SEEDS
Sustainability Environmental Economics and Dynamics Studies

Working Paper Series

Revisiting the literature on the dynamic Environmental Kuznets Curves using a latent structure approach

by

Saptorshee Kanto Chakraborty, Massimiliano Mazzanti

05/2021

SEEDS is an interuniversity research centre. It develops research and higher education projects in the fields of ecological and environmental economics, with a special focus on the role of policy and innovation. Main fields of action are environmental policy, economics of innovation, energy economics and policy, economic evaluation by stated preference techniques, waste management and policy, climate change and development.

The SEEDS Working Paper Series are indexed in RePEc and Google Scholar.
Papers can be downloaded free of charge from the following websites:
http://www.sustainability-seeds.org/.

Enquiries:info@sustainability-seeds.org

SEEDS Working Paper 05/2021
April 2021
By Saptorshee Kanto Chakraborty, Massimiliano Mazzanti

Revisiting the literature on the dynamic Environmental Kuznets Curves using a latent structure approach

Saptorshee Kanto Chakraborty

Paris School of Economics

Massimiliano Mazzanti

University of Ferrara, SEEDS

Abstract

Theories of the association between environmental degradation and economic growth are not new and are very important under current global conditions to understand and tackle challenges like decarbonisation and the circular economy among others. Countries must balance growth with environmental degradation, and in the extensive literature that deals with this association, applied economists have largely used the environmental Kuznets curve (EKC) setting, with different empirical methodologies in various data settings. However, under a panel data framework, researchers often assume that countries are similar when dealing with unobservable heterogeneity. The papers exploits one of the methodologies to unveil heterogeneity to determine groupings from the data. We consider the countries that account for nearly 80% of global carbon dioxide emissions and apply the EKC setting. Using a Classifier Lasso framework that applies latent group methodologies to address unobservable heterogeneity, we find for two distinct groups substantial heterogeneity in types of energy consumption (renewable and total) with both positive and negative effects observed in the data. The results provide a new perspective on potential impacts illustrated in the EKC literature that might be relevant to policy makers.

Keywords: EKC, Group Lasso, Latent structure, Unobserved heterogeneity, decarbonisation

1. Introduction

The Rio Convention of 1992 followed by the Kyoto Summit of 1997 and the recent COP21 Summit of Paris have served as milestones for policymakers in reducing the extent of greenhouse gas (GHG) emissions and maintaining a sustainable future. However, the effect of GHGs on Earth might lead to an increase in temperatures of 3 °C by 2050 (United Nations Climate Change Secretariat, 2015), leading to catastrophic climatic changes globally. This also might lead to a reduction in economic output in developed and developing countries , and annual GDP growth might decrease by 2 to 4% by 2040 and by 10% by 2100. Therefore, a common global agenda is to reduce emission levels in both developed and developing countries without hampering economic progress. Some policies developed in this regard involve funding technology transfer to the developing world and increasing the use of renewable energies (IEA, 2018).

Sustainable development and human development interconnections have developed a fruitful broad view on development strategies. The re-appearance of the Kuznets legacy, with the rise of the Environmental Kuznets curve hypothesis, in the 90s, interconnected environmental sustainability, development and inequality issues. In 1995, Grossman and Krueger (Grossman and Krueger, 1995) extends the theory originally postulated by Simon Kuznets [1] and extents it to environmental degradation, assuming an~~d~~ inverted-U relationship between economic growth and the environment. In the Environmental Kuznets Curve (EKC) [2] , the initial increase in per capita GDP is due to the shift of the workforce from agriculture to industry and causes an increase in pollution. Assuming that after a certain threshold of per capita income all basic needs are being met, contextually to a new shift to the services sector, a demand for environmental quality emerges and develops so that the supply side of the economy must adapt by introducing cleaner technologies. Technology is central also in the theoretical work of Brock and Taylor (2010), which update Solow model of economic growth to incorporate technological progress in abatement. Within this framework, the relationship between climate change and economic development has been well studied (Grossman and Krueger, 1991, Grossman and Krueger, 1995, Holtz-Eakin and Selden, 1995, Carson, 2010), and some extensive reviews can be found in (Borghesi, 2000, Brock and Taylor, 2010, Uchiyama, 2016). Additionally, the relationship between growth and emissions, including innovation, has been extensively used in the policy-making literature especially for developed countries (e.g. the Stern Review (Dietz, 2011)). Researchers often assume that countries are similar when dealing with unobservable heterogeneity. From an econometric point of view , many methodologies have been developed from simple time-series methodologies to very complex ones using GAMS (Generalized Additive Models), Bayesian and heterogeneous estimators to tackle cross-sectional dependence (Musolesi et al., 2010, Mazzanti and Musolesi, 2013, Mazzanti and Musolesi, 2017). All of these works have shown the presence of strong forms of heterogeneity in developed countries [3] .

[1]The Kuznets curve describes the trend of inequality in relation to the rate of development, showing the evolution of income distribution over time Kuznets, 1955.

[2]New EKC applications are among others Massimiliano Mazzanti and Antonio Musolesi, 2014,Mazzanti and Musolesi, 2017,Wagner, 2015,Wagner et al., 2020

[3]Musolesi Antonio and Mazzanti Massimiliano, 2014 approach is the following: disentangling income and time-

This paper adopts a panel data model to account for this crucial heterogeneity. In our panel data model, cross-sectional units form a number of groups; within these groups, the slope coefficients are similar, but they vary across groups, and both the number of groups and individual group membership are unknown. This mode of determining the number of groups and group membership provides a new perspective on the EKC literature, with interesting insights for development and policy analyses. To analyse heterogeneity, we apply a recent classification method, the C-Lasso method, developed by (Su et al., 2016 [SSP (2016) hereafter] and Huang et al., 2020b [HJS (2020) hereafter]). The methodology is pretty novel, as it provides a consistent estimator for unknown group structures and delivers oracle-efficient estimates for the coefficients of each group. We use data from thirty-four countries in our sample for a period of 1971-2015 , and we conclude with two groups, revealing marked heterogeneity. Some diversified developmental patterns are evident from the data. Specifically, the effect of renewable energy consumption on gross domestic product is positive for one group and negative for the other.

The rest of the paper is organized as follows. Section 2 provides a brief literature review, Section 3 explains the panel structure model and C-lasso technique provided by SSP (2016), which are used account for latent group structure across different countries within the time period 1971-2015, and Section 4 describes the data and the model employed. Section 5 provides the results and Section 6 concludes.

2. Background and Literature Review

From the start of the 20^{th} century, there has been an increase in extreme weather-based damage, which is hypothesized to be led by global warming. These environmental issues can be broadly classified into two main categories: local issues, which relate to environmental pollution, and other issues, which are influenced by global warming and ozone depletion. Scientists and economists have agreed on the fact that unrestricted economic activities are a main cause of environmental destruction, one such being the mass consumption of fossil fuels. The *Limits to Growth* report commissioned by the Club of Rome illustrates the foundational interactions between economic activities and environmental issues.

The environmental Kuznets curve hypothesis depicts the relationship between economic growth and the environment. Briefly, it can be said that when one explores per-capita income and per-capita measures for any environmental variable, one might find an inverted U-shaped curve, which can be explained as follows. In early stages of development, environmental degradation increases but falls after per capita income exceeds a certain level (i.e., the turning point). The theory was first proposed by Simon Kuznets (Kuznets, 1955) to understand the relationship between per-capita national income and income inequality. With the introduction of the concept of sustainable development, the EKC literature has gained much momentum among researchers.

Grossman and Krueger (Grossman and Krueger, 1991) and Shafik and Bandyopadhyay (Shafik and Bandyopadhyay, 1992) , were some of the first to address the EKC in the literature. Many survey studies have been published in the EKC area, including (Stern, 1998, Andreoni and Levinson,

related effects (which are possibly heterogeneous across countries) in the study of greenhouse gas dynamics, while allowing for possible residual serial correlation at the same time, using Generalized Additive Mixed Models

2001 Stern, 2004, Dasgupta et al., 2002, Dinda, 2005) a recent one being (Shahbaz and Sinha, 2019). The EKC literature has been heavily debated among scholars for a long time as the availability of new datasets on various dimensions has increased over time. In addition, previous research has left some issues unresolved, which are being studied with new econometric techniques. While we do not explore such issues further here in the interest of conciseness, a very good review of issues (both theoretical and empirical) surrounding the EKC can be found in (Uchiyama, 2016 [Chapter 2]). Some recent literature on the subject of EKC worth mentioning are (Andre et al., 2019, Awaworyi Churchill et al., 2020 and Ik et al., 2019).

The present study proposes a somewhat novel approach to analyse EKC. The analysis does not only deal with time-varying coefficients, that may capture the instability of the EKC, but also uses an unknown latent group structure methodology to partition our sample of countries into groups, to focus on slope heterogeneity. A similar approach was taken by (Mazzanti and Musolesi, 2013), who classify groups first and then address heterogeneity and structural breaks. Li et al., 2016 also consider structural breaks and interactive fixed effects to address heterogeneity. However, both of these studies lack an unknown group structure.

3. Econometric Methodology

The traditional fixed-effects panel data model assumes cross-sectional units are heterogeneous in terms of time-varying intercepts with a homogeneous slope coefficient, but this assumption of slope homogeneity has been debated in the econometric literature. Heterogeneity is theoretically and empirically a cornerstone argument in the literature about growth, development, innovations (Azariadis and Drazen, 1990, Durlauf et al., 2001 for conceptual insights, Musolesi and Mazzanti, 2014; List and Gallet, 1999 on environment, policy development dynamics). As it is well known, panel data presents various alternatives to cope with the analysis of individual heterogeneity and omitted or unobserved effects. The key issue is always to specify a model that is able to account for behaviour heterogeneity across individuals and over time in a common pool of data. The various degrees of Poolability and the management of intrinsic trade-offs between different models are the main issue within panel data analysis. Key specifications are, in a linear additive parametric world: (i) intercept and slopes are constant, error terms capture heterogeneity; (ii) constant slopes but different intercept (one way model, with deterministic or stochastic fixed effects); (iii) constant slopes but different intercepts by individuals and by time (two ways models, again with deterministic or stochastic fixed effects); (iv) intercepts and slopes may vary by individuals; (v) intercepts and slopes may vary by time. Poolability is relaxed moving from i to v. Given that consistency is assured for large T*N dataset, the intrinsic trade off is between higher efficiency in poolable models that estimate a more limited set of parameters and lower efficiency but heterogeneity accounting in models that consider cross section and time related heterogeneity.

Against this background, to address heterogeneity, a traditional approach is to split the data into similar groups and apply standard fixed-effects model so that unobserved heterogeneity enters the model additively. This method has been repeatedly criticized in studies (Hsiao and Tahmiscioglu, 1997, Lee et al., 1997, Phillips and Sul, 2007, Su and Chen, 2013). Over the years, different approaches have emerged to address unknown group structure in inferencing unobserved slope heterogeneity. The first set of models used are finite mixture models . Sun, 2005 proposes a finite parametric linear mixture model, and~~;~~ Kasahara and Shimotsu, 2009and Browning and Carro, 2013 use nonparametric discrete mixture distributions to identify finite number of groups in discrete choice panel data. Another method in use is of cluster analysis using the K-means algorithm, and much progress has been made using this approach . (Lin and Ng, 2012, Sarafidis and Weber, 2015, Bonhomme and Manresa, 2015, Ando and Bai, 2016) use a K-means algorithm to deal with slope based heterogenenity. SSP(2016) used a variant form of the Lasso, the C-Lasso, to identify latent group patterns when slope coefficients exhibit a group structure. HJS (2020) extend this approach to cointegrated panels, and Huang et al., 2020a extended it to non-stationary panel data while dealing with cross-sectional dependence, which is very useful for tackling problems relevant to spillover-based studies. In our analysis, we do not assume the presence of any group structure in our data, but we deal with slope heterogeneity by applying new group structure concepts.

3.1. A. Model

We adopt the estimation technique of SSP (2016) for our empirical purpose and a brief explanation of the technique is given below takes the following form:

$$y_{it} = {\beta^{0}}'_{i} x_{it} + \phi_i + \tau_t + \varepsilon_{it},\ i = 1, ..., N, t = 1, ..., T \tag{1}$$

where i and t denotes country i and period t, respectively , y_{it} is a scalar, x_{it} is a $p \times 1$ vector of exogenous or predetermined variables, β_i is a $p \times 1$ vector of slope parameters, ϕ_i and τ_i are individual fixed and time effects ε_{it} is the idiosyncratic error term with mean zero. Following SSP (2016) we assume that β_i^0 is heterogeneous across groups and homogeneous within a group. A latent country specific group structure is imposed on the β_i^0

$$\beta_i^0 = \begin{cases} \alpha_1^0, \text{if } i \in G_1^0 \\ . \\ . \\ \alpha_{K_0}^0, \text{if } i \in G_{K_0}^0 \end{cases} \tag{2}$$

where, $\alpha_j^0 \neq \alpha_k^0$ for any j≠k, $\bigcup_{k=1}^{K_0} G_k^0 = 1, 2, ...N$ and $G_k^0 \bigcup G_j^0 = \emptyset$ for any j≠k, at this instance the number of groups K_0 is known and fixed is known and fixed but that each individuals group membership is unknown, and we calculate the number using an Information criterion as following SSP (2016) as described below. [4]

By allowing for the latent group structures for the long-run parameters, a right balance between parameter parsimony and model misspecification is expected to be achieved. It is important to note that the key parameters of interest in nonstationary panels are the coefficients of the nonstationary regressors as they characterize the long-run equilibrium relationship between the dependent variables and the nonstationary regressors. The estimation technique allows these parameters to be individual-specific, so an individual time-series regressions can estimate them but their estimators

[4]An important assumption is that individual group membership does not vary over time.

will have nonstandard limiting distributions and can converge to the true values only at the rate T. But if these coefficients are assumed to be common across all individuals,ti will have a convenient yet restrictive assumption that facilitates estimation and inference and meanwhile a very large chance of model mis-specification. The latent group structure adopted in this article is an inter-mediate approach. It allows for a certain degree of heterogeneity in the long-run parameters and helps to overcome some problems associated with nonstationary time series analysis too. In particular, under some conditions one can easily identify the group structure and estimate the group-specific long-run parameters at the rate of $\sqrt{NT}$, SSP (2016) also states that the long-run parameter estimators are asymptotically normal.

3.2. B. Methodology

After eliminating individual fixed and time effects from (1) following (Hsiao, 2003 [Chapter 3.6], Lu and Su, 2017, Wang et al., 2018), we obtain the following[5]

$$\tilde{y}_{it} = \beta_i' \tilde{x}_{it}' + \tilde{\tau}_t + \tilde{\varepsilon}_{it} \tag{3}$$

where $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}$, $\tilde{\tau}_t = \tau_t - \bar{\tau}$ and $\bar{\tau} = T^{-1}\sum_{t=1}^{T}\tau_t$. So we eliminate $\tilde{\tau}_t$ from (3)

$$\ddot{y}_{it} = \beta_i' \tilde{x}_{it} - \frac{1}{N}\sum_{j=1}^{N}\beta_j' \tilde{x}_{jt} + \ddot{\varepsilon}_{it} \tag{4}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}$, $\bar{y}_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} y_{it}$, $\bar{y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$ also $\ddot{\varepsilon}_{it}$, $\bar{\varepsilon}_{\cdot t}$and $\bar{\varepsilon}_t$ are similarly defined. We set the number of groups as K_0 and then calculate the number as described in the following section and can thus estimate $\beta \equiv (\beta_1^0, ..., \beta_N^0)$ and $\alpha_{K_0} \equiv (\alpha_1^0, ..., \alpha_{K_0}^0)$ by minimizing based SSP(2016).

$$Q_{2NT,\lambda}^{K_0}(\boldsymbol{\beta}, \boldsymbol{\alpha}_{k0}) = Q_{2,NT}(\boldsymbol{\beta}) + \frac{\lambda}{N}\sum_{i=1}^{N}\prod_{k=1}^{K_0} ||\beta_i - \alpha_k|| \tag{5}$$

SSP (2016) introduces the classifier Lasso (C-Lasso) framework by extending the group Lasso estimation technique proposed by Yuan and Lin, 2006 and propose to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha_{k0}}$ by minimizing equation (5).

where,

$$Q_{2NT}(\boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{y}_{it} - \beta_i' \tilde{x}_{it} + \frac{1}{N}\sum_{J=1}^{N}\beta_j' \tilde{x}_{jt}\right) \tag{6}$$

Equation (5) is a Penalized Profile Likelihood function (PPL), where λ=λ_{NT} the tuning parameter. The second term on the right hand side of equation (5) is the penalty term, which takes a

[5]For a better explanation please refer to SSP (2016) page 2220, they use a Gaussian quasi-maximum likelihood estimation (QMLE) technique they minimize β_i, ϕ_i and τ_t from eq (1) with $\psi(\omega_{it}, \beta_i, \phi_i, \tau_t) = \frac{1}{2}(y_{it} - \beta_i' x_{it} - \phi_i - \tau_t)^2$ and $\omega_{it} = (y_{it}, x_{it}')'$. Where $\psi(\omega_{it}, \beta_i, \phi_i, \tau_t)$ is assumed to be the logarithm of the pseudo-true conditional density function of y_{it} given x_{it}, the history of (y_{it}, x_{it}), and $(\beta_i, \phi_i, \tau_t)$.

novel mixed additive-multiplicative form. Traditional Lasso includes additive penalty terms to an objective function by differentiating zeros from non-zero-valued parameters to select relevant regressors, whereas the C-Lasso has N additive terms, each of which takes a multiplicative form as the product of K_0 separate penalties. The multiplicative component is needed because for each unknown i a priori to which point β_i should shrink and must allow β_i to shrink to any one of the K_0 unknown values $\alpha_1 \ldots \alpha_K$. Each of the K_0 penalty terms in the multiplicative expression permits β_i to shrink to a particular unknown group-specific parameter vector α_K . The summation component collate information from all N cross-sectional units in order to identify the group-specific parameters and the individual-specific parameters jointly. The tuning parameter is used to control the size of the penalty, a too small value of means that the penalty term would not play an important role so that many of β_is would not shrink toward one of the group-specific values in ($\alpha_1 \ldots \alpha_K$); a too large value of will force all β_is to shrink toward one of the group-specific values in ($\alpha_1 \ldots \alpha_K$), which will definitely result in misclassification. In theory, as put forward by SSP (2016), λ tends to zero at an appropriate rate as $(N, T) \to \infty$. A detailed assumptions and specifications can be found in SSP (2016) and HJS (2020). Then, post-Lasso estimates can be easily obtained by pooling all observations within each estimated group and then estimating the group-specific parameters for each group separately after individuals are demeaned over-time and across individuals. As a result, the standard error for each group-specific estimate can be determined.

3.3. C. The Information criteria

Tuning parameter λ is set following SSP (2016) as $\lambda = c\ s_Y^2\ T^{-1/3}$, where s_Y is the sample standard deviation of Y_{it} and c is constant. We use five different values of c (0.05, 0.10, 0.15, 0.20, 0.25) to examine the sensitivity of the results to c (thus λ). By assuming K is upper-bounded by K$_{MAX}$, we choose K by minimizing the following information criterion (IC)

$$IC(K, \lambda) = ln[(\hat{\sigma}^2_{(K,\lambda)}] + K_p + \frac{1}{\sqrt{NT}} \tag{7}$$

4. Empirics and data

In accordance with the previous literature, we apply a simple dynamic model~~,~~ to explain the relationship for the EKC model.

$$y_{it} = \alpha_i + \beta_{1i} y_{it-1} + \beta_{2i} x_{i,t} + \varepsilon_{it} \tag{8}$$

where y_{it} is denoted by the environmental quality indicator of i-th individual for t-th time period, x_{it} can be denoted by a vector of p$\times$1 explanatory variables, α_i is the fixed effect and ϵ_{it} is an idiosyncratic error term. We use carbon dioxide emission per-capita as our environmental quality indicator while we use per-capita gross domestic product and per-capita renewable energy consumption as our explanatory variables.

$$co_{i,t} = \alpha_i + \beta_{1i} lco_{i,t-1} + \beta_{2i} gdp_{i,t} + \beta_{3i} ren_{i,t} + \varepsilon_{i,t} \tag{9}$$

Where CO stands for log of per-capita carbon dioxide emission in tonnes, GDP is the log of per-capita gross domestic product in PPP terms in constant 2005 United States billion dollars, REN is per-capita renewable energy consumption in thousand tonnes of oil equivalent. Table 2 reports the descriptive statistics. We use annual data for a list of countries (see Table 1 for details) for a period of 1971-2015.

Table 1: List of countries in our sample

Argentina	Greece	Norway
Australia	India	Portugal
Austria	Indonesia	Singapore
Belgium	Ireland	South Africa
Brazil	Israel	Spain
Canada	Italy	Sweden
Chile	Japan	Switzerland
China	Korea	Turkey
Denmark	Malaysia	United Kingdom
Finland	Mexico	United States
France	Netherlands	
Germany	New Zealand	

We also use per-capita total primary energy consumption in million tonnes of oil equivalent represented by TP instead of per-capita renewable energy consumption to determine the difference in between the two and (9) can be re-written with the following form:

$$co_{i,t} = \alpha_i + \beta_{1i} lco_{i,t-1} + \beta_{2i} gdp_{i,t} + \beta_{3i} tp_{i,t} + \varepsilon_{i,t} \tag{10}$$

Data for carbon dioxide emissions, GDP in 2005 PPP USD billions and TP in millions of tonnes of oil equivalent for population measured in millions were acquired from IEA (IEA, 2017) to convert data in per-capita terms, and renewable energy consumption in millions tonnes of oil equivalent was collected from OECD (OECD, 2018). Figure 1, 2 and 3 provides a more comprehensive account of the nexus in between our variables.

Table 2: Descriptive Statistics

Stats	CO	GDP	REN	TP
Mean	1.703	2.944	12.098	0.921
Median	1.906	3.168	12.142	1.084
S.D.	0.822	0.830	1.496	0.758
Skewness	-1.164	-1.604	-1.003	-0.806
Kurtosis	1.382	3.155	1.828	0.101
Minimum	3.096	4.363	14.915	2.139
Maximum	-1.543	-0.730	5.547	-1.317

It is important to mention as in figure 2 the relationship between CO and REN has been flat, while in figure 3 the relationship between CO and TP has been steep upward rising. Over the years, there has been an increase in energy efficiency along with some structural break events like oil-price shocks, events of war this can be a possible explanation of underlying heterogeneity (Mazzanti and Musolesi, 2013, Mazzanti and Musolesi, 2017, Musolesi and Mazzanti, 2014), while renewable energy is a pretty recent concept. Some countries were faster to adopt renewable energy than the others which marks quiet a distinctive feature to explain the possible existence if heterogenity.

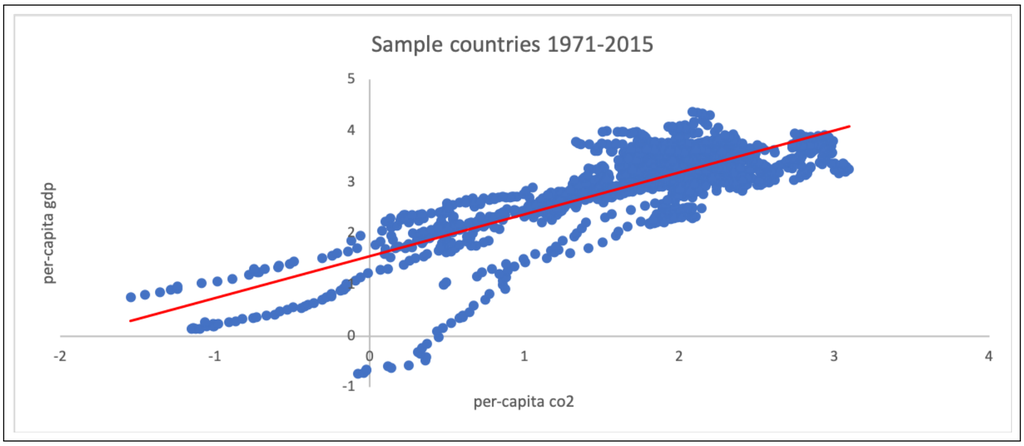

Figure 1: CO2-GDP nexus: 1971-2015 in logarithmic scale

As stated earlier in section 2, existing literature on environmental kuznets curve have always treated slope heterogeneity as to be of known in nature, this was mostly due to unavailability of estimation techniques which can deal otherwise. SSP (2016) is of innovative in this regard, they put forward a new method for econometric estimation and inference in panel models when the regression parameters are heterogenous across groups, individual group membership is unknown, and classification is to be determined empirically. Using an automated data-determined procedure, which do not require the specification of any modeling mechanism for the unknown group structure. We adopt the technique to deal with the unknown extent of slope heterogeneity among other things like cross-sectional dependence, nonstationarity, and existence of combination of both. Given the background of controversy in the kuznets curve literature, this paper argues that the versatility of the panel structure model in accommodating heterogeneity in behavior by means of data-determined grouping offers a new look at this long-standing issue.

5. Results

5.1. A. Cross-sectional dependence

To understand the nature of cross-section of our variables, we mainly follow (Pesaran, 2004, Pesaran, 2015) . The results are shown in Table 3. The null hypothesis is strict cross-sectional independence, which is rejected for all of the concerned variables. The results in Table 3 indicates that panel data have cross-sectional dependence.

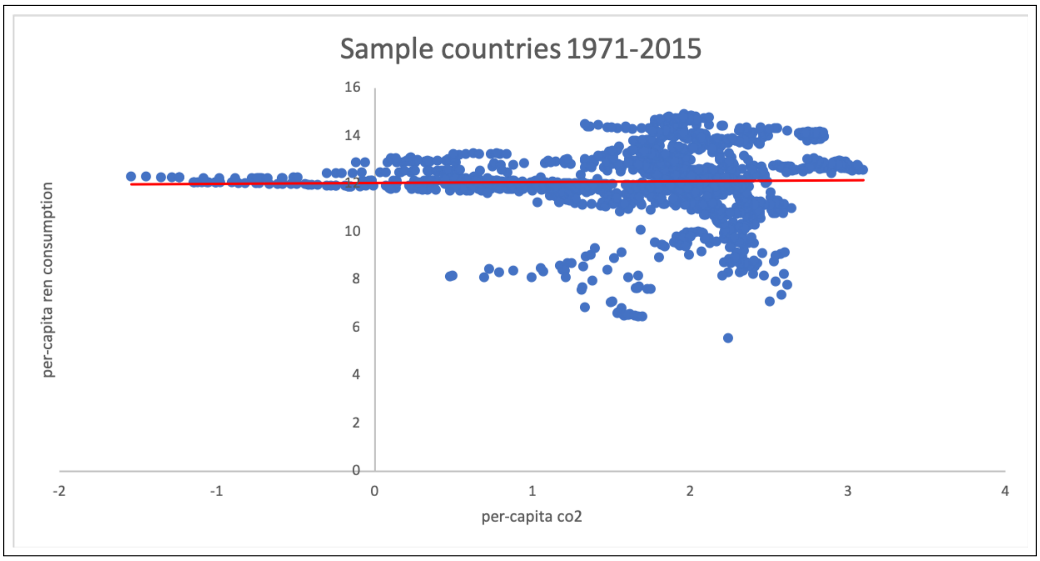

Figure 2: CO2-REN nexus: 1971-2015 in logarithmic scale

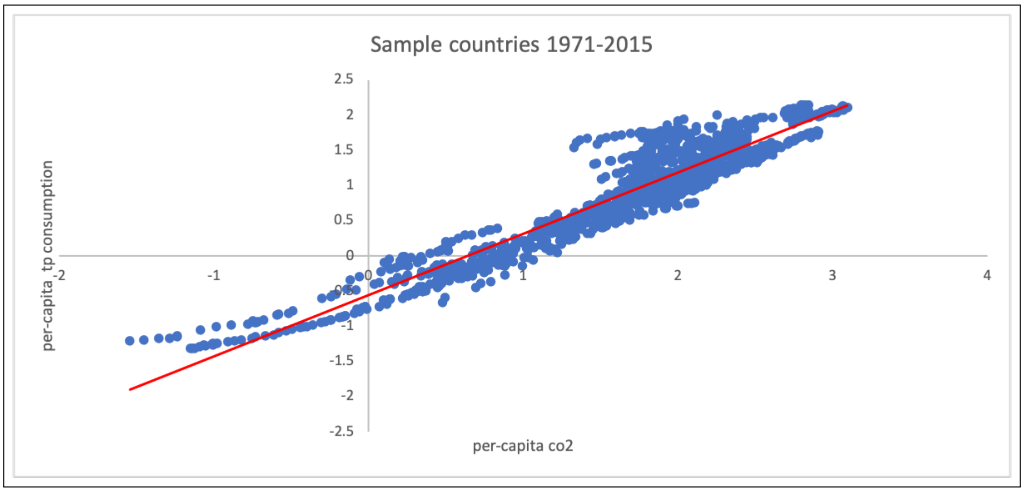

Figure 3: CO2-TP nexus: 1971-2015 in logarithmic scale

Table 3: CD Results- I

Var.	CD-test	p-value	mean ρ	mean abs (ρ)
CO	32.519	0.000	0.20	0.62
GDP	144.463	0.000	0.91	0.91
REN	62.172	0.000	0.39	0.68
TP	91.101	0.000	0.57	0.70

5.2. B. Unit root tests

Table 4 reports the outcomes of three first-generation unit root tests with cross-sectionally demeaned data based on (Im et al., 2003) along with first-difference of each variable~~,~~ . W~~w~~e use BIC lags for each case. The results show the variables are generated by unit root stochastic processes and their first differences are of stationary in nature.

Table 4: First generation Panel Unit root tests- I

Variable	Statistic	p-value
CO	1.472	0.929
GDP	6.344	1.00
REN	3.585	0.998
TP	3.023	0.998
DCO	-33.614	0.000
DGDP	-20.758	0.000
DREN	-37.490	0.000
DTP	-37.253	0.000

To understand the non-stationarity among our variables as in Table 4, we use a much more evolved unit root test using a multi-factor error structure framework, we then follow (Pesaran, 2007 and Pesaran et al., 2013) and as in the previous case, the variables become stationary at first difference. The results are presented in Table 5. We choose 3 lags following the literature that sets lags equal to an integer of $T^{1/3}$ (in our case T = 45), and so the lag length becomes 3.

5.3. C. Cointegration

After checking for stationarity, we focus on cointegration relationships among our variables, and we follow Westerlund, 2007, Persyn and Westerlund, 2008 to understand the order of integration among our variables. Table 6 represents the cointegration relationships among CO, GDP and REN and Table 7 represents cointegration relationship between CO, GDP and TP. We choose 3 lags, 3 leads and 1000 bootstrap replications.

We conclude with presence of cointegration among our variables for both the cases. The mean-group test (G_τ) averages heterogeneous OLS estimates of the speed of adjustment of their standard errors ~~,~~ while the panel test (P_τ) provides estimates of the aggregate speed of adjustment and its standard error.

Table 5: Second generation Panel Unit root tests- II

Variable	t-bar	Z[t-bar]	p-value
CO	-1.354	2.582	0.995
GDP	-1.903	-0.827	0.204
REN	-1.822	-0.322	0.374
TP	-1.713	0.352	0.638
DCO	-3.266	-9.281	0.000
DGDP	-2.778	-6.252	0.000
DREN	-3.319	-9.611	0.000
DTP	-3.058	-7.988	0.000

Table 6: Cointegration Results- co, gdp, ren

Statistic	Value	Z-value	P-value	Robust P-value
G_τ	-1.947	4.032	1.000	0.841
Ga	-5.236	6.678	1.000	1.000
P_τ	-9.563	4.186	1.000	0.655
Pa	-4.273	5.361	1.000	0.963

Table 7: Cointegration Results- co, gdp, tp

Statistic	Value	Z-value	P-value	Robust P-value
G_τ	-1.999	3.673	1.000	0.737
Ga	-6.480	5.689	1.000	0.993
P_τ	-10.095	3.596	1.000	0.516
Pa	-4.474	5.187	1.000	0.807

Because both G_τ and P_τ distributions assume error-correction models and are independently distributed, one can conclude that the tests take into consideration cross-sectional dependence based on bootstrapped standard errors.

5.4. D. Group Selection

Group selection is one of the most important criteria for this kind of estimation technique, and we select the number of groups following, Lin and Ng, 2012, SSP 2016. The exact number of groups is typically unknown, but finite integer K_{max} is assumed, which is considered to be an upper bound to the actually number of groups K_0. The tuning parameter is set as $\lambda = c_\lambda \times T^{-3/4}$ where c_λ takes five candidates 0.05, 0.10, 0.15, 0.20 and 0.25. We arbitrarily fix K_{max} at 7.

For each combination of number of groups and tuning parameter c_λ, we compute the information criterion value according to equation (7). The results are reported in table 8 and 9.

In both cases , we conclude with 2 latent groups following the majority rule, we find that the information criterion suggests two groups for both model, i.e. the minimal value for the I.C (Information Criteria), which tables 8 and 9 shown. Therefore, we set **K=2** and $\mathbf{c}_\lambda$**= 0.15** for equation 9 and **K=2** and $\mathbf{c}_\lambda$**= 0.25** for equation 10 in subsequent analyses.

Table 8: Number of Groups: Equation 9

	c = 0.05	c = 0.10	c = 0.15	c = 0.20	c = 0.25
K = 1	-1.565	-1.565	-1.565	-1.565	-1.565
K = 2	-1.643	-1.643	**-1.697**	-1.680	-1.680
K = 3	-1.636	-1.604	-1.616	-1.624	-1.624
K = 4	-1.575	-1.568	-1.577	-1.462	-1.600
K = 5	-1.486	-1.483	-1.517	-1.521	-1.490
K = 6	-1.471	-1.462	-1.436	-1.4483	-1.454
K = 7	-1.383	-1.394	-1.272	-1.287	-1.482

Table 9: Number of Groups: Equation 10

	c = 0.05	c = 0.10	c = 0.15	c = 0.20	c = 0.25
K = 1	-1.713	-1.713	-1.713	-1.713	-1.713
K = 2	-1.690	-1.690	-1.690	-1.714	**-1.798**
K = 3	-1.636	-1.649	-1.657	-1.674	-1.676
K = 4	-1.580	-1.580	-1.594	-1.600	-1.618
K = 5	-1.497	-1.529	-1.565	-1.565	-1.567
K = 6	-1.518	-1.508	-1.514	-1.514	-1.523
K = 7	-1.446	-1.457	-1.463	-1.463	-1.472

5.5. E. PLS estimation results

We now present the post-Lasso regression results for each group along with fixed effects for both equation (9) and equation (10) as presented in Table 10 and Table 12. The results shown

Table 10: PLS estimation results: Equation 9

Variables	Pooled FE	Group 1	Group2
GDP	0.040 ***	0.5442 ***	0.051**
	(0.030)	(0.0562)	(0.055)
REN	-0.016	0.007*	-0.033
	(0.019)	(0.016)	(0.0467)
LAGGED CO	0.988 ***	0.440***	0.942***
	(0.019)	(0.0464)	(0.0336)

[Values inside parenthesis indicates values for standard errors. Symbols ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.]

in Table 10 suggest that the estimate for the coefficient of GDP is always positive and significant for both the groups and the pooled regression. For lagged carbon dioxide values, the coefficients follow a similar pattern. However, for renewable energy consumption, , the pooled regression and Group 2 coefficients are without significance; for Group 1, the coefficient is positive with 10%

level of significance. Heterogeneity thus arises. Negative relationships between renewable energy and carbon dioxide do not emerge neither in the pooled nor separate regressions.

Table 11: GROUP: Equation 9

Group 1 membership = **21**	**Argentina**, **Australia**, **Austria**, **Brazil**, **Chile**, **China**, **Greece**, **India**, **Indonesia**, **Israel**, **Japan**, **Korea**, **Malaysia**, **Mexico**, **New Zealand**, **Norway**, **Portugal**, **Singapore**, **South Africa**, **Spain**, **Turkey**
Group 2 membership = **13**	**Belgium**, **Canada**, **Denmark**, **Finland**, **France**, **Germany**, **Ireland**, **Italy**, **Netherlands**, **Sweden**, **Switzerland**, **United Kingdom**, **United States**

The classification results based on equation (9) are shown in Table 11. Two groups are computed from the technique : the first group includes 21 members , and the second group includes 13 members. With the exception of some countries, most of the EU countries, along with United States and Canada, belong to Group 2; a group which includes mainly northern EU countries and industrial countries like Italy [6]

Table 12: PLS estimation results: Equation 10

Variables	Pooled FE	Group 1	Group 2
GDP	-0.157 (0.024)	-0.167*** (0.032)	-0.158 (0.055)
TP	0.252 *** (0.039)	0.0715*** (0.048)	-0.010 (0.0414)
LAGGED CO	0.872 *** (0.029)	0.427*** (0.043)	0.945*** (0.0502)

[Values inside parenthesis indicates values for standard errors. Symbols ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.]

The results shown in Table 12 suggest the estimates for the coefficient of GDP are always negative and only significant for Group 1. For total primary energy consumption, the estimates of coefficients are positive with 10% significance level for both Pooled FE and Group 1, but for Group 2 the coefficient value is not significant. The LAGGED CO values echo the results of equation (9) and are always positive and significant at 10% level.

The classification results based on equation (10) are shown in Table 13. Two groups are computed from the technique, and the first group same as before 26 members and the second group of 8 members. There is also a significant amount of change in membership from Table 11, in this

[6]Looking at Eco-innovation scoreboards, Italy and France are close to the Northern EU countries (https://ec.europa.eu/environment/ecoap/indicators/index$_en$).

Table 13: GROUP: Equation 10

Group 1 membership = **26**	**Argentina**, **Australia**, **Austria**, **Brazil**, **Chile**, **China**, **Denmark**, **Finland**,**Greece**,**India**, **Indonesia**, **Ireland**, **Israel**, **Italy**, **Japan**, **Korea**, **Malaysia**, **Mexico**, **Netherlands**, **New Zealand**,**Norway**, **Portugal**, **Singapore**, **Spain**, **Turkey**, **United States**
Group 2 membership = **8**	**Belgium**, **Canada**, **France**, **Germany**, **South Africa**, **Sweden**, **Switzerland**, **United Kingdom**

case South Africa moves to Group 2 from Group 1, but Denmark, Finland, Italy, Netherlands, USA move to Group 1 from Group 2 when compared between the two equations (9) and (10).

The exercise shows that when heterogeneity is a relevant factor to consider, specific analyses should be carried out specification by specification, given the potential effect of the selection of relevant covariates on the group-wise clustering of results. Methodologically speaking, results show that appropriate techniques can be exploited to unveil the unknown heterogeneity, the alternative being defining ex ante the groups (Massimiliano Mazzanti and Antonio Musolesi, 2014). The policy relevance is clear; grouping countries can be useful to assess diversified effects of economic and policy variables not considered in this paper, but food for thought for further research.

6. Conclusion

This paper revisits the EKC literature by applying a pretty novel econometric C-Lasso method to develop a data-determined approach to the classification of countries into common groups. A panel structure model is used to capture inherent heterogeneity across countries, and the C-Lasso mechanism determines group membership and estimates for each group. We find definitive group patterns and substantial heterogeneity in types of energy consumption (renewable and total). Total energy consumption effects are stronger than renewables, as it may be expected. Renewable energy consumption effects might be positive in the macroeconomic setting since the entire life cycle production is captured. All in all, renewable energy consumption is not strongly correlated with CO2 in the long run dynamics; our dataset captures the Kyoto protocol phase in the second part of the 1971-2015 panel. The results provide a new perspective on potential impacts illustrated in the EKC literature that might be relevant to policy makers, since group-wise heterogeneity / clustering may vary across specifications. Econometric Research should always take into account the model selection issue, namely testing what the preferred model is out of a general setting. Functional form and nonlinearities, Latent common factors and cross-sectional dependence, endogeneity of the environmental policy variable and of the knowledge inputs, Model uncertainty are key econometric issues to consider in model selection (Mazzanti and Musolesi, 2020). This papers tries to address some of those crucial aspects, with an application to a classic and relevant for environment development dynamics EKC modelling. A model that addresses all of those clashes with the curse of dimensionality and is not feasible. If heterogeneity is the key issue, a proper model selection and heterogeneity analysis might convey to policy makers more information on the similarities countries/regions show in any specific case. The inclusion of policy indicators in the panel data framework is a next step for this research.

References

Ando, Tomohiro and Jushan Bai (2016). “Panel Data Models with Grouped Factor Structure Under Unknown Group Membership:” in: *Journal of Applied Econometrics* 31.1, pp. 163–191.

Andreoni, James and Arik Levinson (2001). “The simple analytics of the environmental Kuznets curve”. In: *Journal of Public Economics* 80.2, pp. 269–286.

Andre, Bo Pieter Johannes, Andres Chamorro, Phoebe Spencer, Eric Koomen, and Harun Dogo (2019). “Revisiting the relation between economic growth and the environment; a global assessment of deforestation, pollution and carbon emission”. In: *Renewable and Sustainable Energy Reviews* 114, p. 109221.

Awaworyi Churchill, Sefa, John Inekwe, Kris Ivanovski, and Russell Smyth (2020). “The Environmental Kuznets Curve across Australian states and territories”. In: *Energy Economics* 90, p. 104869.

Azariadis, Costas and Allan Drazen (1990). “Threshold Externalities in Economic Development”. In: *The Quarterly Journal of Economics* 105.2, p. 501.

Bonhomme, Stphane and Elena Manresa (2015). “Grouped Patterns of Heterogeneity in Panel Data: Grouped Patterns of Heterogeneity”. In: *Econometrica* 83.3, pp. 1147–1184.

Borghesi, Simone (2000). “The Environmental Kuznets Curve: A Survey of the Literature”. In: *SSRN Electronic Journal*.

Brock, William A. and M. Scott Taylor (2010). “The Green Solow model”. In: *Journal of Economic Growth* 15.2, pp. 127–153.

Browning, Martin and Jesus M. Carro (2013). “The Identification of a Mixture of First-Order Binary Markov Chains*:” in: *Oxford Bulletin of Economics and Statistics* 75.3, pp. 455–459.

Carson, R. T. (2010). “The Environmental Kuznets Curve: Seeking Empirical Regularity and Theoretical Structure”. In: *Review of Environmental Economics and Policy* 4.1, pp. 3–23.

Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler (2002). “Confronting the Environmental Kuznets Curve”. In: *Journal of Economic Perspectives* 16.1, pp. 147–168.

Dietz, Simon (2011). “High impact, low probability? An empirical analysis of risk in the economics of climate change”. In: *Climatic Change* 108.3, pp. 519–541.

Dinda, Soumyananda (2005). “A theoretical basis for the environmental Kuznets curve”. In: *Ecological Economics* 53.3, pp. 403–413.

Durlauf, Steven N, Andros Kourtellos, and Artur Minkin (2001). “The local Solow growth model”. In: *European Economic Review* 45.4, pp. 928–940.

Grossman, G. M. and A. B. Krueger (1995). “Economic Growth and the Environment”. In: *The Quarterly Journal of Economics* 110.2, pp. 353–377.

Grossman, Gene and Alan Krueger (1991). *Environmental Impacts of a North American Free Trade Agreement*. w3914. Cambridge, MA: National Bureau of Economic Research.

Holtz-Eakin, Douglas and Thomas M. Selden (1995). “Stoking the fires? CO2 emissions and economic growth”. In: *Journal of Public Economics* 57.1, pp. 85–101.

Hsiao, Cheng (2003). *Analysis of panel data*. 2nd ed. Econometric Society monographs no. 34. Cambridge ; New York: Cambridge University Press.

Hsiao, Cheng and A. Kamil Tahmiscioglu (1997). “A Panel Analysis of Liquidity Constraints and Firm Investment”. In: *Journal of the American Statistical Association* 92.438, pp. 455–465.

Huang, Wenxin, Sainan Jin, and Liangjun Su (2020a). “IDENTIFYING LATENT GROUPED PATTERNS IN COINTEGRATED PANELS”. In: *Econometric Theory* 36.3, pp. 410–456.

Huang, Wenxin, Sainan Jin, Peter C.B. Phillips, and Liangjun Su (2020b). “Nonstationary panel models with latent group structures and cross-section dependence”. In: *Journal of Econometrics*, S0304407620302165.

IEA (2017). *CO2 emissions from fuel combustion 2017*. OCLC: 1047531056.

— (2018). *Renewables 2018: Analysis and Forecasts to 2023*. Market Report Series: Renewables. OECD.

Im, Kyung So, M.Hashem Pesaran, and Yongcheol Shin (2003). “Testing for unit roots in heterogeneous panels”. In: *Journal of Econometrics* 115.1, pp. 53–74.

Ik, Cem, Serdar Ongan, and Dilek zdemir (2019). “Testing the EKC hypothesis for ten US states: an application of heterogeneous panel estimation method”. In: *Environmental Science and Pollution Research* 26.11, pp. 10846–10853.

Kasahara, Hiroyuki and Katsumi Shimotsu (2009). “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices”. In: *Econometrica* 77.1, pp. 135–175.

Kuznets, Simon (1955). “Economic Growth and Income Inequality”. In: *The American Economic Review* 45.1, pp. 1–28.

Lee, Kevin, M. Hashem Pesaran, and Ron Smith (1997). “Growth and convergence in a multi-country empirical stochastic Solow model”. In: *Journal of Applied Econometrics* 12.4, pp. 357–392.

Li, Degui, Junhui Qian, and Liangjun Su (2016). “Panel Data Models With Interactive Fixed Effects and Multiple Structural Breaks”. In: *Journal of the American Statistical Association* 111.516, pp. 1804–1819.

Lin, Chang-Ching and Serena Ng (2012). “Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown”. In: *Journal of Econometric Methods* 1.1.

List, John A. and Craig A. Gallet (1999). “The environmental Kuznets curve: does one size fit all?” In: *Ecological Economics*. Ecological Economics 31.3, pp. 409–423.

Lu, Xun and Liangjun Su (2017). “Determining the number of groups in latent panel structures with an application to income and democracy: Number of groups in latent panel structures”. In: *Quantitative Economics* 8.3, pp. 729–760.

Massimiliano Mazzanti and Antonio Musolesi (2014). *Nonlinearity, heterogeneity and unobserved effects in the carbon dioxide emissions-economic development relation for advanced countries*. SEEDS, Sustainability Environmental Economics and Dynamics Studies.

Mazzanti, Massimiliano and Antonio Musolesi (2013). “The heterogeneity of carbon Kuznets curves for advanced countries: comparing homogeneous, heterogeneous and shrinkage/Bayesian estimators”. In: *Applied Economics* 45.27, pp. 3827–3842.

— (2017). “The effect of Rio Convention and other structural breaks on long-run economic development-CO2 relationships”. In: *Economia Politica* 34.3, pp. 389–405.

— (2020). *Modeling Green Knowledge Production and Environmental Policies with Semiparametric Panel Data Regression models*. SEEDS Working Papers 1420. SEEDS, Sustainability Environmental Economics and Dynamics Studies.

Musolesi Antonio and Mazzanti Massimiliano (2014). “Nonlinearity, heterogeneity and unobserved effects in the carbon dioxide emissions-economic development relation for advanced countries”. In: *Studies in Nonlinear Dynamics & Econometrics*. Studies in Nonlinear Dynamics & Econometrics 18.5, pp. 1–21.

Musolesi, Antonio and Massimiliano Mazzanti (2014). “Nonlinearity, heterogeneity and unobserved effects in the carbon dioxide emissions-economic development relation for advanced countries”. In: *Studies in Nonlinear Dynamics & Econometrics* 18.5.

Musolesi, Antonio, Massimiliano Mazzanti, and Roberto Zoboli (2010). “A panel data heterogeneous Bayesian estimation of environmental Kuznets curves for CO2 emissions”. In: *Applied Economics* 42.18, pp. 2275–2287.

OECD (2018). *Renewable energy*. OECD.

Persyn, Damiaan and Joakim Westerlund (2008). “Error-correctionbased cointegration tests for panel data”. In: *Stata Journal* 8.2, pp. 232–241.

Pesaran, M. Hashem (2004). *General Diagnostic Tests for Cross Section Dependence in Panels*. CESifo Group Munich.

— (2007). “A simple panel unit root test in the presence of cross-section dependence”. In: *Journal of Applied Econometrics* 22.2, pp. 265–312.

— (2015). “Testing Weak Cross-Sectional Dependence in Large Panels”. In: *Econometric Reviews* 34.6, pp. 1089–1117.

Pesaran, M. Hashem, L. Vanessa Smith, and Takashi Yamagata (2013). “Panel unit root tests in the presence of a multifactor error structure”. In: *Journal of Econometrics* 175.2, pp. 94–115.

Phillips, Peter C.B. and Donggyu Sul (2007). “Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence”. In: *Journal of Econometrics* 137.1, pp. 162–188.

Sarafidis, Vasilis and Neville Weber (2015). “A Partially Heterogeneous Framework for Analyzing Panel Data”. In: *Oxford Bulletin of Economics and Statistics* 77.2, pp. 274–296.

Shafik, Nemat and Sushenjit Bandyopadhyay (1992). *Economic growth and environmental quality : time series and cross-country evidence*. The World Bank.

Shahbaz, Muhammad and Avik Sinha (2019). “Environmental Kuznets curve for CO2 emissions: a literature survey”. In: *Journal of Economic Studies* 46.1, pp. 106–168.
Stern, David I (2004). “The Rise and Fall of the Environmental Kuznets Curve”. In: *World Development* 32.8, pp. 1419–1439.
Stern, David (1998). “Progress on the environmental Kuznets curve?” In: *Environment and Development Economics* 3.2, pp. 173–196.
Su, Liangjun and Qihui Chen (2013). “Testing Homogeneity in PanelData Models with Interactive Fixed Effects”. In: *Econometric Theory* 29.6, pp. 1079–1135.
Su, Liangjun, Zhentao Shi, and Peter C. B. Phillips (2016). “Identifying Latent Structures in Panel Data”. In: *Econometrica* 84.6, pp. 2215–2264.
Sun, Yixiao X (2005). *Estimation and Inference in Panel Structure Models*. Department of Economics, UC San Diego.
Uchiyama, Katsuhisa (2016). *Environmental Kuznets Curve Hypothesis and Carbon Dioxide Emissions*. SpringerBriefs in Economics. Tokyo: Springer Japan.
United Nations Climate Change Secretariat (2015). *Climate action now: summary for policymakers 2015*. OCLC: 938001701.
Wagner, Martin (2015). “The Environmental Kuznets Curve, Cointegration and Nonlinearity: THE ENVIRONMENTAL KUZNETS CURVE”. In: *Journal of Applied Econometrics* 30.6, pp. 948–967.
Wagner, Martin, Peter Grabarczyk, and Seung Hyun Hong (2020). “Fully modified OLS estimation and inference for seemingly unrelated cointegrating polynomial regressions and the environmental Kuznets curve for carbon dioxide emissions”. In: *Journal of Econometrics* 214.1, pp. 216–255.
Wang, Wuyi, Peter C. B. Phillips, and Liangjun Su (2018). “Homogeneity pursuit in panel data models: Theory and application”. In: *Journal of Applied Econometrics* 33.6, pp. 797–815.
Westerlund, Joakim (2007). “Testing for Error Correction in Panel Data”. In: *Oxford Bulletin of Economics and Statistics* 69.6, pp. 709–748.
Yuan, Ming and Yi Lin (2006). “Model selection and estimation in regression with grouped variables”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.1, pp. 49–67.