Carbon dioxide emissions and economic activities: A mean field variational Bayes semiparametric panel data model with random coefficients

Badi H. Baltagi[a,*], Georges Bresson[b] and Jean-Michel Etienne[c]

[a] *Department of Economics and Center for Policy Research 426 Eggers Hall, SyracuseUniversity, Syracuse, NY 13244-1020, USA*

[b] *Department of Economics, Université Paris II, 12 place du Panthéon 75 230 Paris Cedex 05, France.*

[c] *Department of Economics, Université Paris-Sud 11, 54 Boulevard Desgranges 92331 Sceaux, Cedex, France.*

April 14, 2019

Abstract - This paper proposes semiparametric estimation of the relationship between CO_2 emissions and economic activities for a panel of 81 countries observed over the period 1991-2015. The observed differentiated behaviors by country reveal strong heterogeneity as well as different trends across countries and years. This is the motivation behind using a mixed fixed- and random-coefficients panel data model to estimate this relationship. Following Lee and Wand (2016a), we apply a mean field variational Bayes approximation to estimate a log model with structural breaks between CO_2 emissions per capita and GDP per capita including control covariates such as energy intensity and use, energy consumption, population density, urbanization and trade. Results reveal a strong "CO_2 emissions - GDP elasticity", close to one, confirming the increasing but complex link between these two variables. The use of this methodology enriches the estimates of climate change models underlining a large diversity of responses across variables and countries.

Keywords: carbon dioxide emissions, energy intensity, environmental Kuznets curve, GDP, greenhouse gas emissions, mean field variational Bayes approximation, panel data, random coefficients, semiparametric model.

JEL codes: C11, C14, C23, Q53, Q54.

---

*Corresponding author. Tel.: + 1 315 443 1630; Fax: + 1 315 443 1081.

E-mail: bbaltagi@maxwell.syr.edu (B.H. Baltagi), georges.bresson@u-paris2.fr (G. Bresson), jean-michel.etienne@u-psud.fr (J-M. Etienne)

We are grateful to participants at the research seminar of Université Laval (Canada) for helpful comments and suggestions. We also thank the co-editors Stéphane Bonhomme, Laurent Davezies of the Special Issue of the *Annals of Economics and Statistics* on Panel Data and an anonymous referee for valuable comments and suggestions. The usual disclaimers apply.

1 Introduction

Air pollution is emerging as one of the main causes of deaths and serious ailments in the world. Greenhouse gases trap heat and make the planet warmer. As shown by a large majority of scientists (see IPCC (2007, 2013, 2014) for instance), human activities are responsible for almost all of the increase in greenhouse gases in the atmosphere over the last 150 years. The World Meteorological Organization (WMO)[1] confirms that the past 4 years were the warmest on record since 1880. The global average surface temperature in 2018 was approximately 1°C above the pre-industrial baseline (1850-1900). It ranks as the fourth warmest year on record. 2017 and 2018 have experienced many high-impact events including catastrophic hurricanes and floods, strong heatwaves and extreme drought. As a result of a powerful El Niño, 2016 is likely to remain the warmest year on record (1.2°C above pre-industrial baseline). But, the long-term temperature upward trend is far more serious. WMO reports that the 20 warmest years on record have been in the past 22 years. Moreover, long-term indicators of climate change (increasing carbon dioxide concentrations, sea level rise and ocean acidification) continue unabated. Arctic sea ice coverage remains below average and previously stable Antarctic sea ice extent was at a record low. Levels of carbon dioxide (CO_2) in the atmosphere have reached new highs, further fuelling global warming. These CO_2 emissions represent more than 75% of the total greenhouse gas emissions.[2]

Greenhouse gas emissions that cause air pollution are also the main factor causing climate change. Reducing air pollution, both globally and nationally, should be of the highest priority. The World Health Organization (WHO (2016a,b)) and recent research have shown that air pollution is the number one environmental cause of human deaths and kills more people annually than road accidents, violence, fires and wars combined. Air pollution is an international problem. A new WHO air quality model confirms that 92% of the world's population live in places where air quality levels exceed WHO limits. Air pollution has no border[3]. Depending on wind, rainfall and temperature, a country can "import" or "export" pollutants. This is the case, for example, with the importation of pollutants into California from China.

In 2015, two thirds of total global emissions came from five countries and the European Union: China (29%), the United States (14%), the European Union (EU-28) (10%), India (7%), the Russian Federation (5%) and Japan (3.5%). The year 2015 closed with the adoption of the landmark Paris Agreement on Climate Change by 194 countries and the European Union. The top emitter China started to curb its carbon dioxide emissions. China and the United States reduced their emissions by 0.65% (1.16% per capita (pc)) and 2.63% (3.39% pc), respectively, compared to 2014. Emissions in the Russian Federation and Japan also decreased, by 3.35% (3.55% pc) and 2.26% (2.18% pc), respectively. Emissions in the OECD countries have also declined by 1.23% (1.86% pc). However, these decreases were counterbalanced by increases in India (5.18%, 3.88% pc), the European Union (EU-28, 1.28%, 1.01% pc), non OECD countries (0.3%, 0.77% pc) and by increased emissions in a large group of the smallest countries (see section 3 and Olivier et al. (2016)).

[1]See https://public.wmo.int/en.

[2]In 2013 (resp. 2017), the relative shares of the greenhouse gases were: 76% (resp. 75%) for carbon dioxide (CO_2), 16% (resp. 17%) for methane (CH_4), 6% (resp. 6%) for nitrogen dioxide (NO_2) and 2% (resp. 3%) for fluorinated gases (IPCC (2014), Olivier and Peters (2018)).

[3]The Air Quality Index considers 6 common air pollutants (Ozone O_3, suflur dioxide SO_2, nitrogen dioxide NO_2, carbon dioxide CO_2 and particule matter PM_{10}, $PM_{2.5}$). Air Visual (https://airvisual.com/earth) produces a worldwide 3D animation that informs users about air quality wherever they are.

Trends in global CO_2 emissions seem to confirm that the slight slowdown in the emissions growth due to fossil fuel combustion may not be random, but may be due to structural changes in the economy, global energy efficiency improvements and the energy mix of key world players (see Olivier et al. (2016), GCP (2016)). Some have gone as far as decoupling global emissions and economic growth (see the announcement of IEA (2016)). Unfortunately, global carbon emissions jumped to an all-time high in 2018 and the Global Carbon Project report (see Figueres et al. (2018)) estimated that CO_2 emissions have risen by 2.7% in 2018.

The relationship between greenhouse gases (hereafter GHG) and economic activities is complex and the literature on this subject is huge. There has been considerable interest in the relationship between economic growth and environmental pollution since the seminal papers by Grossman and Krueger (1991, 1995). They found an inverted U-shaped pattern, now commonly referred to as the Environmental Kuznets Curve (EKC). This well-known curve suggests that various indicators of environmental degradation tend to get worse as modern economic growth occurs until average income reaches a certain point over the course of development. After this turning point and beyond some level of income per capita, the trend reverses and high-income levels per capita economies lead to environmental improvement. This EKC is essentially an empirical phenomenon and the empirical evidence is mixed. Proponents and opponents of this view are numerous as there is no guarantee that economic growth will see a decline in pollutants. Pollution is not only a function of income, but many other factors as well. These include population levels and densities, urbanization, technology, development of the economy, trade liberalization, effectiveness of government regulation, to mention a few. The statistical evidence for the EKC is not robust and the mechanisms that might drive such patterns are still contested (see for instance Van Alstine and Neumayer (2010) and Stern (2004, 2017)) to mention a few).

Most of the empirical studies in this area imposed relatively restrictive functional forms (such as a quadratic form for the EKC or for the link between CO_2 emissions and urbanization). However, recently some have used semiparametric approaches, which do not impose any *a priori* restrictions on the functional form.

In this paper, we study the relationship between carbon dioxide emissions and GDP using a semiparametric panel data approach with random coefficients. It allows us to take into account the heterogeneity between countries and periods. Inspired by the work of Lee and Wand (2016a), we use a mean field variational Bayesian (MFVB) approach which has great advantages as compared to Markov Chain Monte Carlo (MCMC) technique such as Gibbs sampling. In section 2, we briefly review the literature on EKC and its critics as well as the works justifying a semiparametric approach. In section 3, we describe our panel data using 81 countries observed over the period 1991-2015. Section 4 presents the model and the estimation method based on a MFVB approximation. In Section 5, the estimation results are discussed. Finally, Section 6 concludes.

2 A quick overview of the literature on EKC and its critics

A significant part of the literature on EKC has focused on estimating the turning point(s), the shape imposed, and the associated econometric techniques. Some caveats should be kept in mind when looking at the results of empirical studies in this literature. First, no turning point is in sight for some aspects of the environment. This includes CO_2 emissions, direct material flows[4] and the biodiversity loss (see Van Alstine and Neumayer (2010)).

Cavlovic et al. (2000) and Hui et al. (2007) attempt to draw a conclusion about the existence of an EKC and the level of the income turning point *via* meta-analyses. The empirical evidence shows that EKC is a very sensitive concept depending on data features, methodology, estimation technique and the environmental indicators used. By and large, empirical studies show that most countries exhibit a positive linear relationship between CO_2 emissions (or sulfur dioxide SO_2 emissions) and GDP. Second, even if an EKC is found, multiple turning points may arise (see for instance Binder and Neumayer (2005)). Third, when EKC exists, it could be due to a trade effect, *i.e.*, rich countries may become clean by importing products that are polluting in production from lower-income countries (see for instance Cole and Neumayer (2005)). Fourth, there are several econometric problems with the estimation results obtained from a standard quadratic equation relating a log-environmental indicator and log-GDP per capita. The most important of these are the omitted variables bias, the problem of spurious regressions and the identification of dynamic effects (see Stern (2004, 2017)). Variables such as density of the urban population, energy use and energy consumption, trade and economic liberalization, land use, income inequality, institutional and political behavior, democracy, literacy, environmental non-government organizations should be taken into account to reduce the omitted variables bias (see for instance Dasgupta et al. (2002), Liu (2005), Van Alstine and Neumayer (2010) to mention a few).

Spurious regressions may arise if the environmental indicator and GDP per capita are both trending over time. Time-specific dummies are not enough to control non-stationarity. Estimating the model in first differences or in growth rates might work as a solution. Identification of time effects lead several authors to use panel data unit root tests and panel data cointegration models to control for non-stationarity and common trends (see for instance Perman and Stern (2003), Galeotti et al. (2006), Romero-Ávila (2008), Vollebergh et al. (2009), Stern (2010), Anjum et al. (2016), Baek (2015), Bernard et al. (2015), Apergis (2016), Uchiyama (2016), Wagner and Grabarczyk (2016) to mention a few).
Most of the papers using panel data found a "CO_2 emissions - GDP elasticity" between 0.6 and 1, but the confirmation of the Kuznets hypothesis is mixed. The estimated EKC shapes are found to be either quadratic or lie between what Dasgupta et al. (2002) call the "race to bottom" and the "new toxics".[5] Wagner and Grabarczyk (2016) are somewhat critical of these papers, because most of these papers use standard panel linear cointegration techniques with cross-sectional independence. Moreover, the implicit hypothesis that all the slope coefficients are in fact identical for all countries is too restrictive in many applications. They advocate estimating a quadratic EKC with country-specific intercepts and slope coefficients for the relationship between log CO_2 per capita and log GDP per capita adding a deterministic trend. They estimate seemingly unrelated cointegrating polynomial regression

[4] Direct material flows are defined as the amount of material in physical weight (excluding water and air) available to an economy. These material flows comprise the extraction of materials inside the economy and the physical imports and exports (*i.e.,* the mass weight of goods imported or exported).

[5] "*The curve will rise to a horizontal line at maximum existing pollution level, as globalization promotes a "race to bottom" in environmental standards.... If certain pollutants are reduced as income increases, industrial society continuously creates new, unregulated and potentially toxic pollutants.... The overall environmental risks from these new pollutants may continue to grow even if some sources of pollution are reduced, as shown by the "new toxics" line in Figure 1.*" (Dasgupta et al. (2002) pp. 148).

using fully modified OLS SUR estimators for 7 European countries over a very long period $1870-2009$ but 4 of out 7 countries show "race to bottom" or "new toxics" shapes, the concavity being barely pronounced. Moreover, there has been recent work on the effects of the business cycle on carbon dioxide emissions and specifically on whether the "emissions-income elasticity" differs during downward and upward movements of the business cycle (see Bowen and Stern (2010), Peters et al. (2012), Jotzo et al. (2012), Heutel (2012), Doda (2013, 2014), Burke et al. (2015), Sheldon (2015)). Using a panel data set of 189 countries over the period 1961–2010, Burke et al. (2015) conclude that there is no strong evidence that the emissions-income elasticity is larger during years of an economic expansion as compared to a recession. They also find that economic growth tends to increase emissions not only in the same year, but also in subsequent years. These delayed effects imply that emissions tend to grow more quickly after booms and more slowly after recessions.

Another main aspect of this literature between environment quality and economic development has been over the appropriate underlying functional form (see Dasgupta et al. (2002)). It must be noted that most of the empirical studies have imposed relatively restrictive functional forms, such as the quadratic form for the EKC. Although some researchers allow for non-linearity by introducing higher order terms. More recently, some have used a semiparametric approach, allowing greater freedom in the relationship between the environmental variable and GDP per capita since it does not impose any *a priori* restriction on the functional form of this relationship. For example, Millimet et al. (2003) have shown with data for the US states that such parametric modeling can be rejected in favor of a semiparametric estimator. Using a panel of 122 countries (of which 95 are LDCs) for carbon dioxide and a panel of 108 countries (of which 81 are LDCs) for sulfur, over the period 1950–1990, Bertinelli and Strobl (2005) estimated a semiparametric model. Their findings suggest that the link between environmental pollution and economic growth is actually monotonically increasing for low levels of GDP/capita, and flat thereafter. In their specification against a linear model and using bootstrapped values, they were unable to reject a linear relationship. These results echoe the skepticism raised by Stern (2004, 2017) over the existence of an international EKC in his review of the literature. However, it should be noted that except for real GDP per capita, Bertinelli and Strobl (2005) used only time- and country-specific dummies to control for other explanatory factors of pollution. One must however pay attention to the presence of unit root and its possible impact on the validity of nonparametric or semiparametric approaches. This is an important point that we will study using a small Monte Carlo simulation to assess the performance of our proposed MFVB method when non stationary $I(1)$ series are present in the explanatory variables set (see appendix B in the supplementary material). Our simulation results support the use of the MFVB approach even in the presence of non stationary variables like log CO_2 per capita and log GDP per capita in our application. Our MFVB estimation on a nonlinear semiparametric panel data specification with random coefficients does not lead to non stationary residuals. In a related context, statistical inference for the random coefficient panel data model find that, in a random coefficient autoregression context, there is no unit root problem. Horváth and Trapani (2016) show that the weighted least squares estimator of the autoregressive root is always asymptotically normal, irrespective of the average value of the autoregressive root, of whether the autoregressive coefficient is random or not, and of the presence and degree of cross dependence (see also Ng (2008)).

In the same literature, the relationship between CO_2 emissions and urbanization has been extensively investigated in recent years. But, the empirical results are mixed. For example, Cole and Neumayer (2004) and Liddle and Lung (2010) demonstrate a positive correlation between urbanization and CO_2 emissions, while Fan *et al.* (2006) find a negative correlation between urbanization and CO_2 emissions in developing countries. Poumanyvong and Kaneko (2010) argue that assuming that the relationship between urbanization and CO_2 emissions is homogenous for all countries may be unreasonable. They examine the effects of urbanization on CO_2 emissions for low-, middle-, and high-income groups, and find that while a positive relationship exists for all income groups, it is most prominent in the middle-income group. The vast majority of the existing literature assumed that there exists a linear relationship between urbanization and CO_2 emissions. Ehrhardt-Martinez et al. (2002) argue that urbanization is a good proxy for modernization, and thus the relationship between urbanization and CO_2 emissions may vary across different stages of development. Our results in Table 1 show that the effect of urbanization on CO_2 emissions is positive and significant. Zhu et al. (2012) investigate the relationship between CO_2 emissions and urbanization in a sample of 20 emerging countries over the period 1992–2008 using the semiparametric panel data model with fixed effects. They find a nonlinear relationship between log CO_2 emissions and log-urbanization and little evidence in support of an inverted-U curve. Wang et al. (2015), using a panel of OECD countries over the period 1960-2010, estimate a semiparametric approach with fixed effects of the Stochastic Impacts by Regression on Population, Affluence and Technology model (STIRPAT) proposed by Dietz and Rosa (1997).[6] Using the work of Baltagi and Li (2002), the STIRPAT framework of the carbon emissions is extended by introducing a semiparametric specification of urbanization and with time- and country-specific dummies. The results show that the estimated "CO_2 emission - energy intensity elasticity" and the "CO_2 emission - affluence elasticity" are unity. Wang et al. (2015) also found evidence for an inverse U-shaped curve relationship between urbanization and carbon emissions. However, the shape of the smoothed relationship between urbanization and log CO_2 emissions lies between the conventional EKC and what Dasgupta et al. (2002) call the "race to bottom".

This rapid overview of the empirical validations of the EKC curve reinforce the idea of using a semiparametric form, allowing slope coefficients to vary between countries and avoiding the use of the cointegration approach with fixed coefficients for a quadratic specification. Nevertheless, methods for estimating a semiparametric panel data model with random coefficients are few. In section 4, we use a Bayesian approach that is capable of estimating such a complex specification.

3 The data

The data come from several sources: the World Development Indicators, from the World Bank, The Emission Data base for Global Atmospheric Research (EDGARv4.3.2), from the European Commission, Joint Research Centre (JRC)/PBL Netherlands Environmental Assessment Agency, the Eurostat database, from the European Commission and the Standardized World Income Inequality Data base (SWIID) created by Solt (2016) for the Gini index. Merging the data coming from these sources lead to an unbalanced panel data set with missing values. The percentage of missing values range between 0 and 45% depending on the variable considered. We dropped countries, variables and years for which missing values were larger than 15%. These variables included education, health, institutional and

[6] Affluence relates to the average consumption of each person in the population.

political behavior, democracy, literacy, environmental non-government organizations. The remaining variables with missing data were processed using cubic B-splines to obtain satisfactory imputations.[7] We have chosen a missingness rate of 15% that allows us to keep the maximum number of observations for the maximum number of countries. Moreover, our choice of using a cubic spline smoothing method rather than a multiple imputation method (like MICE or EM) is reinforced by the results of Yoon et al. (2017). Missing data may cause bias in estimation and inference. The missing data mechanisms (missing at random (MAR), missing completely at random (MCAR) and missing not at random (MNAR), see Rubin (1976)) have greater impact on the research results than does the proportion of missing data (see also Tabachnick and Fidell (2001)). Many modern missing data methods (*e.g.*, multiple imputation, FIML, EM, etc.) assume MAR. Yoon et al. (2017), using health longitudinal data, have compared the most familiar methods for estimating missing data. They show that recurrent neural networks (RNN), just followed by cubic splines, give the best results (*i.e.*, smallest rmse) as compared to imputation (MICE (multiple imputation by chained equations (see White et al. (2011)) or EM). Increasing the missingness rate (10%, 20%, ..., 50%) does not change significantly the rmse of RNN-based methods or cubic smoothing while the rmse of MICE or EM increase faster. Likewise, increasing the individual size (from N = 500, 1000, 2000, ..., 16000) or the time length (T=5, 10, 15, ..., 30) does not deteriorate the results of the first two methods as compared to multiple imputation methods. The final data set utilizes a balanced panel data set for 81 countries over 25 years (1991-2015).[8]

The left panel of Figure 1 shows the smoothed trends of CO_2 emissions for some countries from locally weighted regressions. China and India experienced episodes of growth in the overall level of CO_2 emissions. In contrast, the US, OECD countries and the European Union (EU-28) saw their CO_2 emissions levels decline. Nevertheless, looking at the right panel of Figure 1, the shapes of the smoothed CO_2 emissions per capita are different except for China and India. There is a significant decrease in the CO_2 emissions per capita for the USA, and to a lesser extent for OECD, European Union, France, Germany, UK and the Russian Federation. Japan seems to have experienced a stable evolution over these two decades.

Please insert Figure 1 here

Please insert Figure 2 here

The left panel of Figure 2 plots log CO_2 emissions per capita against log-GDP per capita for the $2,025$ observations (81 countries, 1991–2015). There is a positive correlation between these two variables (0.89) with an evolution in the form of a funnel as the log-GDP per capita increases. This seems to reflect high levels of heterogeneity for large log-GDP per capita values. The right panel of Figure 2 zooms on the richest countries. The 6th order local polynomial fit shows a maximum and a pronounced curvature mainly due to the presence of two countries (Norway and Singapore) with low pollution levels and very high levels of income per capita. Figure 3 gives the same results but for the levels of the scatter plot of CO_2 emissions per capita against GDP per capita. The evolution in the form of a funnel

[7] We also tried multiple imputation, using Bootstrap-based EM algorithms proposed by Honaker and King (2010), Honaker et al. (2011) but we got implausible values, mainly for the oldest or most recent years. So, we prefer to use cubic B-spline interpolation.

[8] See Appendix 1 for a detailed description of these variables.

is accentuated as the GDP per capita increases. The local polynomial fit reveals a maximum estimated at 10.2525 metric tons of per capita CO_2 emissions for a GDP per capita of $42,801$ USD. We find a slight concave form for the EKC with a turning point at $42,801$ USD. When we drop both Norway and Singapore from the sample, the curvature seems more accentuated, emphasizing the fact that the local polynomial adjustment gives more weight to rich countries with low levels of pollution (France, Sweden, Switzerland), compared to rich countries with high levels of pollution (Australia, Canada, USA). This remains to be confirmed by the semiparametric model proposed in this paper.

Please insert Figure 3 here

Please insert Figure 4 here

The left panel of Figure 4 plots CO_2 emissions per capita against GDP per capita for the countries means and reveals significant differences between countries. Some countries have peculiar behavior, see for instance USA, Australia and Canada which represent the most polluted rich countries. In the United States, the largest source of greenhouse gas emissions from human activities is from burning fossil fuels for electricity, heat and transportation (EPA (2016)). Australia has one of the highest per capita emissions of carbon dioxide in the world, with 0.3% of the world's population it produces 1.8% of the world's greenhouse gases. Australia uses principally coal power (70%) for electricity, with the remainder mainly gas, with no nuclear, low levels of hydro power, and low, but increasing, levels of solar, wind and wave power. Some of the reasons for Australia's high levels of emissions include Australia smelts Aluminium, a warm climate results in high use of air conditioning and the effect of agriculture. Canadians have a huge appetite for energy. Canada makes up less than 0.5% of the world's population, but is the world's eighth largest producer of greenhouse gases. Canada's greenhouse gas emissions are increasing. Energy consumption has grown about 22% and emissions by 19% since 1990. The energy industry and the transportation sector contribute the greatest share of emissions (ECCC (2016)). Other rich countries such as France, Sweden, Switzerland and, to a lesser extent, Norway and Singapore have energy and environmental policies that generate less pollution (about 6 metric tons per capita for the first three countries and 10 metric tons per capita for the second two countries). The right panel of Figure 4 plots the time means CO_2 emissions per capita against years. The local polynomial smooth curves (or order 6 and 10) show very clearly the existence of a cycle that seems to follow the international business cycle with three distinct periods: a recession (1991-1995), an expansion (1996-2007) and a new recession (2008-2015). The last period (2014-2015) seems to express the end of the depression and a new expansion. However, there is a very important collapse in the per capita CO_2 emissions during the 2008-2009 crisis.

The graphs show differentiated behaviors by country, year and whether level, per capita or logs where used. There is strong heterogeneity as well as different trends across countries and years. We carry out IPS and CIPS unit root tests[9] on these two fundamental variables:

[9]The Im-Pesaran-Shin (IPS) unit root tests (Im et al. (2003)) and the Pesaran (2007) CIPS tests relax the assumption of a common autoregressive parameter in the augmented Dickey-Fuller (ADF) specification contrary to other standard tests such as Levin-Lin-Chu, Harris-Tzavalis or Breitung tests. IPS tests assume cross-sectional independence, but allow for heterogeneity of the form of country deterministic effects (constant and/or linear time trend) and heterogenous serial correlation structure of the error terms. Cross-sectional augmented IPS (CIPS) test (Pesaran (2007)) allow for cross-sectional dependence by augmenting the test with a cross-section average.

per capita log CO_2 and per capita log-GDP (see Table A1 in the supplementary material). With the full sample (81 countries and 25 years), we cannot reject the null of unit roots ($I(1)$ process) for the log CO_2 emissions per capita and for the log-GDP per capita. In contrast, trend stationary (TS) processes are not rejected at the 5% level for the log CO_2 emissions per capita when we drop China and India from the sample. Similarly, when restricting the sample to only the 30 OECD countries, the TS processes are not rejected at the 5% level for the log CO_2 emissions per capita. This is not true for the 49 non-OECD countries alone (excluding China and India)). The existence of a deterministic trend in the variables underlies the phenomenon of spurious regression in many cases. But, when structural breaks occur, the spurious regression between trend stationary series (or between a TS dependent variable and $I(1)$ covariates) can be removed by including the trend variable and the structural breaks as regressors (see Kim et al. (2004), García-Belmonte and Ventosa-Santaulària (2011), Noriega and Ventosa-Santaulària (2005) and Wu et al. (2016) to mention a few). To guard against possible spurious regressions, we will also introduce structural breaks which correspond to the business cycle reversal phases.

These statistics show that there are fundamental differences among countries. The motivation of a mixed fixed- and random-coefficients model is then conditioned on these individual specific effects. Such a specification allows us to draw inference on certain population characteristics through the imposition of *a priori* constraints on the coefficients (see Hsiao and Pesaran (2008), Bresson and Hsiao (2011), Hsiao (2014, 2015) among others). These different elements lead us to choose a semiparametric panel data specification with random intercepts and slopes coefficients which is the subject of the next section.

4 The semiparametric panel data model with random coefficients

4.1 The model

We start with a semiparametric extension of the usual STIRPAT model, *i.e.,* the stochastic version of the impacts of population, affluence and technology (IPAT) model: $I_{it} = aPop_{it}^b A_{it}^c Tech_{it}^d e^{\varepsilon_{it}}$ where I refers to environmental impact, Pop, A, and $Tech$ refer to population, affluence and technology factors, respectively. Generally, authors use the amount of carbon dioxide emitted (in tons per capita) by country i in year t for the environmental impact variable I, A is the GDP per capita, $Tech$ is the energy intensity. As discussed in the brief review of literature on EKC, introducing a semiparametric form between the environmental variable and GDP per capita offers greater freedom in this relationship as it does not impose any *a priori* restriction on the functional form. Furthermore, we introduce random intercepts and slopes coefficients for other explanatory variables to take into account heterogeneity between the countries. In the spirit of Ruppert (2002) and Lee and Wand (2016a), we use a mixed model with a penalized spline for log-GDP per capita. Then, the initial random coefficient model

$$\log\left(\frac{CO_2}{Pop}\right)_{it} = \sum_{j=1}^{q^R}\left(\beta_j^R X_{it,j}^R + u_{i,j}^R Z_{it,j}^R\right) + \varepsilon_{it} \quad (1)$$

is extended by adding a smooth function for the lth predictor, say s_l, $1 \leq l \leq L$

$$f\left(s_{l,it}\right) = \beta_l s_{l,it} + \sum_{k=1}^{q_l^G} u_{lk}^G z_{lk}\left(s_{l,it}\right) \ , \ u_{lk}^G \sim N\left(0, \sigma_{u_l}^2\right) \ , \ 1 \leq l \leq L \tag{2}$$

where $\left\{z_{l1}\left(.\right), ..., z_{lq_l^G}\left(.\right)\right\}$ is a set of a penalized spline functions of size q_l^G and $\sigma_{u_l}^2$ is the penalized parameter for the spline coefficients $\left\{u_{l1}^G, ..., u_{lq_l^G}^G\right\}$. We suppose here that we have only one predictor, the log of GDP per capita, so $L = 1$. Putting these together, the semiparametric panel data model with random coefficients is given by

$$\begin{aligned} \log\left(\frac{CO_2}{Pop}\right)_{it} &= \textstyle\sum_{j=1}^{q^R}\left(\beta_j^R X_{it,j}^R + u_{i,j}^R Z_{it,j}^R\right) + f\left(\log\left(\frac{GDP}{Pop}\right)_{it}\right) + \varepsilon_{it} \\ \text{where } f\left(\log\left(\frac{GDP}{Pop}\right)_{it}\right) &= \beta_l \log\left(\frac{GDP}{Pop}\right)_{it} + \sum_{k=1}^{q_l^G} u_{lk}^G z_{lk}\left(\log\left(\frac{GDP}{Pop}\right)_{it}\right) \\ \text{and } Z^R &= \underset{(1\leq i\leq N)}{\text{blockdiag}}\left(X_i^R\right) \end{aligned} \tag{3}$$

for $i = 1, ..., N(81)$, $t = 1, ..., T(25)$. $X_{it,1}^R$ is the intercept and $X_{it,j}^R$, $2 \leq j \leq q^R$ are the other covariates (in log, percent, dummies or time trend). X^R is an $(NT \times q^R)$ matrix of covariates, Z^R is an $(NT \times Nq^R)$ block-diagonal matrix of the X_i^R submatrices. In the statistics literature, X and Z are called the fixed and random effects design matrices associated with β and u. The latter are called the fixed effects and random effects vectors. This terminology is different from what the panel data literature dubs as “fixed” and “random” effects. The random intercept is defined by the sum $(\beta_1^R + u_{i,1}^R)$, the random slope for variable $X_{i,2}$ is the sum $(\beta_2^R + u_{i,2}^R)$, etc. The disturbances ε_{it} have a normal distribution $\varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right)$. Following Wand and Ormerod (2008) and Lee and Wand (2016a), we use transformed cubic O'Sullivan splines.[10] Penalized splines can be viewed as random effects as one penalizes the spline basis function coefficients by treating them as a random sample from a multivariate normal distribution to avoid overfitting of the data (see Lee (2016)).[11]

This semiparametric panel data model with random coefficients is a complex specification and the methods for estimating such a specification are few. In the panel data literature, random coefficient models (RCM) and linear mixed-effects models frequently use either maximum likelihood estimators involving repeated applications of the penalized least squares method (see Bates et al. (2015)) or Bayesian Gibbs sampling (see Hsiao and Pesaran (2008), Bresson and Hsiao (2011) and Hsiao (2015) among others). Adding semiparametric elements reinforces the complexity. The varying coefficient model considered here is therefore part of the already long tradition of research on semiparametric estimation of partially linear varying coefficient panel data models which allow flexibility to characterize trending phenomenon in nonlinear panel data analysis. Some use semi-parametric profile likelihood methods (Chen et al. (2012), Li et al. (2017)), kernel or averaged local linear estimation (Li

[10] O'Sullivan penalized splines are similar to P-splines, but have the advantage of being a direct generalization of smoothing splines.

[11] That is, $u_{11}^G, ..., u_{Lq_L^G}^G \mid \sigma_{u_1}^2, ..., \sigma_{u_L}^2 \sim N\left(0, \underset{(1\leq l\leq L)}{\text{blockdiag}}\left(\sigma_{u_l}^2 I_{q_l^G}\right)\right)$ where $I_{q_l}^G$ is an $\left(q_l^G \times q_l^G\right)$ identity matrix.

et al. (2011)), series estimation methods (Huang et al. (2002), Li et al. (2003), Fan and Li (2004), Qu and Li (2006), Fan et al. (2007), Li et al. (2015), An et al. (2016)) to mention a few. Adding semiparametric elements gives preference to the Bayesian techniques rather than the frequentist ML estimators. In the spirit of works on semiparametric partially linear model using series estimation methods, we use a Bayesian linear-mixed Gaussian model-based penalized spline specification with random coefficients proposed by Lee and Wand (2016a). This approach is part of a series of Bayesian works that has been developing in recent years, such as the work of Park et al. (2015), Jeong and Park (2016) or Huang and Lu (2017), all of which use Markov Chain Monte Carlo (MCMC) techniques. Unfortunately, the MCMC techniques such as Gibbs sampling could become computationally prohibitive and may suffer from poor mixing and do not scale well when applied to models that require inversion of large sparse covariance matrices (as in our case). This paper uses the Lee and Wand (2016a) approach of a mean field variational Bayes approximation, which has many advantages over MCMC.

4.2 The mean field variational Bayes approximation

In this section, we present the panel data semiparametric model with random coefficients and we outline the mean field variational Bayesian approach (hereafter MFVB) used to estimate this model. We are influenced by the work of Zhao et al. (2006), Lee (2016) and Lee and Wand (2016a) and we follow their notation. The linear-mixed Gaussian specification is given by

$$y \sim N\Bigg(X^R\beta^R + Z^R u^R + f\left(X^G\right), \sigma_\varepsilon^2 I_{NT}\Bigg) \tag{4}$$

where y is a $(NT \times 1)$ vector $(y_{11}, ..., y_{1T}, y_{21}, ..., y_{2T}, ..., y_{N1}, ..., y_{NT})'$ for $i = 1, ..., N$ countries and $t = 1, ..., T$ time periods. I_{NT} is an $(NT \times NT)$ identity matrix. X^R is an $(NT \times q^R)$ matrix of covariates, Z^R is an $(NT \times Nq^R)$ block-diagonal matrix of the X_i^R submatrices. The dimensions of the vectors β^R and u^R are respectively $(q^R \times 1)$ and $(Nq^R \times 1)$. $X_{it,1}^R$ is the intercept and $X_{it,j}^R$, $2 \leq j \leq q^R$ are the other control covariates. The random intercept is defined by the sum $(\beta_1^R + u_{i,1}^R)$, the random slope for variable $X_{i,2}$ is the sum $(\beta_2^R + u_{i,2}^R)$, etc. The semiparametric additive function is given by

$$\begin{aligned} f\left(X^G\right) &= X^G\beta^G + Z^G u^G = \sum_{l=1}^{L} X_l^G \beta_l^G + \sum_{l=1}^{L} Z_l^G u_l^G \\ &= \sum_{l=1}^{L} X_l^G \beta_l^G + \sum_{l=1}^{L}\sum_{k=1}^{q_l^G} u_{lk}^G z_{lk}^G\left(X_l^G\right) \end{aligned} \tag{5}$$

The $(NT \times L)$ matrix X^G contains L covariates that are not already included in X^R. The $(NT \times q^L)$ matrix Z^G, with $q^L = \sum_l^L q_l^G$, would contain spline basis functions Z_l^G of the same covariates, using q_l^G knots and u^G are the $(q^L \times 1)$ spline coefficient vectors. For the covariate X_l, $Z_l^G = \left\{z_{lk}^G\left(X_l^G\right), 1 \leq k \leq q_l^G\right\}$ are spline bases of size q_l^G and $\left\{u_{lk}^G, 1 \leq k \leq q_l^G\right\}$ are the spline coefficients. In our case, X^G contains only the log of GDP per capita (*i.e.*, $L = 1$).

The quantity of spline basis functions has a minimal effect on the adequacy of (4) and, as Ruppert (2002) showed, the number of knots q_l^G is not a crucial parameter because smoothing is controlled by the penalty parameter. A common default for the number of knots in the penalized spline literature is $q_l^G = \min(\nu/4, 35)$ where ν is the number of unique $X_{l,it}^G$'s (see Ruppert (2002), Ruppert et al. (2003)). In our case, as $NT = 2025$, the number of knots could be chosen as 35. Ruppert (2002) discusses 'hi-tech' choice of q_l^G knots and follows the recommendation of Eilers and Marx (1996) to work with equally-spaced knots. *"Because smoothing is controlled by the penalty parameter, the number of knots, is not a crucial parameter"* (Ruppert (2002), p. 740). To select the optimal number of interior knots, some use cross-validation techniques (*e.g.*, Wahba (1985), Li and Racine (2007)). Knot selection could be also made using the Lasso (*e.g.*, Tibshirani (1996)). Some others use the backward elimination procedure (Smith (1982)). But generally, researchers arbitrarily choose the number of knots. We can quote several papers in which a majority of applied or simulated studies use 20, 25, 30 or 50 interior knots whatever the size of N and/or T (or N in the case of cross-sections).[12] $q_l^G = 25$ is sometimes recommended for models of practical interest (see Li and Ruppert (2008)). Moreover, in an extension of our small Monte Carlo simulation study (see appendix C in the supplementary material), we use $N = 100$, $T = 25$, $q_l^G = (10, 25, 50)$ interior knots for the spline bases. Whatever the chosen number of interior knots (10, 25 or 50), MVFB estimates are close to the theoretical values in both the non stationary or stationary cases. It also confirms the fact that there is no optimal number of knots for spline bases and supports the idea of Ruppert (2002) that the number of knots is not a crucial parameter because smoothing is controlled by the penalty parameter. These results also highlight the choice of the previous quoted applied and/or simulated studies. Then,

$$\begin{aligned} y &\sim N\left(X^R\beta^R + Z^Ru^R + X^G\beta^G + Z^Gu^G, \sigma_\varepsilon^2 I_{NT}\right) \\ &\sim N\left(X\beta + Zu, \sigma_\varepsilon^2 I_{NT}\right) \end{aligned} \tag{6}$$

where

$$\begin{aligned} X^R &= vec\left(X_1^R, ..., X_N^R\right), \; Z^R = \underset{(1\le i\le N)}{\text{blockdiag}}\left(X_i^R\right), \; \beta = \left(\beta^{R'}, \beta^{G'}\right)' \\ X &= \left(X^R, X^G\right), \; u = \left(u^{G'}, u^{R'}\right)', \; \text{ and } Z = \left(Z^G, Z^R\right) \end{aligned} \tag{7}$$

X and Z are the fixed effects and random effects design matrices associated with the fixed effects and random effects vectors β and u. The random effects vector u^R has an unstructured $\left(q^R \times q^R\right)$ covariance matrix $I_N \otimes \Sigma^R$, $\otimes$ denoting the Kronecker product. The spline coefficient vector u^G has a block-diagonal covariance matrix $\sigma_{u_l}^2 I_{q_l^G}$ where $\sigma_{u_l}^2$ is the penalized parameter for the spline coefficients $\left\{u_{lk}^G, 1 \le k \le q_l^G\right\}$. Then, the random effects covariance

[12] For instance, Ruppert (2002) shows the weak impact of the number of knots on performance of the penalized spline when he uses 20, 40 or 80 knots. Wand and Ormerod (2011) study penalized wavelets in a semiparametric regression and use 25 knots for $N = 4096$ observations. Pham et al. (2013) use 30 knots in a simulation study on measurement error for N=(50, 500) observations. Lee and Wand (2016a), in a Monte Carlo study, use 25 knots with $N = (100, 500, 2500, 12500)$ and $10 \le t_i \le 20$ for $i = 1, ..., N$. They also use 25 knots in a health study of $N = 3978$ mothers with $1 \le t_i \le 6$ births for $i = 1, ..., N$ leading to an unbalanced panel with a total of 8604 births. Lee and Wand (2016b) use 25 knots in a simulation study with $N = 50$, $T = 50$. They also use 25 for $N = 295,340$ low-risk nulliparous women giving birth for the first time in $T = 99$ australian public or private hospitals. Hajargasht and Griffiths (2018) for estimation and testing stochastic frontier models use 20 knots for $N = 43$ smallholder rice producers in the Philippines between 1990 and 1997, etc.

matrix is homoskedastic and given by

$$Cov\left(u\right)=\left(\begin{array}{cc} Cov\left(u^{G}\right) & 0 \\ 0 & Cov\left(u^{R}\right) \end{array}\right)=\left(\begin{array}{cc} \underset{(1\leq l\leq L)}{\text{blockdiag}}\left(\sigma_{u_l}^{2}I_{q_l^G}\right) & 0 \\ 0 & I_N\otimes\Sigma^R \end{array}\right)$$

In our specification, y will be the log-CO_2 per capita, $\left[X^G, Z^G\right]$ are restricted to the log-GDP per capita and $\left[X^R, Z^R\right]$ are the other covariates including the intercept. Lee and Wand (2016a) used a Bayesian approach to fit the model to the data and for inference. The full Bayesian model (with priors on parameters and hyperparameters) is given by:

$$y \mid \beta, u, \sigma_{\varepsilon}^{2} \sim N\left(X\beta + Zu, \sigma_{\varepsilon}^{2} I_{NT}\right) \tag{8}$$

with

$$\left\{\begin{array}{lcl} \beta & \sim & N\left(0,\sigma_{\beta}^{2}I_P\right),\ I_P=I_{q^R+L} \\ u\mid\Sigma^R,\sigma_{u_l}^{2} & \sim & N\left(0,\left(\begin{array}{cc} \underset{(1\leq l\leq L)}{\text{blockdiag}}\left(\sigma_{u_l}^{2}I_{q_l^G}\right) & 0 \\ 0 & I_N\otimes\Sigma^R \end{array}\right)\right) \\ \Sigma^R\mid a_1^R,..,a_{q^R}^R & \sim & IW\left(\nu+q^R-1, 2\nu\text{diag}\left(1/a_1^R,..,1/a_{q^R}^R\right)\right) \\ a_r^R & \sim & IG\left(\frac{1}{2},A_{R_r}^{-2}\right),\ 1\leq r\leq R \\ \sigma_{\varepsilon}^{2}\mid a_{\varepsilon} & \sim & IG\left(\frac{1}{2},1/a_{\varepsilon}\right) \\ a_{\varepsilon} & \sim & IG\left(\frac{1}{2},A_{\varepsilon}^{-2}\right) \\ \sigma_{u_l}^{2}\mid a_{u_l} & \sim & IG\left(\frac{1}{2},1/a_{u_l}\right),\ 1\leq l\leq L \\ a_{u_l} & \sim & IG\left(\frac{1}{2},A_{u_l}^{-2}\right) \end{array}\right.$$

where $IW\left(.\right)$ and $IG\left(.\right)$ are inverse-Wishart and inverse-Gamma distributions.[13] The likelihood combined with the prior distributions yields a joint posterior distribution which does not have a known tractable distribution and the parameters have to be sampled using MCMC techniques such as Gibbs sampling. But, MCMC becomes computationally prohibitive and do not scale well when applied to massive data sets and/or models that require storage and inversion of large sparse covariance matrices. Inference based on MCMC can be very slow for such models and MCMC methods may suffer from poor mixing.

Variational Bayesian inference can help in tackling the scalability challenge of big data sets and/or models with large sparse covariance matrices as they use a deterministic optimization approach to approximate the posterior distribution. The parameters of the approximate distribution are chosen to minimize some measure of distance (as the Kullback-Leibler divergence) between the approximation and the posterior. Mean field variational Bayes approximation is analogous to Gibbs sampling for conjugate models (see Bishop (2006), Ormerod and Wand (2010), Pham et al. (2013) and Lee and Wand (2016a) to mention a few).

[13] The initial values of the hyperparameters of the priors are: $\sigma_{\beta}^{2}=10^5$, $A_{\varepsilon}=10^5$, $A_u=10^5$, $A_R=10^5$ and $\nu=2$ leading to diffuse priors. We use transformed cubic O'Sullivan splines with 25 knots. The standard deviation parameters have independent Half-Cauchy priors $\sigma_{u_l}\sim$ Half-Cauchy(A_{u_l}), $\sigma_{\varepsilon}\sim$ Half-Cauchy(A_{ε}) which are equivalent to the following statements: $\sigma_{u_l}^{2}\sim IG\left(\frac{1}{2},1/a_{u_l}\right)$ with $a_{u_l}\sim IG\left(\frac{1}{2},A_{u_l}^{-2}\right)$ and $\sigma_{\varepsilon}^{2}\sim IG\left(\frac{1}{2},1/a_{\varepsilon}\right)$ with $a_{\varepsilon}\sim IG\left(\frac{1}{2},A_{\varepsilon}^{-2}\right)$. Indeed, the hierarchical representation enables MCMC and variational methods to be easily carried out because of the conditional conjugacy properties of the Inverse-Gamma distribution. If x and a are random variables such that $x\mid a\sim IG\left(1/2,1/a\right)$ with $a\sim IG\left(1/2,1/A^2\right)$, then $\sqrt{x}\sim$ Half-Cauchy(A) (see Gelman (2006)).

In what follows, we give a quick overview of the MFVB method and its application to the semiparametric panel data model with random coefficients.

Consider a generic Bayesian model with observed vector y and parameter vector θ that is continuous over the parameter space Θ. The Bayes theorem allows one to define the posterior distribution as:

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y \mid \theta)\, p(\theta)}{p(y)} \text{ with } p(y) = \int_{\Theta} p(\theta, y)\, d\theta \tag{9}$$

Let q be an arbitrary density function over Θ. Then, the logarithm of the marginal likelihood satisfies (see Bishop (2006), Ormerod and Wand (2010)):

$$\begin{aligned}
\log p(y) &= \log p(y) \int_{\Theta} q(\theta)\, d\theta = \int_{\Theta} q(\theta) \log p(y)\, d\theta \\
&= \int_{\Theta} q(\theta) \log \left\{ \frac{p(\theta, y)/q(\theta)}{p(\theta \mid y)/q(\theta)} \right\} d\theta \\
&= \int_{\Theta} q(\theta) \log \left\{ \frac{p(\theta, y)}{q(\theta)} \right\} d\theta + \int_{\Theta} q(\theta) \log \left\{ \frac{q(\theta)}{p(\theta \mid y)} \right\} d\theta \\
&= \log \underline{p}(y, q) + KL(q, p)
\end{aligned} \tag{10}$$

where $KL(q,p)$ is the Kullback-Leibler divergence between $q(\theta)$ and $p(\theta \mid y)$. Furthermore, $\log \underline{p}(y, q)$ is a lower bound on the marginal log-likelihood. The Kullback-Leibler divergence becomes

$$\begin{aligned}
KL(q,p) &= E_{q(\theta)}[\log q(\theta)] - E_{q(\theta)}[\log p(\theta \mid y)] \\
&= E_{q(\theta)}[\log q(\theta)] - E_{q(\theta)}[\log p(\theta, y)] + \log p(y)
\end{aligned} \tag{11}$$

where the last term, $\log p(y)$, is a constant. The minimization of the Kullback-Leibler divergence is thus equivalent to maximizing the scalar quantity,

$$\log \underline{p}(y, q) = E_{q(\theta)} \left[\log \left(\frac{p(\theta, y)}{q(\theta)} \right) \right] \tag{12}$$

which is usually referred as the evidence lower bound (ELBO)[14].
Let $\{\theta_1, ..., \theta_M\}$ be a partition of the parameter vector θ. The MFVB approximates the posterior distribution $p(\theta \mid y)$ by the product of the q-densities:[15]

$$q(\theta) = \prod_{j=1}^{M} q_j(\theta_j) \tag{13}$$

The *optimal* q-densities which minimize the Kullback-Leibler divergence are given by

$$q_j^*(\theta_j) \propto \exp\left[E_{q(-\theta_j)} \{ \log p(\theta_j \mid \text{rest}) \} \right] \ , j = 1, ..., M \tag{14}$$

where $E_{q(-\theta_j)}$ denotes expectation with respect to $\prod_{k \neq j} q_k(\theta_k)$.
rest $\equiv \{y, \theta_1, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_M\}$ is the set containing the rest of the random vectors in the

[14]The lower bound is also known as the negative variational free energy and the entropy of the variational distribution is given by $E_{q(\theta)} \log[q(\theta)]$.

[15]This is known as the *mean field restriction.* The term *mean field* originated from physics.

model, except θ_j and the distributions $(\theta_j|$ rest) are the full conditionals in the MCMC literature. The iterative scheme for obtaining the optimal q-densities under product restriction (13) is

---

1. Initialize $q_1^*(\theta_1)$, $q_2^*(\theta_2)$, $\cdots$, $q_M^*(\theta_M)$
2. Cycle through updates:

$$\begin{array}{rcl} q_1^*(\theta_1) & \leftarrow & \dfrac{\exp\left[E_{q(-\theta_1)}\{\log p(y,\theta)\}\right]}{\int \exp\left[E_{q(-\theta_1)}\{\log p(y,\theta)\}\right] d\theta_1} \\ \cdots & \cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ q_M^*(\theta_M) & \leftarrow & \dfrac{\exp\left[E_{q(-\theta_M)}\{\log p(y,\theta)\}\right]}{\int \exp\left[E_{q(-\theta_M)}\{\log p(y,\theta)\}\right] d\theta_M} \end{array}$$

   until the increase in $\log \underline{p}(y,q)$ is negligible.

Compared to the minimization of the KL divergence, the maximization of the ELBO is often a more convenient objective of the optimization over the free distributional parameters. Lee and Wand (2016a) apply this principle and derive the MFVB approximation of the linear-mixed Gaussian model-based penalized spline specification (8) on the following factorization:

$$\begin{aligned} p(\theta \mid y) &= p(\theta_1,...,\theta_M \mid y) = p\left(\beta, u^R, u^G, a^R, a_u, a_\varepsilon, \Sigma^R, \sigma_u^2, \sigma_\varepsilon^2 \mid y\right) \\ &\approx q\left(\beta, u^R, u^G, a^R, a_u, a_\varepsilon, \Sigma^R, \sigma_u^2, \sigma_\varepsilon^2\right) \\ &= q\left(\beta, u^R, u^G\right) q\left(\Sigma^R\right) q\left(\sigma_\varepsilon^2\right) q(a_\varepsilon) \prod_{r=1}^{q^R} q\left(a_r^R\right) \prod_{l=1}^{L} q(a_{u_l}) \prod_{l=1}^{L} q\left(\sigma_{u_l}^2\right) \\ &= q(\beta,u)\, q\left(\Sigma^R\right) q\left(\sigma_\varepsilon^2\right) q(a_\varepsilon) \prod_{r=1}^{q^R} q\left(a_r^R\right) \prod_{l=1}^{L} q(a_{u_l}) \prod_{l=1}^{L} q\left(\sigma_{u_l}^2\right) \end{aligned} \tag{15}$$

They first derive the conditional posterior densities $p(\theta_j \mid \text{rest})$ for $j = 1,...,M$ from the full Bayesian model (8), *i.e.*, the Gibbs sampling algorithm. Then, they derive the optimal q-densities and the associated updated parameters using (13). Updating parameters is stopped when the maximum ELBO (12) is reached. After tedious derivations[16], this leads to the following forms of the optimal q-densities

$$\begin{cases} q^*(\beta,u) & \sim & N\left(\mu_{q(\beta,u)}, \Sigma_{q(\beta,u)}\right) \\ q^*\left(\Sigma^R\right) & \sim & IW\left(\nu + N + q^R - 1, B_{q(\Sigma^R)}\right) \\ q^*\left(\sigma_\varepsilon^2\right) & \sim & IG\left(\frac{1}{2}(T+1), B_{q(\sigma_\varepsilon^2)}\right) \\ q^*(a_\varepsilon) & \sim & IG\left(1, B_{q(a_\varepsilon)}\right) \\ q^*\left(\sigma_{u_l}^2\right) & \sim & IG\left(\frac{1}{2}\left(q_l^G + 1\right), B_{q(\sigma_{u_l}^2)}\right) \\ q^*(a_{u_l}) & \sim & IG\left(1, B_{q(a_{u_l})}\right) \\ q^*\left(a_r^R\right) & \sim & IG\left(\frac{1}{2}\left(\nu + q^R\right), B_{q(a_r^R)}\right) \end{cases} \tag{16}$$

[16]See Lee (2016) for the derivations of the optimal q-densities.

where the parameters are updated according to Algorithm 1 (see below). Convergence in Algorithm 1 is assessed using ELBO on the marginal log-likelihood:

$$\begin{aligned}
\log \underline{p}(y,q) &= E_{q(\theta)}\left[\log\left(\frac{p(\theta,y)}{q(\theta)}\right)\right] \\
&= E_{q(\theta)}\left[\begin{array}{l}\log p\left(y,\beta,u,a^R,a_u,a_\varepsilon,\Sigma^R,\sigma_u^2,\sigma_\varepsilon^2\right)\\ -\log q\left(\beta,u,a^R,a_u,a_\varepsilon,\Sigma^R,\sigma_u^2,\sigma_\varepsilon^2\right)\end{array}\right]
\end{aligned} \tag{17}$$

which is presented below. Convergence of such an algorithm to at least a local optima is guaranteed based on convexity properties. The ELBO is judged to cease increasing when the tolerance criterion is less than 10^{-7}. This algorithm is part of the family of coordinate ascent variational inference (CAVI). It iteratively optimizes each factor of the mean field variational density, while holding the others fixed (see Bishop (2006) and Blei et al. (2017)).

Algorithm 1. Mean field variational Bayes algorithm (see Lee and Wand (2016a), pp. 882).

---

1. Initialize $\mu_{q(1/\sigma_\varepsilon^2)}>0$, $\mu_{q(1/a_\varepsilon)}>0$, $\mu_{q\left(1/\sigma_{u_l}^2\right)}>0$, $\mu_{q\left(1/a_{u_l}\right)}>0$, $1\le l\le L$, $\mu_{q(1/a_r^R)}>0$, $1\le r\le q^R$,$M_{q\left((\Sigma^R)^{-1}\right)}$ positive definite.

2. Cycle through updates:

   (a) $S\leftarrow 0$, $s\leftarrow 0$, for $i=1,...,N$:
   $G_i\leftarrow \mu_{q(1/\sigma_\varepsilon^2)}\left(C_i^G\right)' X_i^R$; $H_i\leftarrow\left[\mu_{q(1/\sigma_\varepsilon^2)}\left(X_i^R\right)'X_i^R+M_{q\left((\Sigma^R)^{-1}\right)}\right]^{-1}$ with $C_i^G=\left[X_i^G,\ Z_i^G\right]$
   $S\leftarrow S+G_iH_i\left(G_i\right)'$; $s\leftarrow s+G_iH_i\left(X_i^R\right)'y_i$

   (b) $$\Sigma_{q(\beta,u^G)}\leftarrow\left[\mu_{q(1/\sigma_\varepsilon^2)}\left(C^G\right)'C^G+\left[\begin{array}{cc}\sigma_\beta^{-2}I_P & 0\\ 0 & \underset{(1\le l\le L)}{\mathrm{blockdiag}}\left(\mu_{q\left(1/\sigma_{u_l}^2\right)}I_{q_l^G}\right)\end{array}\right]-S\right]^{-1}$$

   (c) $\mu_{q(\beta,u^G)}\leftarrow\mu_{q(1/\sigma_\varepsilon^2)}\Sigma_{q(\beta,u^G)}\left[\left(C^G\right)'y-s\right]$ with $C^G=\left[X^G,\ Z^G\right]$

   (d) for $i=1,...,N$:
   $\Sigma_{q\left(u_i^R\right)}\leftarrow H_i+H_i\left(G_i\right)'\Sigma_{q(\beta,u^G)}G_iH_i$
   $\mu_{q\left(u_i^R\right)}\leftarrow H_i\left[\mu_{q(1/\sigma_\varepsilon^2)}\left(X_i^R\right)'y_i-\left(G_i\right)'\mu_{q(\beta,u^G)}\right]$
   $$B_{q(\sigma_\varepsilon^2)}\leftarrow\mu_{q(1/a_\varepsilon)}+\tfrac{1}{2}\left[\begin{array}{c}\parallel D\parallel^2+\mathrm{tr}\left[\left(C^G\right)'C^G\Sigma_{q(\beta,u^G)}\right]\\ +\sum_{i=1}^N\mathrm{tr}\left[\left(X_i^R\right)'X_i^R\Sigma_{q\left(u_i^R\right)}\right]\\ -2\mu_{q(1/\sigma_\varepsilon^2)}^{-1}\sum_{i=1}^N\mathrm{tr}\left[G_iH_i\left(G_i\right)'\Sigma_{q(\beta,u^G)}^R\right]\end{array}\right]$$
   with $D=y-C^G\mu_{q(\beta,u^G)}\left(\begin{array}{c}X_1^R\mu_{q\left(u_1^R\right)}\\ \vdots\\ X_N^R\mu_{q\left(u_N^R\right)}\end{array}\right)$

   (e) $\mu_{q(1/\sigma_\varepsilon^2)}\leftarrow\frac{1}{2}\left(T+1\right)/B_{q(\sigma_\varepsilon^2)}$; $\mu_{q(1/a_\varepsilon)}\leftarrow 1/\left[\mu_{q(1/\sigma_\varepsilon^2)}+A_\varepsilon^{-2}\right]$

(f) for $r=1,...,q^R$:

$B_{q(a_r^R)} \leftarrow \nu \left(M_{q\left((\Sigma^R)^{-1}\right)}\right)_{rr} + A_{R_r}^{-2}$; $\mu_{q(1/a_r^R)} \leftarrow \frac{1}{2}\left(\nu + q^R\right)/B_{q(a_r^R)}$

$B_{q(\Sigma^R)} \leftarrow \sum_{i=1}^N \left(\mu_{q\left(u_i^R\right)}\mu'_{q\left(u_i^R\right)} + \Sigma_{q\left(u_i^R\right)}\right) + 2\nu \text{diag}\left(\mu_{q\left(1/a_1^R\right)}, ..., \mu_{q\left(1/a_{q^R}^R\right)}\right)$

(g) $M_{q\left((\Sigma^R)^{-1}\right)} \leftarrow \left(\nu + N + q^R - 1\right) B_{q(\Sigma^R)}^{-1}$

(h) for $l=1,...,L$:

$\mu_{q\left(1/a_{u_l}\right)} \leftarrow 1/\left[\mu_{q\left(1/\sigma_{u_l}^2\right)} + A_{u_l}^{-2}\right]$; $B_{q(a_{u_l})} \leftarrow 1/\mu_{q\left(1/a_{u_l}\right)}$

$\mu_{q\left(1/\sigma_{u_l}^2\right)} \leftarrow \frac{q^G+1}{2\mu_{q\left(1/a_{u_l}\right)} + \|\mu_{q\left(u_l^G\right)}\|^2 + \text{tr}\left[\Sigma_{q\left(u_l^G\right)}\right]}$; $B_{q(\sigma_{u_l}^2)} \leftarrow \frac{1}{2}\left(q_l^G + 1\right)/\mu_{q\left(1/a_{u_l}\right)}$

(i) for $i=1,...,M$:

$\Lambda_{q\left(\beta,u^G,u_i^R\right)} \equiv E_q\left[\left(\begin{bmatrix}\beta \\ u^G\end{bmatrix} - \mu_{q(\beta,u^G)}\right)\left(u_i^R - \mu_{q\left(u_i^R\right)}\right)'\right] \leftarrow -\Sigma_{q(\beta,u^G)}G_iH_i$

$\Sigma_{q\left(\beta,u^G,u_i^R\right)} \equiv Cov\begin{pmatrix}\beta \\ u^G \\ u_i^R\end{pmatrix} \leftarrow \begin{pmatrix}\Sigma_{q(\beta,u^G)} & \Lambda_{q\left(\beta,u^G,u_i^R\right)} \\ \Lambda'_{q\left(\beta,u^G,u_i^R\right)} & \Sigma_{q\left(u_i^R\right)}\end{pmatrix}$

(j) $\Sigma_{q(\beta,u)} \leftarrow \begin{pmatrix}\Sigma_{q(\beta,u^G)} & \Lambda_{q\left(\beta,u^G,u_1^R,...,u_N^R\right)} \\ \Lambda'_{q\left(\beta,u^G,u_1^R,...,u_N^R\right)} & \Sigma_{q\left(u_i^R\right)}\end{pmatrix}$

until the increase in the ELBO $\log \underline{p}(y,q)$ is negligible.

The variational lower bound on the marginal log-likelihood has the following expression (see Lee and Wand (2016a), pp. 893):

$$\begin{aligned}
\log \underline{p}(y,q) \; = \; & \frac{1}{2}q^R\left(\nu + q^R - 1\right)\log 2\nu - \frac{T}{2}\log 2\pi - \left(\frac{1}{2}q^R + L + 1\right)\log \pi \qquad (18) \\
& -\frac{P}{2}\log \sigma_\beta^2 - \frac{\sigma_\beta^{-2}}{2}\left[\| \mu_{q(\beta)} \|^2 + \text{tr}\left[\Sigma_{q(\beta)}\right]\right] + \frac{1}{2}\left(\textstyle\sum_{l=1}^L q_l^G + P + N\right) \\
& -\frac{1}{2}\textstyle\sum_{i=1}^N \log\left|\mu_{q(1/\sigma_\varepsilon^2)}\left(X_i^R\right)' X_i^R + M_{q\left((\Sigma^R)^{-1}\right)}\right| - \frac{1}{2}\log\left|\Sigma_{q(\beta,u^G)}^{-1}\right| \\
& -\log\left(\mathcal{C}_{q^R,\nu+q^R-1}\right) + \log\left(\mathcal{C}_{q^R,\nu+N+q^R-1}\right) - \frac{1}{2}\left(\nu + N + q^R - 1\right)\log\left|B_{q(\Sigma^R)}\right| \\
& + \textstyle\sum_{l=1}^L \log\Gamma\left(\frac{q_l^G+1}{2}\right) - \frac{1}{2}\sum_{l=1}^L\left(q_l^G + 1\right)\log B_{q\left(\sigma_{u_l}^2\right)} + \sum_{l=1}^L \log\Gamma\left(N\left(T+1\right)\right) \\
& -\frac{1}{2}\left(T+1\right)\log B_{q(\sigma_\varepsilon^2)} - \textstyle\sum_{r=1}^{q^R}\log A_{R_r} + q^R\log\Gamma\left(\frac{q^G+\nu}{2}\right) \\
& + \textstyle\sum_{r=1}^{q^R}\nu M_{q\left((\Sigma^R)^{-1}\right)}\mu_{q(1/a_r^R)} - \frac{1}{2}\left(q^R + \nu\right)\sum_{r=1}^{q^R}\log B_{q(a_r^R)} \\
& - \textstyle\sum_{l=1}^L\left[\log A_{u_l} + \log B_{q\left(a_{u_l}\right)} + \mu_{q\left(1/a_{u_l}\right)}\mu_{q\left(1/\sigma_{u_l}^2\right)}\right] + \log A_\varepsilon \\
& - \log B_{q(a_\varepsilon)} + \mu_{q(1/a_\varepsilon)}\mu_{q(1/\sigma_\varepsilon^2)}
\end{aligned}$$

where $\mathcal{C}_{a,b}$ is the normalizing factor: $\mathcal{C}_{a,b} = 2^{ab/2}\pi^{a(a-1)/4}\prod_{j=1}^a \Gamma\left(\frac{b+1-j}{2}\right)$ and $\Gamma(.)$ is the Gamma function.

The computing time gains afforded by the MFVB algorithm, as compared to Gibbs sampling, are huge (see for instance Pham et al. (2013) and Lee and Wand (2016a)).[17] More importantly, this approximation avoids the pitfalls of poor mixing of MCMC methods on models with large sparse covariance matrices. But variational inference algorithms involve different implementation challenges from sampling algorithms. They are harder, in that they may require lengthy mathematical derivations to determine the updating rules. However, once implemented, variational Bayes can be easier to test, because one can use the standard checks for optimization code (gradient checking, local optimum tests, etc). Most variational inference algorithms converge to optima, which eliminates the need to check convergence diagnostics and the output of most variational inference algorithms is a distribution, rather than samples. Moreover, the accuracy scores of the MFVB approximation (as compared to MCMC) generally exceed $95-97\%$ and rarely drop below 90% in most papers on MFVB (see for instance Bishop (2006), Ormerod and Wand (2010), Faes et al. (2011), Pham et al. (2013), Lee and Wand (2016a) and Blei et al. (2017) to mention a few). More recently, using a variational Bayes Kalman filter, the computationally efficient approximations have also been confirmed by Koop and Korobilis (2018) for efficient posterior and predictive inference in high-dimensional time-varying parameter models.
While there has not been much theory developed around variational inference, there are nevertheless major trends that emerge. Blei et al. (2017) have summarized a variety of results about theoretical guarantees of variational inference but conditional on specific models and families of variational approximations. A number of recent contributions have shed new light on characterizing frequentist properties of variational estimators which is an important question (see Alquier and Ridgway (2017), Zhang and Gao (2017), Bhattacharya et al. (2018), Chérief-Abdellatif (2018) and Wang and Blei (2018) to mention a few).[18]

Thus, the MFVB approach has great advantages as compared to the MCMC technique such as Gibbs sampling. More generally, if MCMC algorithms, like Metropolis-Hastings or the Gibbs sampler, rely on stochastic processes that yield samples from the posterior, variational inference transforms posterior inference into optimization. In variational inference, the posterior distribution over a set of unobserved variables θ, given some data y, is approximated by a variational distribution $p(\theta \mid y) \approx q(\theta)$. The distribution $q(\theta)$ is restricted to belong to a family of distributions of simpler form than $p(\theta \mid y)$. Variational inference posits a class of distributions q over the latent space Θ and tries to find the closest distribution in Kulback-Leibler (KL) divergence to the posterior and minimizing the KL divergence is equivalent to maximizing the evidence lower bound (ELBO). The choice of variational family $q(\theta)$ trades off the fidelity of the posterior approximation with the difficulty of optimizing over the variational parameters. The classical choice of variational family is the mean field family, which factorizes over some partition of the latent variables $q(\theta) = \prod_{j=1}^{M} q_j(\theta_j)$.
The optimal q-density functions, obtained *via* an iterative coordinate ascent variational inference algorithm (which is a generalization of the EM algorithm), has been the workhorse for deploying variational inference in a variety of applications especially when MCMC algorithms are essentially impractical or impossible to implement (see Ranganath et al. (2014), Ranganath (2017)).

[17] The MFVB can be $3,500$ (or more) times faster than the MCMC. Lee and Wand (2016a) used a data set with more than $N = 140,000$ mothers with 1 to 5 children for an application on the link between birthweight of children and infant's gestational age (in weeks). It took 4 days with MCMC and few minutes with MFVB.

[18] We would like to thank a referee for suggesting this point.

Though this model class imposes many restrictions, variational methodology is used in a diversified range of applications, from Bayesian Gaussian mixtures, latent Dirichlet allocation (Blei et al. (2003)), probabilistic matrix factorization (Salakhutdinov and Mnih (2008)), hierarchical linear regression (Gelman (2006)), hierarchical Bayesian nonparametric models (Hjort et al. (2010)), complex models with elaborate distributions (such as asymmetric Laplace and skew normal) to spline and wavelet regression models (Neville et al. (2014)), to mention a few.
Because of their ability to handle complex specifications, large samples, sparse matrices, ..., these methods have developed very rapidly since a decade in physics, in statistics, health studies, neuroimaging, machine learning, ... and make a more timid and more recent appearance in econometrics. However, more and more studies, in various topics, use variational inference (see Wang and Blei (2018)). More recently, for discrete-margined copula models (Loaiza-Maya and Smith (2019)), for spatiotemporal models (Quiroz et al. (2018)), for generalized linear latent variable models (Hui et al. (2017)), for models with specific factor covariance structure (Ong et al. (2018)), for stochastic frontier models using variational Bayes (Hajargasht and Griffiths (2018)) or for inference in high-dimensional time-varying parameter models (Koop and Korobilis (2018)), to mention a few. The algorithm proposed by Lee and Wand (2016a) has a promising future with a remarkable ability to perform high quality Bayesian inference for large panel data models faster than ever before.

5 The results

Results for β_j^R and β^G coefficients are given in Table 1.[19] Convergence is quickly reached after 644 cycles and 42.88 seconds of computing time, which is incredibly faster than MCMC.[20] Convexity properties of the MFVB algorithm guarantees quick convergence of such an algorithm to at least a local optima. In the left panel of Table 1, some variables are not significant and have been dropped in the restricted model of the right panel of Table 1.[21] We have introduced 4 dummies: *high pop density*, *large country*, $large\ forest$ and *non* $OECD$ *countries*. The dummy variable *high pop density* takes the value 1 for the 25% highest population density countries and 0 otherwise. The *large country* dummy variable takes the value 1 for the 25% largest countries and 0 otherwise. The $large\ forest$ dummy variable takes the value 1 for the 50% largest forests and 0 otherwise. The fourth dummy is for the non OECD countries. We can see that most of the coefficients are highly significant whatever the risk level, except for the electricity production from nuclear sources, the alternative and nuclear energy use, the trade of services and the highest forest percent dummy variable. Elasticities of the CO_2 emissions per capita relative to energy intensity and energy use are significant. A 10% increase in energy intensity (resp. energy use per GDP per capita) leads to a 4.35% (resp. 2.07%) increase in the CO_2 emissions per capita. Remember that the energy intensity is the ratio between energy supply and GDP. It is an indication of how much energy is used to produce one unit of economic output. Energy use per GDP is the kilogram of oil equivalent of energy use per GDP. It refers to the use of primary energy

[19] We tested several specifications including other covariates such as surface area (sq. km), land area (sq. km), agricultural land (% of land area that is agricultural), a Gini index of income inequality. But these variables were not statistically significant.

[20] Estimation was conducted using R version 3.3.2 on a MacBook Pro, 2.8 GHz core i7 16Go MGz DDR3 ram. Some elements of the R code are available in the supplementary material of Lee and Wand (2016a)).

[21] For the restricted model, convergence is reached after 607 cycles and 33.70 seconds of computing time (see Figure D1 in the supplementary material for the plot of the ELBO).

before transformation to other end-use fuels. The "CO_2 emission - pump price of diesel fuel elasticity" is -0.03 which confirms a certain sensitivity of greenhouse gases to an increase in fuels prices. An increase of one percent of the fossil fuel energy consumption increases the CO_2 emissions per capita by 1.44%. Similarly, an increase of one percent of electricity production from hydroelectric sources decreases the CO_2 emissions per capita by 0.16%. The rate of urbanization has also a strong effect since an increase of one unit leads to an increase of 0.46% of the CO_2 emissions per capita. The highest population density countries have a significant but moderate effect (as compared to the other countries) since they increase the CO_2 emissions per capita by $\exp(0.059) - 1 = 6.07\%$. In contrast, the largest countries have a strong impact on the greenhouse gases evolution. As compared to smaller countries, they increase the CO_2 emissions per capita by a huge $\exp(0.60) - 1 = 82.2\%$, *ceteris paribus.*

Please insert Table 1 here

Non-OECD countries emit less pollution per capita than OECD countries ($\exp(-0.514) - 1 = -40.19\%$). Merchandise trade as a % of GDP has also a positive effect on CO_2 emissions. An increase of 10 points leads to an increase in CO_2 emissions per capita by 0.6%. We introduced a linear trend with structural breaks which has a positive and significant effect on the regression.[22,23] But more interestingly and more importantly is the link between the CO_2 emissions per capita and the GDP per capita. The estimated coefficient $\widehat{\beta}^G = 0.847$ leads to a strong "CO_2 emission - GDP elasticity" close to unity, the 95% confidence interval being $[0.798; 0.895]$. It confirms the positive link observed in Figures 2 and 3. This result contradicts the assumption of the decoupling of global emissions and economic growth retained in the International Energy Agency announcement (IEA (2016)), an assumption which seems unrealistic.

Figure 5 shows the MFVB spline fit of the CO_2 emissions per capita against the observed GDP per capita and the pointwise 95% credible set using estimated values $\widehat{\beta}_1^R + \widehat{f}(\log(GDP/POP)_{it})$. It reveals an increasing curve of the greenhouse gases as the GDP per capita increases. The end of the MFVB spline fit of the CO_2 emissions per capita forms a bearing, corresponding to a kind of ratchet effect, but does not show any maximum of a concave function that would express the standard form of an EKC. This is a confirmation of the results of Bertinelli and Strobl (2005) which suggest that the link between environmental pollution and economic growth is actually monotonically increasing for low levels of GDP/capita, and flat thereafter. Dropping Norway and Singapore from the sample lead to a continuously increasing relationship which gives a little more weight to the richest and strongest polluters (Australia, Canada, USA) abandoning what Dasgupta et al. (2002) call the "race to bottom" in favor of the "new toxics" shape.

[22] It is defined as:

$$\text{trend}=\begin{cases} t \ (= 1, 2, .., 4, 5) & \text{if} \quad \text{year} <1996 \\ t + 10 \ (= 16, 17, .., 26, 27) & \text{if} \quad \text{year} >1995 \text{ and year } <2008 \\ t - 12 \ (= 6, 7, .., 12, 13) & \text{if} \quad \text{year} > 2007. \end{cases}$$

[23] We have also tested a standard linear trend (without structural breaks) but its coefficient was not significantly different from zero, thus confirming the need to introduce structural breaks corresponding to the cycle reversal phases observed in Figure 4.

$\boxed{\text{Please insert Figure 5 here}}$

Figure 5 confirms that the shape of the estimated EKC gives no evidence for an inverted-U curve between the environmental variable and economic growth. The introduction of additional explanatory variables and the freedom of functional form allowed by the semiparametric method does not confirm the concavity observed with the simple local polynomial adjustment found in Figures 2 and 3. We cannot confirm the existence of a turning point marking the reversal of the trend in the pollution levels. On the contrary, the econometric model favors an increase (or a stabilization) in the relationship between CO_2 emissions per capita and GDP per capita.

The semiparametric approach seems to be fully justified.[24] Note the greatly increasing breadth of the confidence interval, which seems to follow very closely the funnel shape of the cloud of points. This increasing confidence interval can also express the very strong heterogeneity observed previously and demonstrates the ability of the MFVB approach to embrace situations where behaviors are very heterogeneous. For a fixed number of knots q_l^G and a large N, under assumptions of the model, OLS estimation with inference based on clustering at the country level could be an interesting benchmark for fixed coefficients $\left(\beta^R, \beta^G\right)$ in Table 1. However, running an extension of the small Monte Carlo study (in appendix C of the supplementary material) shows that differences in the accuracy of the estimates of the fixed coefficients $\left(\beta^R, \beta^G\right)$ between MFVB and OLS with clustered standard errors, could be huge. Hence, OLS estimation with inference based on clustering at the individual level can not serve as a benchmark for the semiparametric estimation of partially linear varying coefficient panel data models. These results also confirm the recent research of D'Adamo (2018) which shows that inference based on the usual cluster-robust standard errors by White (1984) is invalid in general when the number of controls is a non-vanishing fraction of the sample size or for semiparametric partially linear model with fixed coefficients.

So far, we have discussed average marginal effects. One of the advantages of the random coefficients specification is that it is possible to discriminate between these marginal effects for different countries $(\beta_j^R + u_{i,j}^R)$ for $i = 1, ..., N(81)$. If the average intercept estimate is $\widehat{\beta}_1^R = -6.288$ with a 95% confidence interval of $[-7.350; -5.225]$, intercepts associated with each country $(\widehat{\beta}_1^R + \widehat{u}_{i,1}^R)$ range between -19 (Ethiopia) to 0 (Sweden) highlighting the extreme sensitivity of country specific effects (see the supplementary material). This sensitivity also applies to covariates, which add to the richness of the random coefficients models.

$\boxed{\text{Please insert Figure 6 here}}$

$\boxed{\text{Please insert Figure 7 here}}$

We report only the plot of these marginal effects for energy intensity (Figure 6) and the fossil fuel energy consumption percent (Figure 7) (see the supplementary material for the other marginal effects[25]). Obviously, there is a large sensitivity of these marginal effects by country resulting in much richer information than the average value provided in Table 1. If

[24]Estimation for OECD countries only, leads to similar results. However, there is a much less "race to the bottom" and more of a "new toxics" shape (see Figure E2 in the supplementary material).

[25]See also the supplementary material for the marginal effects from the semiparametric estimation for the 30 OECD countries only.

we compare the marginal effects for energy intensity, its effects on CO_2 emissions per capita is about 164% more important for China than for France. The corresponding figures for Norway, USA and Sweden are 136% 288% and 559%, respectively. In other words, China needs much more energy to produce a one unit of economic output than France, Norway, USA or Sweden. In contrast, the sensitivity of the CO_2 emissions per capita relative to the fossil fuel energy consumption (% of total) is about 140% more important for France as compared to China. The corresponding figures for Norway, USA and Sweden are 38%, 95% and 107%.[26]

Our estimates suggest a positive increasing relationship between CO_2 emissions per capita and GDP per capita. We use bootstrap tests suggested by Cai et al. (2000) and Henderson et al. (2008) which allow one to test the null hypothesis of a parametric specification against the alternative of a semiparametric specification. Using 500 replications, we test our MFVB estimates against a) the random intercept and constant slope coefficients model (*i.e.*, the one-way error component model (OWEC)), b) the random intercept and slope coefficients model (*i.e.*, RCM). The random intercept - constant slope coefficients model (OWEC) and the RCM have been estimated using maximum likelihood. The computational methods for maximum likelihood fitting of the linear mixed-effects model involve penalized quasi-likelihood methods as repeated applications of the penalized least squares method which can be computationally intensive (see Bates et al. (2015)). Second, we specify a quadratic EKC[27] and test again our MFVB semiparametric estimates against these parametric estimates. Table 2 gives the bootstrapped p-values of the tests of Cai et al. (2000) $[CFY]$ and Henderson et al. (2008) $[HCL]$ for the different specifications. Results show that we reject the null hypothesis of parametric specifications in favor of our semiparametric models with random coefficients since all the bootstrapped p-values are less than 4.5%.

Please insert Table 2 here

Our results show that economic growth tends to increase greenhouse gas emissions. Thus the relationship between CO_2 emissions per capita and GDP per capita may vary across different stages of development. On average, the "CO_2 emissions - GDP elasticity" is around 0.85 at the 5% level. This estimated value is close to those found in the literature using parametric panel data specifications with or without cointegrating relations. Moreover, we find a nonlinear relation and no evidence in support of an inverted-U curve between log CO_2 per capita and log-GDP per capita. Thus, the Kuznets hypothesis is not confirmed. We find no turning points with this semiparametric EKC. When Norway and Singapore are in the sample of countries considered, its shape is associated with what Dasgupta et al. (2002) call the "race to bottom". When these two countries are excluded from the sample, its shape is associated with what Dasgupta et al. (2002) call the "new toxics".

[26] This high sensitivity of CO_2 emissions to the fossil fuel energy consumption can be explained by the fact that France is one of the most nuclearized countries and therefore emits less CO_2 per inhabitant. Indeed, in 2015, France had 58 nuclear reactors against 100 for the USA, 36 for China and 10 for Norway. Relative to the number of inhabitants, the ratio of France/China nuclear reactors is 3474.92 (respectively 290.19 and 86.52 for France/USA and France/Norway).

[27] Our parametric RCM have the following form: $y_{it} = X_{it}\beta_i + w_{it}\gamma_{1i} + w_{it}^2\gamma_{2i} + \varepsilon_{it}$ where y_{it} is the CO_2 per capita (in log), w_{it} is the GDP per capita (in log) and X_{it} are the other covariates including the intercept. We test a linear EKC ($\gamma_{2i} = 0$, $\forall i$) and a quadratic EKC ($\gamma_{2i} \neq 0$). For the one-way error component model: $\beta_i = \beta$, $\forall i$ except for the intercept and $\gamma_{1i} = \gamma_1$, $\gamma_{2i} = \gamma_2$, $\forall i$.

Compared to other panel data studies, this approach has several advantages. First, it rejects the restrictive assumption that all slope coefficients are identical for all countries. It allows one to estimate country-specific intercepts and slope coefficients and can be applied to a large cross-sectional dimension (*i.e.*, a large number of countries). Second, the use of a semiparametric function in addition to the presence of random intercepts and slope coefficients allow us flexibility in the functional form between CO_2 and GDP.

In a field such as climate change where the relationship between environmental variables and economic variables are very complex, such a semiparametric specification can be promising.

6 Conclusion

This paper proposed a semiparametric estimation of the relationship between CO_2 emissions and economics activities for 81 countries observed over the period $1991-2015$. The data set reveals differentiated behaviors by country, year and whether level, per capita or logs were used. There is strong heterogeneity as well as different trends across countries and years. We reject the null hypothesis of parametric specifications in favor of our semiparametric model with random coefficients. The motivaion of a mixed fixed- and random-coefficients model has been conditioned on these country specific effects. Following a recent approach proposed by Lee and Wand (2016a), we specified and estimated a MFVB semiparametric panel data model with random coefficients. This approach has numerous advantages as compared to the MCMC techniques such as Gibbs sampling. We specify and estimate a log model with structural breaks between CO_2 emissions per capita and GDP per capita. Results reveal a strong "CO_2 emissions - GDP elasticity", close to one, confirming the increasing but complex link between these two variables. These results hold when considering OECD countries only. The inclusion of random coefficients in a mixed model-based penalized spline basis function enriches the estimates and their interpretations, given the large diversity of responses by variables and countries. Our study focuses on the CO_2-GDP link, but one could also track other greenhouse gases emissions and economic activities using the same methodology.

7 Appendix - Definition of the variables

Variable	Description
CO_2	Global CO_2 emissions from fossil fuel use and cement production (metric tons)
Pop	Population, total
GDP	GDP, PPP (constant 2011 international US$)
Land area	Land area (sq. km)
Pop density	Population density (people per sq. km of land area)
Urban	Urban population (% of total)
Agricult	Agricultural land (% of land area)
Forest	Forest area (% of land area)
Energy intensity	Energy intensity (% of GDP)
Energy use per GDP	Energy use per GDP (in kg of oil equivalent per $1,000 GDP)
Energy imports	Energy imports, net (% of energy use)
Fossil fuel consumption	Fossil fuel energy consumption (% of total)
Electric consumption	Electric power consumption (kWh per capita)
Nuclear energy use	Alternative and nuclear energy (% of total energy use)
Electric prod oil gas coal	Electricity production from oil, gas and coal sources (% of total)
Electric prod hydroelectric	Electricity production from hydroelectric sources (% of total)
Electric prod nuclear	Electricity production from nuclear sources (% of total)
Electric transm losses	Electric power transmission and distribution losses (% of output)
Openness	Trade (% of GDP)
Merchandise trade	Merchandise trade (% of GDP)
Services trade	Trade in services (% of GDP)
Pump price diesel	Pump price for diesel fuel (US$ per liter)
Gini index	Gini index of income inequality

References

Alquier, P. and Ridgway, J. (2017). Concentration of tempered posteriors and of their variational approximations. *Cornell University Library arXiv:1706.09293.*

An, Y., Hsiao, C., and Li, D. (2016). Semiparametric estimation of partially linear varying coefficient panel data models. In González-Rivera, G., Carter Hill, R., and Lee, T., editors, *Advances in Econometrics, Essays in Honor of Aman Ullah*, volume 36, pages 47–65. Emerald Group Publishing Limited.

Anjum, Z., Burke, P., Gerlagh, R., and Stern, D. (2016). Modeling the emissions-income relationship using long-run growth rates. Technical report, Centre for Climate Economics & Policy, Crawford School of Public Policy, The Australian National University, Canberra, Australia. CCEP Working Papers 1403.

Baek, J. (2015). Environmental Kuznets curve for CO2 emissions: The case of Arctic countries. *Energy Economics*, 50:13–17.

Baltagi, B. H. and Li, D. (2002). Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance*, 3(1):103–116.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 76:1–48.

Bernard, J.-T., Gavin, M., Khalaf, L., and Voia, M. (2015). Environmental Kuznets curve: Tipping points, uncertainty and weak identification. *Environmental and Resource Economics*, 60(2):285–315.

Bertinelli, L. and Strobl, E. (2005). The environmental Kuznets curve semi-parametrically revisited. *Economics Letters*, 88(3):350–357.

Bhattacharya, A., Pati, D., and Yang, Y. (2018). On statistical optimality of variational Bayes. In *Proceedings of Machine Learning Research*, volume 84, pages 1579–1588. PMLR.

Binder, S. and Neumayer, E. (2005). Environmental pressure group strength and air pollution: An empirical analysis. *Ecological Economics*, 55(4):527–538.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bowen, A. and Stern, N. (2010). Environmental policy and the economic downturn. *Oxford Review of Economic Policy*, 26(2):137–163.

Bresson, G. and Hsiao, C. (2011). A functional connectivity approach for modeling cross-sectional dependence with an application to the estimation of hedonic housing prices in Paris. *Advances in Statistical Analysis*, 95(4):501–529.

Burke, P. J., Shahiduzzaman, M., and Stern, D. I. (2015). Carbon dioxide emissions in the short run: The rate and sources of economic growth matter. *Global Environmental Change*, 33:109–121.

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Cavlovic, T. A., Baker, K. H., Berrens, R. P., and Gawande, K. (2000). A meta-analysis of environmental Kuznets curve studies. *Agricultural and Resource Economics Review*, 29(1):32–42.

Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chérief-Abdellatif, B.-E. (2018). Consistency of ELBO maximization for model selection. *Cornell University Library arXiv:1810.11859.*

Cole, M. A. and Neumayer, E. (2004). Examining the impact of demographic factors on air pollution. *Population and Environment*, 26(1):5–21.

Cole, M. A. and Neumayer, E. (2005). Environmental policy and the environmental Kuznets curve: Can developing countries escape the detrimental consequences of economic growth? In Dauvergne, P., editor, *International Handbook of Environmental Politics*, pages 298–318. Edward Elgar.

D'Adamo, R. (2018). Cluster-robust standard errors for linear regression models with many controls. *Working Paper, Department of Economics, University College London, UK.*

Dasgupta, S., Laplante, B., Wang, H., and Wheeler, D. (2002). Confronting the environmental Kuznets curve. *Journal of Economic Perspectives*, 16(1):147–168.

Dietz, T. and Rosa, E. A. (1997). Effects of population and affluence on CO2 emissions. *Proceedings of the National Academy of Sciences*, 94(1):175–179.

Doda, B. (2013). Emissions-GDP relationship in times of growth and decline. *Working Paper No. 136, Centre for Climate Change Economics and Policy, London School of Economics and Political Science, UK.*

Doda, B. (2014). Evidence on business cycles and CO2 emissions. *Journal of Macroeconomics*, 40:214–227.

ECCC (2016). Canadian environmental sustainability indicators: Global greenhouse gas emissions. Technical report, Environment and Climate Change Canada, Gatineau, Québec.

Ehrhardt-Martinez, K., Crenshaw, E. M., and Jenkins, J. C. (2002). Deforestation and the environmental Kuznets curve: A cross-national investigation of intervening mechanisms. *Social Science Quarterly*, 83(1):226–243.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–102.

EPA (2016). Greenhouse gas inventory report: 1990-2014. Technical report, United States Environmental Protection Agency, Washington, DC.

Faes, C., Ormerod, J. T., and Wand, M. P. (2011). Variational Bayesian inference for parametric and nonparametric regression with missing data. *Journal of the American Statistical Association*, 106(495):959–971.

Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102(478):632–641.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467):710–723.

Figueres, C., Le Quéré, C., Mahindra, A., Bäte, O., Whiteman, G., Peters, G., and Guan, D. (2018). Emissions are still rising: Ramp up the cuts. *Nature*, 564:27–30.

Galeotti, M., Lanza, A., and Pauli, F. (2006). Reassessing the environmental Kuznets curve for CO2 emissions: A robustness exercise. *Ecological Economics*, 57(1):152–163.

García-Belmonte, L. and Ventosa-Santaulària, D. (2011). Spurious regression and lurking variables. *Statistics & Probability Letters*, 81(12):2004–2010.

GCP (2016). The Global Carbon Project. Technical report, CSIRO Marine and Atmospheric Research, Canberra, Australia. Internet: http://www.globalcarbonproject.org/.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534.

Grossman, G. M. and Krueger, A. B. (1991). Environmental impacts of a North American free trade agreement. *Working Paper, vol. 3914, National Bureau of Economic Research. NBER, Cambridge, MA, USA.*

Grossman, G. M. and Krueger, A. B. (1995). Economic growth and the environment. *The Quarterly Journal of Economics*, 110(2):353–377.

Hajargasht, G. and Griffiths, W. E. (2018). Estimation and testing of stochastic frontier models using variational Bayes. *Journal of Productivity Analysis*, 50(1-2):1–24.

Henderson, D. J., Carroll, R. J., and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1):257–275.

Heutel, G. (2012). How should environmental policy respond to business cycles? Optimal policy under persistent productivity shocks. *Review of Economic Dynamics*, 15(2):244–264.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics: Principles and Practice*. Cambridge University Press, Cambridge, UK.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581.

Honaker, J., King, G., and Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

Horváth, L. and Trapani, L. (2016). Statistical inference in a random coefficient panel model. *Journal of Econometrics*, 193(1):54–75.

Hsiao, C. (2014). *Analysis of Panel Data*. Cambridge University Press, Cambridge, UK, third edition.

Hsiao, C. (2015). Random coefficients models in panels. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*, pages 402–417. Oxford University Press, Oxford.

Hsiao, C. and Pesaran, M. H. (2008). Random coefficients models. In Sevestre, P. and Mátyás, L., editors, *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, pages 187–216. Springer, Berlin, third edition.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.

Huang, Y. and Lu, T. (2017). Bayesian inference on partially linear mixed-effects joint models for longitudinal data with multiple features. *Computational Statistics*, 32(1):179–196.

Hui, F. K. C., Warton, D. I., Ormerod, J. T., Haapaniemi, V., and Taskinen, S. (2017). Variational approximations for generalized linear latent variable models. *Journal of Computational and Graphical Statistics*, 26(1):35–43.

Hui, L., Grijalva, T., and Berrens, R. P. (2007). Economic growth and environmental quality: A meta-analysis of environmental Kuznets curve studies. *Economics Bulletin*, 17(5):1–11.

IEA (2016). Decoupling of global emissions and economic growth confirmed. Technical report, International Energy Agency, March, 16 2016, Paris. Internet: https://www.iea.org/.

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1):53–74.

IPCC (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

IPCC (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland.

Jeong, S. and Park, T. (2016). Bayesian semiparametric inference on functional relationships in linear mixed models. *Bayesian Analysis*, 11(4):1137–1163.

Jotzo, F., Burke, P. J., Wood, P. J., Macintosh, A., and Stern, D. I. (2012). Decomposing the 2010 global carbon dioxide emissions rebound. *Nature Climate Change*, 2(4):213.

Kim, T.-H., Lee, Y.-S., and Newbold, P. (2004). Spurious regressions with stationary processes around linear trends. *Economics Letters*, 83(2):257–262.

Koop, G. and Korobilis, D. (2018). Variational Bayes inference in high-dimensional time-varying parameter models. *Paper No. 87972, MPRA (Munich Personal RePEc Archive).*

Lee, C. Y. Y. (2016). *Fast Approximate Inference for Longitudinal and Multilevel Data Analysis*. PhD thesis, University of Technology Sydney, Australia.

Lee, C. Y. Y. and Wand, M. P. (2016a). Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. *Biometrical Journal*, 58(4):868–895.

Lee, C. Y. Y. and Wand, M. P. (2016b). Variational methods for fitting complex Bayesian mixed effects models to health data. *Statistics in Medicine*, 35(2):165–188.

Li, D., Chen, J., and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, 14(3):387–408.

Li, H., Lin, Z., and Hsiao, C. (2015). Testing purchasing power parity hypothesis: A semi-parametric varying coefficient approach. *Empirical Economics*, 48(1):427–438.

Li, K., Li, D., Liang, Z., and Hsiao, C. (2017). Estimation of semi-varying coefficient models with nonstationary regressors. *Econometric Reviews*, 36(1-3):354–369.

Li, Q., Hsiao, C., and Zinn, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics*, 112(2):295–325.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika*, 95(2):415–436.

Liddle, B. and Lung, S. (2010). Age-structure, urbanization, and climate change in developed countries: Revisiting STIRPAT for disaggregated population and consumption-related environmental impacts. *Population and Environment*, 31(5):317–343.

Liu, X. (2005). Explaining the relationship between CO2 emissions and national income — The role of energy consumption. *Economics Letters*, 87(3):325–328.

Loaiza-Maya, R. and Smith, M. S. (2019). Variational Bayes estimation of discrete-margined copula models with application to time series. *Journal of Computational and Graphical Statistics*, 1:1–32.

Millimet, D. L., List, J. A., and Stengos, T. (2003). The environmental Kuznets curve: Real progress or misspecified models? *Review of Economics and Statistics*, 85(4):1038–1047.

Neville, S. E., Ormerod, J. T., and Wand, M. P. (2014). Mean field variational Bayes for continuous sparse signal shrinkage: Pitfalls and remedies. *Electronic Journal of Statistics*, 8(1):1113–1151.

Ng, S. (2008). A simple test for nonstationarity in mixed panels. *Journal of Business & Economic Statistics*, 26(1):113–127.

Noriega, A. E. and Ventosa-Santaulària, D. (2005). Spurious regression under deterministic and stochastic trends. *Working paper, Department of Economics and Finance, Universidad de Guanajuato, Mexico.*

Olivier, J. G. J., Janssens-Maenhout, G., Muntean, M., and Peters, J. A. H. W. (2016). Trends in global CO2 emissions: 2016 report. Technical report, PBL Netherlands Environmental Assessment Agency, European Commission, Joint Research Centre, Directorate Energy, Transport & Climate JRC Science for Policy Report, The Hague, The Netherlands.

Olivier, J. G. J. and Peters, J. A. H. W. (2018). Trends in global CO2 emissions: 2018 report. Technical report, PBL Netherlands Environmental Assessment Agency, European Commission, Joint Research Centre, Directorate Energy, Transport & Climate JRC Science for Policy Report, The Hague, The Netherlands.

Ong, V. M.-H., Nott, D. J., and Smith, M. S. (2018). Gaussian variational approximation with a factor covariance structure. *Journal of Computational and Graphical Statistics*, 27(3):465–478.

Ormerod, J. T. and Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, 64(2):140–153.

Park, B. U., Mammen, E., Lee, Y. K., and Lee, E. R. (2015). Varying coefficient regression models: A review and new developments. *International Statistical Review*, 83(1):36–64.

Perman, R. and Stern, D. I. (2003). Evidence from panel unit root and cointegration tests that the environmental Kuznets curve does not exist. *Australian Journal of Agricultural and Resource Economics*, 47(3):325–347.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2):265–312.

Peters, G. P., Marland, G., Le Quéré, C., Boden, T., Canadell, J. G., and Raupach, M. R. (2012). Rapid growth in CO2 emissions after the 2008–2009 global financial crisis. *Nature Climate Change*, 2(1):2–4.

Pham, T. H., Ormerod, J. T., and Wand, M. P. (2013). Mean field variational Bayesian inference for nonparametric regression with measurement error. *Computational Statistics & Data Analysis*, 68:375–387.

Poumanyvong, P. and Kaneko, S. (2010). Does urbanization lead to less energy use and lower CO2 emissions? A cross-country analysis. *Ecological Economics*, 70(2):434–444.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62(2):379–391.

Quiroz, M., Nott, D. J., and Kohn, R. (2018). Gaussian variational approximation for high-dimensional state space models. *Cornell University Library arXiv:1801.07873.*

Ranganath, R. (2017). *Black Box Variational Inference: Scalable, Generic Bayesian Computation and its Applications.* PhD thesis, Princeton University, NJ.

Ranganath, R., Gerrish, S., and Blei, D. (2014). Black box variational inference. In Kaski, S. and Corander, J., editors, *Artificial Intelligence and Statistics*, pages 814–822. Proceedings of Machine Learning Research (PMLR).

Romero-Ávila, D. (2008). Questioning the empirical basis of the environmental Kuznets curve for CO2: New evidence from a panel stationarity test robust to multiple breaks and cross-dependence. *Ecological Economics*, 64(3):559–574.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11(4):735–757.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press.

Salakhutdinov, R. and Mnih, A. (2008). Probabilistic matrix factorization. In Platt, J. C., Koller, D., Singer, Y., and Roweis, S. T., editors, *Advances in Neural Information Processing Systems*, volume 20, pages 1257–1264. Curran Associates, Inc., NY.

Sheldon, T. (2015). Asymmetric effects of the business cycle on carbon dioxide emissions: A new layer of climate change uncertainty. In *The Dynamic Energy Landscape*. 33rd USAEE/IAEE North American Conference, International Association for Energy Economics.

Smith, P. L. (1982). Curve fitting and modeling with splines using statistical variable selection techniques. Technical report, NASA, Langley Research center, Hampton. Report NASA 166034.

Solt, F. (2016). The standardized world income inequality database. *Social Science Quarterly*, 97(5):1267–1281.

Stern, D. I. (2004). Environmental Kuznets curve. In Cleveland, C., editor, *The Encyclopedia of Energy*, volume 2, pages 517–525. Academic Press, San Diego, CA.

Stern, D. I. (2010). Between estimates of the emissions-income elasticity. *Ecological Economics*, 69(11):2173–2182.

Stern, D. I. (2017). The environmental Kuznets curve after 25 years. *Journal of Bioeconomics*, 19(1):7–28.

Tabachnick, B. G. and Fidell, L. S. (2001). *Using Multivariate Statistics*. Allyn & Bacon, Nedham Heights, MA, 6th edition edition.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Van Alstine, J. and Neumayer, E. (2010). The environmental Kuznets curve. In Gallagher, K. P., editor, *Handbook on Trade and the Environment*, pages 49–59. Edward Elgar, Cheltenham, UK.

Vollebergh, H. R. J., Melenberg, B., and Dijkgraaf, E. (2009). Identifying reduced-form relations with panel data: The case of pollution and income. *Journal of Environmental Economics and Management*, 58(1):27–42.

Wagner, M. and Grabarczyk, P. (2016). The environmental Kuznets curve for carbon dioxide emissions: A seemingly unrelated cointegrating polynomial regressions approach. *Discussion paper 75/2016, SFB 823, "Statistical modelling of nonlinear dynamic processes", TU Dortmund University, Dortmund.*

Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, 13(4):1378–1402.

Wand, M. P. and Ormerod, J. (2008). On semiparametric regression with O'Sullivan penalized splines. *Australian & New Zealand Journal of Statistics*, 50(2):179–198.

Wand, M. P. and Ormerod, J. T. (2011). Penalized wavelets: Embedding wavelets into semiparametric regression. *Electronic Journal of Statistics*, 5:1654–1717.

Wang, Y. and Blei, D. M. (2018). Frequentist consistency of variational Bayes. *Journal of the American Statistical Association*, pages 1–15. Published online: 06 Aug 2018.

Wang, Y., Zhang, X., Kubota, J., Zhu, X., and Lu, G. (2015). A semi-parametric panel data analysis on the urbanization-carbon emissions nexus for OECD countries. *Renewable and Sustainable Energy Reviews*, 48:704–709.

White, H. L. (1984). *Asymptotic Theory for Econometricians.* Academic Press.

White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399.

WHO (2016a). Ambient air pollution: A global assessment of exposure and burden of disease. Technical report, World Heath Organization, Geneva, Switzerland. Available at http://www.who.int/phe/health topics/.

WHO (2016b). Fact sheet on ambient air quality and health. Technical report, World Heath Organization, Geneva, Switzerland. Available at http://www.who.int/mediacentre/factsheets/.

Wu, M., You, P., and Zou, N. (2016). On spurious regressions with trending variables. *Cornell University Library arXiv:1606.05049.*

Yoon, J., Zame, W. R., and van der Schaar, M. (2017). Multi-directional recurrent neural networks: A novel method for estimating missing data. In Precup, D. and Teh, Y. W., editors, *Proceedings of Machine Learning Research*, volume 70, pages 3958–3966. Proceedings of Machine Learning Research (PMLR).

Zhang, F. and Gao, C. (2017). Convergence rates of variational posterior distributions. *Cornell University Library arXiv:1712.02519.*

Zhao, Y., Staudenmayer, J., Coull, B. A., and Wand, M. P. (2006). General design Bayesian generalized linear mixed models. *Statistical Science*, pages 35–51.

Zhu, H.-M., You, W.-H., and Zeng, Z.-F. (2012). Urbanization and CO2 emissions: A semi-parametric panel data analysis. *Economics Letters*, 117(3):848–850.

Table 1 - Log-linear model for CO_2 emissions per capita.

$\log(CO_2/Pop)$	coef.	std. err.	t	coef.	std. err.	t
intercept	-6.1441	0.5246	-11.7116	-6.2881	0.5420	-11.6012
log(energy intensity)	0.4231	0.0419	10.1101	0.4353	0.0440	9.8937
log(energy use)	0.2188	0.0414	5.2808	0.2074	0.0451	4.6020
log(pump price diesel)	-0.0309	0.0071	-4.3334	-0.0308	0.0074	-4.1659
fossil fuel consumption	1.3043	0.1465	8.9051	1.4435	0.1212	11.9090
elect prod hydroelectric	-0.1598	0.0352	-4.5432	-0.1591	0.0354	-4.4901
elect prod nuclear	-0.1049	0.1180	-0.8891			
nuclear energy use	-0.1008	0.1309	-0.7695			
urban	0.4915	0.2012	2.4430	0.4637	0.1988	2.3328
high pop density	0.0572	0.0375	1.5267	0.0593	0.0372	1.5944
large country	0.6136	0.2772	2.2135	0.6049	0.2567	2.3562
large forest	0.0016	0.0140	0.1156			
merchandise trade	0.0752	0.0189	3.9808	0.0643	0.0182	3.5250
services trade	-0.1022	0.0635	-1.6106			
non OECD	-0.5015	0.2292	-2.1882	-0.5139	0.2064	-2.4904
trend with structural breaks	0.0005	0.0002	2.7465	0.0005	0.0002	2.4706
log(GDP/Pop)	0.8585	0.0249	34.4200	0.8473	0.0245	34.5928
σ^2_ε	0.0008			0.0009		

Table 1 gives estimated coefficients, standard errors and t-stats for β^R and β^G and the estimated residual variance $\widehat{\sigma}^2_\varepsilon$.

Table 2 - Bootstrapped p-values for the specification tests of Cai et al. (2000) [CFY] and Henderson et al. (2008) [HCL].

		CFY	HCL
		p-value	p-value
linear EKC	OWEC	0.012	0.040
	RCM	0.016	0.044
quadratic EKC	OWEC	0.008	0.036
	RCM	0.014	0.042

Table 2 gives bootstrapped p-values of the bootstrap tests of the null hypothesis of a parametric specification against the alternative of a semiparametric MFVB specification. EKC: environmental Kuznets curve. OWEC: one-way error component model. RCM: random coefficient model.

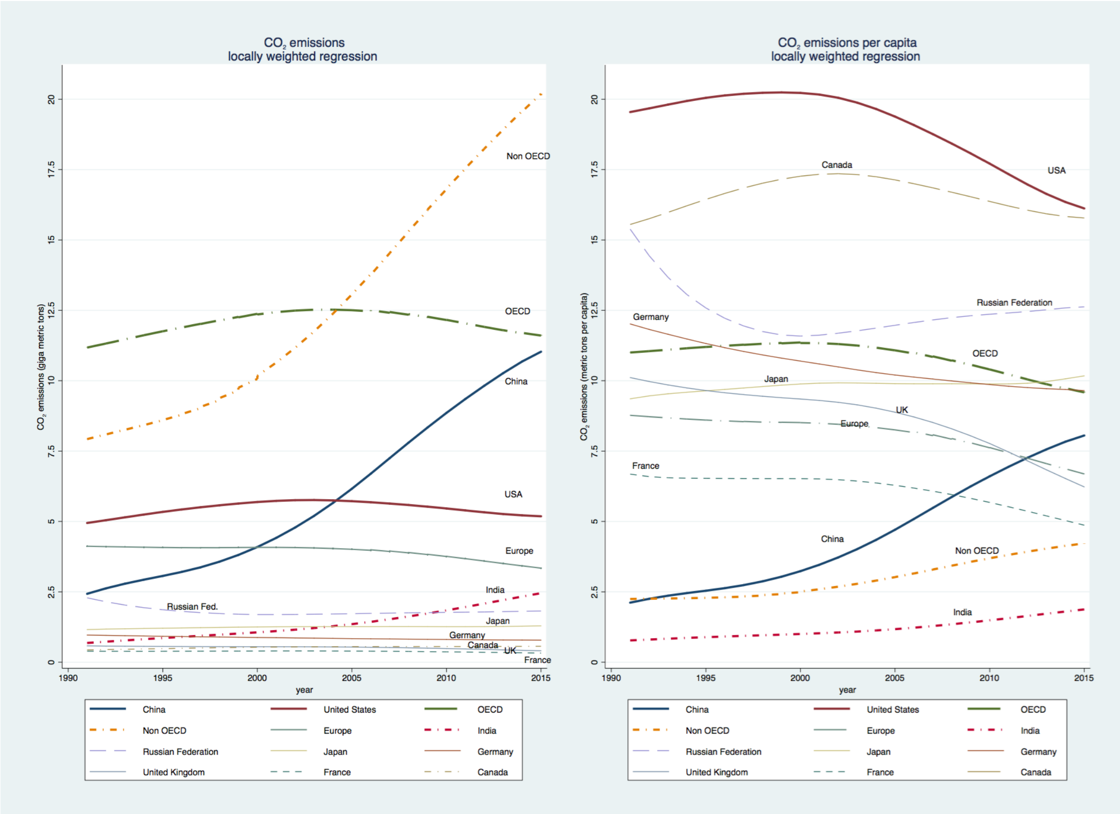

Figure 1. Smoothed trends of CO_2 emissions for some countries from locally weighted regressions.

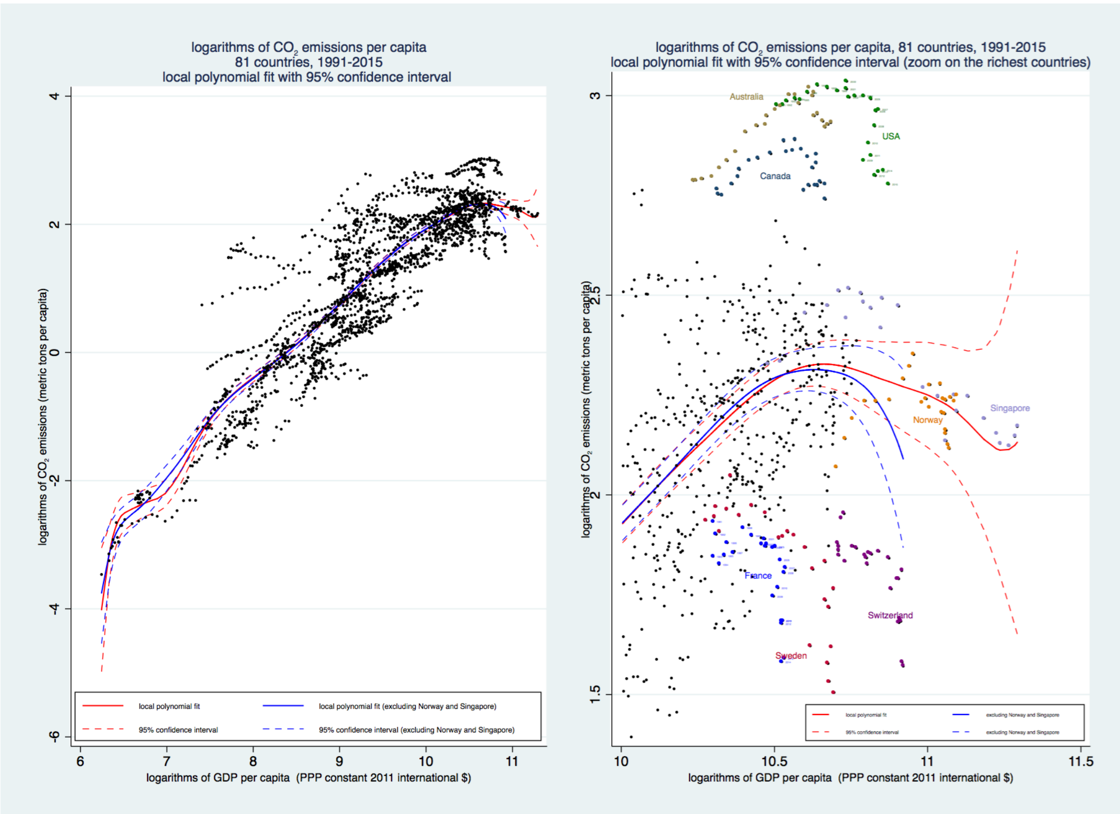

Figure 2. log CO_2 emissions per capita against log GDP per capita (left panel) and zoom on the richest countries (right panel).

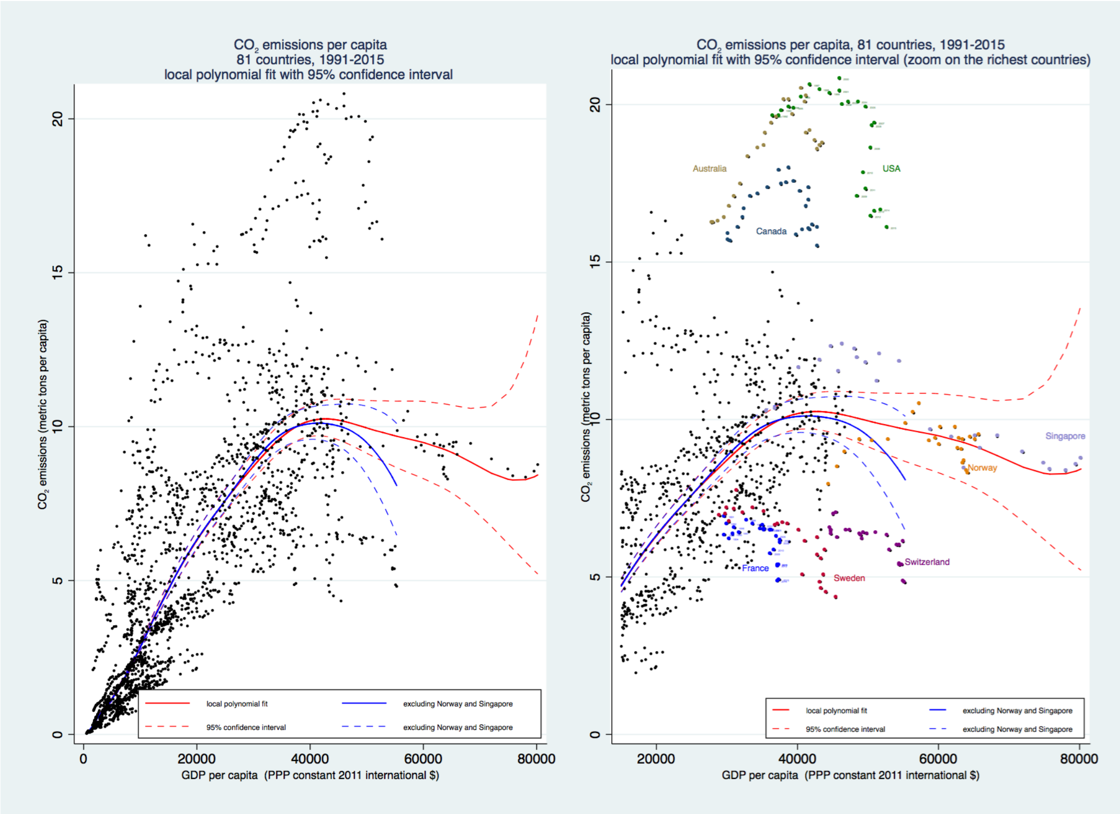

Figure 3. CO_2 emissions per capita against GDP per capita (left panel) and zoom on the richest countries (right panel).

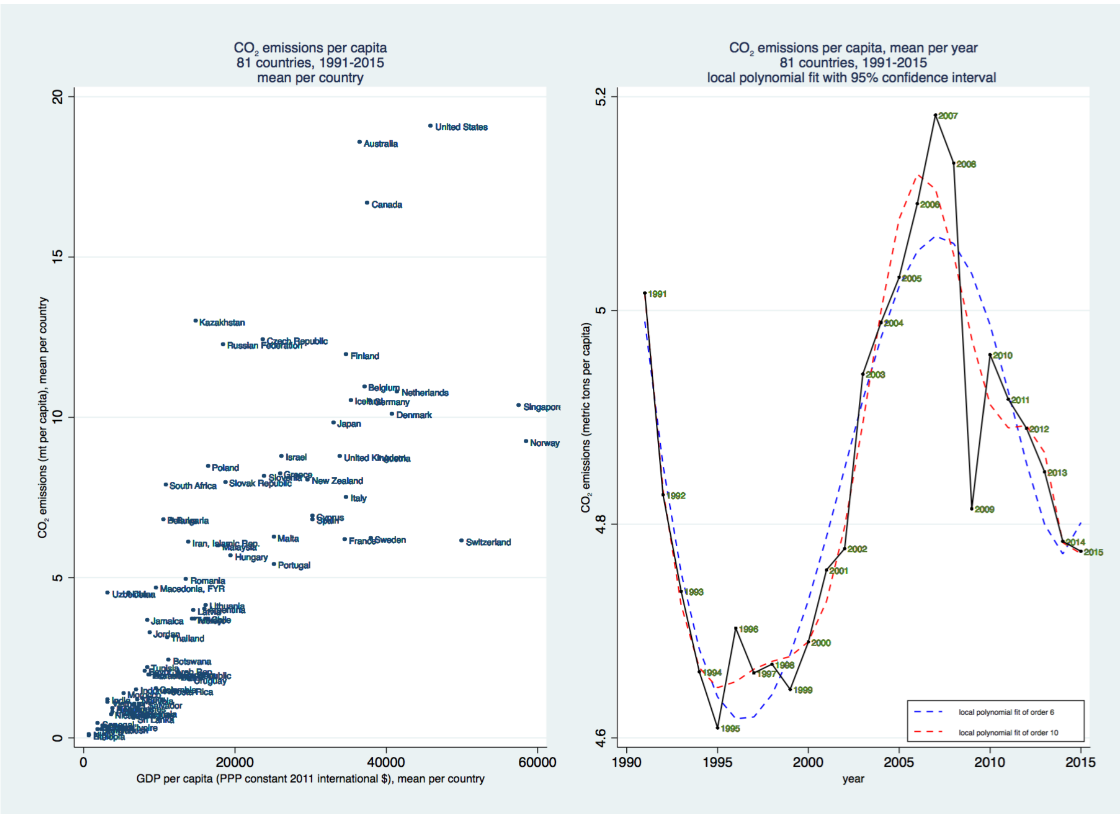

Figure 4. CO_2 emissions per capita against GDP per capita for the country means (left panel) and for the time means (right panel).

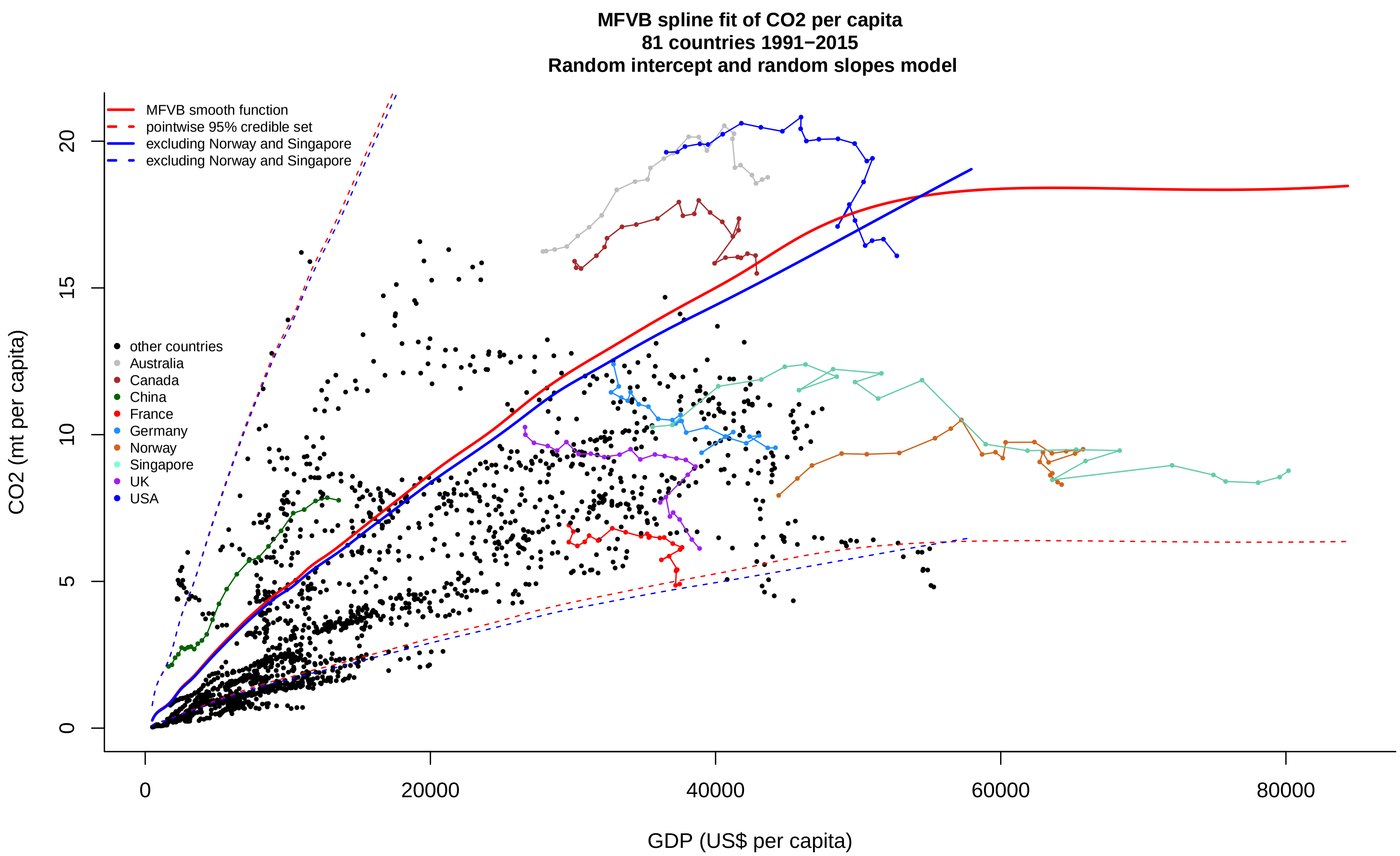

Figure 5. MFVB spline fit of the CO_2 emissions per capita against the observed GDP per capita and the pointwise 95% credible set using estimated values $\exp\left(\widehat{\beta}_1^R + \widehat{f}\left(\log\left(GDP/POP\right)\right)\right)$. The colored broken lines represent the observed per capita pairs (CO_2, GDP) for some countries.

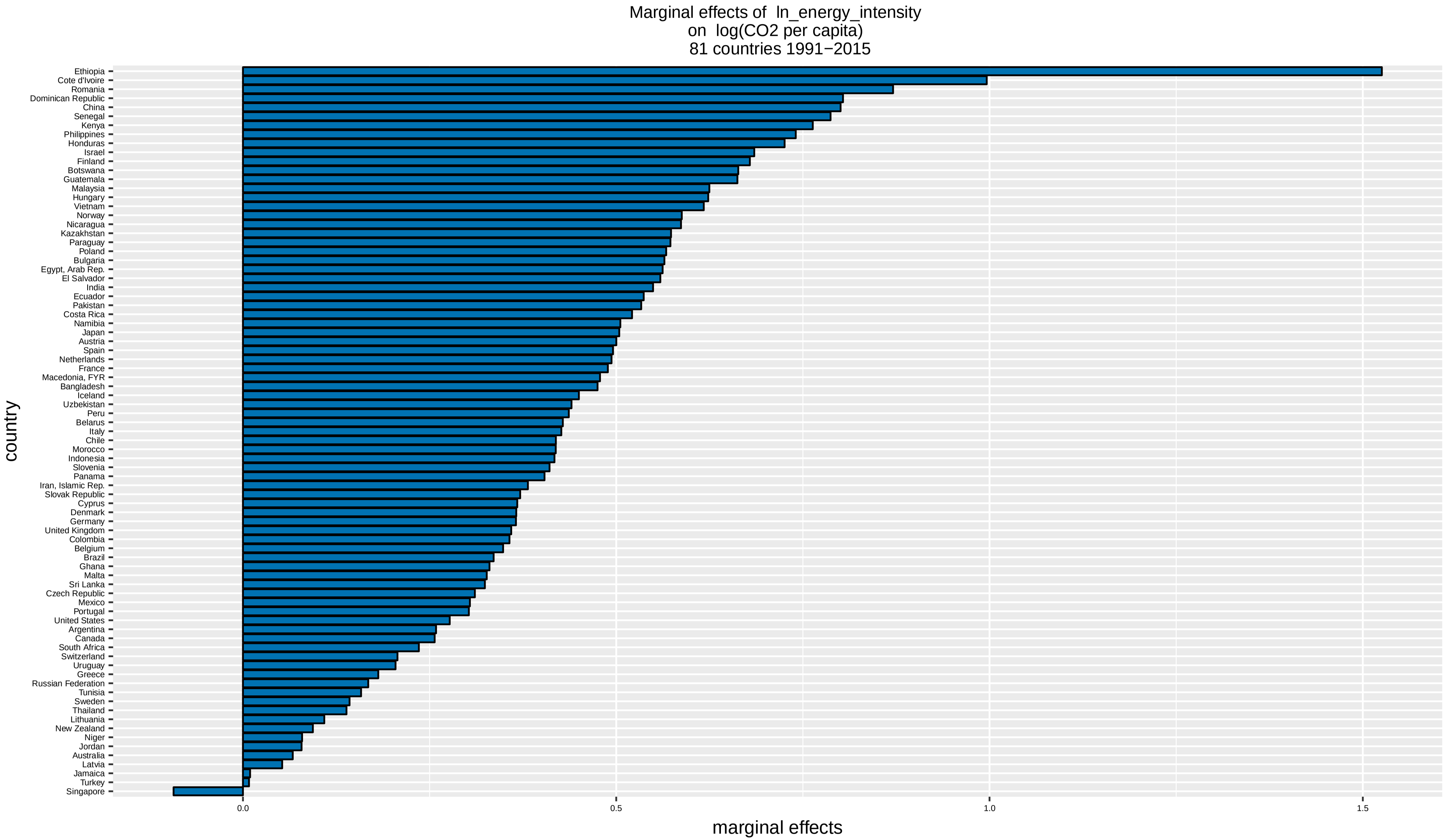

Figure 6. Estimated marginal effects $\left(\widehat{\beta}^R + \widehat{u}_i^R\right)$ for log-energy intensity per country $(i = 1, ..., 81)$.

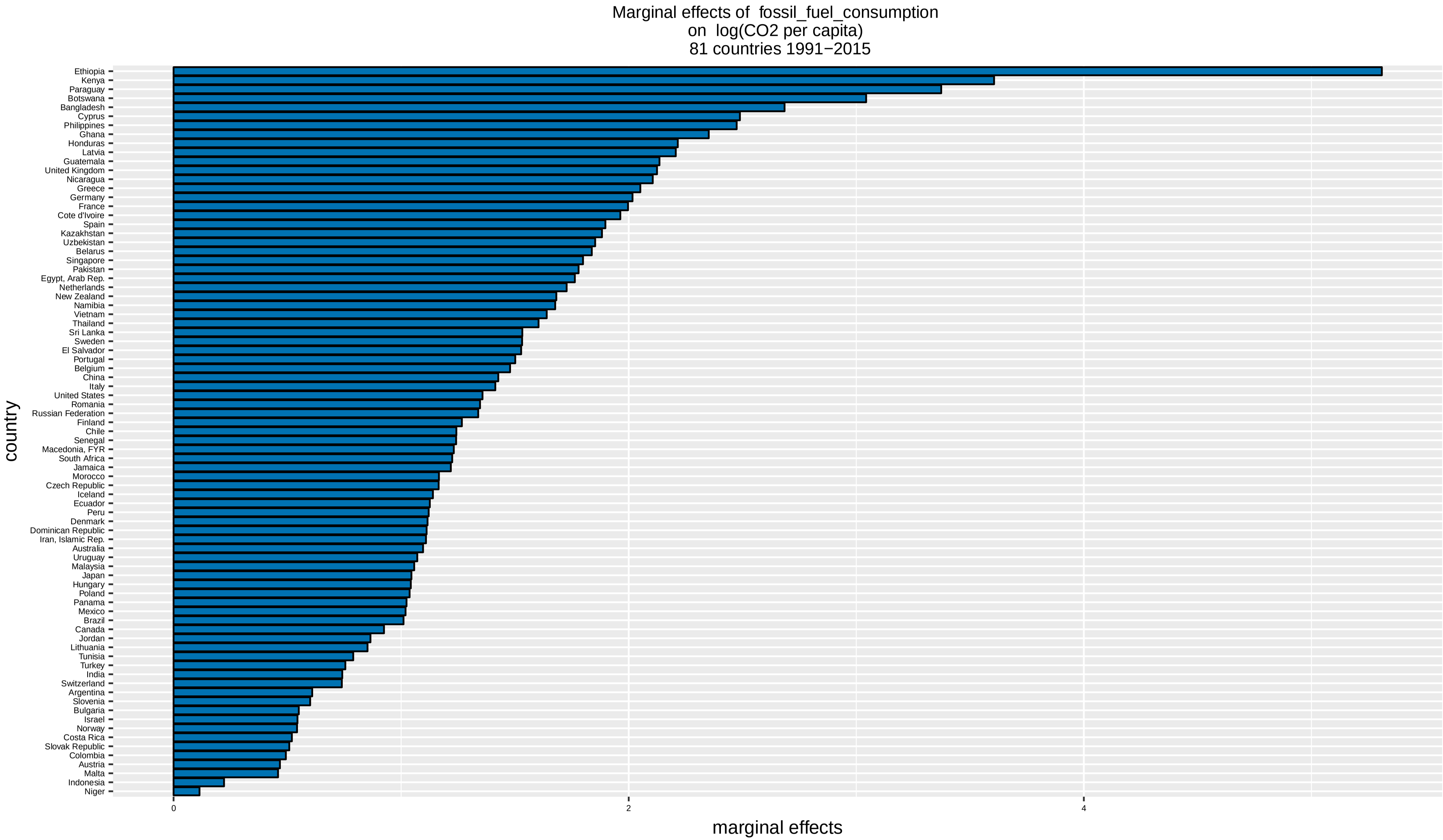

Figure 7. Estimated marginal effects $\left(\widehat{\beta}^{R} + \widehat{u}_i^{R}\right)$ for fossil fuel energy consumption per country $(i = 1, ..., 81)$.