

Article

# The Environmental Kuznets Curve: A Semiparametric Approach with Cross-Sectional Dependence

Alexandra Soberon \*  and Irene D'Hers \*

Department of Economics, University of Cantabria, Avda. de los Castros 56, 39005 Santander, Spain

\* Correspondence: alexandra.soberon@unican.es (A.S.); irenedherssierra31@gmail.com (I.D.)

Received: 21 October 2020; Accepted: 16 November 2020; Published: 23 November 2020



**Abstract:** This paper proposes a new approach to examine the relationship between CO<sub>2</sub> emissions and economic developing. In particular, we propose to test the Environmental Kuznets Curve (EKC) hypothesis for a panel of 24 OECD countries and 32 non-OECD countries by developing a more flexible estimation technique which enables to account for functional form misspecification, cross-sectional dependence, and heterogeneous relationships among variables, simultaneously. We propose a new nonparametric estimator that extends the well-known Common Correlated Effect (CCE) approach from a fully parametric framework to a semiparametric panel data model. Our results corroborates that the nature and validity of the income–pollution relationship based on the EKC hypothesis depends on the model assumptions about the functional form specification. For all the countries analyzed, the proposed semiparametric estimator leads to non-monotonically increasing or decreasing relationships for CO<sub>2</sub> emissions, depending on the level of economic development of the country.

**Keywords:** environmental Kuznets curve; CO<sub>2</sub> emissions; cross-sectional dependence; semi-parametric regression; panel data

**JEL Classification:** C33; C14; Q5

---

## 1. Introduction

The growing interest of citizens and governments in environmental degradation, generated mainly by economic growth and the excessive use of natural resources, has led economies around the world to take measures to mitigate the effects of global warming and climate change, according to what was established in the Kyoto Protocol and the 2015 Paris Agreement on Climate Change Mitigation. However, despite the efforts made, the burning fossil fuels (i.e., carbon, oil and gases) used for the production of energy necessary for economic development continues to significantly contribute to CO<sub>2</sub> emissions.

This situation has led the main policymakers to focus their efforts on promoting policies that pretend to achieve sustainable economic growth, that is, to achieve an increase in economic growth that is compatible with the environment. Since the relationship between economy and environment is highly complex and controversial, much of the empirical studies carried out in the literature are based on the so-called Environmental Kuznets Curves (EKC), which are still subject to an intense debate nowadays.

The EKC is based on the concept of the Kuznets curve, initially proposed by [Kuznets \(1955\)](#), which describes an inverted U-shaped relationship between income inequality and income per capita. This hypothesis argues that income inequality initially increases as per capita income rises and then begins to decrease from a certain threshold point. In the 1990s, the concept of the Kuznets curve was applied for the first time to the environmental quality trying to corroborate whether the relationship

between income per capita and environmental degradation follows a similar relationship in terms of inverted U-shaped pattern (see [Grossman and Krueger 1993, 1995](#) and [Stern 1998](#), for example). In this framework, the EKC postulates that low income levels are correlated with a growing deterioration of the environment, but after a certain threshold point of the income per capita, the relationship between the two variables becomes negative again. In this way, this hypothesis is supported by the argument that higher levels of development imply a change in the economic structure in favor of industry and services, where production processes turn to more efficient and environmentally friendly technologies that help to preserve natural resources and reduce significantly the environmental deterioration. Similarly, the EKC shows the development of an economy over time. In a first stage, the economy is based on the agricultural sector with a strong impact on the quality of the environment. In a second stage, the industry is developed and, although it generates a higher level of wealth, it causes a great damage in the environmental's quality. After a threshold point, the economy sustains its growth in efficient and cleaner technologies, mainly in the services sector. Thus, the EKC hypothesis emphasizes that the economic growth is a precondition for reducing environmental pollution (see [Beckerman 1992](#)).

Since the pioneering work of [Grossman and Krueger \(1991\)](#), several empirical studies have emerged in the literature with the aim of corroborating the EKC hypothesis in the income–pollution relationship. However, so far there is no consensus on this relationship (see [Dinda \(2004\)](#); [Galeotti \(2007\)](#) and [Kaika and Zervas \(2013a, 2013b\)](#) for intensive reviews about this issue). On the one hand, [Selden and Song \(1994\)](#), [Grossman and Krueger \(1995\)](#), [List and Gallet \(1999\)](#) and [Stern and Common \(2001\)](#), among others, point out that the intensity of pollutant emissions initially increases with the income per capita in the early stages of the economic development, but eventually falls as the income per capita rises above certain threshold point, at least in the case of the developed countries. On the other hand, [Harbaugh et al. \(2002\)](#) and [Effiong and Oriabije \(2018\)](#) highlight that there is no evidence that this relationship is robust for a number of emission pollutants.

A possible reason for this lack of consensus may be the presence of several specification errors in the models used to corroborate the EKC hypothesis. Firstly, it is common in this literature to assume ex-ante specific functional forms (i.e., quadratic or cubic polynomials) that may not fully explain the complexity of the EKC relationship. In the last two decades, nonparametric and semiparametric models are gaining popularity in order to deal with this problem (see [Harbaugh et al. 2002](#); [Bertinelli and Strobl 2005](#), and [Azomahou et al. 2006](#) among others). Secondly, most of these studies completely ignore the unobservable heterogeneity of countries or regions due to economic, social, political, structural, and biophysical differences that can have variable effects on environmental quality (see [Dinda 2004](#) or [Purcel 2020](#) and the references therein, among others). Hence, assuming homogeneity of parameters (i.e., suggesting that the income–pollution relationship is the same for all countries) can lead to misleading inferences. Thirdly, most studies with longitudinal data use standard panel data techniques that crucially depend on the independence assumption between the individuals. However, the growing process of globalization and the importance of economic and social interconnections between economic agents make necessary to control for the unobserved cross-sectional dependence (see [Apergis et al. 2017](#) and [Martinez-Zarzoso and Bengochea-Morancho 2004](#), for example).

Given the lack of consensus about the income–pollution relationship based on the results obtained with fully parametric estimation techniques, this article reviews the EKC hypothesis trying to overcome the three potential misspecification problems discussed above: (i) functional misspecification; (ii) cross-sectional dependence from common factors; (iii) heterogeneous relations among variables. More precisely, in this article a new nonparametric estimation procedure is developed that allows to detect the true shape of the income–pollution relationship, allowing heterogeneous relationships among countries and controlling for the possible cross-section dependence related to the presence of a finite number of unobservable common factors (see [Bai 2009](#) and [Pesaran 2006](#), among others). To the best of our knowledge, this is the first paper that try to solve the three misspecification problems simultaneously with the aim of modelling the income–pollution relationship in a more efficient way and leading to more suitable political prescriptions. Finally, with the aim of being able to fully

characterize the Kuznets curve, we will carry out an empirical study with both OECD and non-OECD countries. In this way, we will be able to draw conclusions for both developing countries and those that are developing.

The paper is organized as follows. In Section 2, we present a review of the literature on EKC, focusing mainly on issues related to econometric specifications. In Section 3, the model specification and the data description are presented. In Section 4, we present the main results and in Section 5 we discuss the main empirical findings. Finally, in Section 6 we offer concluding remarks. In the Appendices, we present details about the econometric methodology of the semiparametric model that we propose and the necessary assumptions to obtain the main asymptotic properties of the proposed estimators.

## 2. Literature Overview on the EKC

After the seminal work of [Grossman and Krueger \(1991, 1995\)](#); [Shafik and Bandyopadhyay \(1992\)](#) and [Panayotou \(1993\)](#) on the EKC, the relationship between pollution and economic growth has been intensively analyzed over the past two decades from both theoretical and empirical point of view.

Despite the enormous amount of studies devoted in this topic, the agreement of the EKC hypothesis are far. Evidence of an EKC has been found for several indicators, but these findings are not unanimously accepted in the literature as it can be seen in the case of the carbon dioxide emissions. For this particular environmental measure, an EKC was corroborated in the studies of [Holtz-Eakin and Selden \(1995\)](#), [Galeotti and Lanza \(1999\)](#), [Heil and Selden \(2001\)](#), [Cole \(2004\)](#), [Galeotti et al. \(2006\)](#), and [Parajuli et al. \(2019\)](#), among others, in contradiction with the results obtained in [Shafik and Bandyopadhyay \(1992\)](#), [Galeotti and Lanza \(2001\)](#), [York et al. \(2003\)](#), and [Nutakor et al. \(2020\)](#), for example. Other researches such as [Sengupta \(1996\)](#) and [Martinez-Zarzoso and Bengochea-Morancho \(2004\)](#) find a cubic N-shaped relationship, whereas [Barra and Zotti \(2018\)](#) points out that preliminary evidence validates the Kuznets's hypothesis, although this relationship turned out to be misleading once the issue of (non)-stationarity has been taken into account.

Most of these empirical studies have relied on parametric specifications (i.e., quadratic or cubic specifications). However, although most of them resort to specific panel data technique to deal with the presence of unobservable individual heterogeneity. In this situation, the risk of misspecification is considerably high for two reasons. Firstly, most of the above studies crucially depend on the independence assumption between the individuals. Hence, the presence of cross-sectional independence may invalidate most of their conclusions. Secondly, fully parametric models are subject to restrictive conditions about the functional form. When the assumptions of the empirical model and parameters are inconsistent with the real data generating process, there exists model specification problem and it can lead us to misleading inference and conclusions. In order to avoid this problem, nonparametric and semiparametric regression models have become very popular in the recent literature since they do not require the specification of a specific functional form in order to investigate the existence of EKC in the relationship between pollution and economic growth.

Among the papers that use nonparametric estimation techniques is that of [Harbaugh et al. \(2002\)](#), where they use a nonparametric pooled regression to examine the robustness of the evidence for the existence of an inverted U-shaped relationship between national income and pollution for a panel of countries. They find that the results are highly sensitive both to slight variations in the data and to reasonable permutations of the econometric specifications. Hence, they conclude that there is little empirical support for an inverted U-shaped relationship between pollution and national income in these data. The work in [Bertinelli and Strobl \(2005\)](#) proposes a more flexible semi-parametric specification to overcome the above problems and obtain that they are unable to reject a linear relationship between economic growth and environmental degradation for a panel of 122 countries over the period 1950–1990.

[Azomahou et al. \(2006\)](#) examine the empirical relationship between CO<sub>2</sub> emissions per capita and GDP per capita during the period 1960–1996 using a panel of 100 countries. In order to avoid any ad hoc choice of a parametric functional form, they consider a fully nonparametric model. Further, in order

to avoid any bias related with the presence of unobserved individual heterogeneity, the propose to estimate this regression using marginal integration technique, since they propose to use the first differencing transformation in order to avoid any bias problem related to the presence of unobserved individual heterogeneity. In this framework they obtain a monotonically-increasing relationship between CO<sub>2</sub> emissions and economic development of the existence of an EKC for CO<sub>2</sub> emissions has been clearly contradicted in this more flexible specification.

Bertinelli et al. (2012) investigated the relationship between the CO<sub>2</sub> emissions per capita and the GDP per capita using a kernel regression estimation in a set of countries individually. They obtain that the relationship between output and pollution between countries is different. For some developed countries the relationship has been heterogeneous after 1960, whereas for almost all the developing countries it was always upward sloping. Chen and Chen (2015) analyze the nonlinear relationship between industrial pollution and economic development level based on nonparametric method for testing and verifying the EKC of carbon dioxide in China. Using standard nonparametric techniques they verify the pattern of EKC in the relationship between the industrial carbon dioxide emissions and the level of economic development. Kalaitzidakis et al. (2018) examine the relationship between total factor productivity growth and emissions using a semiparametric smooth coefficient model that allow to directly estimate the output elasticity of emissions using a panel of 17 countries for the period of 1981–1998. They are able to show that there exists a monotonically-increasing relationship between emissions and total factor productivity growth. Further, they find that the CO<sub>2</sub> emissions effect varies depending on a country’s emissions level.

Although the literature on the relationship between pollution and economic growth is extensive, there is still no consensus on the EKC hypothesis even when we resort to more flexible specifications such as nonparametric or semiparametric specifications. Further, more of the above studies ignores the presence of cross-sectional dependence. In the following section, we present an alternative estimation technique that enables us to overcome all these difficulties simultaneously.

### 3. Methodology and Data Source

#### 3.1. Model Specification

Following the standard approach in the literature, we use the Stochastic Impacts by Regression on Population, Affluence and Technology (STIRPAT) model proposed in Dietz and Rosa (1997) as the reference analytical framework for evaluating the anthropogenic forces behind environmental change. Hence, the environmental impacts (*I*) can be understood as a multiplicative function of population size (*P*), affluence described per capita of economic activity (*A*), and the level of technology per unit of consumption and production (*T*) whose model specification is given by

$$I_i = aP_i^b A_i^c T_i^d \varepsilon_i, \tag{1}$$

where (*a, b, c, d*) are the parameters of the model,  $\varepsilon$  denotes the idiosyncratic error term, and the subscript *i* are observational units (i.e., countries, regions) in a cross-section data.

Taking the natural logarithm of (1) and adding a quadratic term of the affluence (*A*) variable in line with the EKC hypothesis to capture possible existence of an inverted U-shaped relationship, the resulting panel data model is specified as follows:

$$\ln(E_{it}) = \beta_0 + \beta_1 \ln(gdpc_{it}) + \beta_2 \ln(gdpc_{it}^2) + \beta_3 \ln(pop_{it}) + \beta_4 \ln(enit_{it}) + \alpha'_i d_t + \varepsilon_{it}, \tag{2}$$

where  $E_{it}$  is a measure of environmental quality of country *i* at time *t*; *pop* denotes the population size; *gdpc* is GDP per capita; *enit* denotes technology which is proxied by energy intensity to capture technology damaging effect on the environment.  $\alpha_i$  represents country-specific effect that is constat with time, and a time-specific effect  $d_t$  to account for time-varying stochastic shocks that are common to all countries (including deterministics such as intercepts or seasonal dummies). All the variables

in Equation (2) are expressed in natural logarithms so the estimated coefficients are interpreted as elasticities.

Further, the above setup is sufficiently general and renders a variety of panel data models as special cases. For example, if  $d_t = 1$ , the familiar fixed or random effects model is obtained. As the reader can notice, a very interesting information is obtained depending on the sign and statistical significance of the slope parameters of the income variable (*gdpc*). On the one hand, if  $\beta_1 > 0$  and  $\beta_2 = 0$ , then the relationship income–pollution is monotonically increasing (or decreasing if  $\beta_1 < 0$  and  $\beta_2 = 0$ ). On the other hand, if  $\beta_1 > 0$  and  $\beta_2 < 0$ , then an inverted U-shaped curve is observed for that relationship with the turning point given as  $E^* = \frac{-\beta_1}{2\beta_2}$ .

In this framework, standard panel data techniques can be used in order to obtain consistent estimates for the slope parameters. Nevertheless, as it was discussed previously, the empirical model (2) exhibits several weaknesses. As noted in Yatchew (1998), it is quite common that Economic Theory does not provide enough information regarding the functional form existing between the dependent variable and the covariates of the model. Therefore, assuming an ex-ante specific functional form and ignoring the possible parameters heterogeneity among countries can lead to misleading inference. Furthermore, the regression model (2) completely ignores the potential cross-sectional dependence.

In order to avoid the possible functional form misspecification problem in the above parametric framework, we propose an alternative approach that enables to relax the functional form assumptions and allows the data generating process to determine the true shape of the income–pollution relationship. Furthermore, with the aim of controlling for the potential cross-sectional dependence, we assume that it is due to the presence of a finite number of unobservable common factors (see Pesaran 2006 and Bai 2009, for example). Therefore, the semiparametric heterogeneous panel data model with cross-sectional dependence that we specify is as follows:

$$\ln(E_{it}) = \beta_{0i} + m_i(\ln gdpc_{it}) + \beta_{3i} \ln(pop_{it}) + \beta_{4i} \ln(enit_{it}) + \alpha'_i d_t + \gamma'_i f_t + \epsilon_{it}, \tag{3}$$

where  $m_i(\cdot)$  is an unknown smooth function to estimate,  $\ln gdpc = \ln(gdpc)$ ,  $f_t \equiv (f_{1t}, f_{2t}, \dots, f_{rt})'$  is a  $r \times 1$  vector of unobserved common effects,  $\gamma_i \equiv (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{ir})'$  is the associated vector of factor loadings, and  $\epsilon_{it}$  are the individual-specific (idiosyncratic) errors assumed to be independently distributed of  $(d_t, x_{it}, z_{it})$ . As the reader can notice, this specification suggests that the income–pollution relationship and the slope parameters can be different among individuals.

In order to obtain consistent estimates for model (3), we propose a new estimation procedure which combines nonparametric techniques with the Common Correlated Effects (CCE) approach proposed in Pesaran (2006). In the Appendix A, we present the proposed estimation procedure in detail.

### 3.2. Data Source

To investigate the empirical relationship between wealth–pollution, we used panel data sets consisting of 24 OECD countries and 32 non-OECD countries for the period 1980 to 2016. Countries with insufficient data on CO<sub>2</sub> emissions are dropped from the database. The detailed list of countries used for the estimation is collected in the Appendix C.

The data used in this study come from two main sources. Environmental degradation captured using the CO<sub>2</sub> emissions obtained from the International Energy Statistics of the U.S. Energy Information Administration (EIA)<sup>1</sup>. The CO<sub>2</sub> emissions (metric tones per capita) include burning of fossil fuels and cement manufacturing, but excludes emissions from land use such as deforestation.

---

<sup>1</sup> Available at <http://www.eia.gov/cfaps/ipdbproject/IEDIndex3.cfm>.



The data for all other variables (population, affluence, and technology) are obtained from the World Development Indicators (WDI) of the World Bank. Taking the study in [Effiong and Oriabije \(2018\)](#) as benchmark, population is measured as total population, while affluence which captures economic prosperity is measured as real GDP per capita (constant 2015 US dollars). For its part, technology is measured using energy intensity. In this literature, energy intensity is usually expressed as total primary energy consumption per dollar GDP (1000 Btu per year 2015 999 US dollars).

#### 4. Results

In order to begin our empirical analysis about the EKC in the OECD and non-OECD countries, we first examine if our panel variables contain cross-sectional dependence (CSD) using the CSD test proposed by [Pesaran \(2004\)](#) which follows a  $N(0, 1)$  distribution under the null hypothesis of cross-section independence. The results of this test, which are collected in [Table 1](#), indicate the null hypothesis of CSD is rejected at the 1% level of significance. Therefore, it is possible to point out that all the series of the database used in this study contains CSD. At the light of this result, we also implemented the [Pesaran \(2007\)](#) panel unit root test, which is robust to the presence of CSD. The results are reported in [Table 2](#) and indicate that all variables have a unit root in their levels and are stationary in their first differences. Hence, the unite root test results indicate the presence of CSD and non-stationary of the variables for both OECD and non-OECD countries.

**Table 1.** [Pesaran \(2004\)](#)'s test for cross-section correlation of the errors in the augmented Dickey–Fuller, ADF(p), regression.

Variable	OECD				Non-OECD			
	CD-Test	p-Value	Corr	abs (corr)	CD-test	p-Value	Corr	abs (corr)
co <sup>2</sup>	18.850 ***	0.000	0.187	0.449	43.329 ***	0.000	0.320	0.520
enit	64.327 ***	0.000	0.643	0.802	23.423 ***	0.000	0.197	0.452
pop	95.964 ***	0.000	0.949	0.949	133.84 ***	0.000	0.988	0.988
gdpc	95.096 ***	0.000	0.941	0.941	119.88 ***	0.000	0.886	0.886
gdpc <sup>2</sup>	92.918 ***	0.000	0.919	0.919	119.95 ***	0.000	0.887	0.887

Note: \*\*\* indicates statistical significance at the 1% level.

**Table 2.** [Pesaran \(2007\)](#)'s panel unit root tests accommodating cross-sectional dependence.

Variable	OECD				Non-OECD			
	Level		First diff.		Level		First diff.	
	CIPS	p-Value	CIPS	p-Value	CIPS	p-Value	CIPS	p-Value
co <sup>2</sup>	−2.720 **	0.040	−4.609 ***	0.000	−2.118 ***	0.000	−4.639 ***	0.000
enit	−2.806 **	0.021	−4.408 ***	0.000	−2.819 ***	0.000	−4.848 ***	0.000
pop	−2.777 **	0.029	−2.780 **	0.028	−6.005 ***	0.000	−5.112 ***	0.000
gdpc	−2.748 **	0.037	−3.822 ***	0.000	−1.929 ***	0.000	−3.589 ***	0.000
gdpc <sup>2</sup>	−3.377 ***	0.000	−3.571 ***	0.000	−2.189 ***	0.000	−3.086 ***	0.000

Note: \*\*\* and \*\* statistical significance at the 1% and 5% level.

Finally, we carry out the [Friedman \(1937\)](#); [Frees \(1995\)](#) and [Pesaran \(2004\)](#) tests for cross-sectional dependence. The results are reported in [Table 3](#) and in all the cases they reject the null hypothesis of cross-sectional independence. Therefore, the presence of CSD has been corroborated for this study and ignoring this fact can seriously damage our results.

**Table 3.** Pesaran (2004); Frees (1995) and Friedman (1937) cross sectional dependence tests.

Countries	Model	Pesaran Test		Frees Test		Friedman Test	
		CD Test	p-Value	CD (Q)	p-Value	CD	p-Value
OECD	RE model	14.237 ***	0.000	4.356 ***	0.000	164.69 ***	0.000
	FE model	14.132 ***	0.000	4.350 ***	0.000	164.22 ***	0.000
non-OECD	RE model	2.897 ***	0.004	3.913 ***	0.001	60.497 ***	0.001
	FE model	4.143 ***	0.000	3.965 ***	0.000	69.389 ***	0.000

Notes: RE and FE denote random and fixed effects models. \*\*\* indicates statistical significance at the 1% level.

In order to account for CSD, a fully parametric model with heterogeneous slope coefficients as the following one is going to be estimated,

$$\ln(co2_{it}) = \beta_{0i} + \beta_{1i}\ln(gdpc_{it}) + \beta_{2i}\ln(gdpc_{it}^2) + \beta_{3i}\ln(pop_{it}) + \beta_{4i}\ln(enit_{it}) + \alpha'_i d_t + \gamma'_i f_t + \epsilon_{it}, \quad (4)$$

where we are working with the natural logarithm of *co2*, *gdpc*, *pop*, and *enit*. Therefore, if we are interesting in calculating the income level at which the turning point occurs to corroborate the EKC hypothesis, it can be done using the following expression:  $gdpc^* = e^{-\frac{\hat{\beta}_1}{2\hat{\beta}_2}}$ .

Tables 4 and 5 illustrates the results of fully parametric models with heterogeneous slopes using the Common Correlated Estimator (CCE) proposed in Pesaran (2006) for OECD and non-OECD countries, respectively. As it has been discussed previously, each country has its own economic structure so it is expected that CO<sub>2</sub> emissions-income relationship may differ across countries. One of the advantages of the CCE estimator is that they facilitate testing the EKC hypothesis for individual countries so we are able to identify country-specific determinants of environmental degradation.

However, in this paper we are not willing to impose restrictive conditions about the functional form, so we propose to estimate an alternate semiparametric regression model with heterogeneous slope coefficients and CSD as the following

$$\ln(co2_{it}) = \beta_{0i} + m_i(\ln gdpc_{it}) + \beta_{3i}\ln(pop_{it}) + \beta_{4i}\ln(enit_{it}) + \alpha'_i d_t + \gamma'_i f_t + \epsilon_{it}. \quad (5)$$

We estimate this regression model using the alternative nonparametric estimator proposed in Appendix A, which enables us to combine nonparametric techniques and controlling for CSD, simultaneously.

Unlike the above fully parametric models which yields a unique coefficient estimator for the population parameter, one of the advantage of the non- and semi-parametric techniques is that they provide a regression plot that describes the true shape of the relationship between the dependent variable and the explanatory variable of interest while holding other regressors at a fixed point such as their means. Tables 6 and 7 illustrates the results for the slope coefficients of OECD and non-OECD countries, respectively, while Figures 1 and 2 presents the nonparametric estimates of the income-CO<sub>2</sub> emissions for the OECD countries and Figures 3 and 4 for the non-OECD countries.

**Table 4.** Parametric estimates for the OECD countries. Fully parametric model.

	cte	ln(enit)	ln(pop)	ln(gdpc)	ln(gdpc <sup>2</sup> )	EKC Holds	Turning Points (cte 2015 USD)
Australia	−12.527 (121.47)	0.481 (0.462)	−0.902 (2.382)	0.468 (0.862)	−0.019 (0.043)	NO	
Austria	186.134 (130.299)	0.840 *** (0.230)	1.261 (1.487)	0.207 (2.412)	−0.010 (0.125)	NO	
Belgium	−172.445 (107.917)	1.009 *** (0.213)	3.666 ** (1.785)	−3.232 * (1.790)	0.160 * (0.092)	NO	24,343
Canada	−1.229 (140.520)	−0.210 (0.252)	−8.469 *** (2.006)	1.762 (1.836)	−0.092 (0.089)	NO	
Chile	589.628 *** (180.748)	0.723 ** (0.324)	−0.269 (3.064)	−0.758 (1.163)	0.081 (0.070)	NO	
Denmark	675.823 *** (119.383)	1.761 *** (0.187)	−9.773 *** (2.423)	−5.434 *** (1.839)	0.286 *** (0.094)	NO	13,359
Finland	−327.650 (241.997)	1.055 *** (0.368)	−14.819 *** (4.757)	11.980 *** (2.871)	−0.581 *** (0.147)	YES	30,001
France	−347.018 (361.849)	−2.689 * (1.576)	−7.240 (11.131)	6.640 (9.308)	−0.328 (0.495)	NO	
Ireland	141.127 ** (58.307)	0.476 ** (0.207)	0.092 (0.409)	0.222 (0.855)	0.010 (0.041)	NO	
Italy	−544.884 (440.651)	−4.263 ** (1.899)	9.911 (11.729)	3.300 (7.809)	−0.163 (0.404)	NO	
Japan	256.529 * (147.414)	0.165 (0.287)	−1.233 (1.874)	−2.366 * (1.361)	0.114 * (0.066)	NO	32,112
Korea	52.178 (110.724)	1.086 *** (0.302)	0.223 (1.648)	1.692 ** (0.783)	−0.078 * (0.042)	YES	51,328
Luxembourg	−586.258 (406.912)	0.251 ** (0.107)	−1.559 (1.529)	5.030 (0.544)	−0.212 (0.033)	NO	
Mexico	335858 ** (127.078)	0.892 ** (0.333)	−3.399 *** (1.105)	0.542 (0.544)	−0.025 (0.033)	NO	
The Netherlands	−224.719 (138.904)	0.446 * (0.228)	−3.203 (3.476)	0.401 (2.886)	−0.019 (0.141)	NO	
New Zealand	355.141 *** (88.296)	0.318 * (0.160)	0.264 (0.696)	1.892 ** (0.809)	−0.091 ** (0.042)	YES	32,728
Norway	139.525 (440.084)	0.194 (0.632)	3.234 (4.205)	4.013 (6.519)	−0.195 (0.292)	NO	
Portugal	398.511 (320.103)	0.223 (0.302)	−1.767 (2.739)	−6.166 *** (1.360)	0.371 *** (0.076)	NO	4064
Spain	37.108 (168.705)	1.201 *** (0.223)	−1.287 (1.825)	−4.932 ** (2.057)	0.285 *** (0.104)	NO	5727
Sweden	−144.840 (199.627)	0.140 (0.176)	4.652 (3.066)	6.014 (3.631)	−0.285 (0.180)	NO	
Switzerland	7.599 (135.368)	0.560 ** (0.269)	−5.011 * (2.505)	−0.146 (2.046)	0.007 (0.098)	NO	
Turkey	−252.856 ** (115.558)	0.598 *** (0.077)	6.866 *** (0.984)	−2.688 *** (0.751)	0.179 *** (0.046)	NO	1822
United Kingdom	265.025 *** (46.884)	0.319 ** (0.133)	−0.075 (0.907)	−1.162 (1.007)	0.057 (0.048)	NO	
United States	158.253 *** (51.021)	0.567 *** (0.152)	0.893 (0.702)	−4.785 *** (0.828)	0.262 *** (0.043)	NO	9246

Note: \*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively.



**Table 5.** Parametric estimates for the non-OECD countries. Fully parametric model.

	cte	ln(enit)	ln(pop)	ln(gdpc)	ln(gdpc <sup>2</sup> )	EKC Holds	Turning Points (cte 2015 USD)
Argentina	34.066 (52.373)	−0.178 (0.337)	−0.652 (4.664)	1.675 ** (0.790)	−0.119 ** (0.054)	YES	1139
Bahamas	279.150 *** (81.493)	0.590 ** (0.265)	4.187 (6.680)	−0.373 (5.799)	0.018 (0.298)	NO	
Belarus	36.797 (23.936)	0.311 (0.300)	0.181 (4.381)	−1.080 * (0.573)	0.084 ** (0.036)	NO	618
Brazil	−161.371 ** (74.899)	0.744 ** (0.336)	22.657 *** (4.840)	0.513 (0.439)	−0.009 (0.032)	NO	
China	−27.347 (17.252)	0.733 *** (0.200)	−5.084 (3.809)	0.783 (0.872)	−0.028 (0.058)	NO	
Colombia	172.102 (116.556)	0.365 (0.257)	13.691 (9.669)	2.880 ** (1.355)	−0.186 * (0.093)	NO	
Costa Rica	−135.116 *** (20.728)	−0.112 *** (0.021)	2.854 *** (0.946)	0.043 (0.993)	0.017 (0.052)	NO	
Cyprus	−130.219 *** (31.464)	−0.129 (0.144)	−10.908 (13.741)	−1.930 (1.171)	0.140 (0.085)	NO	
Dominican Republic	28.406 (119.033)	0.321 (0.887)	−43.363 ** (18.838)	0.066 (3.103)	0.009 (0.225)	NO	
Ecuador	−3.044 (34.738)	0.380 * (0.196)	3.749 (7.211)	11.223 *** (2.603)	−0.740 *** (0.178)	YES	1964
El Salvador	14.612 (98.948)	0.036 (0.191)	2.229 (2.025)	2.025 (3.298)	−0.125 (0.201)	NO	
Gabon	−19.909 (36.395)	1.267 *** (0.290)	14.730 ** (5.443)	−0.565 (3.800)	0.036 (0.276)	NO	
Guatemala	41.916 * (22.727)	0.032 (0.186)	7.452 ** (2.734)	10.890 ** (4.497)	−0.886 ** (0.363)	YES	466
Honduras	15.240 (55.567)	0.174 (0.382)	−1.055 (1.915)	2.510 (2.087)	−0.120 (0.112)	NO	
Hong Kong of China	−127.346 *** (20.151)	0.123 (0.165)	0.161 (0.764)	−0.208 (1.456)	0.012 (0.082)	NO	
Israel	54.842 (42.606)	0.680 ** (0.260)	−30.795 *** (10.826)	−3.395 * (1.880)	0.275 * (0.140)	NO	479
Jamaica	−93.768 *** (34.106)	−0.120 (0.228)	−0.498 (2.847)	3.595 (2.236)	−0.188 (0.133)	NO	
Malaysia	−510.806 *** (72.052)	0.280 (0.367)	−15.069 *** (2.607)	1.129 (2.558)	−0.043 (0.257)	NO	
Nepal	−47.838 (52.107)	0.282 * (0.141)	−18.493 *** (4.298)	−0.884 (1.971)	0.107 (0.179)	NO	
Nicaragua	1051.098 * (583.419)	0.313 (0.323)	−39.380 (28.689)	−0.134 (1.075)	0.016 (0.080)	NO	
Nigeria	−52.200 *** (20.390)	0.187 (0.161)	4.516 * (2.391)	2.123 (1.339)	−0.186 (0.128)	NO	
Pakistan	122.608 (180.81)	−0.015 (0.097)	−16.178 (28.561)	2.923 (2.939)	−0.142 (0.165)	NO	
Panama	−178.660 *** (51.818)	0.147 (0.322)	−6.672 (5.616)	−2.268 *** (0.732)	0.133 ** (0.052)	NO	5044
Paraguay	40.414 (106.309)	0.777 * (0.396)	0.324 (11.100)	0.845 (1.625)	−0.090 (0.122)	NO	
Peru	61.526 (46.032)	0.362 (0.388)	−12.416 *** (11.100)	1.348 (0.925)	−0.089 (0.070)	NO	
Philippines	186.920 *** (49.095)	0.339 (0.257)	3.772 (2.323)	4.490 (5.178)	−0.349 (0.389)	NO	
Senegal	−88.332 (149.700)	−0.882 (1.477)	4.648 ** (2.076)	−4.300 (6.670)	0.229 (0.353)	NO	
Singapore	49.580 (42.931)	1.353 *** (0.443)	−6.855 *** (1.537)	1.163 (0.886)	−0.077 (0.056)	NO	
South Africa	44.601 (56.440)	0.507 * (0.290)	−14.216 *** (4.032)	−0.279 (1.492)	−0.028 (0.100)	NO	
Sri Lanka	−39.911 *** (14.340)	1.001 *** (0.150)	−0.831 (1.258)	0.948 (0.603)	−0.031 (0.041)	NO	
Thailand	205.845 *** (57.738)	−0.276 (0.276)	−18.279 (12.234)	0.705 (0.915)	−0.024 (0.068)	NO	
Uruguay	−2.234 (32.848)	0.566 ** (0.275)	2.641 (1.902)	−0.219 ** (0.095)	0.050 ** (0.018)	NO	8935

Note: \*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 6.** Parametric estimates for the OECD countries. Semiparametric model.

	<b>ln(enit)</b>	<b>ln(pop)</b>		<b>ln(enit)</b>	<b>ln(pop)</b>
Australia	1.303 *** (0.322)	4.605 *** (0.921)	Luxembourg	0.172 ** (0.063)	−0.320 *** (0.014)
Austria	0.924 *** (0.205)	−0.092 *** (0.025)	Mexico	0.043 (0.203)	−0.373 *** (0.030)
Belgium	1.132 *** (0.228)	0.073 (0.132)	The Netherlands	0.614 *** (0.179)	−2.198 *** (0.307)
Canada	−0.018 (0.339)	1.317 *** (0.467)	New Zealand	0.197 (0.173)	0.088 ** (0.034)
Chile	0.720 ** (0.353)	5.277 *** (0.467)	Norway	0.100 (0.535)	−0.074 (0.134)
Denmark	1.445 *** (0.295)	−0.441 *** (0.065)	Portugal	−0.452 * (0.248)	1.560 *** (0.185)
Finland	0.734 ** (0.292)	−0.033 (0.074)	Spain	1.492 *** (0.325)	−0.012 (0.064)
France	−3.371 ** (1.340)	−0.381 *** (0.037)	Sweden	0.227 (0.206)	0.531 *** (0.100)
Ireland	−0.453 * (0.240)	0.027 (0.044)	Switzerland	0.532 ** (0.240)	0.178 (0.106)
Italy	−5.387 ** (1.992)	−1.695 *** (0.529)	Turkey	0.578 *** (0.157)	−0.532 *** (0.062)
Japan	0.582 *** (0.121)	0.067 *** (0.003)	United Kingdom	0.194 * (0.102)	0.090 *** (0.004)
Korea	0.270 (0.207)	−0.287 *** (0.089)	United States	0.712 *** (0.238)	0.202 *** (0.035)

Note: \*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

**Table 7.** Parametric estimates for the non-OECD countries. Semiparametric model.

	<b>ln(enit)</b>	<b>ln(pop)</b>		<b>ln(enit)</b>	<b>ln(pop)</b>
Argentina	−0.023 (0.194)	0.780 *** (0.079)	Jamaica	−0.563 *** (0.210)	2.169 *** (0.147)
Bahamas	0.212 (0.271)	−0.409 *** (0.012)	Malaysia	−0.221 (0.328)	−5.886 *** (0.521)
Belarus	0.196 (0.268)	−0.023 (0.021)	Nepal	0.108 (0.192)	1.014 *** (0.043)
Brazil	0.609 *** (0.140)	−0.003 (0.042)	Nicaragua	−0.480 (0.365)	−0.655 *** (0.115)
China	1.300 *** (0.243)	0.267 *** (0.066)	Nigeria	0.161 (0.104)	−0.389 *** (0.008)
Colombia	0.544 ** (0.253)	0.166 *** (0.048)	Pakistan	−0.075 (0.091)	−0.058 *** (0.017)
Costa Rica	−0.098 *** (0.028)	−0.449 *** (0.005)	Panama	−0.201 (0.461)	0.989 *** (0.234)
Cyprus	−0.202 (0.157)	0.025 (0.083)	Paraguay	0.676 *** (0.312)	−0.269 ** (0.133)
Dominican Republic	0.176 (0.680)	−5.300 *** (1.710)	Peru	1.133 *** (0.275)	−0.334 *** (0.025)
Ecuador	0.075 (0.188)	1.023 *** (0.075)	Philippines	0.433 * (0.249)	4.315 *** (0.360)
El Salvador	0.122 (0.165)	−0.408 *** (0.039)	Senegal	−1.185 (1.066)	−2.545 ** (0.937)
Gabon	0.887 *** (0.278)	3.458 *** (1.208)	Singapore	−0.051 (0.494)	1.366 *** (0.440)
Guatemala	−0.004 (0.323)	1.253 *** (0.034)	South Africa	0.037 (0.259)	−3.551 *** (0.505)
Honduras	0.257 (0.226)	−1.465 *** (0.086)	Sri Lanka	0.128 (0.287)	0.107 ** (0.045)
Hong Kong of China	0.022 (0.151)	−1.909 *** (0.024)	Thailand	0.093 (0.291)	0.047 (0.030)
Israel	1.003 *** (0.275)	−0.184 *** (0.031)	Uruguay	0.408 (0.272)	1.180 *** (0.239)

Note: \*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

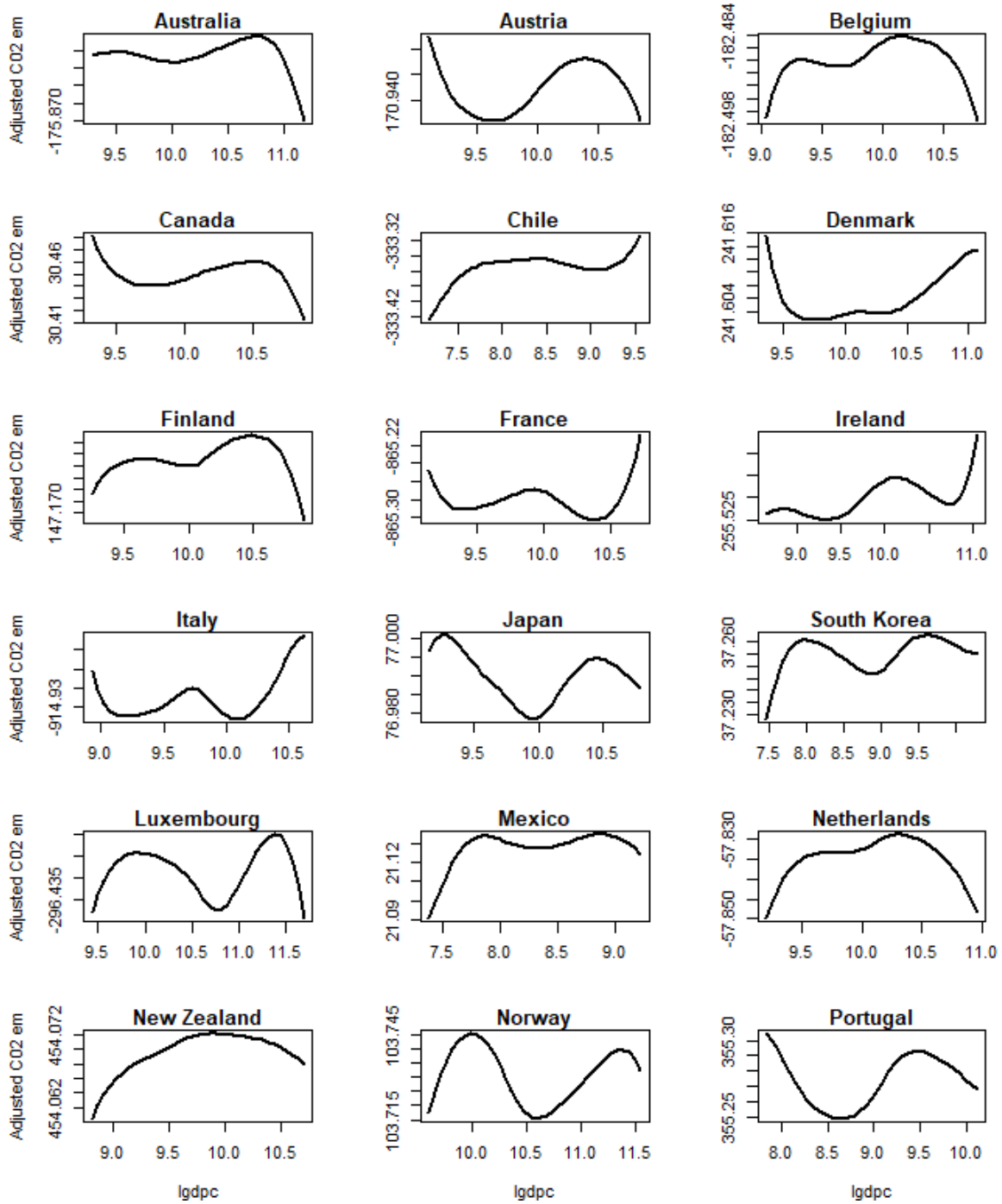


Figure 1. Nonparametric estimates of income-CO<sub>2</sub> emissions for the OCDE countries.

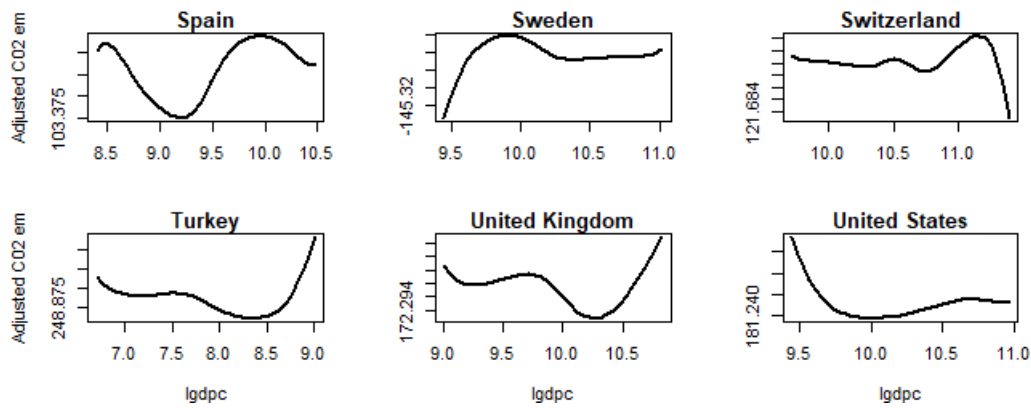


Figure 2. Nonparametric estimates of income-CO<sub>2</sub> emissions for the OCDE countries (Continued).

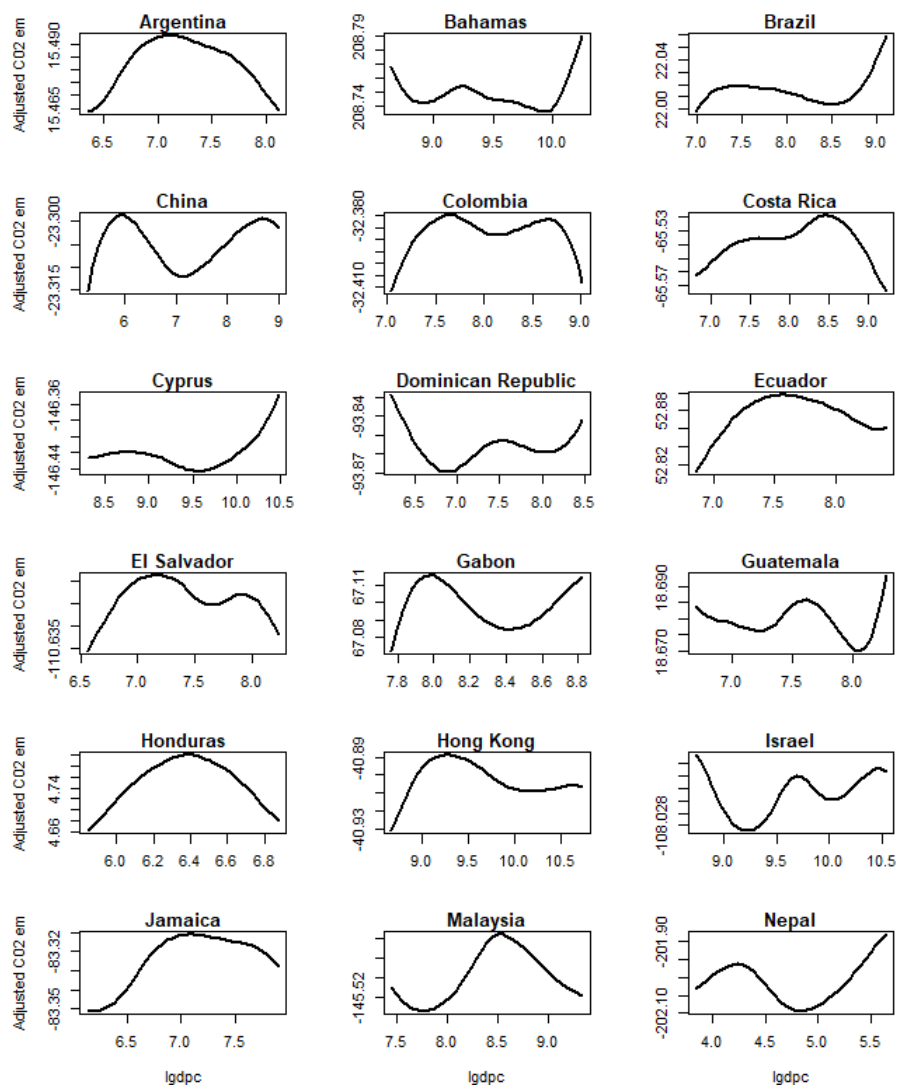
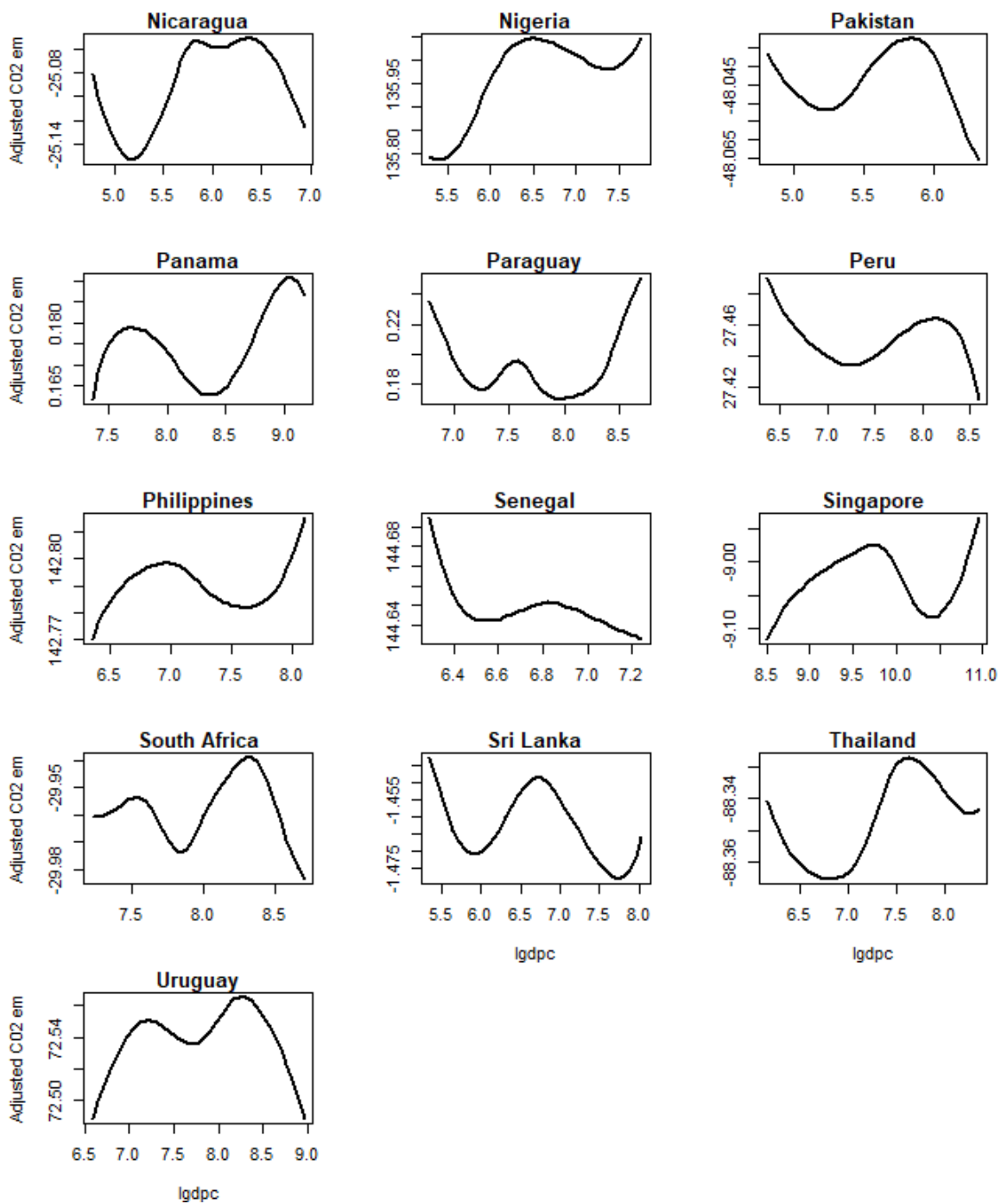


Figure 3. Nonparametric estimates of income-CO<sub>2</sub> emissions for the non-OCDE countries.



**Figure 4.** Nonparametric estimates of income-CO<sub>2</sub> emissions for the non-OECD countries (Continued).

## 5. Discussion

Analyzing the empirical estimates in Tables 4 and 5, we can point out that the energy intensity variable is positive and statistically significant in most of the OECD countries. This implies that higher consumption of fossil fuels in the production process will increase CO<sub>2</sub> emissions which in turn, will put further pressure on environmental quality. On its part, we obtain that this variable is not statistically significant in most of the non-OECD countries. For the population variable we do not get a definite conclusion. On the one hand, there are some OECD countries such as Canada, Denmark or Finland, for example, which exhibit a negative and significant effect, which implies that

higher population reduces the pressure on environmental quality. On the other hand, there are some non-OECD countries such as Brazil, Costa Rica or Guatemala, among others, that show a positive and significant effect, which means that higher population exacerbates pressure on environmental quality.

Focusing now on the EKC hypothesis, it is expected that the coefficient on GDP will be positive because of a scale effect, whereas the coefficient on GDP<sup>2</sup> will be negative since emissions fall via the composition effects in the latter stages of development. When we analyze the results of Tables 4 and 5 we find evidence that supports the EKC hypothesis only in six of the countries analyzed. In particular, the EKC hypothesis is supported in Finland, Korea, New Zealand (in the OECD database) and in Argentina, Ecuador, and Guatemala (in the non-OECD database). In particular, we find a peak turning point for Finland occurring at USD 30,001. Korea possesses a higher peak of USD 51,328 while New Zealand has a turning point at USD 32,728. On its part, Argentina exhibits a turning point at USD 1139, Ecuador has a peak of USD 1964, and Guatemala has a turning point of USD 466.

These findings are generally comparable to other empirical studies who find an EKC in the OECD. In particular, Galeotti (2007), Apergis et al. (2017), Churchill et al. (2018), and Dijkgraaf and Volleberg (2005), among others, find evidence supporting an inverted EKC relationship for some OECD countries, but not others. Our failure to find an EKC for Italy is consistent with Apergis (2016) and Churchill et al. (2018), but differs from Shabaz et al. (2017). Furthermore, Shabaz et al. (2017) find evidence of an EKC for Germany, which we do not. Other authors such as Churchill et al. (2018) try to examine the possibility of a N-shaped EKC for which they suggest to include the cube term of the per capita income. In this case, they obtain that there is a second turning point (i.e., the emissions begin to rise again when rich countries reach a second income tipping point<sup>2</sup>) in Australia, Canada and Japan, whereas an inverted N-shaped EKC was found for Denmark.

Despite what was obtained in Table 4, where only Finland, Korea, and New Zealand corroborated the EKC hypothesis, completely different conclusions are obtained if the plots in Figures 1 and 2 are analyzed. In the 24 OECD countries analyzed, the presence of a nonlinear relationship between the CO<sub>2</sub> emissions and the economic development is clearly corroborated. Further, there is evidence supporting the validity of the EKC hypothesis in the majority of these OECD countries. However, for Denmark, France, Ireland, Italy, Turkey, and the United Kingdom the EKC hypothesis is rejected. In these particular cases our findings are consistent with the scale, technological and composition effects of economic growth (see Dinda 2004). These countries are characterized by a large agricultural sector which only have a marginal, or no, impact on the environment. However, globalization has allowed them to carry out the transition process from the agriculture to the industrial sector very quickly, and probably that is where it comes from the exponential growth that can be seen in the CO<sub>2</sub> emissions as the income increases. Furthermore, for Japan, South Korea, Luxembourg, Norway and Spain a N-shaped EKC was found.

Regarding the non-OECD countries, the nonlinear relationship between pollution and income is also corroborated. Likewise, countries can be grouped into two categories. On the one hand, the countries that offer insight into the existence of the EKC hypothesis such as Argentina, Costa Rica, Ecuador, El Salvador, Honduras, Hong Kong, Jamaica, Malaysia, Nicaragua, Nigeria, Pakistan, and Peru. On the other hand, there is another group of countries which exhibit an exponential growth as the income per capita increases such as The Bahamas, Brazil, Cyprus, Republica Dominicana, Gabon, Guatemala, Israel, Nepal, and Paraguay. This result suggests that these countries are still in an intermediate stage of development where the agricultural sector remains dominant and the industrial sector is less sophisticated, so that economic growth will normally have a scale effect on

---

<sup>2</sup> According to the EKC, there is an initial increase in environmental pollution when a country's per capita income is low, but this falls over time as income per capita grows. However, it is possible that when income increases beyond a particular threshold (i.e., the second turning point), emissions begin to rise again, suggesting an N-shaped relationship between income and environmental quality.



the environment. Further, a N-shaped EKC was found for China, Colombia, Panama, South Africa, and Uruguay.

Analyzing the estimators obtained for the slope coefficients of the semiparametric model, we can point that the significance of the population and energy intensity variables has increased considerably for both OECD and non-OECD countries (see Tables 6 and 7), compared to the results for the fully parametric models (see Tables 5). Furthermore, in most of the countries the effect of the energy intensity is positive, so we can conclude that higher consumption of fossil fuels in the production process will increase CO<sub>2</sub> emissions impacting negatively on the environmental quality. On its part, the effect of the population variable is more ambiguous and it will be depend on the particularities of the country.

In summary, the empirical results show that the nature and validity of the income–pollution relationship based on the EKC hypothesis depends on the model assumptions on functional form specification, corroborating the results obtained in [Effiong and Oriabije \(2018\)](#). The fully parametric estimate, that avoid the possible bias related to the CSD, only offers insight into the existence of an inverted U-shaped EKC curve in six of the 50 analyzed countries. On the contrary, the semiparametric analysis with CSD provides a more definite shape of the income–pollution relationship with flexibility in functional form specification as a non-monotonically increasing or decreasing relationship for CO<sub>2</sub> emissions in all the countries, depending on the level of economic development of the country. This inconsistency reiterates the econometric caveats in the literature surrounding ex-ante restrictions on the functional form specification and robustness issues. Finally, allowing heterogeneous unknown smooth functions and slope parameters enables us to characterize the individual behaviour of each of the countries directly. This result is very useful in this literature given the heterogeneity that characterizes the economic structure of the countries.

The conclusions from this study can be very useful for the policy makers in order to decide where they have to focus their effort to achieve sustainable economic growth. Our results suggest that not only poor countries, but also richer countries, face environmental pollution. It implies that economic development is not a sufficient condition to reduce CO<sub>2</sub> emissions. In this situation, all countries, especially developed countries because of their important resources (financial, technological, etc.), should make an effort to reduce these emissions in order to reduce global warming.

## 6. Conclusions

CO<sub>2</sub> emissions play an important role in global warming and much effort is being pout into reducing such emissions all over the world. Consequently, studying the relationship between CO<sub>2</sub> emissions and economic development is important in order to be able to correctly advice policy makers on how best to make economic growth compatible with the environment. With this aim, a huge international literature has emerged trying to test of the validity of the EKC hypothesis although there is still a great lack of consensus about the income–pollution relationship.

A possible reason for this lack of consensus may be the presence of several specification errors in the models used to corroborate the EKC hypothesis. In this paper, we contribute to the literature revisiting the EKC hypothesis trough an alternative more flexible estimation technique which enables to account for functional form misspecification, cross-sectional dependence, and heterogeneous relationships among variables, simultaneously. The empirical results show that the nature and validity of the income–pollution relationship based on the EKC hypothesis crucially depends the model assumptions on functional form specification.

The semiparametric estimators proposed in this paper enables to provide a more definite shape of the income–pollution relationship non-monotonically increasing or decreasing for the OECD and non-OECD countries, depending on the level of economic development of the country. On the contrary, the fully parametric estimate, only offers insight into the existence of an inverted U-shaped EKC curve in six of the 50 analyzed countries. This inconsistency reiterates the econometric caveats in the literature surrounding ex-ante restrictions on the functional form specification and robustness issues.

**Author Contributions:** Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, A.S.; writing—original draft preparation, writing—review and editing, supervision, I.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors gratefully acknowledge financial support from the Programa Estatal de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del Sistema de I+D+i y del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad/Spanish Ministry of Science and Innovation (Ref. PID2019-105986GB-C22) for the partial support of this work.

**Acknowledgments:** The authors are thankful for the constructive suggestions provided by the members of the evaluation board of the final project of the Economics Degree of the University of Cantabria, which improved the quality of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Econometric Estimation. Semiparametric Approach

A semi-parametric model specifies the conditional mean of the dependent variable as two separate components, one parametric and one non-parametric. These types of models are very attractive from an empirical point of view due to their flexibility to balance precision and robustness. On the one hand, it allows us to incorporate prior information from economic theory or past experience while maintaining more flexibility in the specification of the model. On the other hand, although there is a nonparametric part that shows a slower convergence rate, the estimators obtained for the parametric part exhibit the same statistical properties as if the whole model were totally parametric. This is the so-called  $\sqrt{N}$ -consistency property (see [Robinson \(1988\)](#) and [Speckman \(1988\)](#), for example) for cross-sectional models. Finally, semiparametric models allow to provide a solution to the well-known “curse of dimensionality” of the fully nonparametric models. Several reviews on this topic exist and we suggest the interested reader consult [Ai and Li \(2008\)](#), [Henderson and Parmeter \(2015\)](#), [Parmeter and Racine \(2019\)](#), [Rodriguez-Poo and Soberon \(2017\)](#), and [Su and Ullah \(2011\)](#), among others.

A semi-parametric panel data model with heterogeneous slopes and unknown functions and cross-sectional dependence is given by

$$y_{it} = m_i(z_{it}) + x'_{it}\beta_i + \alpha'_i d_t + \gamma'_i f_t + \epsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{A1}$$

where  $y_{it}$  denotes the dependent variable (i.e., the environmental quality measure in Equation (5)),  $x_{it}$  and  $z_{it}$  are  $p \times 1$  and  $q \times 1$  vectors of the explanatory variables of interest (i.e., in the case of Equation (5),  $x_{it} = (\ln(pop_{it}), \ln(enit_{it}))$  and  $z_{it} = \ln(gdpc_{it})$ ), respectively,  $m_i(\cdot)$  is an unknown smooth function and  $\beta_i$  is a  $p \times 1$  vector of unknown population parameters.

The aim of the researchers is to obtain consistent estimators of  $m_i(\cdot)$  and  $\beta_i$  knowing that  $d_t \equiv (d_{1t}, d_{2t}, \dots, d_{Nt})'$  is a  $N \times 1$  vector of observed common effects (including deterministic regressors such as intercepts or seasonal dummies),  $\alpha_i$  is a  $N \times 1$  vector of unknown parameters,  $f_t \equiv (f_{1t}, f_{2t}, \dots, f_{rt})'$  is a  $r \times 1$  vector of unobserved common factors,  $\gamma_i \equiv (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{ir})'$  is the corresponding vector of the factor loadings, and  $\epsilon_{it}$  are the individual-specific (idiosyncratic) errors assumed to be independently distributed of  $(d_t, x_{it}, z_{it})$ . In general, however, the unobserved factors  $f_t$  could be correlated with  $(d_t, x_{it}, z_{it})$ , and to allow for such a possibility, we adopt the fairly general models for the individual specific regressors,

$$x_{it} = A'_i d_t + \Gamma'_{1i} f_t + v_{1it}, \tag{A2}$$

$$z_{it} = B'_i d_t + \Gamma'_{2i} f_t + v_{2it}, \tag{A3}$$

where  $A_i$ ,  $B_i$ ,  $\Gamma_{1i}$ , and  $\Gamma_{2i}$  are  $N \times p$ ,  $N \times q$ ,  $r \times p$ , and  $r \times q$ , factor loading matrices with fixed components, and  $v_{1it}$  and  $v_{2it}$  are the specific components of  $x_{it}$  and  $z_{it}$ , respectively, distributed independently of the common effects and across  $i$ , but assumed to follow general covariance stationary processes.

With the aim of obtaining consistent estimators for  $\beta_i$  and  $m_i(\cdot)$ , in the following we will show how, with modifications, the Common Correlated Effects (CCE) approach proposed in Pesaran (2006) can be applied to a semiparametric regression model.

Let  $0_p$  and  $0_q$  be matrices of zero of  $p \times p$  and  $q \times q$  dimension, respectively,  $I_p$  and  $I_q$  identity matrices of  $p \times p$  and  $q \times q$  dimension. If we combine (A1)–(A3) and rearrange terms, we can write

$$\begin{pmatrix} y_{it} \\ x_{it} \\ z_{it} \end{pmatrix} = C_i' d_t + D_i' f_t + \begin{pmatrix} m_i(z_{it}) + v_{1it} + \epsilon_{it} \\ v_{1it} \\ v_{2it} \end{pmatrix}, \tag{A4}$$

where  $C_i$  and  $D_i$  are matrices of  $N \times (1 + p + q)$  and  $r \times (1 + p + q)$  dimension, respectively, of the form

$$C_i = (\alpha_i, A_i, B_i) \begin{pmatrix} 1 & 0 & 0 \\ \beta_i & I_p & 0_p \\ 0_q & 0_q & I_q \end{pmatrix} \quad \text{and} \quad D_i = (\gamma_i, \Gamma_{1i}, \Gamma_{2i}) \begin{pmatrix} 1 & 0 & 0 \\ \beta_i & I_p & 0_p \\ 0_q & 0_q & I_q \end{pmatrix}.$$

In order to show that using suitable proxies for the unobserved factors is enough to avoid having to use initial estimates of  $\beta_i$ , we take the cross-sectional sample averages of (A4) obtaining

$$\begin{pmatrix} \bar{y}_{At} \\ \bar{x}_{At} \\ \bar{z}_{At} \end{pmatrix} = \bar{C}'_{At} d_t + \bar{D}'_{At} f_t + \begin{pmatrix} \bar{m}_{At} + \bar{v}_{1At} + \bar{\epsilon}_{At} \\ \bar{v}_{1At} \\ \bar{v}_{2At} \end{pmatrix}, \tag{A5}$$

where  $\bar{y}_{At} = N^{-1} \sum_{i=1}^N y_{it}$ ,  $\bar{x}_{At} = N^{-1} \sum_{i=1}^N x_{it}$ , and  $\bar{z}_{At} = N^{-1} \sum_{i=1}^N z_{it}$ . Furthermore,  $\bar{C}_{At}$ ,  $\bar{D}_{At}$ ,  $\bar{v}_{1At}$ , and  $\bar{v}_{2At}$  are the cross-sectional averages of  $C_{it}$ ,  $D_{it}$ ,  $v_{1it}$ , and  $v_{2it}$ , respectively, and  $\bar{m}_{At} = N^{-1} \sum_{i=1}^N m(z_{it})$ . Let  $\tilde{\Gamma} \equiv (\bar{\gamma}, \bar{\Gamma}_1, \bar{\Gamma}_2)$ , where  $\bar{\gamma} = E(\gamma_i)$ ,  $\bar{\Gamma}_1 = E(\Gamma_{1i})$ , and  $\bar{\Gamma}_2 = E(\Gamma_{2i})$ . Following Pesaran (2006), we can premultiply both sides of (A5) by  $\tilde{\Gamma}$  and solve for  $f_t$ ,

$$f_t = (\tilde{\Gamma}'\tilde{\Gamma})^{-1} \tilde{\Gamma}' \left[ \begin{pmatrix} \bar{y}_{At} \\ \bar{x}_{At} \\ \bar{z}_{At} \end{pmatrix} - \bar{C}'_{At} d_t - \begin{pmatrix} \bar{m}_{At} + \bar{v}_{1At} + \bar{\epsilon}_{At} \\ \bar{v}_{1At} \\ \bar{v}_{2At} \end{pmatrix} \right] \tag{A6}$$

provided that

$$\text{rank}(\tilde{\Gamma}) = r \leq (1 + p + q) \quad \text{for sufficiently large } N. \tag{A7}$$

As  $N \rightarrow \infty$ ,  $\bar{v}_{1At} \xrightarrow{p} 0$ ,  $\bar{v}_{2At} \xrightarrow{p} 0$ ,  $\bar{\epsilon}_{At} \xrightarrow{p} 0$ , and  $\bar{m}_{At} \xrightarrow{p} 0$  for each  $t$  under weak conditions. It follows,

$$f_t - (\tilde{\Gamma}'\tilde{\Gamma})^{-1} \tilde{\Gamma}' \left[ \begin{pmatrix} \bar{y}_{At} \\ \bar{x}_{At} \\ \bar{z}_{At} \end{pmatrix} - \bar{C}'_{At} d_t \right] \xrightarrow{p} 0 \quad \text{as } N \rightarrow \infty. \tag{A8}$$

The result of this last line suggests that we can use  $\lambda_t \equiv (\bar{y}_{At}, \bar{x}_{At}, \bar{z}_{At}, d_t)'$  as observable proxies of the unobservable factors,  $f_t$ . Therefore, we can conclude that effectively the Common Correlated Effect (CCE) approach proposed in Pesaran (2006) for fully parametric models can be applied in a semi-parametric setting with slight changes.

In this situation, we can estimate  $\beta_i$  and  $m_i(\cdot)$  by augmenting the semiparametric regression of  $y_{it}$  on  $x_{it}$  and  $z_{it}$  with  $h_t$  obtaining the following regression model

$$y_{it} = m(z_{it}) + x'_{it} \beta_i + \lambda'_i \delta_i + \epsilon_{it} + o_p(1), \tag{A9}$$

where  $o_p(1)$  captures possible approximation errors of the proxies. In addition,  $\lambda_t \equiv (\bar{y}_{At}, \bar{x}_{At}, \bar{z}_{At}, d_t)'$  is a  $\ell \times 1$  vector of proxies, where  $\ell = (1 + p + q + N)$ .

In order to get a  $\sqrt{T}$ -consistent estimator of  $\beta_i$ , we follow Robinson (1988) to eliminate the unknown functional  $m_i(\cdot)$ . Taking conditional expectations of (A9) yields

$$E(y_{it}|z_{it}) = m_i(z_{it}) + E(x'_{it}|z_{it})\beta_i + E(\lambda'_t|z_{it})\delta_i + o_p(1) \tag{A10}$$

and subtracting (A10) from (A9) yields

$$y_{it} - E(y_{it}|z_{it}) = [x_{it} - E(x_{it}|z_{it})]' \beta_i + [\lambda_t - E(\lambda_t|z_{it})]' \delta_i + \epsilon_{it} + o_p(1). \tag{A11}$$

In order to obtain feasible estimators for  $\beta_i$ , it is well-known that these conditional expectations are unknown and need to be estimated. With this aim, Robinson (1988) proposes to use (higher-order) Nadaraya-Watson kernel estimators. Later, Linton (1995) and Hamilton and Truong (1997), among others, pointed out that partial regression methods can be improved further by using local linear smoothers (see Fan and Gijbels (1996) to a deeper discussion about the desirable properties of these estimators). At the light of these results, we propose to use a local linear smoothers to estimate these conditional expectations.

Let  $\mu_1 = E(\mathcal{A}_{it}|z_{it} = z)$ , where ( $\mathcal{A}_{it} = y_{it}$ , or  $x_{it}$ , or  $\lambda_t$ ). For a given point  $z \in \mathbb{R}^q$  and for  $z_{it}$  in a neighbourhood of  $z$ , we propose to minimize the following weighted local linear least-squares (LLS) problem for  $\mu_1$ ,

$$\min \sum_{t=1}^T [\mathcal{A}_{it} - \mu_1 - (z_{it} - z)' \mu_2]^2 K_a(z_{it} - z), \tag{A12}$$

where  $K_a(\cdot) = a^{-q}K((z_{it} - z)/a)$  is a product kernel function such as  $K(u) = \prod_{l=1}^q k(u_l)$ ,  $u_l$  is the  $l$ th component of  $u$ , and  $a$  is a positive bandwidth term. Of course, a general diagonal or non-diagonal bandwidth matrix could be employed, but for the sake of simplicity, a single scalar bandwidth is used.

Using the resulting estimators for these conditional estimators in (A11) and writing the resulting expression in vectorial form yields

$$(I_T - S_i)Y_i = (I_T - S_i)X_i\beta_i + (I_T - S_i)\Lambda + \epsilon_i^* \tag{A13}$$

where  $Y_i \equiv (Y_{i1}, \dots, Y_{iT})'$  and  $\epsilon_i \equiv (\epsilon_{i1}^*, \dots, \epsilon_{iT}^*)'$  are  $T$ -dimensional vectors,  $X_i \equiv (X_{i1}, \dots, X_{iT})'$  and  $\Lambda \equiv (\lambda_1, \dots, \lambda_T)'$  are matrices of dimension  $T \times d$  and  $T \times \ell$ , respectively, and  $I_T$  is a  $T \times T$  diagonal matrix. Further, assuming that  $Z'_{z_i}K_a(z_{it})Z_{z_i}$  is invertible,  $S_i$  is a  $T \times T$  smoothing matrix associated to the individual  $i$  of the form

$$S_i \equiv \begin{pmatrix} e'_1(Z'_{z_i}K_a(z_{i1})Z_{z_i})^{-1}Z'_{z_i}K_a(z_{i1}) \\ \vdots \\ e'_1(Z'_{z_i}K_a(z_{iT})Z_{z_i})^{-1}Z'_{z_i}K_a(z_{iT}) \end{pmatrix}, \quad \text{for } Z_{z_i} = \begin{pmatrix} 1 & (z_{i1} - z) \\ \vdots & \vdots \\ 1 & (z_{iT} - z) \end{pmatrix},$$

where  $Z_{z_i}$  is a  $T \times (1 + q)$  matrix,  $e_1$  is a  $(1 + q) \times 1$  vector having 1 in the first entry and all other entries 0,  $K_a(z) = \text{diag}(K_a(z_{i1} - z), \dots, K_a(z_{iT} - z))$  is a  $T \times T$  diagonal matrix. Note that  $\epsilon_{it}^*$  is the new error term which consists of three elements: (i) original error term, (ii) approximation error of the proxies, (iii) approximation error of the Taylor expansion.

By the formula for partitioned regression, the estimator of  $\beta_i$  in (A13) is given by

$$\hat{\beta}_i = \left( \hat{X}'_i M_{\hat{\Lambda}} \hat{X}_i \right)^{-1} \hat{X}'_i M_{\hat{\Lambda}} \hat{Y}_i. \tag{A14}$$

where  $M_{\hat{\Lambda}} = I_T - \hat{\Lambda}'(\hat{\Lambda}\hat{\Lambda}')^{-1}\hat{\Lambda}$  is a projection matrix, for  $\hat{\Lambda} = (I_T - S_i)\Lambda$ ,  $\hat{X}_i = (I_T - S_i)X_i$ , and  $\hat{Y}_i = (I_T - S_i)Y_i$ .

Following a similar reasoning, the estimator of  $\delta_i$  in (A13) is given by

$$\hat{\delta}_i = \left(\hat{\Lambda}'M_{\hat{X}_i}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'M_{\hat{X}_i}\hat{Y}_i, \tag{A15}$$

where  $M_{\hat{X}_i} = I_T - \hat{X}_i'(\hat{X}_i\hat{X}_i')^{-1}\hat{X}_i$ .

Focusing now on the nonparametric estimation of the smooth unknown function  $m_i(\cdot)$ , we use the above estimator so the corresponding weighted local linear least-squares problem to minimize is of the following form

$$\sum_{t=1}^T \left( (y_{it} - x'_{it}\hat{\beta}_i - \lambda'_t\hat{\delta}_i) - m_i(z) - (z_{it} - z)'Dm_i(z) \right)^2 K_h(z_{it} - z), \tag{A16}$$

where  $K_h(\cdot)$  is a product kernel defined as in (A12) and  $h$  is the new bandwidth term.

Then, assuming that  $(Z'_{z_i}K_h(z_i)Z_{z_i})$  is invertible, the resulting CCE nonparametric estimator of  $m_i(\cdot)$  is given by

$$\hat{m}_i(z, h) = e'_1(Z'_{z_i}K_h(z_i)Z_{z_i})^{-1}Z'_{z_i}K_h(z_i)[Y_i - X_i\hat{\beta}_i - \Lambda\hat{\delta}_i], \tag{A17}$$

where  $K_h(z)$  is a  $T \times T$  diagonal matrix defined as  $K_a(z)$ , with  $h$  instead of  $a$ .

Under the conditions in Appendix B one can show that the semiparametric CCE estimator,  $\hat{\beta}_i$  is consistent and asymptotically normal as  $N$  and  $T$  tends to infinity. More precisely, following a similar proof scheme as in Pesaran (2006), the following result is obtained.

**Theorem A1.** Consider the panel data model (A1), and suppose that  $\|\beta_i\| < C$ ,  $\|\Gamma_{1i}\| < C$ ,  $\|\Gamma_{2i}\| < C$ , Assumptions A1–A3 and A4–A7 hold,  $(N, T) \rightarrow \infty$  (in no particular order), and the rank condition (A7) is satisfied. Then,  $\hat{\beta}_i$  and  $\hat{\delta}_i$  are consistent estimators of  $\beta_i$  and  $\delta_i$ , respectively. If it is further assumed that  $\sqrt{T}/N \rightarrow 0$  as  $(N, T) \rightarrow \infty$ , then

$$\begin{aligned} \sqrt{T}(\hat{\beta}_i - \beta_i) &\xrightarrow{d} \mathcal{N}(0, \Sigma_{\beta_i}), \quad \Sigma_{\beta_i} = \Sigma_{v_{1i}}^{-1}Y_{\beta_i}\Sigma_{v_{1i}}^{-1}, \\ \sqrt{T}(\hat{\delta}_i - \delta_i) &\xrightarrow{d} \mathcal{N}(0, \Sigma_{\delta_i}), \quad \Sigma_{\delta_i} = \Sigma_{\Lambda}^{-1}Y_{\beta_i}\Sigma_{\Lambda}^{-1}, \end{aligned}$$

where  $Y_{\beta_i} = \text{plim}_{T \rightarrow \infty} \left[ \frac{1}{T} \tilde{X}'_i M_{\tilde{G}} \Omega_{\epsilon_i} M_{\tilde{G}} \tilde{X}_i \right]$ ,  $Y_{\delta_i} = \text{plim}_{T \rightarrow \infty} \left[ \frac{1}{T} \tilde{\Lambda}' M_{\tilde{G}} \Omega_{\epsilon_i} M_{\tilde{G}} \tilde{\Lambda} \right]$ ,  $\Sigma_{\Lambda} = \text{Var}(\lambda_t)$  and  $\Omega_{\epsilon_i} = E(\epsilon_i \epsilon'_i)$  are covariance matrices. Furthermore,  $M_{\tilde{G}} = I_T - \tilde{G}(\tilde{G}'\tilde{G})^{-1}\tilde{G}'$  for  $\tilde{G} = (\tilde{D}, \tilde{F})$ , where  $\tilde{D}$  and  $\tilde{F}$  are vectors whose  $t$ th element are such as  $\tilde{d}_t = d_t - E(d_t|z_{it} = z)f_{z_{it}}(z)$  and  $\tilde{f}_t = f_t - E(f_t|z_{it} = z)f_{z_{it}}(z)$ .

Similarly, under the conditions in Appendix B one can show that the nonparametric CCE estimator,  $\hat{m}_i(z; h)$  is consistent and asymptotically normal as  $N$  and  $T$  tends to infinity.

**Theorem A2.** Consider the panel data model (A1), and suppose that Assumptions A1–A9 hold. Given the  $\sqrt{T}$ -consistency of  $\hat{\beta}_i$  and  $\hat{\delta}_i$ , as  $T \rightarrow \infty$ ,

$$\sqrt{Th^q} \left( \hat{m}_i(z; h) - m_i(z) - \frac{h^2}{2} \mu_2^q(K) \text{tr}\{\mathcal{H}_{m_i}(z)\} \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{R^q(K)\sigma_{\epsilon_i}^2}{f_{z_{it}}(z)} \right),$$

where  $\sigma_{\epsilon_i}^2 = E(\epsilon_{it}^2)$  and  $\mathcal{H}_{m_i}(\cdot)$  is the Hessian matrix of  $m_i(\cdot)$ .

That theorem is proved following a similar proof scheme as in Musolesi et al. (2020), so it is omitted. The detailed proof of the theorem can be provided upon request.

Finally note that the estimate of the variances of the above theorems can be used to construct standard errors for  $\hat{\beta}_i$  or confidence bands for  $\hat{m}_i(\cdot)$ . We use a standard multivariate kernel density estimator with a Epanechnikov kernel and the Silverman’s rule-of-thumb to chose the bandwidth.

**Appendix B. Assumptions**

In order to derive the asymptotic distribution of  $\hat{\beta}_i$  and  $\hat{m}_i(\cdot)$  obtained in Appendix A, the following notation is used. Denoting  $\tilde{X}_i = X_i - \mathcal{B}_X(z)$  and  $\tilde{\Lambda} = \Lambda - \mathcal{B}_\Lambda(z)$ , where  $\mathcal{B}_X(z) = E(X_i|z_{it} = z)f_{z_{it}}(z)$  and  $\mathcal{B}_\Lambda(z) = E(\Lambda|z_{it} = z)f_{z_{it}}(z)$ . Furthermore, the following conditions are required.

**Assumption A1. (Common Effects).** The  $(N + r) \times 1$  vector for common effects  $(d'_t, f'_t)$ , is covariance stationary with absolute summable autocovariances, distributed independently of the individual-specific errors  $\epsilon_{it}$ ,  $v_{1it}$ , and  $v_{2it}$  for all  $i$ ,  $t$ , and  $t'$ .

**Assumption A2. (Individual-Specific Errors).** The individual-specific errors  $\epsilon_{it}$ ,  $v_{1jt}$ , and  $v_{2jt}$  are distributed independently for all  $i$ ,  $j$ ,  $t$ , and  $t'$ . Furthermore, for each  $i$ ,  $\epsilon_{it}$ ,  $v_{1it}$ , and  $v_{2it}$  follow linear stationary processes with absolute summable autocovariances  $\epsilon_{it} = \sum_{\varphi=0}^{\infty} b_{i\varphi}\zeta_{i,t-\varphi}$ ,  $v_{1it} = \sum_{\varphi=0}^{\infty} S_{1i\varphi}v_{1i,t-\varphi}$ , and  $v_{2it} = \sum_{\varphi=0}^{\infty} S_{2i\varphi}v_{2i,t-\varphi}$ , where  $(\zeta_{it}, v'_{1it}, v'_{2it})'$  are  $(p + q + 1) \times 1$  vectors of identically, independently distributed (i.i.d.) random variables with mean zero, variance matrix  $I_{p+q+1}$ , and finite fourth-order cumulants. In particular,

$$\begin{aligned} \text{Var}(\epsilon_{it}) &= \sum_{\varphi=0}^{\infty} b_{i\varphi}^2 = \sigma_i^2 \leq \bar{\sigma}^2 < \infty, \\ \text{Var}(v_{1it}) &= \sum_{\varphi=0}^{\infty} S_{1i\varphi}S'_{1i\varphi} = \Sigma_{1i} \leq \bar{\Sigma}_1 < \infty, \\ \text{Var}(v_{2it}) &= \sum_{\varphi=0}^{\infty} S_{2i\varphi}S'_{2i\varphi} = \Sigma_{2i} \leq \bar{\Sigma}_2 < \infty, \end{aligned}$$

for all  $i$  and some constants  $\bar{\sigma}^2$ ,  $\bar{\Sigma}_1$ , and  $\bar{\Sigma}_2$ , where  $\sigma_i^2 > 0$  and  $\Sigma_{1i}$  and  $\Sigma_{2i}$  are positive definite matrices.

**Assumption A3. (Identification of  $\beta_i$ )** For each  $i$ ,  $\hat{\Psi}_{iA} = T^{-1}\tilde{X}'_i M_{\tilde{\Lambda}} \tilde{X}_i$ . and  $\hat{\Psi}_{iG} = T^{-1}\tilde{X}'_i M_{\tilde{G}} \tilde{X}_i$ . are nonsingular  $p \times p$  matrices and have finite second-order moments for all  $i$ . Furthermore,  $\sup_i E \left\| T^{-1}\tilde{X}'_i M_{\tilde{G}} \tilde{X}_i \right\|^2 < C < \infty$ .

**Assumption A4. (Density function).** The density of  $z_{it}$  satisfies  $0 < f_{z_{it}}(\cdot) < \infty$  and is twice continuously differentiable in all its arguments with bounded second-order derivatives at any point of its support.

**Assumption A5. (Smoothness condition).** Let  $\mathcal{Z} \in \mathbb{R}^q$  be the support of  $z_{it}$ . The unknown functions  $E(\lambda_t|z_{it})$ ,  $E(x_{it}|z_{it})$ , and  $E(y_{it}|z_{it})$  are bounded and twice continuously differentiable at  $z$  in the interior of  $\mathcal{Z}$  with second-order derivatives bounded.

**Assumption A6. (Kernel function).**  $K(u) = \prod_{l=1}^q k(u_l)$  is a product kernel, and the univariate kernel function  $k(\cdot)$  is compactly supported and bounded such that  $\int k(u)du = 1$ ,  $\int uu'k(u)du = \mu_2(K)I_q$ , and  $\int k^2(u)du = R(K)$ , where  $\mu_2(K) \neq 0$  and  $R(K) \neq 0$  are scalars and  $I_q$  is a  $q \times q$  identity matrix. All odd-order moments of  $k$  vanish, that is,  $\int u_1^{l_1}, \dots, u_q^{l_q} k(u)du = 0$ , for all non-negative integers  $l_1, \dots, l_q$  such that their sum is odd.

**Assumption A7. (Bandwidth).**  $a$  and  $h$  are positive bandwidths such that as  $T \rightarrow \infty$ ,  $a + (Ta^q)^{-1} \rightarrow 0$ ,  $Ta^4 \rightarrow 0$  and  $Ta^{2q}/\log(T) \rightarrow \infty$ . Furthermore, as  $T \rightarrow \infty$ ,  $h + (Th^q)^{-1} \rightarrow 0$ .



**Assumption A8.** The map  $m(\cdot) : \mathcal{Z} \rightarrow \mathbf{R}$  is twice continuously differentiable at  $z$  in the interior of  $\mathcal{Z}$  with second-order derivatives bounded.

**Assumption A9.** (Lyapounov). For some  $\varsigma > 0$ ,  $E[|\epsilon_{it}|^{(2+\varsigma)}]$  exists and is bounded.

### Appendix C. List of Countries

**Table A1.** List of OCDE countries.

Australia	Denmark	Japan	New Zealand	Switzerland
Austria	Finland	South Korea	Norway	Turkey
Belgium	France	Luxembourg	Portugal	United Kingdom
Canada	Ireland	Mexico	Spain	United States
Chile	Italy	The Netherlands	Sweden	

**Table A2.** List of non-OCDE countries.

Argentina	Cyprus	Hong Kong of China	Pakistan	South Africa
Bahamas, The	Dominican Republic	Israel	Panama	Sri Lanka
Belarus	Ecuador	Jamaica	Paraguay	Thailand
Brazil	El Salvador	Malaysia	Peru	Uruguay
China	Gabon	Nepal	Philippines	
Colombia	Guatemala	Nicaragua	Senegal	
Costa Rica	Honduras	Nigeria	Singapore	

### References

- Ai, Chunrong, and Qi Li. 2008. *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Chapter Semi-Parametric and Non-Parametric Methods in Panel Data Models. Berlin/Heidelberg: Springer-Verlag, pp. 451–78.
- Apergis, Nicholas 2016. Environmental Kuznets curves: New evidence on both panel and country-level CO<sub>2</sub> emissions. *Energy Economics* 54: 263–71. [[CrossRef](#)]
- Apergis, Nicholas, Christina Christou, and Rangan Gupta. 2017. Are there Environmental Kuznets Curves for US state-level CO<sub>2</sub> emissions? *Renewable and Sustainable Energy Reviews* 69: 551–58. [[CrossRef](#)]
- Azomahou, Théophile, François Laisney, and Phu Nguyen Van. 2006. Economic development and CO<sub>2</sub> emissions: A nonparametric panel approach. *Journal of Public Economics* 90: 1347–63. [[CrossRef](#)]
- Bai, Jushan. 2009. Panel data models with interactive fixed effects. *Econometrica* 77: 1229–79.
- Barra, Cristian, and Roberto Zotti. 2018. Investigating the non-linearity between national income and environmental pollution: International evidence of kuznets curve. *Environmental Economics and Policy Studies* 20: 179–210. [[CrossRef](#)]
- Beckerman, Wilfred 1992. Economic growth and the environment: Whose growth? Whose environment? *World Development* 20: 481–96. [[CrossRef](#)]
- Bertinelli, Luisito, and Eric Strobl. 2005. The environmental Kuznets curve semi-parametrically revisited. *Economics Letters* 88: 350–57. [[CrossRef](#)]
- Bertinelli, Luisito, Eric Strobl, and Benteng Zou. 2012. Sustainable economic development and the environment: Theory and evidence. *Energy Economics* 34: 1105–14. [[CrossRef](#)]
- Chen, Linna, and Shiyi Chen. 2015. The estimation of environmental Kuznets curve in China: Nonparametric panel approach. *Computational Economics* 46: 405–20. [[CrossRef](#)]
- Churchill, Sefa Awaworyi, John Inekwe, Kris Ivanovski, and Russell Smyth. 2018. The environmental Kuznets Curve in the OECD: 1870–2014. *Energy Economics* 75: 389–99. [[CrossRef](#)]
- Cole, Matthew A. 2004. Us environmental load displacement: examining consumption, regulations and the role of nafta. *Ecological Economics* 48: 439–50. [[CrossRef](#)]
- Dietz, Thomas, and Eugene A. Rosa. 1997. Effects of population and affluence on CO<sub>2</sub> emissions. *Proceedings of the National Academy of Sciences USA* 94: 175–79. [[CrossRef](#)] [[PubMed](#)]

- Dijkgraaf, Elbert, and Herman R.J. Volleberg. 2005. A test for parameter homogeneity in CO<sub>2</sub> emissions. *Environmental and Resource Economics* 32: 229–39. [\[CrossRef\]](#)
- Dinda, Soumyananda 2004. Environmental Kuznets curve hypothesis: A survey. *Ecology Economics* 49: 431–55. [\[CrossRef\]](#)
- Effiong, Ekpeno L., and Alex O. Oriabije. 2018. Let the data speak: Semiparametric evidence on the environmental Kuznets curve in africa. *Quality and Quantity* 52: 771–82. [\[CrossRef\]](#)
- Fan, Jinqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
- Frees, Edward W. 1995. Assessing cross-sectional correlation in panel data. *Journal of Econometrics* 69: 393–414. [\[CrossRef\]](#)
- Friedman, Milton 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* 32: 675–701. [\[CrossRef\]](#)
- Galeotti, Marzio 2007. Economic growth and the quality of the environment: Taking stock. *Environment, Development and Sustainability* 9: 427–54. [\[CrossRef\]](#)
- Galeotti, Marzio, and Alessandro Lanza. 1999. Richer and cleaner? A study on carbon dioxide emissions in developing countries. *Energy Policy* 27: 565–73. [\[CrossRef\]](#)
- Roca, Jordi, Emilio Padilla, Mariona Farré, and Vittorio Galletto. 2001. Economic growth and atmospheric pollution in Spain: Discussion the environmental kuznets hypothesis. *Ecological Economics* 39: 85–99. [\[CrossRef\]](#)
- Galeotti, Marzio, Alessandro Lanza, and Francesco Pauli. 2006. Reassessing the environmental kuznets curve for CO<sub>2</sub> emission: A robustness exercise. *Ecological Economics* 57: 152–63. [\[CrossRef\]](#)
- Grossman, Gene M., and Alan B. Krueger. 1991. Environmental impacts of a North American free trade agreement. [\[CrossRef\]](#)
- Grossman, Gene M., and Alan B. Krueger. 1993. *The U.S-Mexico Free Trade Agreement*. Chapter Environmental Impacts of the North American Free Trade Agreement. Cambridge: MIT Press, pp. 13–56.
- Grossman, Gene M., and Alan B. Krueger. 1995. Economic growth and the environment. *The Quarterly Journal of Economics* 110: 353–77. [\[CrossRef\]](#)
- Hamilton, Scott A., and Young K. Truong. 1997. Local linear estimation in partly linear models. *Journal of Multivariate Analysis* 60: 1–19. [\[CrossRef\]](#)
- Harbaugh, William T., Arik Levingson, and David M. Wilson. 2002. Reexamining the empirical evidence for an environmental Kuznets curve. *Review of Economics and Statistics* 84: 541–51. [\[CrossRef\]](#)
- Heil, Mark T., and Thomas M. Selden. 2001. Carbon emissions and economic development: Future trajectories based on historical experience. *Environment and Development Economics* 6: 63–83. [\[CrossRef\]](#)
- Henderson, Daniel J., and Christopher F. Parmeter. 2015. *Applied Nonparametric Econometrics*. New York: Cambridge University Press.
- Holtz-Eakin, Douglas, and Thomas M. Selden. 1995. Stoking the fires? CO<sub>2</sub> emissions and economic growth. *Journal of Public Economics* 57: 85–101. [\[CrossRef\]](#)
- Kaika, Dimitra, and Efthimios Zervas. 2013a. The environmental Kuznets curve (EKC) theory—Part A: Concept, causes and the co2 emissions case. *Energy Policy* 62: 1392–402. [\[CrossRef\]](#)
- Kaika, Dimitra, and Efthimios Zervas. 2013b. The environmental Kuznets curve (EKC) theory—Part B: Critical issues. *Energy Policy* 62: 1403–11. [\[CrossRef\]](#)
- Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos. 2018. Greenhouse emissions and productivity growth. *Journal of Risk and Financial Management* 11: 38. [\[CrossRef\]](#)
- Kuznets, Simon. 1955. Economic growth and income inequality. *American Economic Review* 45: 1–28.
- Linton, Oliver. B. 1995. Second order approximation in the partially linear regression model. *Econometrica* 63: 1079–112. [\[CrossRef\]](#)
- List, John A., and Craig A. Gallet. 1999. The environmental Kuznets curve: Does one size fit all? *Ecological Economics* 31: 409–23. [\[CrossRef\]](#)
- Martinez-Zarzoso, Inmaculada, and Aurelia Bengochea-Morancho. 2004. Pooled mean group estimation of an environmental Kuznets curve for CO<sub>2</sub>. *Economics Letters* 82: 121–26. [\[CrossRef\]](#)
- Musolesi, Antonio, Juan M. Rodriguez-Poo, and Alexandra Soberon. 2020. A Semiparametric Panel Data Model With Common Factors and Spatial Dependence: Theory and Empirical Application to the Knowledge Production Function. Working paper, Department of Economics, University of Cantabria, Santander, Spain.

- Nutakor, Felix, Sylvestre Bizumuremyi, Jinke Li, and Wei Liu. 2020. Kuznets curve for CO<sub>2</sub> emissions exits for Rwanda? Evidence from bootstrapped rolling-window Granger causality test. *Sustainability* 12: 8636. [CrossRef]
- Panayotou, Theodore. 1993. Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development. Working Paper WP238 Technology and Employment Programme, International Labor Office, Geneva, Switzerland.
- Parajuli, Rajan, Omkar Joshi, and Tek Maraseni. 2019. Incorporating forest, agriculture and energy consumption in the framework of the Environmental Kuznets Curve: A dynamic panel data approach. *Sustainability* 11: 2688. [CrossRef]
- Parmeter, Christopher F., and Jeffrey S. Racine. 2019. *Panel Data Econometrics: Theory*. Chapter Nonparametric Estimation and Inference for Panel Data Models. Cambridge, MA: Academic Press, pp. 97–129.
- Pesaran, M. Hashem. 2004. General Diagnostic Tests for Cross Section Dependence in Panels. Cambridge Working Papers in Economics, University of Cambridge, Cambridge, UK.
- Pesaran, M. Hashem. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012. [CrossRef]
- Pesaran, M. Hashem. 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22: 265–312. [CrossRef]
- Purcel, Alexandra-Anca. 2020. New insights into the environmental kuznets curve hypothesis in developping and transition economies: A literature survey. *Environmental Economics and Policy Studies* 22: 585–631. [CrossRef]
- Robinson, Peter M. 1988. Root-n-consistent semiparametric regression. *Econometrica* 56: 931–54. [CrossRef]
- Rodriguez-Poo, Juan M., and Alexandra Soberon. 2017. Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys* 31: 923–60. [CrossRef]
- Selden, Thomas M., and Daqing Song. 1994. Environmental quality and development: Is there a Kuznets curve for air pollution emission? *Journal of Environmental Economics and Management* 27: 147–62. [CrossRef]
- Sengupta, Ramprasad. 1996. Co2 emission-income relationship: Policy approach for climate control. *Pacific and Asian Journal of Energy* 7: 207–29.
- Shabaz, Muhammad, Muhammad Shafiullah, Vassilios G. Papavassiliou, and Shawkat Hammoudh. 2017. The CO<sub>2</sub>-growth nexus revisited: A nonparametric analysis for the G7 economies over nearly two centuries. *Energy Economics* 65: 183–93. [CrossRef]
- Shafik, Nemat, and Sushenjit Bandyopadhyay. 1992. Econometric Growth and Environmental Quality: Time Series and Cross Country Evidence. World Development Report Working Paper WPS 904, The World Bank, Washington, DC, USA.
- Speckman, Paul. 1988. Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society: Series B* 50: 413–36. [CrossRef]
- Stern, David I. 1998. Progress on the environmental Kuznets curve? *Environment and Development Economics* 3: 173–96. [CrossRef]
- Stern, David I., and Michael S. Common. 2001. Is there an environmental Kuznets curve for shulphur? *Environmental Economics and Management* 41: 162–78. [CrossRef]
- Su, Liangjun, and Aman Ullah. 2011. Nonparametric and semiparametric panel econometric models: Estimation and testing. In *Handbook of Empirical Economics and Finance*. Edited by Aman Ullah and David E. A. Giles. New York: Taylor and Francis Group.
- Yatchew, Adonis. 1998. Nonparametric regression techniques in economics. *Journal of Economic Literature* 36: 669–721.
- York, Richard, Eugene A. Rosa, and Thomas Dietz. 2003. STIRPAT, IPAT and ImPACT: Analytic tools for unpacking the driving forces of environmental impacts. *Ecological Economics* 46: 351–65. [CrossRef]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).