A Nonparametric and Semiparametric Analysis on the Inequality-Development Relationship

Kui-Wai Li [a, *] and Xianbo Zhou [b]
[a] City University of Hong Kong and University of Geneva
[b] Lingnan College, Sun Yat-sen University, China

Abstract: This paper studies the income inequality and economic development relationship by using unbalanced panel data on OECD and non-OECD countries for the period 1962 - 2003. Nonparametric estimation results show that income inequality in OECD countries are almost on the backside of the inverted-U, while non-OECD countries are approximately on the foreside, except that the relationship in both country groups shows an upturn at a high level of development. Development has an indirect effect on inequality through control variables, but the modes are different in the two groups. Model specification tests show that the relationship is not necessarily captured by the conventional quadratic function. Cubic and fourth-degree polynomials, respectively, fit the OECD and non-OECD country groups better. Our finding is robust regardless whether the specification uses control variables. Development plays a dominant role in mitigating inequality.

Keywords: Kuznets inverted-U; Nonparametric and semiparametric models; Unbalanced panel data

JEL classification: C14; C33; O11.

---

* Corresponding author: City University of Hong Kong, Tel.: 852 3442 8805, E-mail: efkwli@cityu.edu.hk, and University of Geneva, Tel.: +41 (0) 22 379 9596.

I Introduction

There are different forms of inequality in human history, including aristocratic, racial, sexual, religious, political, social and territorial inequalities. Some inequalities are irrevocable. While the Gini coefficient shows an inter-personal comparison and provides a static snapshot measure of income inequality, improvement in income inequality can often be made intra-personally, as a person's income improves through experience, skill, job diversity and personal endowment (Li, 2002). Indeed, given that modern societies train and educate people for employment with different rewards, income inequality is inevitable (Letwin, 1983).

The relationship between income inequality and economic development has been characterized by the Kuznets inverted-U curve (Kuznets, 1955) which argued that income inequality tends to increase at an initial stage of development and then decrease as the economy develops, implying that income inequality will eventually fall as income continues to rise in developing countries. Studies conducted along the line of the inverted-U relationship include Sen (1991, 1992 and 1993) who discussed inequality through individual capability and functioning. Some studies concentrate on the causes of income inequality which include human capital, technological advancement, job diversity and political stability, while other studies examine the long run income inequality convergence (Galor and Zeire, 1993; Galor and Moav, 2000; Gould *et al*. 2001; Acemoglu, 2001; Desai *et al*. 2005; Bėnabou, 1996; Ravallion, 2003).

The Kuznets inverted-U relationship between inequality and economic development has attracted both supporters and critics. In particular, whether the relationship is considered as a law or can be improved through appropriate economic policies (Kanbur, 2000). Nevertheless, the Kuznets inverted-U relationship has not been fully confirmed and validated in studies with parametric quadratic models (Li *et al*

This paper presents a nonparametric (without control variables) and semiparametric (with control variables) investigation on the inequality-development relationship by using unbalanced panel data from the devel

We consider two kinds of controls. The policy control is indicated by the variables of openness (indicated by the percentage trade share of GDP in 2005 constant prices), urbanization (indicated by the percentage of urban population in total population), and investment (indicated by the percentage of investment share in real GDP per capita), denoted as openk, urbanize and ki, respectively. The other control variables of GDP growth and inflation (indicated by the annual percentage of GDP deflator) reflect the economic characteristics of the sample country. These data are obtained from the Penn World Table and World Development Indicators. Table 1 reports the basic statistics of these variables for both OECD and non-OECD countries. One observation is that non-OECD countries on average have a larger inequality and variation than OECD countries, while OECD countries have higher level of development with more variations than non-OECD countries.

Table 1 Basic Statistics

	Gini	gdppc	growth	openk	urbanize	ki	deflator
OECD							
minimum	17.80	2,028.78	-19.29	5.29	2.90	10.41	-2.00
maximum	58.00	63,419.40	12.49	264.14	67.60	48.83	208.00
mean	34.34	18,658.44	2.53	45.59	28.84	26.85	9.35

between inequality and development for linear parametric regression models, but is invalid in the nonlinear and nonparametric relationships. Inequality and development are also moderately correlated with other variables. For example, the correlations of log GDP per capita with openness and urbanization are larger when compared to the other variables. Both inequality and level of development are moderately correlated with openness in OECD countries and with investment in non-OECD countries. These reciprocal relations imply that, in addition to the direct effect on inequality, the level of development may have an indirect effect on inequality through other channels.[2]

Table 2 Correlations between Variables

	Gini	log(gdppc)	growth(-1)	openk	urbanize	ki	inflation
OECD							
Gini	1.0000	-0.3506	0.1994	-0.3245	-0.0789	-0.1309	0.0394
log(gdppc)	-0.3506	1.0000	-0.0465	0.3689	-0.4740	0.1870	-0.3698
growth(-1)	0.1994	-0.0465	1.0000	-0.0565	0.1285	0.3530	-0.2705
openk	-0.3245	0.3689	-0.0565	1.0000	-0.1844	0.0376	-0.0196
urbanize	-0.0789	-0.4740	0.1285	-0.1844	1.0000	0.3676	0.1828
ki	-0.1309	0.1870	0.3530	0.0376	0.3676	1.0000	-0.2246
inflation	0.0394	-0.3698	-0.2705	-0.0196	0.1828	-0.2246	1.0000
Non-OECD							
Gini	1.0000	0.1266	0.0255	-0.0144	-0.0835	-0.2156	0.1248
Log(gdppc)	0.1266	1.0000	-0.0807	0.4232	-0.7964	0.2183	0.0532
growth(-1)	0.0255	-0.0807	1.0000	0.0810	0.1103	0.2938	-0.1586
openk	-0.0144	0.4232	0.0810	1.0000	-0.3339	0.4074	-0.0652
urbanize	-0.0835	-0.7964	0.1103	-0.3339	1.0000	-0.1201	-0.1577
ki	-0.2156	0.2183	0.2938	0.4074	-0.1201	1.0000	-0.1393
inflation	0.1248	0.0532	-0.1586	-0.0652	-0.1577	-0.1393	1.0000

To study the relationship between inequality and economic development, we first specify the following nonparametric (unconditional) panel data model with fixed effects

[2] The indirect effects via channels can also be found in growth studies (Barro, 2000; Frankel and Rose, 2002).

without control variables:

$$gini_{it} = g(lgdp_{it}) + u_i + v_{it}, \; t = 1,2,\ldots,m_i; \; i = 1,2,\ldots,n, \qquad (1)$$

where the functional form of $g(\,)$ is not specified and $lgdp_{it}$

inequality model (Huang *et al*. 2009). In Model (2) the indirect effect of development on inequality is controlled by the term $x_{it}^{'}$, hence $g(\,)$ reflects the inequality from development directly.

The mechanism in Models (1) and (2) and their relationships are intuitively illustrated in Figure 1. The g(z) in nonparametric Model (1) gives the gross contribution of development to inequality, while the g(z) in semiparametric Model (2) gives the net contribution of development to inequality, given x . The difference between the two g(z) is the indirect contribution of development to inequality via control variables x .

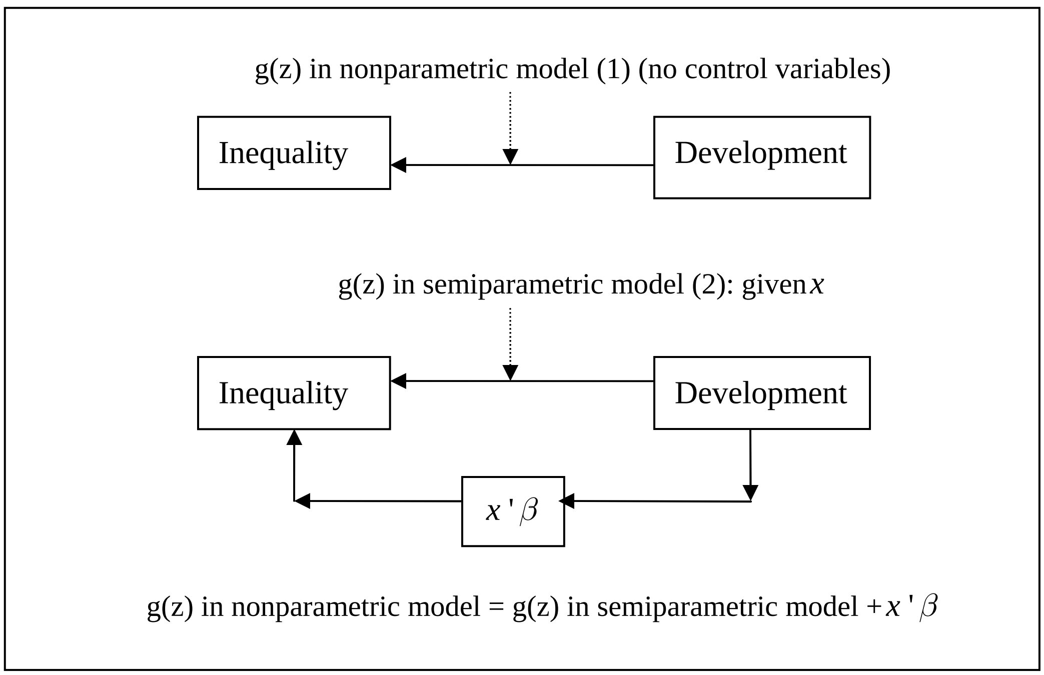


Fig. 1 The Mechanism in the Nonparametric and Semiparametric Models

When $g(\,)$ is specified as a parametric quadratic, cubic or fourth-degree polynomial function of $lgdp_{it}$, Model (1) and Model (2) become parametric unbalanced panel data models with fixed effects, which can be estimated by the conventional method

(Baltagi, 2008). However, in order to keep the approach comparable to the nonparametric counterpart, we use the difference of $y_{it} - y_{i1}$ instead of the transformation of $y_{it} - \bar{y}_{i\cdot}$ or the difference of $y_{it} - y_{i,t-1}$ in removing the fixed effects.

Table 3 contains the parametric estimation results for the two samples of OECD and non-OECD countries. The conventional quadratic specification is used to test the Kuznets hypothesis, and the coefficients on the linear and quadratic terms are expected to be positive and negative, respectively. The estimates for the non-OECD countries have the expected signs and are highly significant, while those for the OECD countries do not have the expected signs, regardless whether control variables are added into the model.

We estimated models with higher-degree polynomials of the logarithm of GDP per capita, as shown by the "cubic" and "4-th degree" columns in Table 3. For OECD, the

Table 3 Parametric Estimation Results

III Nonparametric Estimation and Testing Method with Unbalanced Panel Data

We use the same notation as those in Henderson *et al*. (2008) to illustrate our model estimation in the unbalanced panel data case. For simplicity, we denote $y \quad gini$ and $z \quad lgdp$. Models (1) and (2) can be estimated by the iterative procedures modified where the argument ${}_{i,tg}$ is $\hat{g}_{[l\ \ 1]}(z_{is})$ for $s\ \ \ t$ and $G_{it}(\ {}_{0},\ \ {}_{1})'$ when $s\ \ \ t$, and

$\hat{g}_{[l\ \ 1]}(z_{is})$ is the $(l\ \ \ 1)^{\text{th}}$ iterative estimates of $(\ {}_{0},\ \ {}_{1})'$. Here

$$\sum_{i=1}^{n}\frac{1}{m_i}\sum_{t=2}^{m_i}\left(\hat{g}_{[l]}(z_{it})-\hat{g}_{[l-1]}(z_{it})\right)^2 \Big/ \sum_{i=1}^{n}\frac{1}{m_i}\sum_{t=2}^{m_i}\hat{g}^2_{[l-1]}(z_{it}) < 0.01.$$

Further, the variance σ_v^2 is estimated by

$$\hat{\sigma}_v^2 = \frac{1}{2n}\sum_{i=1}^{n}\frac{1}{m_i-1}\sum_{t=2}^{m_i}(y_{it}-y_{i1}-(\hat{g}(z_{it})-\hat{g}(z_{i1})))^2 .$$

The variance of the iterative estimator $\hat{g}(z)$ is calculated as $(nh\hat{f}(z))^{-1}\int k^2(v)dv$, where $\int k^2(v)dv$, and $\hat{f}(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{m_i-1}{m_i}\sum_{t=2}^{m_i}K_h(z_{it}-z)/\hat{\sigma}_v^2$.

For the estimation of semiparametric Model (2), we denote the nonparametric estimator of the regression functions of the dependent variable y

$I_n^{(1)} \quad \frac{1}{n}\sum_{i\ 1}^{n} \frac{1}{m_i}\sum_{t\ 1}^{m_i} (g_0(z_{it}, \hat{\ }) \quad \hat{g}(z_{it}))^2$, where $\hat{\ }$ is a consistent estimator of the parametric panel data model with fixed effects; $\hat{g}(\ )$ is the iterative consistent estimator of Model (1).

The second specification test is to choose in Model (2) between parametric and semiparametric models with control variables. The null hypothesis H_0 is parametric model with $g(z) \quad g_0(z, \ )$. The alternative is that $g(z)$ is nonparametric in Model (2). The test statistic for testing this null is $I_n^{(2)} \quad \frac{1}{n}\sum_{i\ 1}^{n} \frac{1}{m_i}\sum_{t\ 1}^{m_i} (g_0(z_{it}, \ ) \quad x_{it}' \quad \quad \hat{g}(z_{it}) \quad x_{it}' \hat{\ })^2$, where and are consistent estimators in the parametric panel data model with fixed effects; $\hat{g}(\ )$ and $\hat{\ }$ are the iterative consistent estimator of model (2).

In the following empirical study, we apply bootstrap procedures in Henderson *et al.* (2008) to approximate the finite sample null distribution of test statistics and obtain the bootstrap probability values for the test statistics.

IV Empirical Results

The kernel in both the estimation and the testing is the Gaussian function and the bandwidth is chosen according to the rule of thumb[3]: $h = 1.06\sigma_z \left(\sum_{i=1}^{n} m_i\right)^{-1/5}$, where σ_z is the sample standard deviation of $\{z_{it}\}$. All the bootstrap replications are set to be 400. The last column in Table 3 reports the coefficient estimation for the control variables in the parametric part of semiparametric Model (2). For the OECD countries, with the exception of "openk" and "urbanize", the coefficient estimates of all other control

[3] We also slightly change the constant instead of 1.06, and find that the estimation and test results are not significantly affected.

variables have the same signs and similar values in both parametric and semiparametric

Table 4 Nonparametric Estimation of $g(\,)$ at Different Points of lgdp

Quantile of lgdp	Nonparametric model (1)	Semiparametric model (2)

curves of $g(\,)$ in Figures 2 and 3 look very similar in shape, which implies that the increasing, albeit short, portion at lower levels of development, a turning point (where the effect of development on inequality changes from positive to negative) at 8.2 (about $3,640 in 2005 dollars), followed by a longer decreasing portion of the curve. While the

Table 5 Nonparametric and Semiparametric Model Specification Tests: OECD

Model		Hypotheses	I_n (p-value)	Model selected
without control variables	Parametric or Nonparametric ?	H_0: Quadratic parametric H_1: Nonparametric	4.006 (0.070)	Nonparametric (10%) Parametric (5%)
		H_0: Cubic parametric H_1: Nonparametric	3.964 (0.060)	Nonparametric (10%) Parametric (5%)
with control variables	Parametric or Semiparametric ?	H_0: Quadratic parametric H_1: Semiparametric	10.205 (0.013)	Semiparametric
		H_0: Cubic parametric H_1: Semiparametric	11.823 (0.008)	Semiparametric

Since the cubic parametric model without control variables is accepted at the conventional 5 percent significant level (note that the quadratic curve is not 53 c998 refBT12.043er28..24

Table 6 Parametric Model Tests for Inclusion of Polynomial Terms

	Without control variables	With control variables
Degree of polynomial	F statistic: OECD / Non-OECD	F statistic: OECD / Non-OECD
H_0:Second vs H_1:Third	11.2043* / 5.3481*	0.3217 / 1.0814
H_0:Third vs H_1:Fourth	1.5146 / 9.2195*	2.8767 / 5.5398*
H_0:Fourth vs H_1:Fifth	0.2849 / 1.0849	2.9167 / 0.0058

Note: * = 5% significance.

In the case with control variables, the tests in Table 6 present no obvious evidences to show which parametric specification is best. One can conclude from Table 3 (OECD) that the cubic form is preferred since all the estimated coefficients of the cubic polynomial statistically prevail over those of the quadratic and fourth-degree counterparts. However, as the test with control variables in Table 5 shows, parametric cubic specification is rejected and semiparametric model is accepted at the 5 percent significant level. Hence, the insignificance of the F tests in the case of control variables for OECD countries is expected since the F tests based on the estimation of parametric model may not be valid for semiparametric models.

Non-OECD Countries

Figures 5 and 6 respectively present the nonparametric estimation of $g(\,)$ in Models (1) and (2) for the non-OECD sample countries. The estimates provide a result stronger than those in Figures 1 and 2 since the boundary effect for nonparametric estimation is less significant. There is much resemblance between the shapes of nonlinearity of $g(\,)$ in Figures 5 and 6. The results reflect a rapid increasing, albeit short, portion at lower levels of development and a first turning point at 7 (about $1,100), then another increasing, albeit long and flat, portion at the middle income level of development, then followed by a slowly decreasing portion of the curve with the second turning point at 8.7 (about $6,000). Finally, the curves also hint at an upturn at a higher income level (around 10, about $22,026), similar to the processes shown in OECD countries and the findings in Ram (1991) and Mushinski (2001). The second turning point is higher than the first one and the final upturn occurs at even higher inequality level. This process presents a "roller coaster" mode, albeit flat and long in the middle of the process.

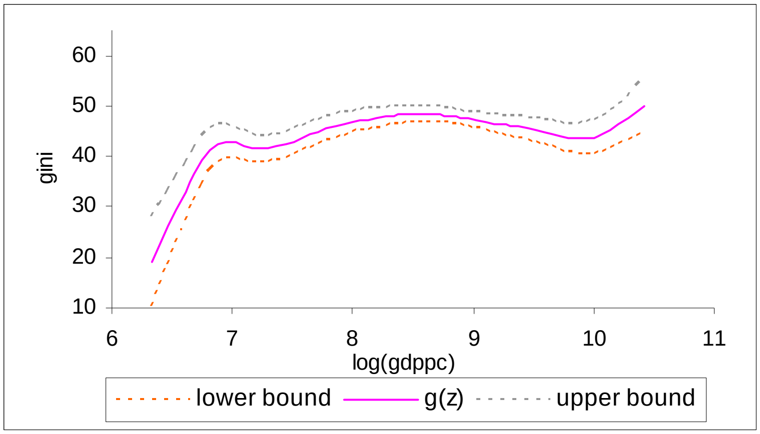
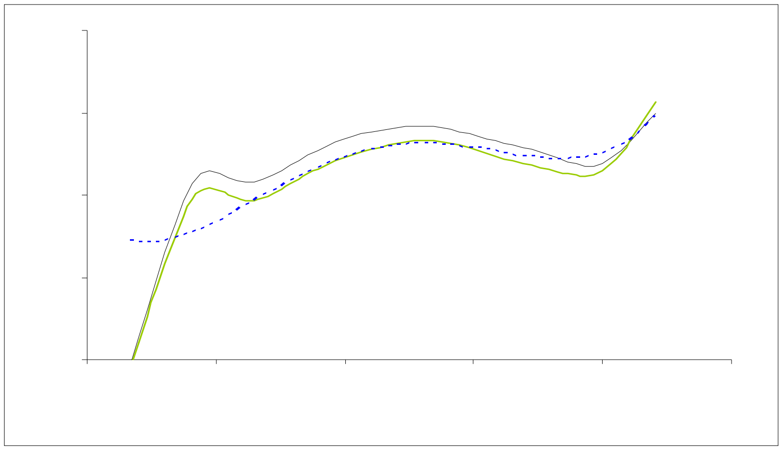


Fig. 6 Semiparametric Estimation in Model 2: Non-OECD

Figure 7 contains both the nonparametric and semiparametic curves estimated for the non-OECD countries. The vertical difference reflects the contribution of control variables to inequality. The integrated effect of development on inequality via control variables is always negative, which is different from that in the OECD countries. In short, control variables generally mitigate inequality in non-OECD countries, except when the logarithm of income level is very large (greater than 10.2). This evidence shows that the channel effect of development on inequality via the control variables as a whole is negative in non-OECD countries.

Table 7 presents the test results for choosing between nonparametric or semiparametric and parametric specifications in the non-OECD countries. The quadratic and cubic parametric specifications cannot be rejected at any usual significant level whether or not the model includes control variables as regressors. Among the parametric models, the F tests for parametric specifications in Table 6 show, when compared to the third- and fifth-degree polynomial specifications, that the fourth-degree polynomial of the logarithm of GDP per capita gives a sufficiently accurate description of the non-OECD countries, whether or not the control variables are added into the models. Recall in Table 3 (non-OECD) that all coefficient estimates in the fourth-degree polynomials are



The estimation and test results from nonparametric model without control variables imply that the conventional quadratic concave function may not necessarily capture the relationship between inequality and development in both OECD and non-OECD countries. The tests show that the cubic polynomial of development levels can capture the relationship more accurately than quadratic and fourth-degree polynomials in OECD countries, while the fourth-degree polynomial can give a better description of the data than other polynomials in non-OECD countries, whether or not the control variables are added as regressors in the models. In both OECD and non-OECD countries, analysis based on a quadratic specification for the relationship between inequality and development is misleading.

The estimation and test results from the semiparametric model with control variables show that the data-driven model selection for OECD countries requires a semiparametric specification while the non-OECD countries require a fourth-degree polynomial parametric specification. Given the integrated contribution by control variables to inequality shown above, we next study the effects of the control variables on inequality by comparing the estimates in the parametric part of the semiparametrc model for the OECD countries (shown in the last column in Table 3) and the 4-th degree parametric model for the non-OECD countries (shown in the "4-th degree" column in Table 3). The implications of the effects of the control variables are different in the two sample country groups, except the variable "growth (-1)" which has a positive effect on inequality.

Specifically, the effect of openness on inequality in OECD countries is negative, albeit insignificant both economically and statistically, while openness has a positive and significant effect on inequality in non-OECD countries. Openness generally will aggravate income inequality in non-OECD countries, but has an emollient effect on inequality in OECD countries. Although integration to the world market is expected to help non-OECD countries to promote prosperity, increasing opportunities to trade are also likely to affect income distribution. Whether or not increasing openness to trade is accompanied by a reduction or an increase inequality has strongly been debated (Julien, 2007; Wood, 1997).

The effect of urbanization on inequality is negative (-0.171) and significant for OECD countries, but positive (0.007) and insignificant for non-OECD countries. Urbanization helps to mitigate inequality in OECD countries, but increases inequality in non-OECD countries, albeit insignificantly. According to Anand (1993), the urban-rural difference generally results in larger inequality in total income distribution due to urbanization. Hence, in the process of urbanization, income inequality will first increase and then decrease with urbanization or the migration of rural population to cities. In our case, OECD countries have much higher urbanization than non-OECD countries. Hence the negative effect of urbanization on inequality in OECD and the positive effect in non-OECD accord with this general urbanization-inequality relationship.

The finding that investment share aggravates inequality in OECD countries but reduces inequality in non-OECD countries contrasts with the result in Barro (1999) that showed little overall relationship between income inequality and investment. One explanation is that investment may have potential endogeneity in inequality models.

Inflation has a negative albeit insignificant effect on inequality in OECD countries but has a positive and significant effect on inequality in non-OECD countries. Generally, cross-country evidence on inflation and income inequality suggests that they are positively related. For example, Albanesi (20019.451395.n7e

g3f064 Twanesi (

determination of government policies, and inflation is positively related to the degree of inequality as low income households are more vulnerable to inflation. Since non-OECD countries have a high average inflation, their monetary authorities should reduce inflation to alleviate income inequality. However, the impact of inflation on income distribution may be nonlinear (Bulíř, 2001), and the positive and significant effect of inflation on inequality would need to be explained with caution.

V Conclusion

This paper provides evidences on the relationship between inequality and development from estimations and tests of nonparametric and semiparametric panel data models with fixed effects. Based on an unbalanced panel dataset, this study contributes to the literature by presenting new evidences about the inequality-development relationship in OECD and non-OECD countries and provides additional information on the mechanics of the effect of development on inequality.

For the OECD countries, inequality generally decreases with development, with the exception of an upturn at a higher income level. The control variables will help reduce income inequality at lower income levels (below about $9,900), but they tend to increase inequality when development exceeds that level. For the non-OECD countries, the inequality-development relationship appears in a “roller coaster” mode with two turning points, and one upturn appears at a very high income level. When compared to the performance in OECD countries, the effect of development on inequality via the control variables is always negative in non-OECD countries, except after an upturn at a high income level. Non-OECD countries seem to face serious inequality at the middle or high income level.

Iradian, Garbis, 2005. Inequality, poverty, and growth: cross-country evidence. Working Paper No. 05/28. International Monetary Fund, Washington.

Julien, Gourdon, 2007. Openness and inequality in developing countries: a new look at the evidence. MPRA Paper No. 4176, November.

Kanbur, Ravi, 2000. Income distribution and development. In: Atkinson, Anthony B., Bourguignon, Francious (Eds.), Handbook of Income Distribution. Elsevier, 791-841.

Keane, M.P. and Prasad, E., 2002. Inequality, transfer, and growth: new evidence from the economic transition in Poland. Review of Economics and Statistics 84(2): 324-341.

Kuznets, Simon, 1955. Economic growth and income inequality. American Economic Review 45: 1-28.

Letwin, W., 1983. The case against equality, In Letwin, W. (Ed.), Against Equality, Macmillan Press.

Li, Hongyi, Squire, Lyn, and Zou, Heng-Fu, 1998. Explaining international and intertemporal variations in income inequality. Economic Journal 108: 26-43.

Li, Kui-Wai, 2002. Capitalist Development and Economism in East Asia: The Rise of Hong Kong, Singapore, Taiwan and South Korea, Routledge.

Lin, Shu-Chin, Huang, Ho-Chuan, and Weng, Hsiao-Wen, 2006. A semiparametric partially linear investigation of the Kuznets' hypothesis. Journal of Comparative Economics 34, 634-647.

Milanovic, Branko, 2005. Worlds Apart: Measuring International and Global Inequality, Princeton University Press.

Mushinski, David W., 2001. Using non-parametrics to inform parametric tests of Kuznets' Hypothesis. Applied Economics Letters 8: 77-79.

Ram, R., 1991. Kuznets' inverted U hypothesis: evidence from a highly developed country. Southern Economic Journal 57, 1112-23.

Ravallion, M., 2003. Inequality convergence. Economics Letters 80: 351-356.

Sen, Amartya K., 1991. Presidential address: the nature of inequality. In Arrow, Kenneth (Ed.), Issues in Contemporary Economics, Volume 1: Markets and Welfare, MacMillan Academic and Professional Ltd., 3-21.

Sen, Amartya K., 1992. Inequality Reexamined, Harvard University Press.

Sen, Amartya K., 1993. Capability and well-being. In Nussbaum, Martha, and Sen, A. K. (Eds), The Quality of Life, Clarendon Press, 30-53.

Wan, Guang Hua, 2002. Income Inequality and Growth in Transition Economies: Are Nonlinear Models Needed? Discussion Paper No. 2002/104, World Institute for