A KERNEL REGRESSION WITH MIXED DATA TYPE INVESTIGATION OF THE KUZNETS HYPOTHESIS

Erik Alencar de Figueiredo*

Abstract: This paper reexamines the Kuznets hypothesis by taking into account a pool of data provided by Iradian (2005). The empirical strategy is based on: i) the use of the test for parametric specification developed by Hisao et al. (2007); ii) the use of the nonparametric estimation method with mixed data proposed by Racine & Li (2004) and iii) the use of the likelihood ratio test devised by Fan et al. (2001). Results indicate inconsistency of the linear parametric model for the dataset under analysis. Nonparametric inferences produced arguable results. While the bivariate model corroborates the Kuznets hypothesis, the multivariate estimation does not support this hypothesis. Finally, the likelihood ratio tests showed statistical superiority of nonparametric models over linear ones.

Keywords: Kuznets hypothesis. Specification tests, Kernel regression with mixed data. Likelihood ratio test.

JEL code: C14; C21; O11; O15

Resumo: Este trabalho faz uma releitura da hipótese de Kuznets, tendo em conta um conjunto de dados fornecidos por Iradian (2005). A estratégia empírica baseia-se: i) no uso do teste para a especificação paramétrica desenvolvida por Hisao et al. (2007); ii) na utilização do método de estimação paramétrico com dados mistos propostos por Racine & Li (2004); e iii) o uso da função de verossimilhança desenvolvida por Fan et al. (2001). Os resultados indicam inconsistência do modelo paramétrico para o conjunto sob análise. Inferências náo paramétricas produziram resultados discutíveis. Enquanto o modelo bivariado corrobora a hipótese de Kuznets, a estimação multivariada não sustenta esta hipótese. Finalmente, as razões de verossimilhança mostraram-se superioridade estatísticamente superiores aos modelos não-paramétricos em detrimento dos lineares.

* Professor do Departamento de Economia da UFPB e do Programa de Pós-Graduação em Economia da UFPB.

Palavras-chave: Hipótese de Kuznets. Testes de especificação. Regressão de Kernel com dados mistos. Testes de razão de verossemelhança.

Código JEL: C14; C21; O11; O15

1. Introduction

The assumption of nonlinear relationship between income inequality levels and economic growth, put forward by Kuznets (1955), has been widely debated in the specialized literature. Roughly speaking, this nonlinearity is described using an inverted U curve, indicating that the inequality pattern initially increases in the short run with economic growth and decreases in the long run from a turning point. The theoretical explanations for this pattern are diverse, being based on the existing transition between the agricultural and industrial sectors, among others (Robinson, 1976); on the change of financing systems (Greenwood & Jovanovic, 1990); and on technological progress (Galor & Tsidonn, 1997).

Conventionally, empirical studies seek to capture the nonlinearity of the Kuznets curve using a parametric equation suggested by Ahluwalia (1976), where inequality is explained by a second-degree polynomial of the per capita income. Due to its simplicity, this functional form became the favorite specification in studies conducted to validate or not the existence of a Kuznets curve. With the purpose of knowing what occurred after the Kuznets curve, List & Gallet (1999) included a third-degree polynomial of per capita income. Their results suggested that, from a given per capita income level, inequality increases again.[13] In brief, there is a wide variety of functional

[13] Other specifications were suggested. For further details, see Anand & Kambur (1993).

forms encouraging the debate on the validity or not of the Kuznets hypothesis.[14]

In this scenario, nonparametric econometrics comes as a natural path, given that it is free of impositions of the data generating process. Several studies were based on this tool, and to cite a few we have the ones by Hang (2004), Lin et al. (2006) and Huang et al. (2007). However, the results described in the literature are subjected to at least two problems. First, the semiparametric and nonparametric estimations used tend to deal with discrete and continuous variables in a similar way [see, for instance, Lin et al. (2006)]. Another recurrent problem concerns the adoption of nonparametric models based on their presumed superiority, without tests that confirm their robustness. In other words, no tests are run for parametric specifications, and there is no statistical comparison between the instruments used.

Considering these pieces of evidence, this paper aims to reexamine the Kuznets hypothesis by analyzing a group of countries in a pool of data provided by Iradian (2005). In relation to previous studies, the present paper seeks to make three major improvements: a) tests will be run for parametric specifications; b) in case of misspecification, a nonparametric tool will be used for mixed data and; c) finally, comparative tests will be run between parametric and nonparametric inferences. To achieve that, we are going to consider tests for kernel-based quantile regression proposed by Hisao et al. (2007), the nonparametric estimation method with mixed data suggested by Racine & Li (2004) and the comparative analysis between parametric and nonparametric models, using the generalized likelihood ratio test developed by Fan et al. (2001).

In addition to this introduction, the paper is organized as follows. Section 2 introduces and discusses the results and Section 3 concludes.

[14] A list of empirical studies is given in Fields (2001).

2. Results

The empirical strategy of the study consists of three stages. First, different parametric specifications will be tested for the Kuznets curve. In this stage, special attention is given to the kernel-based tests devised by Hisao et al. (2007). If parametric specifications are not statistically significant, a nonparametric model will be used for the data. This stage includes the kernel regression model with mixed data types proposed by Racine & Li (2004).[15] Finally, a comparative test between the two methods (parametric and nonparametric) will confirm the robustness of results.

The Kuznets hypothesis will be tested using a pool of data provided by Iradian (2005). The data include information on 82 countries for the 1965-2003 period. Variables include the Gini coefficient as a proxy for income inequality and the per capita income, measuring the level of economic growth. Other variables are available, namely: government spending as a proportion of the GDP, population growth, level of education and dummy variables for Latin American and African countries. All continents are represented in the sample.

Three functional forms described in the literature are considered:

Model A $$G = \alpha + \beta_1 Y + \beta_2 Y^2 + \theta_i Z_i + u,$$

Model B $$G = \alpha + \beta_1 Y + \beta_2 Y^2 + \beta_3 Y^3 + \theta_i Z_i + u,$$

Model C $$G = \alpha + \beta_1 Y + \beta_2 (1/Y) + \theta_i Z_i + u.$$

Where G is the Gini coefficient, Y is the log of the per capita income, Z_i are the other control variables and ε is the normally

[15] Strategy similar to the one adopted by Maasoumi et al. (2007).

distributed stochastic term with zero mean and constant variance.

The results for the ordinary least squares (OLS) regressions are summarized in Table 1. The coefficients related to the control variables were omitted due to space restrictions. In summary, the estimates indicate a good fit, with R^2 adjusted always greater than 0.62. In general, the Kuznets hypothesis is confirmed in models A and C. The nonsignificance of the coefficient related to the cubed per capita income, Model B, is consistent with the behavior described by List & Gallet (1999), that is: from a given per capital income level, inequality increases again.

Table 1: Linear Regressions

	Model A	**Model B**	**Model C**
Y	18.651* (5.830)	98.489*** (52.039)	-9.561* (2.824)
Y^2	-1.110* (0.344)	-10.759*** (6.258)	---
Y^3	---	0.383 (0.248)	---
$(1/Y)$	---	---	-647.735* (188.274)
R^2 -adjusted	0.622	0.624	0.633
F-statistics	64.11	56.69	64.64

Note: Control variables: the government expenditure in GDP, the percent change in population, the secondary school enrollment rate, and two dummies for African and Latin American countries. Standard errors in parentheses. *, **, ***, denote significant at 1%, 5% and 10% level, respectively.

Nevertheless, it is important to make the following question: are these results obtained from correct specifications? It is well known that parametric results are inconsistent in case of a misspecification (Li & Racine (2007)). Given this

challenge, the present paper chooses to use the test for kernel-based specifications developed by Hisao et al. (2007).

In sum, define the parametric model as $m(x_i, \gamma)$ and the unknown conditional mean as $E(y_i \mid x_i)$. Thus, the test consists of $H_0 : E(y_i \mid x_i) = m(x_i, \gamma)$ vis-à-vis alternative hypothesis $H_1 : E(y_i \mid x_i) \neq m(x_i, \gamma)$. By defining $u_i = y_i - m(x_i, \gamma)$, the correct specification demands that $[E(u_i \mid x_i)]^2 = 0$, otherwise $[E(u_i \mid x_i)]^2 \geq 0$. On account of that, define $I \equiv E\{[E(u_i \mid x_i)]^2 f(x_i)\}$, where $f(x_i)$ is a nonparametric density function. Note that $I = 0$ if and only if H_0 holds true. The sample I is obtained from the replacement of u_i with the residuals derived from the parametric estimation, $\hat{u}_i = y_i - m(x_i, \hat{\gamma})$, and by considering the consistent kernel estimators for $E(y_i \mid x_i)$ and $m(x_i, \gamma)$. Therefore, define:

$$I_n = n^{-2} \sum_i \sum_{j \neq i} \hat{u}_i \hat{u}_j K_{\eta,ij}, \qquad \textbf{(1)}$$

where $K_{\eta,ij} = W_{h,ij} L_{\lambda,ij}$, $\eta = \hat{h}, \hat{\lambda}$ are the bandwidths calculated by cross-validation,[16] $W_{h,ij}$ and $L_{\lambda,ij}$ are the multivariate kernel functions for the discrete and continuous data, respectively. The statistical distribution under the null is calculated by wild bootstrap.

Statistic (2.1) was calculated for models A, B and C. Bootstraps with 1,000 replicates were considered. The results, summarized in Table 2, indicate rejection of the hypothesis of correct specification for all parametric models considered, with a 1% significance level. Hence, it may be concluded that the parametric inference is not suitable for this finite sample.

[16] For further details, see Härdle (1990).

Table 2: Test for parametric specifications

	Statistic	p-value
Model A	5.4728*	0.0000
Model B	5.5102*	0.0000
Model C	5.4850*	0.0000

Note: Control variables: the government expenditure in GDP, the percent change in population, the secondary school enrollment rate, and two dummies for African and Latin American countries. Standard errors in parentheses. *, **, ***, denote significant at 1%, 5% and 10% level, respectively.

Given the inconsistency of parametric models, we propose the use of a nonparametric tool. Similar strategies have already been used in the literature [see Hang (2004), Lin et al. (2006), Huang et al. (2007), among others]. However, the approach used herein differs from all others because we use a data structure containing discrete and continuous variables (mixed data). Therefore, the nonparametric regression will be:

$$y_i = g(x_i) + u_i, \qquad \textbf{(2)}$$

where $g(x_i)$ is an unknown function. The estimator for $g(x_i)$ is given by

$$\hat{g}(x) = \frac{n^{-1}\sum_{i=1}^{n} y_i K_{\eta,ij}}{\hat{f}(x)} \qquad \textbf{(3)}$$

Again, the estimation includes multivariate kernel functions for discrete and continuous data, as previously outlined. As with the test for the specification, the smoothing parameters are estimated by cross-validation.

First of all, a bivariate estimation is considered, i.e., inequality is explained by the per capita GDP. Figure 1 summarizes this estimation process. Visual inspection indicates the existence of a Kuznets curve for this dataset. In this case, the results are consonant with the semiparametric and nonparametric estimations described in Lin et al. (2006), Huang et al. (2007), for instance.

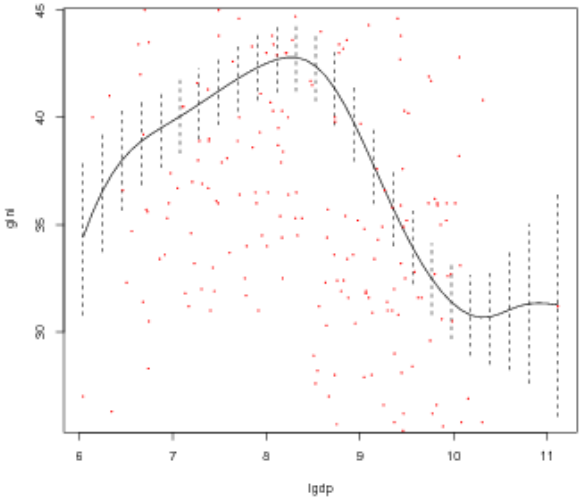

Figure 1: Kuznets curve – bivariate estimation

Nonetheless, the inclusion of other covariables (discrete and continuous) indicates a change in this scenario. This behavior can be seen in Figure 2. Nonlinear behaviors between the dependent variable and the covariables related to government spending, level of education and population growth are highlighted. In their empirical studies, Lin et al. (2006) and Huang et al. (2007) established a linear relationship between the aforementioned covariables and the level of inequality, considering the nonparametric component only in the per capita income variable. However, the results portrayed in Figure 2

demonstrate misspecification of this structure. With regard to the association between inequality and economic growth, it is not possible to characterize it using an inverted U curve. That is, the estimation does not indicate the existence of a Kuznets curve.

Finally, as a way to compare parametric inferences (three models) and the nonparametric one, we use the generalized likelihood ratio test developed by Fan et al. (2001):

$$GLR = \frac{N}{2}\frac{SQR - SQIR}{SQR}, \qquad \textbf{(4)}$$

where SQR is the sum of squares of the residuals of the linear model and $SQIR$ is the sum of squares of the residuals of the nonparametric model. Under the null hypothesis of equality between the two methods, Fan & Yao (2003) calculate the asymptotic distribution for (2.2) from the bootstrap.

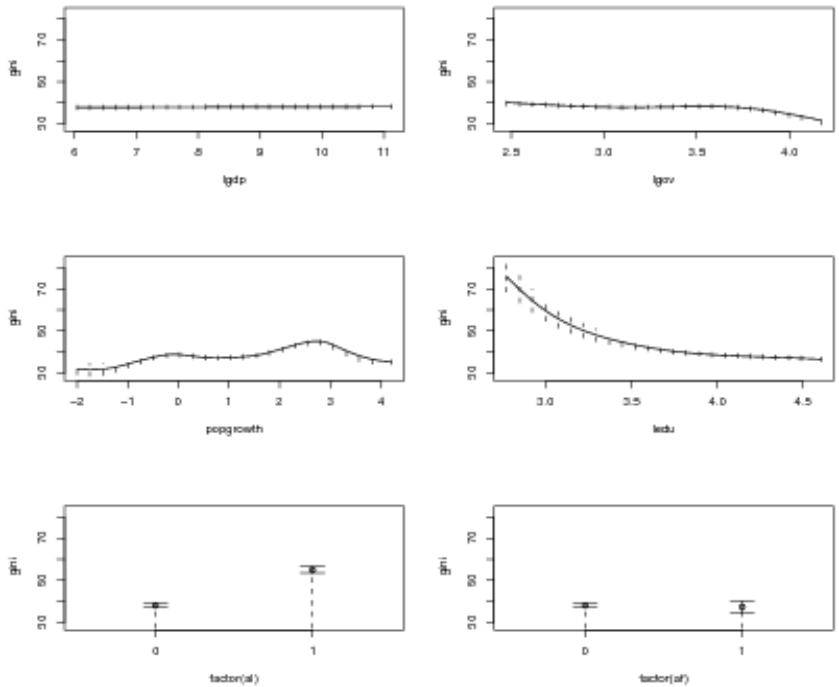

Figure 2: Kuznets curve – multivariate estimation

The results outlined in Table 3 consider bootstraps with 1,000 replicates. Observe that the nonparametric model is superior in all estimated models, given that the null hypothesis of the test could be rejected at a 1% significance level.

Table 3: Generalized Likelihood Ratio Test

	Model A	**Model B**	**Model C**
p-value	0.0037*	0.0037*	0.0038*

Nota: *, **, ***, denote significant at 1%, 5% and 10% level, respectively.

In brief, the results suggest that: a) the linear inference is not consistent; b) the nonparametric model provides two results: in the bivariate case, there is an inverted U behavior; and the

multivariate result does not relate to the Kuznets hypothesis and; c) the nonparametric inference is more adequate if compared to the three parametric models estimated.

3. Final Remarks

This paper sought to reexamine the Kuznets hypothesis. The empirical strategy employed a wide series of nonparametric instruments in order to a) test the linear functional forms; b) to make a nonparametric inference by considering mixed data and; c) to compare linear and nonparametric methods.

The results revealed inconsistency of the linear parametric structure for the dataset used. Nonparametric inferences yielded arguable results. While the bivariate model supports the Kuznets hypothesis, the multivariate estimation does not. Finally, the likelihood ratio tests showed statistical superiority of nonparametric models over linear ones.

4. References

Ahluwalia, M. (1976). Inequality, poverty and development. *Journal of Development Economics*, 3: 307–342.

Anand, S. & Kanbur, S. (1993). The Kuznets process and the inequality-development relationship. *Journal of Development Economics*, 40: 25-52.

Fan, J. & Yao, Q. (2003). *Nonlinear time series*: nonparametric and parametric methods. Springer.

Fields, G. (2001). *Distribution and development: a new look at the developing world*. Cambridge: MIT Press.

Galor, O. & Tsiddon, D. (1997). Technological progress, mobility, and economic growth. *American Economic Review*, 87: 363–382.
Greenwood, J. & Jovanovic, B. (1990). Financial development, growth and the distribution of income. *Journal of Political Economy*, 98: 1076–1107.
Iradian, G. (2005). *Inequality, poverty, and growth: cross-country evidence*. IMF working paper.
Härdle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.
Hsiao, C., Li Q. & Racine, J. (2007). A consistent model specification test with mixed categorical and continuous data. *Journal of Econometrics*, 140: 802-826.
Huang, H. (2004). A flexible nonlinear inference to the Kuznets hypothesis. *Economics Letters*, 84: 289–296.
Huang, H., Lin, S., Suen, Y. & Yeh, C. (2007). A quantile inference of the Kuznets hypothesis. *Economic Modelling*, 24: 559–570.
Li, Q. & Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
Lin, S, Huang, H & Weng, H. (2006). A semiparametric partially linear inference of the kuznets hypothesis. *Journal of Comparative Economics*, 34: 634–647.
List, J. & Gallet, C. (1999). The Kuznets' curve: what happens after the inverted-U? *Review of Development Economics*, 3: 200-206.
Maasoumi, E., Racine, J. & Stengos, T. (2007). Growth and convergence: a profile of distribution dynamics and mobility. *Journal of Econometrics*, 136: 483-508.
Racine, J. & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1): 99-130.

Robinson, S. (1976). A note on the U-hypothesis relating income inequality and economic development. *American Economic Review*, 66: 437–440.