A semiparametric analysis of determinants of protected area[*]

Phu NGUYEN VAN[†]

Bureau d'Économie Théorique et Appliquée (BETA)
Université Louis Pasteur, Strasbourg

March 27, 2003

Abstract

We use a semiparametric additive model to study the relationship between protected area, income, trade, population, education, and political institutions in a sample of 89 countries. The results show the nonexistence of environmental Kuznets curve in the data sample. The study also points out the existence of nonlinearity in the relationship between protected area and the ratio of net secondary school enrollment.

Key words: Education; environmental Kuznets curve; protected area; semiparametric additive models

JEL classification: C14; O13

[*] I would like to thank Théophile Azomahou and François Cochard for helpful comments. All remaining errors are mine.

[†] BETA, Université Louis Pasteur, 61 avenue de la Forêt Noire, F-67085 Strasbourg Cedex, France; Tel.: +33 (0)3 90 24 21 00; Fax.: +33 (0)3 90 24 20 71; E-mail: `nvphu@cournot.u-strasbg.fr`

1 Introduction

The literature about the determinants of environment (environmental quality, pollutant emissions, etc.) is abundant. In recent years, most studies have used parametric models to analyse the significance of key variables representing economic development (income per capita, income inequality, etc.), population (population growth and population density), social situation (educational level, etc.), and political institutions (political rights, civil liberties, etc.). However, parametric models have a major inconvenient that impose a priori functional forms on the relationship between the dependent variables representing the environment and its determinants.

This restriction is relaxed in semiparametric and nonparametric models, which have been used, for example, by Schmalensee, Stoker, and Judson (1998), Taskin and Zaim (2000), Millimet and Stengos (2000), and Millimet, List, and Stengos (2001). In the study done by Schmalensee, Stoker, and Judson (1998), the authors used a piecewise linear model to analyse the relationship between national carbon dioxide emissions and income. Taskin and Zaim (2000) estimated the relationship between environmental efficiency and income. Millimet and Stengos (2000) and Millimet, List, and Stengos (2001) used a semiparametric partial linear model to estimate the relationship between US state-level emissions of several pollutants and income.

As recognised in the literature, other factors such as trade, population, education, and political institutions might affect the environment. In this paper, we propose a semiparametric additive partially linear model to investigate the relationships between the demand for environmental quality (represented by the percentage of protected area within national territory), economic growth, trade, population, education, and political institutions.

The paper is organised as follows. Section 2 presents the data and variables used in this paper. The econometric specification and estimation results are discussed in Section 3. Section 4 concludes the study.

2 Data and variables

In this paper, we use the percentage of protected area within national territory as an indicator of the demand for environmental quality. Protected area is defined by the

International Union for Conservation of Nature and Natural Resources (IUCN) as 'an area of land and/or sea especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, through legal or other effective means'. Protected area is then a direct measure of environmental expenditure and policies, and accounts for the stock effect (contrary to flow variables such as carbon emissions and deforestation rate, etc., which account for the flow effect). Moreover, it is a measure of the country's environmental preferences (Bimonte (2002)). An increase in the surface of protected area is viewed as an increased demand for environmental conservation, therefore it represents an increase in the demand for environmental quality.

The literature on the Environmental Kuznets Curve (EKC), which states that environmental degradation increases with income but decreases when income per capita exceeds a certain level, is abundant (see Panayotou (2000) for an overview). There was evidence of EKC for some environmental indicators. Concerning protected area, using a parametric specification with linear and squared terms of log-income, Bimonte (2002) found the existence of EKC in a European dataset in 1996.

Other determinants of the environment are also discussed in the literature. International trade has been seen as an explanatory factor of environment. Rich countries might spin-off pollution-intensive products to developing countries with lower environmental standards, either through trade or direct investment in these countries (see, for example, Panayotou (2000)). We characterise this variable by the ratio between total trade (imports + exports) and GDP.

Population is an important factor, especially for local environment. Cropper and Griffiths (1994) and Koop and Tole (1999) found that population density and population growth rate have a positive effect on deforestation. Bhattarai and Hammig (2001) found that rural population density is a significant factor contributing to the deforestation process in Latin America and Africa. We only use population density in regressions.[1] Population factor was not used in Bimonte (2002) for protected area.

Human capital or education might also have an important role in environmen-

[1] The reason is that protected area and population density are stock variables whereas population growth rate is a flow variable.

tal degradation because it facilitates information accessibility (awareness of consequences of environmental damage, etc.) and the degree of participation of people in the development process (participation in the decision-making process for the sustainability of development, etc.). We use the ratio of net secondary school enrollment as a measure of education.[2] Torras and Boyce (1998) found that a higher literacy is significantly associated with better environmental quality in the case of concentrations of sulfur dioxide, heavy particle, dissolved oxygen and in the case of percent of access to sanitation. Bimonte (2002) used the number of newspapers per 1000 people sold yearly in each country and found that it has a positive impact on protected area. It might be thought that the ratio of net secondary school enrollment is close to the indicator used by Bimonte (2002).

Political institutions of a country can also affect the process of environmental degradation. We use two indicators: political rights and civil liberties. For each indicator, countries are classified according to an ordinal scale from 1 (free) to 7 (not free). As in the study made by Bhattarai and Hammig (2001), we aggregate these two indicators to obtain an index of political institutions, the values of which vary from 2 to 14. Torras and Boyce (1998) found a strong effect of political institutions, in low-income countries, on concentrations of sulfur dioxide, smoke, heavy particles, dissolved oxygen and fecal coliform. Bhattarai and Hammig (2001) found that political institutions have a significant effect on the tropical deforestation process. Variables on political institutions were not present in the study of Bimonte (2002).

Explanatory variables in our model are real GDP per capita (measured in thousands 1985$), trade (total trade/GDP), population density (people/hectare), political institutions, and education (ratio of net secondary school enrollment). To control for regional heterogeneity, we include regional dummies for Asia (excluding Middle East, 13 countries), Europe (21), Middle East & North Africa (8), Sub-

[2] The Gini index, measuring income inequality, might also represent the participation degree. We do not use the Gini index here because of data limitations. Indeed, the data on the Gini index (the largest dataset is from the UNDP at `http://www.undp.org/poverty/initiatives/wider/wiid.htm`) contains many missing values. Moreover, the Gini index is not comparable across countries because of different measures of income (gross/net income, earnings, expenditure, etc.) and different sampling bases (entire population, employed population, urban/rural population, age limitation, etc.).

Saharan Africa (23, considered as the reference group), North America (2), Latin America (19), and Oceania (3).[3] All the data used in this paper is obtained from the World Resources Institute (2000), except the data on political institutions, which is obtained from the Freedom House.[4] Most data is from 1997, only the data on population density is from 1996 because of its availability. Table 1 reports descriptive statistics of variables.[5] On average, protected area covers about 8.4% of national territory of the countries in our sample. United Arab Emirates has no protected area whereas 42.1% of the territory of Ecuador is protected area. The ratio of net secondary school enrollment has a mean value of 65%. Concerning political institutions, the mean value is quite high (around 6.5), which shows that the majority of countries in the sample are not entirely free.

Table 1

3 Estimation

The econometric model consists of a semiparametric additive partially linear model, which is described in Hastie and Tibshirani (1990):[6]

$$Y = \alpha + \sum_{j=1}^{p} f_j\left(X_j\right) + \mathbf{Z}'\boldsymbol{\gamma} + \epsilon, \tag{1}$$

with $E\left(\epsilon \mid X_1, \ldots, X_p, \mathbf{Z}\right) = 0$ and $V\left(\epsilon \mid X_1, \ldots, X_p, \mathbf{Z}\right) = \sigma^2$. Y is the dependent variable representing environmental indicator. $X_j, j = 1, ...p$, and $\mathbf{Z}$ are explanatory variables. The f_js are unknown univariate functions, one for each predictor. For identification purpose, it is assumed that $E\left[f_j\left(X_j\right)\right] = 0$. $\mathbf{Z}$ is a vector of discrete variables that enter in (1) linearly. In this paper, X_j are real GDP per capita, trade, population density, and the ratio of net secondary school enrollment ($p = 4$). We remark that all these variables are continuous. $\mathbf{Z}$ includes discrete variables: political institutions and regional dummies.

Two major arguments are in favour of the model (1) in this paper. First, it helps us to avoid the 'curse of dimensionality', which appears in nonparametric regressions

[3] This regional classification is used by the World Resources Institute (WRI) and the Food and Agriculture Organization (FAO).

[4] `http://www.freedomhouse.org/`

[5] The list of countries is provided in Appendix.

[6] Bold characters represent matrix notations.

when functional dimension is high, i.e. several explanatory variables are present. Secondly, it enables us to capture eventual nonlinearities or heterogeneities in the effects of explanatory variables on environmental quality. The latter argument allows us to apply this model to datasets which might include heterogenous countries, i.e. countries in different stages of development and notably to test whether an EKC exists.

Estimation of the model in (1) might be implemented by using the 'backfitting algorithm' described in Hastie and Tibshirani (1990) (see Appendix). Another method of estimation is marginal integration but it is more time consuming. Estimation results are reported in Table 2. We also present an estimation of the parametric coefficients associated to X_j in the parametric linear specification $Y = \alpha + \sum_{j=1}^{p} \beta_j X_j + \mathbf{Z}'\boldsymbol{\gamma} + \epsilon$, which is performed by Ordinary Least Squares (OLS).

Table 2

To compare the nonparametric function of a variable with the corresponding parametric function, we compute a 'gain' statistic which follows approximately a χ^2 (see Appendix). The individual gain statistics show that the nonparametric function for the net secondary school enrollment is highly preferred against the linear function at the 5% level. The total gain statistic, which is the sum of individual gains and follows a χ^2 with degrees of freedom equal to the sum of individual degrees of freedom, is equal to $23.835 > \chi^2\,(13.625) = 23.190$ at the 5% level. As a result, the parametric model is rejected against the semiparametric model.

The relation between protected area and the ratio of net secondary school enrollment has a nonlinear pattern, as shown in Figure 1. This relation is significant because the 95% confidence interval does not include the horizontal line at zero, which represents a zero effect. We can conclude that the demand for environmental quality increases with the ratio of net secondary school enrollment, but decreases when this ratio exceeds a certain level. This might be explained by the following argument. When the educational level (information accessibility and degree of participation) of people increases, they have a higher demand for environmental quality. As a result, environmental quality increases. However, when the educational level exceeds a certain level, this demand diminishes because people do not need a higher environmental quality as its level is already high. These results suggest that information accessibility and degree of participation are important in environmental protection.

Figure 1

The parametric coefficient of GDP per capita is insignificant. Moreover, when we use a nonparametric function for this variable, $f_1(.)$, there is no significant improvement. Therefore, there is no correlation between protected area and GDP per capita: EKC does not exist for protected area, contrary to the result of Bimonte (2002). Trade, population density, and political institutions have no significant effect on protected area. Regional heterogeneity (comparing to Sub-Saharan countries) exists, in particular Latin America has a positive and significant effect on protected area.

4 Concluding remarks

We use a semiparametric additive model to study the relationship between protected area, income, trade, population, education, and political institutions in a dataset of 89 countries. Semiparametric techniques help us to account for nonlinearities in the relationship between environmental quality and its determinants (here the ratio of net secondary school enrollment is the case). The results show the nonexistence of EKC in the data sample but show evidence of the effects of education on the demand for environmental quality. Therefore, this study suggests that policy makers should pay more attention on the important role of education in environmental protection.

References

Bhattarai, M., and M. Hammig (2001): “Institutions and the Environmental Kuznets Curve for Deforestation: A Crosscountry Analysis for Latin America, Africa and Asia,” *World Development*, 29, 995–1010.

Bimonte, S. (2002): “Information Access, Income Distribution, and the Environmental Kuznets Curve,” *Ecological Economics*, 41, 145–156.

CROPPER, M., AND C. GRIFFITHS (1994): "The Interaction of Population Growth and Environmental Quality," *American Economic Review*, 82, 250–254.

HASTIE, T. J., AND R. J. TIBSHIRANI (1990): *Generalized Additive Models*. Chapman and Hall, London, New York.

KOOP, G., AND L. TOLE (1999): "Is There an Environmental Kuznets Curve for Deforestation?," *Journal of Development Economics*, 58, 231–244.

MILLIMET, D. L., J. A. LIST, AND T. STENGOS (2001): "The Environmental Kuznets Curve: Real Progress or Misspecified Models?," Department of Economics working paper, Southern Methodist University.

MILLIMET, D. L., AND T. STENGOS (2000): "A Semiparametric Approach to Modelling the Environmental Kuznets Curve Across US States," Department of Economics working paper, Southern Methodist University.

PANAYOTOU, T. (2000): "Economic Growth and the Environment," CID working paper no. 56, Harvard University.

SCHMALENSEE, R., T. M. STOKER, AND R. A. JUDSON (1998): "World Carbon Dioxide Emissions: 1950-2050," *Review of Economics and Statistics*, 80, 15–27.

TASKIN, F., AND O. ZAIM (2000): "Searching for a Kuznets Curve in Environmental Efficiency Using Kernel Estimation," *Economics Letters*, 68, 217–223.

TORRAS, M., AND J. K. BOYCE (1998): "Income, Inequality, and Pollution: a Reassessment of the Environmental Kuznets Curve," *Ecological Economics*, 25, 147–160.

WORLD RESOURCES INSTITUTE (2000): *World Resources Database 2000-2001*. World Resources Institute, Washington DC.

5 Appendix

5.1 List of countries

Algeria, Angola, Argentina, Australia, Austria, Bangladesh, Belize, Benin, Bolivia, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cameroon, Canada, Central

African Rep., Chad, Chile, China, Colombia, Congo (Dem. Rep.), Congo Rep., Czech Rep., Denmark, Dominican Rep., Ecuador, Egypt, El Salvador, Ethiopia, Fiji, Finland, France, Gambia, Georgia, Greece, Guatemala, Guinea, Honduras, Hungary, Iceland, India, Indonesia, Ireland, Italy, Ivory Coast, Jamaica, Japan, Kenya, Korea Rep., Latvia, Lesotho, Malawi, Malaysia, Mali, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Namibia, Nepal, Netherlands, New Zealand, Niger, Norway, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Saudi Arabia, Senegal, Singapore, South Africa, Spain, Sri Lanka, Swaziland, Sweden, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, United Arab Emirates, United Kingdom, United States, Uruguay, Venezuela, Zambia, Zimbabwe.

5.2 Backfitting algorithm and specification test

The estimation of the model (1) might be implemented by the following steps (see Hastie and Tibshirani (1990)):

(i) Center the data.

(ii) Regress the residuals on $X_j, j = 1, ..., p$, by using the backfitting algorithm, described below. The resulting smooth is the first estimate of $f_j(.)$, $\hat{f}_j(.)$.

(iii) Obtain the estimate of $\boldsymbol{\gamma}$ by Ordinary Least Squares (OLS): $\hat{\boldsymbol{\gamma}} = E\left(Y - \hat{\alpha} - \sum_{j=1}^{p} \hat{f}_j(X_j) |\mathbf{Z}\right)$, where $\hat{\alpha} = \frac{1}{n}\sum_i^n Y_i$. Center the data again, and the process continues until convergence.

Note that in step (i), an initial estimate of $f_j(.)$ has to be used. For this purpose, we can use the parametric OLS estimator $\hat{\beta}_j X_j$.

The backfitting algorithm consists of the following steps:

(a) Initialize: $\hat{\alpha} = \frac{1}{n}\sum_i^n Y_i$, $f_j(X_j) = f_j^0(X_j)$, $j = 1, ..., p$.

(b) Cycle: $j = 1, ..., p, 1, ..., p, ...$

$$\hat{f}_j(X_j) = S_j\left(Y - \hat{\alpha} - \sum_{k \neq j} \hat{f}_k(X_k) \mid X_j\right).$$

Continue (b) until the individual functions don't change. S_j is the smoother, using $k-$nearest symmetric neighborhood, for $f_j(.)$. Note that in the step (a), we can use linear estimators for f_j^0.

The degree of freedom of the fit $\hat{f}_j$, df_j – considered as the effective number of parameters – might be approximated by the trace of $2\mathbf{S}_j - \mathbf{S}_j\mathbf{S}_j'$, where $\mathbf{S}_j$ is the smoothing matrix so that $\hat{\mathbf{f}}_j = \mathbf{S}_j\mathbf{w}$ (note that $\hat{\mathbf{f}}_j$ is the vector of $\hat{f}_j$ and $\mathbf{w}$ is the vector corresponding to $Y - \hat{\alpha} - \sum_{k\neq j} \hat{f}_k(X_k)$ in the step (b)). Therefore, df_j might be fractional. In case of linear estimator (OLS), we have $\mathbf{S}_j = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, where $\mathbf{X}$ is the matrix of regressors, and $df_j = 1$.

To compare two individual smooths $\hat{\mathbf{f}}_{j,1} = \mathbf{S}_{j,1}\mathbf{w}$ and $\hat{\mathbf{f}}_{j,2} = \mathbf{S}_{j,2}\mathbf{w}$, for example $\hat{\mathbf{f}}_{j,1}$ is linear, we can use the following approximative statistic (see Hastie and Tibshirani (1990)):

$$J = \frac{(RSS_1 - RSS_2)/(df_2 - df_1)}{RSS_2/(n - df_2)} \sim F_{df_2 - df_1, n - df_2},$$

where RSS_1 and RSS_2 are respectively the deviance (or the residual sum of squares) of the models corresponding to $\hat{f}_{j,1}$ and $\hat{f}_{j,2}$. This distribution of the statistic 'gain', $= J \times (df_2 - df_1)$, might be approximated by $\chi^2(df_2 - df_1)$.

Table 1: Descriptive statistiques

Variable	Mean	Std.Err.	Min.	Max.	#obs.
Protected area within national territory	0.084	0.080	0	0.421	89
Real GDP per capita (thousands 1985$)	5.623	5.624	0.198	20.049	89
Trade ((imports + exports)/GDP)	0.555	0.368	0.129	2.660	89
Population density (people/hectare)	1.609	5.942	0.016	55.475	89
Net secondary schooling enrollment	0.654	0.267	0.094	0.999	89
Political rights	3.124	2.120	1	7	89
Civil liberties	3.360	1.660	1	7	89
Political inst. (political rights + civil liberties)	6.483	3.687	2	14	89

Table 2: Estimation results

Variable	Coef.	Std.Err.	*df*	Gain
GDP per capita	0.002	0.002	2.0	0.617
Trade	-0.023	0.027	2.6	2.632
Population density	-0.001	0.002	3.0	1.790
Net secondary schooling	-0.029	0.049	10.0	18.796**
Political institutions	-0.003	0.003	1	–
Asia (excluding Middle East)	0.009	0.029	1	–
Europe	0.002	0.036	1	–
Middle East &North Africa	-0.040	0.033	1	–
North America	-0.016	0.067	1	–
Latin America	0.050*	0.027	1	–
Oceania	0.006	0.053	1	–
Intercept	0.077**	0.021	1	–
#obs.	89			

Notes: *df* is the effective number of parameters (or degrees of freedom). * and ** represent the significance at the 10% and the 5% levels, respectively.

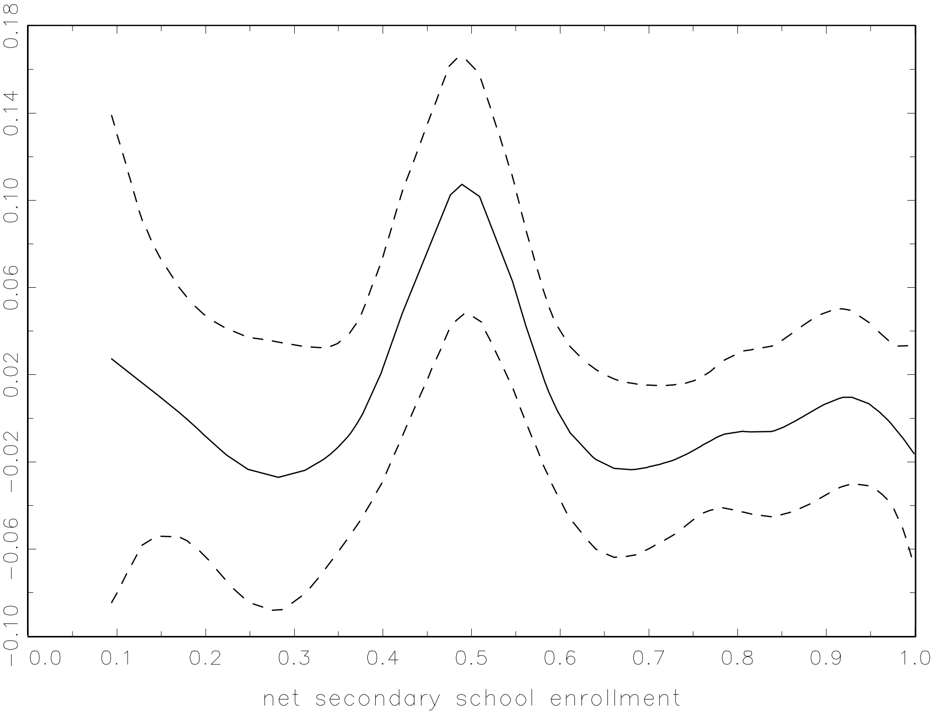

Figure 1: Nonparametric estimation of the effect of the ratio of net secondary school enrollment on protected area. The solid curve is the estimate. The dash curves present the upper and lower bands of the 95% pointwise confidence interval. The data is normalized such that $E\left[f\left(.\right)\right]=0$.