it} = b_0 + b_1 (\ln INC_{it}) + b_2 (\ln INC_{it})^2 + \varepsilon_{it} \quad (6)$$

where $INEQ_{it}$ is the income inequality of country i at time period t , and INC_{it} is the per capita income. The specification in (5) is recommended by Anand and Kanbur (1993) as appropriate when testing the Kuznets hypothesis with the Gini coefficient. We consider specification (6) because it is one of the most popular specifications for testing the Kuznets hypothesis. In both cases, the threshold variable, q_t , is the per capita income.

We use the high-quality panel estimates of the Gini coefficient of Deininger and Squire (1998) as our measure of income inequality. This data set is designed to ensure intertemporal and international comparability of income inequality. As in all previous tests of the Kuznets hypothesis the level of economic development is measured by per capita income. We use the Summers–Heston estimates of per capita income in international dollars at 1985 prices to allow for differences in purchasing power across countries and thus avoid arbitrary conversions via official exchange rates. Our sample consists of the 618 observations in the Deininger–Squire data set for which Summers–Heston provide estimates of per capita income. Given that some of the 95 countries in this sample are represented by one or a very small number of observations, we follow previous researchers [for example, Bruno et al. (1998), Deininger and Squire (1998)] and consider an alternative sample that contains only countries for which at least four observations are available. This alternative sample contains 547 observations from 52 countries. In what follows we test the sensitivity of our results to the alternative sample and note any differences in results between the two samples.

3. Empirical results

First we test whether the TR model is statistically significant relative to a linear specification, i.e. we test the null hypothesis in (4). The values of the Lagrange Multiplier (LM) test are 43.3 and 39.4, for specifications (5) and (6), respectively. Based on 1000 bootstrap replications, the null hypothesis in (4) can be decisively rejected regardless of specification (the p value for this test is less than 0.01 in both cases). Subsequently, we subdivide the sample and test for additional threshold levels. For the subsamples of observations that exceed the threshold level, the LM tests for specifications (5) and (6) yield p values of 0.35 and 0.43, respectively. Thus, we conclude that no additional threshold levels are present. In summary, our results provide strong evidence that the inequality–development relationship is described by a two-regime split of the sample based on per-capita income.

Given the presence of only one threshold level in the inequality–development relationship, it is necessary to estimate it and split the sample accordingly. The estimate of the threshold is \$PPP 2199

Table 1
OLS and threshold regression estimates of Eq. (5)^a

Explanatory variables	OLS Estimates		TR model estimates: > threshold		TR model estimates: < threshold	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	42.20 (32.98)		22.91 (7.59)		21.86 (2.64)	
<i>INC</i>	−0.740 (7.080)	0.016 (0.218)	0.463 (2.416)	0.114 (1.226)	8.931 (2.356)	−1.669 (0.731)
(1/ <i>INC</i>)	−2.716 (2.194)	0.751 (0.781)	54.44 (6.515)	8.902 (2.055)	6.097 (1.516)	−0.158 (0.056)
\bar{R}^2	0.085	0.918	0.167	0.932	0.050	0.887
LM-heteroskedasticity	66.98	21.01	23.73	15.33	4.16	0.10
[significance level]	[0.00]	[0.00]	[0.00]	[0.00]	[0.04]	[0.76]
No. observations	618	618	449	449	169	169

^a Columns (1), (3) and (5) report estimates without country-specific effects and columns (2), (4) and (6) with country-specific effects. Numbers in parentheses below coefficient estimate are *t*-statistics. LM-heteroskedasticity in the Breusch–Pagan Lagrange Multiplier statistic for heteroskedasticity (*p* values are shown in brackets). Where the hypothesis of homoskedasticity is rejected, *t*-statistics are based on heteroskedasticity-consistent standard errors.

for specification (5) and \$PPP 2102 for (6). Tables 1 and 2 contain the empirical results for models (5) and (6), respectively. For comparison purposes, the first two columns in each table display estimates from a linear specification that ignores threshold effects: column (1) contains the OLS results (with a common intercept across countries) and column (2) the results from a panel model that includes country-specific intercepts. The results of column (1) indicate an inverted-U relationship between per capita income and inequality. As several authors have pointed out, however, once country-specific effects are allowed for in column (2), there is no evident relation between the two variables.

Table 2
OLS and threshold regression estimates of Eq. (6)^a

Explanatory variables	OLS Estimates		TR model estimates: > threshold		TR model estimates: < threshold	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	−80.78 (2.848)		588.5 (6.20)		122.1 (0.594)	
ln <i>INC</i>	31.86 (4.512)	−2.020 (0.435)	−120.2 (5.524)	−21.77 (2.0.63)	−26.24 (0.443)	9.321 (0.249)
(ln <i>INC</i>) ²	−2.105 (4.875)	0.105 (0.378)	6.504 (5.238)	1.193 (1.994)	2.024 (0.475)	−0.784 (0.294)
\bar{R}^2	0.095	0.918	0.201	0.932	0.009	0.870
LM-heteroskedasticity	68.89	20.88	19.84	20.41	1.18	0.76
[significance level]	[0.00]	[0.00]	[0.00]	[0.00]	[0.28]	[0.38]
No. observations	618	618	458	458	160	160

^a See note to Table 1.

The remaining columns of Tables 1 and 2 provide estimates of the *TR* model: columns (3) and (4) correspond to observations above the threshold (without and with country-specific effects) and columns (5) and (6) below the threshold. It is clear that no relation between per capita income and inequality exists for observations below the threshold. When it comes to observations above the threshold [columns (3) and (4)], an uninverted- rather than inverted-U pattern is evident. Our conclusion, however, depends on the functional form chosen for the underlying model. It holds in the case of the popular specification in (6) but does not extend to specification (5) recommended by Anand and Kanbur (1993): in column (4) of Table 1, the estimated coefficient for INC_{it} is positive but insignificant.¹ Finally, the estimate of the turning point (the minimum income level) for the functional form in (6) is \$PPP 9140 and is highly significant (t -statistic = 3.74).

Next, we adopt the strategy of all previous studies and classify countries into two groups. In contrast to previous work, however, we use the threshold level from the *TR* model as our classification criterion. We classify a country into the higher income category if all (or in a very small number of cases all but one) observations for a specific country exceed the previously identified threshold levels (2199 and 2102). The results are in Table 3. Columns (1) and (2) report estimates (without and with country-specific effects) for specification (5) and columns (3) and (4) for specification (6). The uninverted-U pattern reported previously is no longer evident. There is no evidence of a well-identified relation (inverted- or uninverted-U) between per capita income and income inequality. This

Table 3
Estimates of the Kuznets curve for countries classified as higher income^a

Explanatory variables	(1)	(2)	(3)	(4)
Constant	20.93 (5.975)		101.12 (1.060)	
<i>INC</i>	0.627 (3.009)	0.125 (1.110)		
(1/ <i>INC</i>)	54.13 (4.624)	10.57 (1.365)		
$\ln INC$			−5.657 (0.257)	−1.729 (0.171)
$(\ln INC)^2$			−0.168 (0.134)	0.085 (0.147)
\bar{R}^2	0.094	0.894	0.330	0.920
LM-heteroskedasticity	11.08	17.57	38.28	24.21
[significance level]	[0.00]	[0.00]	[0.00]	[0.00]
No. countries/observations	46/358	46/358	52/402	52/402

^a Columns (1) and (3) report estimates without country-specific effects, and columns (2) and (4) with country-specific effects. Countries are classified as higher income if all observations on per capita income exceed the threshold level identified by the *TR* model. See the text for additional details.

¹This conclusion holds for the full sample of 618 observations that includes all countries regardless of the number of observations per country. We also estimated the *TR* model for the limited sample with at least four observations per country. The results are similar to those reported for the full sample: the estimates of the threshold level are \$PPP 5449 and \$PPP 2849 for the two specifications. In this case, however, a significant uninverted-U pattern is observed for observations above the threshold level for both specifications (5) and (6).

finding points to the importance of employing an appropriately specified model such as the *TR*, rather than relying on ad hoc classification criteria for dividing countries into relevant groups.

As an additional check on our results, we examine the possibility of a linear relation between per capita income and inequality for observations above and below the threshold level. This is the methodology of Eusufzai (1997) who used the Anand and Kanbur (1993) cross-section data set and the Quandt log-likelihood ratio test to identify a break in the data. The study reported a positive correlation between income per capita and inequality for countries prior to the break and a negative for countries subsequent to the break. The author offered this finding as evidence in favor of the inverted-U hypothesis. We test for a linear negative (positive) relation between per capita income and inequality for observations above (below) the threshold. No relationship conforming to the inverted-U pattern is found. If anything, a (weakly significant) negative relationship can be observed for the low-income observations.

Finally, we examine the robustness of our results to alternative inequality measures. The Deininger–Squire data set includes estimates of the quintile shares of income for a smaller sample of observations (558). We use these to compute four indices of inequality: Theil's entropy index, the income share of the lowest 40% of the population, the income share of the top 20%, and the ratio of the income share of the top 20% to the lowest 40%. These indicators were chosen because of their widespread usage. We find evidence (results available on request) of a significant threshold effect for all four indices (the *p* value is less than 0.01 in each case). The estimate of the threshold is \$PPP 2102 in each case. No additional thresholds are present (the *p* value is 0.41, 0.45, 0.38 and 0.48, respectively). Estimates from the *TR* model show that for observations below the threshold, there is no evidence of any relationship between income inequality and per capita income, regardless of inequality index. For observations above the threshold the evidence is mixed: an uninverted-U relationship is evident for Theil's entropy index and the income share of the top 20%, but there is no evidence of any relationship between income inequality and per capita income for the other two measures of inequality.

Acknowledgements

T. Stengos would like to thank the SSHRC of Canada for financial support.

References

- Anand, S., Kanbur, S.M.R., 1993. The Kuznets process and the inequality–development relationship. *Journal of Development Economics* 40, 25–52.
- Bruno, M., Ravallion, M., Squire, L., 1998. Equity and growth in developing countries: old and new perspectives on the policy issues. In: Tanzi, V., Chu, K.-Y. (Eds.), *Income Distribution and High-Quality Growth*. MIT Press, Cambridge, MA.
- Deininger, K., Squire, L., 1998. New ways of looking at old issues: inequality and growth. *Journal of Development Economics* 57, 259–287.
- Eusufzai, Z., 1997. The Kuznets hypothesis: an indirect test. *Economics Letters* 54, 81–85.
- Hansen, B.E., 1999. Sample splitting and threshold estimation, *Econometrica*, in press.
- Kuznets, S., 1955. Economic growth and income inequality. *American Economic Review* 45, 1–28.