RESEARCH ARTICLE

CrossMark

Markov-switching vector autoregressive neural networks and sensitivity analysis of environment, economic growth and petrol prices

Melike Bildirici[1] · Özgür Ersin[2]

Abstract

The paper aims at evaluating the nonlinear and complex relations between CO_2 emissions, economic development, and petrol prices to obtain new insights regarding the shape of the environmental Kuznets curve (EKC) in the USA and in the UK in addition to introducing a newly proposed nonlinear approach. Within this respect, the paper has three purposes: the first one is to combine the multilayer perceptron neural networks (MLP) with Markov-switching vector autoregressive (MS-VAR) type nonlinear models to obtain the MS-VAR-MLP model. The second is to utilize one of the largest datasets in the literature covering the 1871–2016 period, a long span of data starting from the late eighteenth century. Since the emission, economic development, and petrol price relation is subject to nonlinearity and trajectory changes due to many historical events, the development of the MS-VAR-MLP model is a necessity to contribute to the ongoing debate regarding the shape of the EKC curve and the stability of the relation. The third purpose is to develop the MS-VAR-MLP-based regime-dependent sensitivity analysis, which eases the visual interpretation of the nonlinear causal relationships, which are allowed to have asymmetric interactions in different phases of the expansionary and recessionary periods of the business cycles. Our results provide clear deviations from the findings in the literature: (i) the shape of the EKC curve cannot be assumed to be stable and is subject to regime dependency, nonlinearity, and magnitude dependency; (ii) the forecast results suggest that incorporation of regime switching and neural networks provide significant improvement over the MS-VAR counterpart; and (iii) for both USA and UK and for the 1871–2016 period, the positive impacts of economic growth on emissions cannot be rejected for the majority of the phases of the business cycles; however, the magnitude of this effect is at various degrees. In addition, the incorporation of petrol price provides significant findings considering its effects on emission and economic growth rates. The analysis suggest clear deviations from the expected shape of the EKC curve and puts forth the necessity to utilize more complex empirical methodologies to evaluate the EKC since the emissions-economic development relation is more complex than it was assumed. Following these findings, several policy recommendations are provided. Lastly, the proposed MS-VAR-MLP methodology is compared with the MS-VAR model and various advantages and disadvantages are enumerated.

Keywords Neural networks · Regime switching · Nonlinear econometrics · Granger causality · Environmental economics

Responsible editor: Muhammad Shahbaz

✉ Melike Bildirici
melikebildirici@gmail.com

Özgür Ersin
ozgurersin@gmail.com

[1] Department of Economics, Yildiz Technical University, Davutpasa, Istanbul, Turkey

[2] Department of Economics, Beykent University, Ayazaga-Sariyer, Istanbul, Turkey

Introduction

Environmental degradation and the role of humanity's carbon footprint create great concern on sustainable economic development. However, due to the unavailability of long-run and reliable data, literature has been rather scarce in terms of testing the impact of economic development on environmental degradation for data covering centuries long. Following the seminal papers of Grossman and Krueger (1991) and Shafik and Bandyopadhyay (1992), the environmental Kuznets curve (EKC) has become highly popularized and a growing body of research has emerged. According to the EKC hypothesis, the environmental degradation and economic development

relation is rather stable and follows an inverted U-shape, i.e., environmental degradation increases with a decreasing rate up to a certain threshold level of economic development after which the positive relation becomes reversed through various channels. The additional seminal papers that led to the emergence of the investigation of the EKC include Selden and Song (1994), Stern et al. (1996), and Cole et al. (1997).

One point that cannot be overlooked is that without a dataset that covers a large period that includes centuries, the possible existence of an inverted U-shaped relation between the environmental degradation and economic development as put forth by the EKC hypothesis is highly questionable. For this end, the paper contributes by focusing on one of the largest datasets starting from the eighteenth century covering the 1871–2016 period for the USA and for the UK. The main advantage of this approach is its coincidence with the historically long-run characteristic of the EKC hypothesis. The EKC hypothesis also requires a dataset that covers various stages of economic development in Rostow's terms, in addition to evaluating countries that underwent those stages. Another advantage of this approach is, if the complex relation between the analyzed variables could be modeled with an approach that captures the regime shifts and the varying nonlinear causal relations due to many historical events, the historically long data could be harnessed to provide significant information regarding the relations between the emissions, economic prosperity, and petrol prices. Nevertheless, such long periods of historical data also has a disadvantage, since it is also subject to trajectory changes that result in deviations from a stable relation between the analyzed variables. These trajectory changes include regime changes due to many historical events, such as policy changes, economic crises, wars, disputes, trade relations, and policies not to mention the industrial revolutions that took place in the history. Due to the trajectory changes in such long period of historical data, the relation between the economic development and emissions become more complex compared to samples covering one or two decades. Overall, the historically long data has significant advantages in terms of meeting the long-run characteristic of the EKC curve in addition to providing important insights by allowing the researcher to evaluate the relationship within a broad perspective by incorporating the impacts of the trajectory changes in the relationship. However, it is also a necessity to augment the econometric tools to control regime shifts, nonlinearity, and asymmetry in the nonlinear causal relations among the analyzed variables. Otherwise, the regime shifts and nonlinearity could result in biased estimates if regression analysis is utilized without controlling such effects.

As to be shown in the literature section, if an overlook to the debate regarding the EKC is to be presented, the reliance to linear or nonlinear in variables econometric models that assume stable relations between the variables constitutes the largest share. Nonlinear in variables models are based on polynomial regressions, i.e., squared and or cubic GDP per capita or other national income variables to evaluate the inverted U-shaped EKC in the pursuit of the determination of turning point or points. More recently, a few number of papers focus on nonlinear in parameters models, such as the threshold autoregressive (TAR) and Markov-switching auto-regressions (MS) in the spirit of Tong (1990) and Hamilton (1990). One of the earliest papers to our knowledge is Esteve and Tamarit (2012), who employ the TAR approach to evaluate the carbon dioxide emission-economic growth relation, since the TAR approach provides statistical determination of optimum thresholds or turning points within the spirit of EKC. Other attempts include the introduction of control variables, such as the carbon option pricing and energy consumption (Arouri et al. 2012a, b; Aslanidis and Iranzo 2009; Fezzi and Bunn 2009; Fosten et al. 2012; Kim et al. 2010; Yavuz and Yilanci 2013) and energy consumption and production to address the indeterminacy issue of the EKC (Richmond and Kaufmann 2006; Tucker 1995; Unruh and Moomaw 1998; Ghosh 2010; Ang 2007; Jalil and Mahmud 2009; Zhang and Cheng 2009; Sanjari and Delangizan 2010; Fodha and Zaghdoud 2010; Menyah and Wolde-Rufael 2010; Arouri et al. 2012a, b; Wang et al. 2011; Pao and Tsai 2010; Alam et al. 2012; Bildirici 2013; Bildirici and Gökmenoğlu 2017). In sum, the majority of the literature could be summarized as following linear econometric methodologies, including Granger causality, VAR, Johansen cointegration, Engle and Granger cointegration, structural VAR, linear panel regressions, and panel cointegration regressions, which fail to capture the nonlinearity in the emissions-economic development relationship, not to mention the trajectory changes that alter the stability of the EKC relation. More recent papers rely on linear autoregressive distributed lag (ARDL) methodology (Pesaran et al. 2001) with squared and cubic terms; however, the inclusion of such terms to cointegration vectors is rather problematic. The evaluation suggests the usage of more advanced empirical methods to evaluate the emissions and economic development relation since complexity is inherent in the relation. As a result, one linear relation or one stable inverted U effect of economic development on emissions fail to exist if a long period of historical data that is subject to continuously changing production technologies, economic policies, and technological advancements is taken into analysis.

To provide a contribution and to suggest a solution to the problem of indeterminacy of EKC, the study proposes the Markov-switching vector autoregressive multilayer perceptron (MS-VAR-MLP) model to capture the complex nonlinear relations between the emissions, economic development, and the petrol prices. Petroleum is a fossil fuel-based non-renewable commodity, and its consumption has strong environmental implications on the emissions. Further, petrol prices are shown to have significant nonlinear impacts on economic growth rates and on the business cycles in addition to being influential on the historical economic crises

(Hamilton 1990, 2011; Krolzig and Clements 2002). Pan et al. (2017) shows that petrol prices have time-varying nonlinear effects on the U.S.'s GDP and the predictability of the GDP fluctuations. Rehman (2018) shows the strong association of petrol price fluctuations and global economic policy changes, including the developed and developing nations. Winchester and Ledvina (2017) provide a global economy wide model and by simulating oil price fluctuations, they show that high (low) petrol prices lead to lower (higher) GHG emissions.

One additional contribution of the paper is the extension of the sensitivity analysis used for neural networks for the analysis of regime-dependent sensitivity analysis that enables the investigation of nonlinear causal relations. Further, the paper utilizes one of the largest datasets covering the 1871–2016 period for UK and USA based on the following: (i) both countries have passed the necessary stages of development as in the spirit of Rostow (1960), and (ii) both are at the 5th stage of economic development, the high mass consumption. (iii) To summarize, the study has three aims to contribute to the literature. The first is to investigate the relationship between economic growth, CO_2 emissions, and petrol prices for one of the largest datasets available in the spirit of the EKC hypothesis with the newly proposed MS-VAR-MLP- and MS-VAR-MLP-based regime-dependent sensitivity methodology. The second aim is to investigate the complex relationship between the petrol prices, environmental pollution, and economic growth, which cannot be captured adequately, especially by polynomial regressions. The MS-VAR approach assumes regime-switching determined with Markov chains between to sub-sample spaces that follow VAR processes. In contrast, the MS-VAR-MLP allows MLP-type nonlinear neural network processes within each sub-sample space. Further, due to the trajectory changes resulting from oil shocks, economic crises, policy changes, and wars in such a historical data, the MS-VAR model could fail to capture only the abrupt shifts in the time series and could result in non-rejection of certain regimes that include the large decrease or increases in magnitudes of variables, such as GDP per capita. The third aim of the paper is to propose MS-VAR-MLP-based regime-dependent sensitivity analysis that could be used to provide regime-dependent causal interpretations between the analyzed variables. As to be shown, the MS-VAR-MLP and its regime-dependent sensitivity analysis provide nonlinear causal interpretations of the analyzed variables within each distinct regime, and the causal impact is nonlinear even in each regime allowing the researcher to investigate complexity in the nonlinear causality. The usage of the large span of data is also fruitful in such a complex nonlinear model, since the sample covers many important factors, including world wars and economic crises especially those caused by the first and second oil shocks in 1970s. This type of regime and magnitude dependent sensitivity analysis allows the researcher to easily

analyze the causal relations between the variables at different stages of business cycles in addition to overcoming the difficulties of the MS-VAR models.

In the MS-VAR literature, the usage of filter selection, i.e., the Kalman or the Hamilton filters, has been discussed in many studies. An additional difference between the MS-VAR and MS-VAR-MLP is, with the use of the neural network learning algorithms in the estimation, the MS-VAR-MLP model is capable of modeling such large span of data without the application of any filters, since the hidden Markov chains are inherent in the MLP models. As a result, the dataset could be directly estimated as their raw form, since the NN learning algorithms are efficient in overcoming this difficulty.

The literature review is given in "Literature review." The econometric methodologies utilized in this study are provided in "Empirical methodology." The data, empirical results and discussions are given in "Dataset and the empirical results." The section "Conclusion" concludes.

Literature review

Similar to our study in the sense of using a large set of data, Lindmark (2002), Stern and Enflo (2013), Fosten et al. (2012), Ersin (2016), and Bildirici and Ersin (2018) analyze a long range of dataset starting from the early nineteenth century. Among these studies, Fosten et al. (2012) is one of the early studies utilizing nonlinear in parameters time series models, namely TAR and momentum-TAR models with data starting from 1830 for the UK. With the use of the TAR models, Fosten et al. (2012) conclude that the positive relation between economic growth and emissions in the UK is reversed after a certain threshold and any temporary disequilibrium is corrected. Further, by evaluating the candidate variables that optimize this threshold effect, one interesting finding of Fosten et al. (2012) is, the reversal is not a function of or a threshold of per capita GDP, but a function of per capita CO_2 emissions and the results are justified as the emissions are reduced by legislation rather than being a function of economic growth. Mensah (2014) uses the threshold analysis to evaluate the CO_2, energy consumption and economic growth relationship and displays that structural changes could result in alterations in the economic activity-environment relation in addition to nonlinearity. Zi et al. (2016) employ TAR models to inspect the effects of urbanization on CO_2 emissions in China by taking the level of urbanization, residential income, urban population density, and industrial production as threshold variables. Ersin (2016) studies the nonlinear effects of GDP on CO_2 emissions for a long span of data during 1870–2011 for 13 developed countries with heterogenous panel smooth transition autoregressive (Panel-STAR) framework, and his empirical findings suggest that the

effects of economic growth on emissions growth had been positive in both regimes, a finding being the opposite of the EKC or a weak form of EKC. Bildirici and Ersin (2018) suggested the STAR distributed lag (STARDL) model that augments the STAR models (Terasvirta 1994) with nonlinear ARDL cointegration framework to investigate long-run and short-run relations between CO_2 emissions and economic growth in the USA.

A literature search on the most recent empirical analysis on the EKC type environment-income relationship could be summarized as being evolving under three strands in addition to the recent STAR approaches to EKC given above. The first strand utilizes the polynomial regressions and this strand constitutes to the majority of the literature. Among the first strand, Atasoy (2017) investigates the EKC for 50 states of USA with panel regressions by using the CO_2, GDP per capita, and its square, energy consumption, and population variables for the 1960–2010 period. Atasoy (2017) concludes that only a weak form of EKC exists, since the inverted U-shape could be achieved only in 10 out of 50 states. Another recent study, which investigates the possibility of EKC in the states of the USA for the 1960–2010 period is Apergis (2016), whose findings are also in favor of no EKC for 38 out of 48 states in the USA. Özokçu and Özdemir (2017) estimate panel models with squared and cubic variables of CO2 emissions and economic growth, and their findings suggest N-shaped and inverted N-shaped relations for 26 OECD and 52 emerging countries, i.e., findings against an EKC type relation. Özokçu and Özdemir (2017) suggest that the results show that environmental pollution could not be inverted by economic growth. Xu (2018) examines the EKC with sulfur dioxide, GDP, and squared GDP variables for China at the aggregate level and at the state level. Xu (2018) concludes that though EKC is shown at the aggregate (country) level, however, the disaggregate data (state level) do not support the EKC; therefore, the regression models suffer from biased parameter estimates due to aggregation. Zambrano-Monserrate et al. (2018) test the EKC hypothesis with ARDL and additionally with causality tests and conclude that inverted U-shape cannot be observed in Peru for the 1980–2011 period. Rasli et al. (2018) examine the EKC relation with "new toxics" (various forms of pollutants including CO2) for a panel of 36 developed and developing countries for the period of 1995–2013, and they conclude that EKC does not hold for models with CO2 and nitrous oxide (N_2O) models, EKC holds for traffic volume and CO2 emissions and for the majority of evaluated models, trade openness and energy consumption worsen the environment. Sinha and Shahbaz (2018) examine the EKC in India by the use of CO2 emissions, renewable energy and its quadratic form, and they suggest non-rejection of EKC resulting from the inclusion of renewables, which underlines the importance of the need to assent renewable energy.

The second strand focuses on MS and MS-VAR models to augment the EKC with regime dependency and business cycles. In the second strand, Halkos and Tsionas (2001) evaluate the EKC with the Bayesian regime switching along with cross-sectional PPP adjusted per capita GDP, population density, infant mortality rates, urban population distribution of GDP in manufacturing, deforestation, and CO_2 data for 61 countries. Halkos and Tsionas (2001) also point at the complexity of the emission-income relation, since they conclude that a monotonic relation between emissions and income, such as the EKC cannot be accepted. Halkos and Tsionas (2001) also point at the inability of the demographic and income variables in identifying the differences between high and low polluters for the analyzed 61 countries. Within this respect, Plassmann and Khanna (2006) criticize the nonmonotonic EKC result obtained, and they note problems associated with the use of aggregated and pooled data under three sources of bias: (i) GDP data is country level, while emissions data is for certain cases area-specific; (ii) usage of panels is highly problematic, since countries are not homogenous; (iii) changes in the state of technology and consumer awareness cause shifts in the income-pollution relation, and panels and regressions could not capture this effect. Though Plassmann and Khanna (2006) suggest micro/household level data to overcome difficulties, they underline the impact of assumptions of the model selected in the literature, and they suggest the utilization of Markov chain Monte Carlo (MCMC) and Gibbs sampler to overcome the problems encountered. Martinez-Zarzoso and Maruotti (2013) investigate the EKC curve with hidden Markov regression models for a sample of 28 OECD countries for the 1968–2006 period, and they conclude that the emissions-income relation varies depending on the analyzed countries, the counties could be assumed as falling into five distinct groups and distributional mixture models provide flexibility instead of assuming a single distribution. Roach (2015) shows results in favor of regime dependency with MS models for the real GDP, retail prices, and motor-gasoline CO_2. Park and Hong (2013) investigate the EKC relation with two regime MS and MS-RW (MS-random walk) models for South Korea. Chevallier (2011a) obtains MS-VAR type nonlinearity in the EKC type relation between macroeconomic variables and CO_2. Chevallier (2011b) extends the analysis to evaluate the nonlinear and regime-dependent relations between the carbon prices and macroeconomy for EU-27 member countries with MS-VAR models. The findings of Chevallier (2011a, b) suggest that the MS-VAR models capture the relation more effectively, the regime dependency characteristics could not be rejected, the two states of boom-bust cycle exist for the analyzed countries, and the industrial production has a positive impact on the carbon prices during expansions, while the relation becomes reverted for the recessions. Bildirici and Gökmenoğlu (2017) investigate environment-income relation

in the context of CO_2 and energy consumption from hydropower for the G7 countries with MS-VAR models and their findings are in favor of decelerating effect of hydropower energy on environmental degradation. Charfeddine (2017) applies the MS-VEC (MS-vector error correction) model to examine the economic development-CO_2 relation, and the results are in favor of EKC hypothesis. Charfeddine (2017) also notes that if the regime-switching is not controlled, the structural breaks result in deviations from true relations; hence, the utilization of nonlinear model is a necessity.

The differences of our approach compared to Halkos and Tsionas (2001) and Plassmann and Khanna (2006), instead of using cross-sectional data, utilize a large set of historical time series data, and to overcome the difficulties resulting from technology and awareness changes, we apply the MS-VAR-MLP model. The model achieves the regime dependency by benefiting from the MS-VAR and the flexibility with the use of MLP type NNs. As to be mentioned, the MS-VAR-MLP is highly flexible in terms of producing within-regime nonlinear causal relations by assuming MLP processes in each regime without prior assumption regarding the distribution compared to the MS-VAR model that assumes linear VAR relations in each regime. Similar to Charfeddine (2017), regime-switching could be utilized to capture structural changes; however, as to be shown in the last section, the forecast performance of the proposed MS-VAR-MLP improves significantly over its MS-VAR counterpart. In contrast to Chevallier (2011a, b) and Bildirici and Gökmenoğlu (2017) that benefit from the MS-VAR model, in the MS-VAR-MLP model, the relationship between the analyzed variables and their nonlinear effects are maintained to be nonlinear even within each regime following MLP processes. As a result, the model captures the nonlinearity within the expansionary and recessionary regimes occurring as a nonlinear function of the size and magnitude of each variable under the investigation.

A third strand in the literature follows more complex methodologies, such as the neural networks and machine learning approaches instead of the previous models, including polynomial regressions, panel regressions, and nonlinear econometric models, such as the MS or TAR. However, the number of the papers in this strand is rather limited. Sun and Liu (2016) investigate the EKC in three major industries in addition to residential consumption in China for the 1978–2012 period based on the support vector machine (SVM) and back-propagation neural networks (BP-NN) in addition to applying cointegration and Granger causality tests. The findings of Sun and Liu (2016) reveal that SVM model performs better in forecast accuracy compared to BP-NN. Liu et al. (2018) conducts a BP-NN -based modeling to augment the forecast capabilities, in addition to applying ridge regression and the evaluation of systematic dynamics to determine key factors of regional carbon emissions in Beijing. Following their empirical results, Liu et al. (2018) suggest various important policies, including urbanization and green transportation strategies. One drawback

of these studies is that their sample size is very limited and the estimation of neural networks and machine learning models require large number of time series observations. Compared to our paper, though nonlinear, their models do not utilize regime-switching, and they implicitly ignore the impact of business cycles, in addition to assuming the proposed models to be capturing the complex relations by ignoring the trajectory changes even within the limited time span of their data. Similar to these papers, our model also necessitates the use of large datasets, and one of the largest datasets is introduced in the study. However, as to be discussed in the last section, the requirement of large datasets is also a limit for the MS-VAR-MLP and for the future applications.

Considering the abovementioned strands, our paper could be taken as following the first and second strands in terms of the usage of historically long span of data and by taking the effects of business cycles and regime-switching in the spirit of MS-VAR approach. By providing a hybrid methodology, the MS-VAR-MLP model and its generalization to MS-VAR-MLP-based regime-dependent sensitivity analysis our methodology has important advantages. The model assumes Markov-switching-based splitting of the regression space into two or possibly more sub-regression spaces, where the regime switching is governed by unobserved Markov chains. In contrast to the MS-VAR models, the relations between the analyzed variables are highly nonlinear, since each regime follows VAR-MLP processes in the MS-VAR-MLP model. As noted in the previous section, this type of modeling strategy becomes relevant, especially due to the complexity of the emissions-income-petrol relation in such historically long period of data corresponding to the 1871–2016 period. Further, the MS-VAR-MLP and the sensitivity approach benefits not only from the hybridization of regime dependency and neural networks but also from the neural network learning algorithms. The overlook is that our paper follows the historically long data approach to EKC in the spirit of Lindmark (2002), Stern and Enflo (2013), Fosten et al. (2012), and Ersin (2016); the paper provides a bridge between the literature suggesting the need to take nonlinearity and MS-type regime switching into account (Roach 2015; Park and Hong 2013; Chevallier (2011a, b); Bildirici and Gökmenoğlu 2017), and our paper provides regime-switching VAR-MLP neural networks modeling of emissions and economic growth relation in relation to its single regime variants of BP-NN models as in Sun and Liu (2016) and Liu et al. (2018). Further, the proposed MS-VAR-MLP model follows the statistical and stochastic neural network modeling approach of Cheng and Titterington (1994) and Ersin (2009) in addition to extending the single regime VAR-NN model proposed in Wutsqa et al. (2006).

Empirical methodology

At the first stage, a general outlook to MS-VAR models will be provided. At the second stage, the MLP neural networks

and neural networks sensitivity analysis will be extended to the MS-VAR-MLP model and MS-VAR-MLP-based regime-dependent sensitivity analysis.

MS-VAR model

This perspective is also in the spirit of Hamilton (1990) that influenced many studies. Hamilton (1990) showed that petrol prices and sudden changes in petrol prices had significant importance on dating the expansions, recessions, peak-through dating, and durations of these cycles based on the Markov-switching (MS) models. Krolzig (1998, 2000), Krolzig and Toro (2005), and Krolzig and Clements (2002) extended the analysis of national income cycles to MS-VAR models to provide visual representations with impulse response functions.

The MSI(r)-VAR(l) model is defined as follows:

$$y_t = \mu^{(s_t)} + \sum_{i=0}^{i} A_i^{(s_t)} x_t + u_t^{(s_t)} \quad (1)$$

where $u_t/s_t \sim N(0, \delta^2(s_t))$ and A_i (.) show the coefficients of the lagged values of the variable in different regimes. $\delta^2(s_t)$ shows the variance of the residuals in each regime. $\mu(s_t)$ defines the dependence of mean μ of the K-dimensional time series vector on the regime variable s_t. The input variables are defined in matrix form as $x_t = [x_t']' = (y_{t-1}, ..., y_{t-p}, x_{t-1}, ..., x_{t-p})'$ for $t = 1,2,...,n$ number of observations. r is the number of regimes and l is the optimum lag length selected with Akaike or Schwarz information criteria (AIC and SIC). s_t is governed by a Markov chain as follows:

$$P_r\left[s_t | \{s_{t-1}\}_{i=1}^{\infty}, \{y_{t-1}\}_{i=1}^{\infty}\right] = P_r\{s_t | s_{t-1}; \rho\}, \quad (2)$$

where p includes the probability parameters. The conditional probabilities are stated as $P(y_t | Y_{t-1}, s_{t-1}) = P_r(y_t | Y_{t-1})$.

The Markov chain is ergodic and irreducible, and an absorbing state does not exist, i.e., $\overline{\xi}_p \in (0, 1)$ for all $m = 1,...,M$ and $\overline{\xi}_p$ which itself is an ergodic or unconditional probability of regime q. p_{ij} has unconditional distribution given by the following:

$$\Pr(s_t = 1) = \frac{1-p_{22}}{2-p_{11}-p_{22}}, \Pr(s_t = 2) = \frac{1-p_{11}}{2-p_{11}-p_{22}} \quad (3)$$

For the estimation of the Markov-switching models, the maximum likelihood estimators (MLE) and the expectation maximization (EM) algorithm are possible approaches. To make an inference, an iterative procedure for $t = 1,2,...,T$ is followed, while the previous value of the probability $\xi_{it-1} = P_r[s_{t-1} = i | \Omega_{t-1}; \theta]$ is taken as an input. $\varepsilon_{t|t}$ denotes the vector of forecast probabilities. Optimal forecast probabilities are obtained using the following:

$$\varepsilon_{t|t} = \frac{\varepsilon_{t|t-1}\varphi_t}{1'\left(\varepsilon_{t|t-1}\varphi_t\right)'} \quad (4)$$

where $\varepsilon_{t+1|t} = P'\varepsilon_{t|t}$ and φ_t is the vector of conditional densities and 1 is a unit column vector with element-by-element multiplication. The estimation is conducted with the following:

$$E_t(y_{t+1}) = \sum_{j=1}^{s} \sum_{i=1}^{s} \Pr_t(S_t = j) P_{ij} \left(w_0^{(j)} + \sum_{l=1}^{p(j)} \beta_l^{(j)} y_{t-l+1} \right) \quad (5)$$

MS-VAR-MLP model

The Hamilton (1990) model is a nonlinear mixture of autoregressive functions, such as the MLP and is evaluated under the hybrid of the MLP and the hidden Markov chain (HMC) model (Olteanu et al. 2004). However, the estimation of the MS-VAR model is generally based on the EM and MLE approaches without benefiting from the neural networks learning algorithms. In fact, the hybrid MLP accepts that the input layer is linked to the output nodes with weighted connections to form a linear model that is parallel to the nonlinear multi-layer perceptrons as shown by Bildirici and Ersin (2013). Moreover, Wutsqa et al. (2006) extended the Sims's (1980) VAR model to NN by proposing a hybrid approach, the VAR-NN model. The model architecture of the VAR-NN model of Wutsqa et al. (2006) is defined with linear activations to mimic the VAR model, and the model is estimated with the BP, one of the most common algorithms in the NN literature.

The MLP is a nonlinear function $N(.)$ that links the variables in the input layer $\tilde{x}_t$ to the output layer y_t of the form as follows:

$$y_t = N\left(\tilde{x}_t; \varphi, f\right) + \varepsilon_t \quad (6)$$

which could be defined as a single hidden layer MLP model as follows:

$$y_t = \varphi\left(\gamma_0 + \sum_{i=1}^{h} \lambda_i \psi\left(\tilde{\omega}' x_t\right)\right) + \varepsilon_t \quad (7)$$

Through a set of the different activation functions $\psi(\tilde{\omega}' x_t)$ is possible (Bishop 1995), the study assumes taking the sigmoid function as follows:

$$\psi\left(\tilde{\omega}' \tilde{x}_t\right) = \left(1 + e^{-\left(\tilde{\omega}' x_t - \omega_0\right)}\right)^{-1} \quad (8)$$

Hence, Eq. (8) is continuous and twice differentiable and is bound in the range of [0,1]. Other possible activation functions include the hyperbolic tangent, sine, heavy side, and the

identity functions (Bishop 1995). For the universal approximation property of MLPs with sigmoidal functions, readers are referred to Cybenko (1989). Following Kuan and White (1994), Swanson and White (1997), White (1992), and Granger and Terasvirta (1993), the ε_t random shocks are incorporated as $\varepsilon_t \sim \mathrm{iid}N(0, \delta^2)$ white noise process. It should be noted that Eq.(7) is defined as a single regime VAR-MLP model as long as the output variable vector and input variable matrix are defined as $y_t = (y_t, x_t)'$, and the input variable matrix is $x_t = [x_t']' = (y_{t-1}, ..., y_{t-l}, x_{t-1}, ..., x_{t-l})'$, in addition to the connection parameters given as $\tilde{\omega}' = [\omega_{0,1}, ..., \omega_{0,h}, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q]'$

By extending the model to regime switching, following the methodology given for the MS-VAR model, the MS-VAR-MLP model is defined as follows:

$$y_t = \varphi^{(s_t)}\left(\gamma_0^{(s_t)} + \sum_{i=1}^{h} \lambda_i^{(s_t)} \psi\left(\tilde{\omega}'^{(s_t)} \tilde{x}_t\right)\right) + \varepsilon_t^{(s_t)} \quad (9)$$

where

$$\psi\left(\tilde{\omega}'^{(s_t)} \tilde{x}_t\right) = \left(1 + \exp\left(-\left\{\tilde{\omega}'^{(s_t)} \tilde{x}_t\right\}\right)\right)^{-1} \quad (10)$$

Similar to the VAR-NN model, which could be considered as a single regime MS-VAR-NN model, the output activation function in the hidden layer in each regime is defined as a linear activation function for the sake of simplicity. As shown

in the literature, this type of specification decreases the number of parameters to be estimated without a loss in the generalization capability of the NN model (Bishop 1995). The input vector is defined as $\tilde{x}_t = [1, x_t']'$, and a vector of unity is added to obtain the bias coefficient; the input variables are defined as $x_t = [x_t']' = (y_{t-1}, ..., y_{t-l}, x_{t-1}, ..., x_{t-l})'$ and $t = 1, 2, ..., n$, with n representing the number of observations. The MS-VAR-MLP model in Eq. (9) extends the model to m number of regimes with transition probabilities being governed by Markov chains and the connection parameters are defined as $\tilde{\omega}' = \left[\omega_{0,v,1}^{(s_t)}, ..., \omega_{0,v,h}^{(s_t)}, \alpha_{1,v}^{(s_t)}, ..., \alpha_{l,v}^{(s_t)}, \beta_{1,v}^{(s_t)}, ..., \beta_{l,v}^{(s_t)}\right]'$, v defining the vector of coefficients as in the VAR and MS-VAR models. l is the optimum lag length determined by the Schwarz information criterion (SIC) for parsimony. The hidden unit is linked to the output layer with the hidden unit parameters $\lambda_{i,v}^{(s_t)} = \lambda_{1,v}^{(s_t)}, ..., \lambda_{h,v}^{(s_t)}$, where h is the number of hidden units and $\gamma_0^{(s_t)}$ is the bias parameter. Similarly, the residuals of the model are assumed to follow $\varepsilon_{t,v}^{(s_t)} \sim \mathrm{iid}N$ $\left(0, \delta_v^{2,(s_t)}\right)$ white noise processes with zero conditional mean and constant variance in each regime, while the conditional variances are allowed to be regime dependent.

Another representation of the MS-VAR-MLP model in Eq. (9) and Eq. (10) is stated as follows:

$$y_{t,v}^{(s_t)} = \varphi^{(s_t)}\left(\gamma_0^{(s_t)} + \sum_{i=1}^{h} \lambda_i^{(s_t)} \times \left(1 + \exp\left\{-\left(\alpha_{1,v}^{(s_t)} y_{t-1} + \alpha_{2,v}^{(s_t)} y_{t-2} +, ..., +\alpha_{l,v}^{(s_t)} y_{t-l} + \beta_{1,v}^{(s_t)} x_{t-1} + \beta_{2,v}^{(s_t)} x_{t-2} +, ..., +\beta_{l,v}^{(s_t)} x_{t-l} + \omega_{0,v}^{(s_t)}\right)\right\}\right)^{-1}\right) + \varepsilon_{t,v}^{(s_t)} \quad (11)$$

the regimes are governed by the unobservable Markov process are as follows:

$$\sum_{i=1}^{m} y_{t(i)} P(S_t = i | z_t), i = 1, ...m. \quad (12)$$

Within each regime, the logistic-type sigmoid function is defined as in Eq. (10). The parameter vector is subject to uniform transformation as follows:

$$\left(^1/_2\right) \lambda_h^{(s_t)} \sim uniform[-1, +1] \quad (13)$$

and $P(S_t = i | z_t)$, the filtered probability function has the following representation as follows:

$$(P(s_t = i | z_t) \alpha f(P(\sigma_{t-1} | z_t, s_{t-1} = 1))) \quad (14)$$

The transition probability $n_{j,i}$ is assumed to be $P(S_t = i | S_{t-1} = j)$. Following Bildirici and Ersin (2009), the input variables are normalized as follows:

$$z_t = [z_t - E(z_t)] / \sqrt{E(z_t^2)} \quad (15)$$

The $s \rightarrow \max\{p, q\}$ recursive procedure is started by constructing $P(z_s = i | z_{s-1})$ where $\psi(Z_t \lambda_h)$ takes the form $1/(1 + \exp(-x))$ with twice-differentiability, continuity properties, and is bounded [0, 1]. The weight vector is defined as $\xi = \omega$, and as before, ψ is the logistic activation function. The input variables are defined as $Z_t \lambda_h = X_i$ with λ_h defined in Eq. (13). If $n_{j,i}$ transition probability $P(z_t = i | z_{t-1} = j)$ is assumed, the following representations are obtained as follows:

$$f(y_t | x_t, z_t = i)$$

$$= \frac{1}{\sqrt{2\pi h_{t(i)}}} \exp\left\{-\left(y_t - x_t'\varphi - \sum_{j=1}^{H} \theta_j p\left(x_t'\gamma_j\right)\right)^2 / 2h_{t(j)}\right\} \quad (16)$$

$$L = \prod_{t=1}^{T} f(y_t | s_t = i, Y_{t-1}) \Pr[s_t = i | Y_{t-1}] \quad (17)$$

The probability $\Pr[s_t = i|Y_{t-1}]$ is calculated through iteration as follows:

$$\pi_{jt} = \Pr[s_t = j|Y_{t-1}]$$
$$= \sum_{i=0}^{1} \Pr[s_t = j|s_{t-1} = i]\Pr[s_t = j|Y_{t-1}]\sum_{i=0}^{1}\eta_{ji}\pi_{it-1}^* \quad (18)$$

In terms of the variables under investigation of the study, namely per capita CO2 emissions, per capita GDP, and petrol prices, assume that their growth rates are denoted as $dlco2_t$, dly_t, and dlp_t after taking logarithms and after taking first differences. For the purpose of the study, the MS-VAR-MLP model given in Eq.(9) and (10) could be written as a three-vector MS-VAR model in which each vector follows MLP processes as follows:

$$dlco2_t = \varphi^{(s_t)}\left(\gamma_0^{(s_t)} + \sum_{i=1}^{h}\lambda_{i,v}^{(s_t)}\psi\left(\tilde{\omega}^{\prime(s_t)}\tilde{x}_t\right)\right) + \varepsilon_{t,v}^{(s_t)}$$

$$dly_t = \varphi^{(s_t)}\left(\gamma_0^{(s_t)} + \sum_{i=1}^{h}\lambda_{i,v}^{(s_t)}\psi\left(\tilde{\omega}^{\prime(s_t)}\tilde{x}_t\right)\right) + \varepsilon_{t,v}^{(s_t)} \quad (19)$$

$$dlp_t = \varphi^{(s_t)}\left(\gamma_0^{(s_t)} + \sum_{i=1}^{h}\lambda_{i,v}^{(s_t)}\psi\left(\tilde{\omega}^{\prime(s_t)}\tilde{x}_t\right)\right) + \varepsilon_{t,v}^{(s_t)}$$

where the sigmoid activation function is defined as follows:

$$\psi\left(\tilde{\omega}^{\prime(s_t)}\tilde{x}_t\right) = \left(1 + \exp\left(-\left\{\tilde{\omega}^{\prime(s_t)}\tilde{x}_t\right\}\right)\right)^{-1} \quad (20)$$

Hence, for Eq.'s (19) and (20), the input vector is defined as $\tilde{x}_t = [1, x_t']'$ where $x_t = [x_t']' = (dlco2_{t-1}, ..., dlco2_{t-l}, dly_{t-1}, ..., dly_{t-l}, dlp_{t-1}, ..., dlp_{t-l})'$. The MS-VAR-MLP model is the theoretic representation for r number of regimes, where the transition probabilities being governed by Markov chains. Given the fact that the model requires the estimation of a large number of connection parameters, $\tilde{\omega}' = \left[\omega_{0,v,1}^{(s_t)}, ..., \omega_{0,v,h}^{(s_t)}, \alpha_{1,v}^{(s_t)}, ..., \alpha_{l,v}^{(s_t)}, \beta_{1,v}^{(s_t)}, ..., \beta_{l,v}^{(s_t)}, \theta_{l,v}^{(s_t)}, ..., \theta_{l,v}^{(s_t)}\right]'$, the hidden unit parameters $\lambda_{i,v}^{(s_t)} = \lambda_{1,v}^{(s_t)}, ..., \lambda_{h,v}^{(s_t)}$, where h is the number of hidden units and v defines each vector for each state s_t. A simplification is made in the study by restricting the number of regimes to $r = 1,2$; two states representing the recessionary and expansionary regimes. The residuals are assumed to follow $\varepsilon_{t,v}^{(s_t)} \sim iidN\left(0, \delta^{2,(s_t)}\right)$ white noise processes with regime-dependent conditional variances in the spirit of the MSIAH-VAR model. The model given in Eq.'s (19) and (20) could also be written as follows:

$$dlco2_t^{(s_t)} = \varphi_v^{(s_t)}\left(\gamma_{0,v}^{(s_t)} + \sum_{i=1}^{h}\lambda_{i,v}^{(s_t)} \times \left(1 + \exp\left\{-\left(\alpha_{1,v}^{(s_t)}dlco2_{t-1} + \alpha_{2,v}^{(s_t)}dlco2_{t-2}+, ..., +\alpha_{l,v}^{(s_t)}dlco2_{t-l} + \beta_{1,v}^{(s_t)}dly_{t-1} + \beta_{2,v}^{(s_t)}dly_{t-2}+, ..., +\beta_{l,v}^{(s_t)}dly_{t-l} + \theta_{1,v}^{(s_t)}dlp_{t-1} + \theta_{2,v}^{(s_t)}dlp_{t-2}+, ..., +\theta_{l,v}^{(s_t)}dlp_{t-l} + \omega_{0,v}^{(s_t)}\right)\right\}\right)^{-1}\right) + \varepsilon_{t,v}^{(s_t)}$$

$$dly_t^{(s_t)} = \varphi_v^{(s_t)}\left(\gamma_{0,v}^{(s_t)} + \sum_{i=1}^{h}\lambda_{i,v}^{(s_t)} \times \left(1 + \exp\left\{-\left(\alpha_{1,v}^{(s_t)}dlco2_{t-1} + \alpha_{2,v}^{(s_t)}dlco2_{t-2}+, ..., +\alpha_{l,v}^{(s_t)}dlco2_{t-l} + \beta_{1,v}^{(s_t)}dly_{t-1} + \beta_{2,v}^{(s_t)}dly_{t-2}+, ..., +\beta_{l,v}^{(s_t)}dly_{t-l} + \theta_{1,v}^{(s_t)}dlp_{t-1} + \theta_{2,v}^{(s_t)}dlp_{t-2}+, ..., +\theta_{l,v}^{(s_t)}dlp_{t-l} + \omega_{0,v}^{(s_t)}\right)\right\}\right)^{-1}\right) + \varepsilon_{t,v}^{(s_t)}$$

$$dlp_t^{s_t} = \varphi_v^{(s_t)}\left(\gamma_{0,v}^{(s_t)} + \sum_{i=1}^{h}\lambda_{i,v}^{(s_t)} \times \left(1 + \exp\left\{-\left(\alpha_{1,v}^{(s_t)}dlco2_{t-1} + \alpha_{2,v}^{(s_t)}dlco2_{t-2}+, ..., +\alpha_{l,v}^{(s_t)}dlco2_{t-l} + \beta_{1,v}^{(s_t)}dly_{t-1} + \beta_{2,v}^{(s_t)}dly_{t-2}+, ..., +\beta_{l,v}^{(s_t)}dly_{t-l} + \theta_{1,v}^{(s_t)}dlp_{t-1} + \theta_{2,v}^{(s_t)}dlp_{t-2}+, ..., +\theta_{l,v}^{(s_t)}dlp_{t-l} + \omega_{0,v}^{(s_t)}\right)\right\}\right)^{-1}\right) + \varepsilon_{t,v}^{(s_t)}$$

$$(21)$$

The estimation of the model is conducted with learning algorithm cooperation of *back-propagation* and *conjugate gradient descent algorithms* with *early stopping* and *weight elimination* to obtain the parameter estimates and to prune the network. For details, readers are referred to Bildirici and Ersin (2009, 2013, 2014) and Bishop (1995).

The modeling of the MS-VAR-MLP with more than two regimes is possible; however, the model requires the estimation of $rv(h + 3p + l + 3 + 3 + 3)$ number of parameters where r,v, and h represent the number of regimes, number of vectors, and number of hidden neuron weights and p represents the optimum lag length selected with SIC. Due to the loss in the degrees of freedom, large datasets are a necessity. The EKC-type environment-economic growth relation requires the investigation of a large span of data covering distinct stages of economic development; however, due to the availability of the

historical data in yearly basis, the model is evaluated for two regimes only. Contrarily, the next section covers the extension of the sensitivity analysis to the MS-VAR-MLP model, which provides important insights regarding the complex nonlinear relations under two distinct regimes of the business cycles.

Sensitivity analysis under the MS-VAR-MLP model

Another approach designed to evaluate the importance of the input variables in ANN literature is sensitivity analysis. Engelbrecht et al. (1999) discussed sensitivity analysis as a way of determining the significance of variables in an ANN model. Dimopoulos et al. (1995) and Dimopoulos et al. (1999) proposed sensitivity criteria to choose between significant variables in ANN models. In addition, Gevrey et al. (2003) investigated various types of analyses, including sensitivity

analysis, and provided a thorough investigation of methods focusing on the contributions of the variables.

Sensitivity analysis is based on partial derivatives of the dependent variable with respect to input variables. By evaluating the gradient vector, the NN model could be utilized to obtain not only nonlinear causal relationships, but also the magnitude of the input variables. Guo et al. (2011) showed that through sensitivity analysis, information could be extracted using partial derivatives for a small neighborhood of data to derive a local visual pattern among variables.

In the MS-VAR-MLP model, the sensitivity analysis is regime dependent: within each regime, local patterns for a small neighborhood of data could be visually expressed for different regimes. As a result, the regime-switching sensitivity analysis provides different asymmetric and regime-dependent nonlinear relations.

In line with the *PaD* and *profile* approaches (Lek et al. 1996; Gevrey et al. 2003), the study proposes the MS-VAR-MLP-based regime-dependent sensitivity analysis, i.e., *regime-dependent causality*, which could also provide visual representations of the nonlinear relations among the analyzed variables at different phases of the business cycles. (For the sensitivity analysis, readers are referred to Engelbrecht et al. 1995, 1999; Molas and Yamazaki 1995; Gevrey et al. 2003; Hadzima-Nyarko et al. 2011). For a logistic function the partial derivatives of the output y_t with respect to the qth input variable for each VAR-MLP contains h number of neurons with *v.r.l* number of input variables. By applying the chain rule for the logistic function of the form $\psi(x) = \frac{1}{1+e^{-(x)}}$, the partial derivative with respect to x is given by $\psi'(x) = \psi(x)(1 - \psi(x))$. Hence, the application of the chain rule to the single regime VAR-MLP yields as follows:

$$d_{t,q} = \vartheta_t \sum_{h=1}^{h} \lambda_h \psi_{ht}(1-\psi_{ht})\omega_{qh} \qquad (22)$$

where $d_{t,\,q} = \partial y_t/\partial x_{t,\,q}$ and ϑ_t are the partial derivatives of the output neuron with respect to the input it received from the hidden layer, ψ_{ht} is the response of the hth neuron, λ_h is the connection parameter of the hidden unit, and ω_{qh} is the connection parameter of the qth variable in the hth hidden unit. Analogous to Eq.(9), for more than one regime MS-VAR-MLP model, the regime-dependent sensitivities is calculated as follows:

$$d_{t,q}^{(s_t)} = \vartheta_t^{(s_t)} \sum_{h=1}^{h} \lambda_h^{(s_t)} \psi_{ht}(1-\psi_{ht})\omega_{qh}^{(s_t)} \qquad (23)$$

which denotes the regime-dependent response of the dependent variable in each vector to the independent variables in each vector. Further, the regime-dependent sensitivity in Eq. (23) represents the *regime-dependent nonlinear causality* among the analyzed variables.

The proposed approach requires the following steps. In the first step, the regime probabilities are obtained through the Markov chain procedure and the data set is divided into sub-regression spaces within the MS-VAR-MLP framework. In the second step, the connection weights and biases (constant terms) for each regime are estimated. In the third step, the partial derivatives are calculated for each input variable for each regime, while giving the mean values to each of the other variables. Fourth, since the model is a time series model with lagged terms, the partial derivatives of each lag are summed to obtain the *overall effect*. In "Conclusion," a decile range with ten equal points for the input variables is calculated based on the *min-max* of each variable. The range corresponds to a one to ten scale that extends between the minimum and the maximum levels for each of the inputs for each regime. In the sixth step, the response of the selected output variable in each regime is plotted against the decile range, while holding other variables at their mean values to gather a visual representation of the effect of each variable to the others. This process is repeated for each regime to obtain *regime-dependent nonlinear causality* between the input and output variables within the MS-VAR-MLP framework.

Dataset and the empirical results

Data

To obtain a large span of data, this study utilizes annual data for the UK and the USA. The CO_2 data for both countries are taken from the Carbon Dioxide Information Analysis Center (CDIAC) and Gapminder databases. The CO_2 data used in the study represent the total per capita CO_2 emissions from fossil fuels in 1000 metric tons. The CDIAC per capita CO_2 dataset covers a large span of data corresponding to the 1751–2016 period for UK and the 1800–2016 period for USA. However, according to our own inspections, though the data starts from 1750s, it is subject to rolling for certain years due to missing data. Yearly, variations in the process followed by the data are a necessity for econometric techniques. As a result, the CO_2 data is restricted to the 1831–2016 period. The crude petrol price data also resulted in a restriction in our analysis since the most historical data could be obtained for 1871. The petrol prices are obtained from the British Petrol *Statistical Review of World Energy*, which represents the crude oil prices in US dollars per barrel and covers the 1871–2016 period. At the third stage, the GDP per capita data for USA is taken from the FRED database of Central Bank St. Louis and is given in nominal dollars. The GDP per capita data for the UK is obtained from the Central Bank of England and is given in nominal pounds. Therefore, due to the availability of the emissions' and petrol data, the dataset analyzed in the study is restricted to the 1871–2016 period for UK and USA. In the

study, all variables are subject to logarithmic transformation. The logarithmic GDP per capita, crude petrol prices, and CO_2 emissions per capita are denoted as ly_t, lp_t, and $lco2_t$ respectively. Further, the first differences of the variables are denoted as dly_t, dlp_t, and $dlco2_t$, which also show their respective yearly growth rates.

The selection of UK and USA is not solely based on the availability of data. It is well-known that both countries had passed the stages of economic development in terms of Rostow's stages of growth and reached the final stage of high mass consumption (Rostow 1960). Additionally, though China has reached the top polluter status in the last two decades in terms of metric tons of CO_2, China is followed closely by the USA. Further, as the recent studies show, in per capita terms, the USA is still the top polluter of CO_2 emissions in the world (Olivier et al. 2015 and EDGAR 4.3 database). Between 1990 and 2012, the per capita CO_2 emissions in China increased by 333%. However, USA is promising, since the per capita increase was only 7% in the USA for this respective period. This argument is also in favor of a Rostow-type portrait and hints the possibility of an inverted U-type EKC relationship. However, this argument is more revealing for the UK, in contrast, the per capita emissions decreased by 28% in this period in the UK (Olivier et al. 2015). In 2012, per capita CO_2 emissions in China, USA, and UK are 7.11, 16.21, and 7.50 metric tons; showing that in per capita basis, USA is followed by the UK, which is closely followed by China.

The CO_2 per capita emissions in the UK and USA are given for the 1751–2012 period in Fig. 1. The figure corresponds to a large time span that includes the first and second industrialization periods in eighteenth and nineteenth centuries, the Great Depression in 1929, the WW I and II in 1914–1918 and 1939–1945, the oil shocks in 1970s, and the 2008 global crisis.

During the first and the second industrialization period, the UK had been the top per capita polluter. After 1960s, the USA experienced an increase in its per capita emissions that led the USA to catch UK in 1900. Following the 1929 Great Depression, both the UK and USA experienced a sharp drop in CO_2 emissions per capita. Following the beginning of WWII, while the USA experienced a drastic increase in per capita CO_2 emissions, the UK experienced a drastic decline. Before the first and second oil shocks, where the petrol experienced a sharp increase in its price per barrel, the CO_2 emissions per capita also reached its historical peak in the USA at 22.17 tons, which is a clear sign hinting the positive impact of GDP growth before the crisis, followed by a decreasing pattern after 1973. After the 2008 Global Crisis, the CO_2 per capita was recorded as 19.3 tons in the USA, almost half of that in the UK (CDIAC 2016).

One could assume that the USA is a significant petrol producer in addition to being a petrol consumer, therefore, expected to have strong influence on petrol prices while the UK does not. As a result, the inclusion of petrol prices to the CO_2–GDP model of UK might be thought as being unnecessary. However, this is not the case. UK is not only a petroleum consumer, but also a significant producer given its petroleum extraction operations in the North Sea for the last four decays, in addition to its role in the petroleum production starting from the late nineteenth century. In addition, UK-originated multinational corporations, such as the Exxon Mobil and British Petrol (BP) are the world's largest and the sixth largest petroleum companies (Bergin 2008). BP has been operational since 1908, the year it established its petroleum extraction operations when petroleum was found in Persia (BP 2018). According to U.S. Energy Information Administration (EIA), except for the OPEC oil shocks in 1970s, the UK has been a net exporter of petroleum until 2013, showing that UK
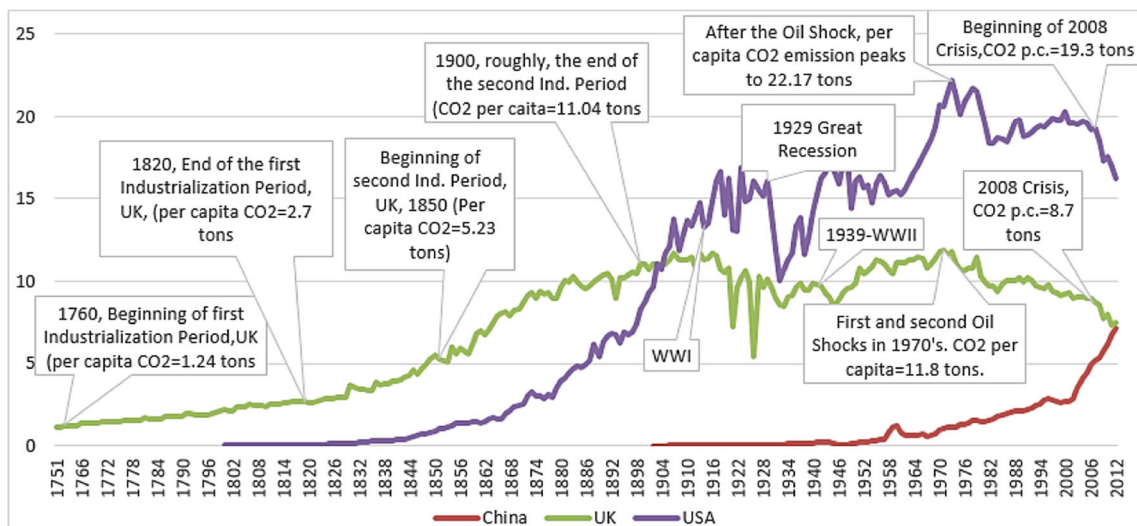

Fig. 1 Historical CO_2 per capita emissions (in tons) in the UK, the USA, and China, 1751–2012. Source: CDIAC Database, Carbon Dioxide Information Analysis Center. http://cdiac.ornl.gov (accessed 11 Nov. 2016)

is capable of meeting its domestic petroleum demand, and historically, the UK has been a net exporter (EIA 2018). The UK also has large amount of onshore reserves; however, 98% of the petroleum comes from 113 petroleum installations off-shore in North Sea Central Graben where 2two rich clusters (the Forties and the Brent oilfields) are located in addition to its single onshore oilfield in Dorset (EIA 2018). According to the BP Statistical World Energy Report, the USA is account-able for the 7.8% of the total crude petrol production in the world and is the major oil producer in the American mainland, compared to Norway being the largest (2.4% of the world) and the UK being the second largest (1.8%) in the Europe (EIA 2018). The formerly International Petroleum Exchange, now called the ICE futures, is a leading petrol market located in the London, UK. Regarding the multinational corporations of the UK, BP has operations in 72 countries worldwide, one third of global operations of BP is located in the USA, from which daily crude oil production equals to 335,000 barrels per day (Bloomberg Businessweek 2018). Other international opera-tions include Egypt, where BP produces 15% of the coun-tries' total crude petrol in Angola, a total of 9 oil exploration blocks in Iraq, where BP operates in Rumaila field, which produces over 1 million barrel per day in 2010, which was estimated by the UK Trade and Investment to be increased to 2.5 million per day between 2010 and 2015 (UKTI Report 2012).

The percentage yearly changes in excess demand (ED) of petroleum, calculated as domestic consumption minus domes-tic production are given in the figures below. Regarding the argument above, the close positive association between petro-leum production and consumption with crude petrol prices could be easily observed for both countries. Note that the figures for the USA and the UK include the barrel prices, which are determined in the financial markets located in each country, namely Brent for the UK and WTI for the USA. Our calculations show that raw US dollars, in Brent and WTI prices, have a positive correlation of 0.968, whereas their yearly percentage growth rates have a correlation of 0.979 (Fig. 2a). A comparison of Fig. 2b, c shows that the petroleum

production and consumption follow a strong co-movement with crude petroleum prices. One interesting finding is op-posed to the argument directed towards the UK, the correla-tions between ED and crude petroleum prices are 0.256 and 0.205 for the UK and the USA, respectively, showing the association between the UK and petroleum prices not only exists, but is also comparatively stronger.

These findings clearly reveal that the petrol prices could be taken as not only an important control variable but also as an important explanatory variable to the GDP and CO_2 emissions-based EKC for the UK and the USA. The influence of the UK on petrol prices is based on three pillars, the size of its petroleum extraction operations mainly located offshore; the multinational operations of the UK originated petroleum firms and having an important energy market, where the in-ternationally accepted Brent is traded.

Empirical results

The study focuses on the following steps in the empirical section.

1. The MLP augmented MS-VAR-MLP is estimated in para-metric form, as a result, the model could be interpreted directly without the sensitivity analysis to interpret the nonlinear impacts of variables. Once the MS-VAR-MLP model is estimated, the direction of causality is evaluated.
2. The MS-VAR-MLP model is augmented with sensitivity analysis to get hold of the shape of the environmental Kuznets curve. The regime-dependent sensitivity analysis eases the evaluation of the causal impact of each variable by providing visual interpretations. The third step should be coupled with the second in terms of providing causal interpretations.
3. MS-VAR-MLP models are compared with the MS-VAR in terms of forecasting and modeling capabilities. This stage is important to analyze the efficiency of the models evaluated.

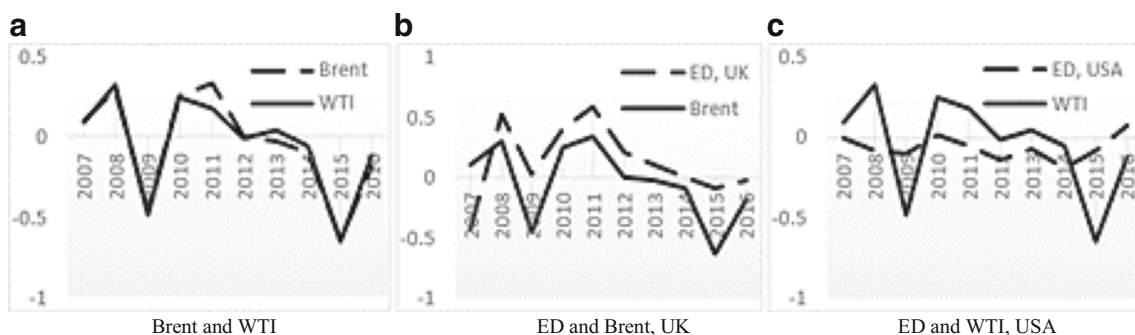**Fig. 2** Brent and WTI Crude Petrol Prices, Excess Petroleum Demand (ED) in the UK and the USA, Yearly % Growth Rates, 2007–201. Source: BP Statistical Review and World Energy, June 2017. Notes: excess demand of domestic petrol is calculated as ED = consumption-produc-tion, in million tones. WTI and Brent represent yearly % change in West Texas Intermediate and Brent crude petrol prices per barrel ($US)

MS-VAR-MLP neural networks estimations

In this section, the neural network-based models discussed in the previous section will be utilized to evaluate the possible relationships between the growth rates of GDP and CO_2 per capita and petrol prices, where the variables are denoted as dly_t, $dlco2_t$, and dlp_t after log-first difference transformation. The MS-VAR-MLP methodology follows the following steps:

i. *Input selection*: The optimum number of input variables is selected with SIC information criteria in the input layer to minimize loss of degrees of freedom and to maintain parsimony.
ii. *EM algorithm*: The model's regime-switching structure is estimated using the EM algorithm and the Markov chain regime probabilities are obtained.
iii. *Estimation and evaluation*: The model is estimated using back-propagation, conjugate gradient descent and Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm cooperation in addition to early stopping and weight decay. The model architecture of the models is selected through estimation of 200 MS-VAR-MLP models with varying model architectures and the best model is selected in terms of the generalization accuracy based on MSE and rho statistics.

MS-VAR-MLP and sensitivity results: the UK The MS-VAR-MLP estimation results obtained for the UK and are reported in Table 1. The model for the UK is an MS-VAR-MLP (12-2-1) in which the input layer consists of three variables dly_t, $dlco2_t$, and dlp_t, each variable has four lagged values, resulting in a total of 12 inputs in each regime. The input layer is connected to two sigmoid activation functions. The hidden layer is connected to the output layer with a linear identity function to predict the dependent variable of each VAR vector in each regime.

Training and test performances are reported as the Pearson correlation (*rho*) coefficient between the predicted and the target variables, training, and test errors are with MSE error criteria. Learning is conducted by the conjugate graduate descent algorithm with the BFGS algorithm for which the epoch of convergence is reported. The MSE errors of the first vector, dly_t, were calculated as 0.007750 and 0.019446 for the training and test samples, respectively, in the first regime and as 0.012399 and 0.012399 for the second regime. The correlations were calculated as 0.845727 and 0.529508 for the training and test samples for the first regime and as 0.544468 and 0.294674 for the second regime of the GDP growth rate vector. The highest training performance was obtained for the dlp_t vector with MSE statistics, being calculated as 0.001038 and 0.002241 with rho coefficients of 0.947366 and 0.505228 for regimes 1 and 2 for the petrol price growth rate vector. It

should be noted that, in the second regime of the CO_2 growth rate vector (i.e., the results in the test sample), the rho statistic is 0.126 which the lowest rho overall. However, the rho statistic reported for the training sample is 0.71, showing strong results in terms of the goodness of fit in the model. If we evaluate the models with all regimes and vectors, the results show good generalization capabilities and promising results in terms of goodness of fit.

The calculated parameter estimates of the MS-VAR-MLP model for the UK are reported in the second part of Table 1. The architecture selected for the model is in the form of MS(2)-VAR(4)-MLP(12-2-6), a model with two regimes, four lags for each variable as in the VAR models. Each input variable has 4 lags totaling to 12 input variables in the input layer. In the neural network structure of each regime, the input layer with 12 input variables are linked to the hidden layer that consists of 2 nonlinear sigmoid functions. The hidden layer is linked to the output layer that consisting of identity functions, which produce three target variables for each regime, totaling to six output variables in the output layer. The signs of weights that link the input variables to the hidden layer play a vital role regarding the impact of the variables in the input layer. Therefore, the important factors that define causality between input and output are based on the directional impact among the variables and on the signs and sizes of the neuronal links to the output layer, while the learning algorithm applied with early stopping also allows weight decay and pruning during estimation.

Instead of analyzing each estimated parameter, one common way to provide interpretable results within the neural network framework is through local sensitivity analysis. This process involves substituting the average of each variable that is to be held constant, while calculating how sensitive a neural network vector is for a given domain of input variables, thus yielding interpretable causal relationships. Accordingly, for each input, the partial derivatives are measured at ten evenly spaced locations starting from the minimum level and ending at the maximum level of the input variable, while giving the mean values to the other variables. As a result, the partial derivative procedure provides an interpretation of the directional causality for various levels of GDP, CO_2, and petrol prices. The pointwise sensitivity results are given in Fig. 3a, b for regime 1 and regime 2 (the recessionary and expansionary regimes, respectively).

It should be noted that the direction of causality between variables is nonlinear and it is governed by the direction (sign) magnitude, i.e., negative and positive contributions and their size, as a result the overall effect of each input variable on the dependent variable is nonlinear within each regime. Firstly, the causal impact of petrol price growth rates on GDP growth rates will be analyzed for different magnitudes in the recessionary and expansionary periods. In the recessionary regime depicted in Fig. 3a, it is evident that the size of the impact of

Fig. 3 The causal effects of the input variables on GDP growth rates in theUK
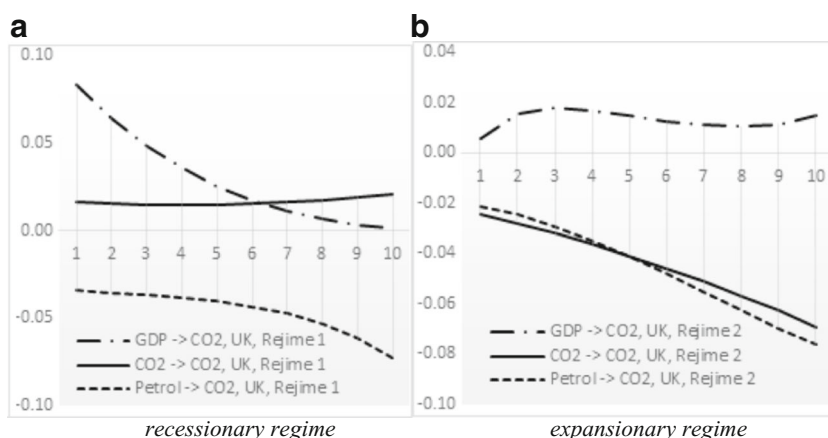

recessionary regime   *expansionary regime*
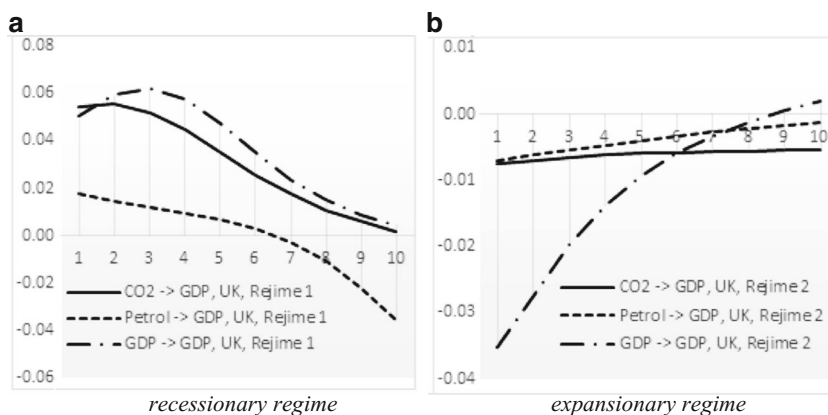
petrol price growth rates on GDP growth rates varies from its minimum towards the maximum values (i.e., along the decile range, 1–10, in the horizontal axis). The impact is comparatively high and positive for the low values of GDP growth, though it is continuously decreasing. The impact becomes negative after the sixth decile. The results show that the increases in petrol prices have a positive but a decreasing impact on GDP for very low and negative GDP growth rates and it becomes negative after the seventh decile. In the second regime, given in Fig. 3b, the impact of petrol prices on GDP growth rates is negative for the majority of cases, it increases towards zero and becomes positive at the highest decile of expansions. The results should be evaluated with care, since the recessionary regime also includes the oil price shocks that resulted in deep recessions by causing large and negative supply side shocks. However, the results are as expected, since the overall causal effect of petrol prices is negative on GDP growth rates.

Secondly, the analysis of the impact of CO_2 on GDP growth rates indicates that the overall causal effect on GDP is positive, although the positive impact attained its peak at the third decile and maintained a decreasing path towards the tenth decile. In the recessionary periods, during which negative growth rates were recorded, there exists a positive relation

between CO_2 and GDP growth rates. If evaluated with the GDP's impact on itself (dashed-dotted line), the impact of GDP on CO_2 follows a similar pattern. This type of co-movement is expected; however, it cannot be interpreted as a directional impact of CO_2 on GDP growth rates, since CO_2 emissions should be taken as a product of industrial production, though the impact of CO_2 on GDP could be taken as a sign of policies directed towards the pollutants. The co-movement is expected and is evaluated as a sign of the strong association of emissions to the level of GDP. The analysis will direct its focus towards evaluating the impact of GDP on CO_2 growth rates. The pointwise sensitivity results are shown in Fig. 4a, b.

The causal impacts of GDP growth rates are given for decile ranges starting from the minimum towards its maximum (one through ten) for the UK's GDP growth rates in the recessionary and expansionary regimes. During recessions (Fig. 4a), the impact of GDP growth rates on CO_2 emissions (dashed-dotted line) had been above zero and was thus positive. As GDP growth rates increase from the first decile towards the tenth, the impact on CO_2 emissions follows a decreasing path, up to the tenth decile in the recessions where it reaches zero. Overall evaluation shows that even at the very low negative economic growth rates, the effect of GDP on

Fig. 4 The causal effects of the input variables on CO_2 emissions in the UK

recessionary regime   *expansionary regime*

emissions had been positive for 90% of the deciles. Further, the impact of GDP growth rates on CO_2 emissions growth rates (Fig. 4b) follows a positive and increasing pattern, mimicking an S-shaped relation for the expansionary regimes, and the positive effect accelerates for moderate to high economic growth rates. The overall evaluation of Fig. 4a, b suggests that the impact of GDP growth rates follows an asymmetric path of altering magnitudes, and the impact is positive for the majority of the recessionary regime and for the whole of the expansionary regime. As noted in the data section, evaluation of the impact of petrol prices on CO_2 could also reveal interesting findings. According to the results (dashed line), increases in petrol prices had a negative impact on emissions, which could be justified with the historical impact of petrol prices on economic growth rates, since petrol price increases are associated with declines in production (GDP), which could lead to increased emissions. As a result, for the UK, the effect of petrol prices on emissions cannot be disregarded, and the overall evaluation confirms the business cycle effects of petrol prices, in addition to the impact of petrol on emissions. High petrol prices affect production negatively in addition to discouraging petroleum consumption, which leads to lower emissions.

One point that cannot be overlooked is the fact that the determination of the shape of the EKC bears significance in the study. Compared to the literature, which focus on the calculation of a single turning point as is for the inverted U-shaped EKC, the results reveal significant differences. In Fig. 4a, b, CO_2 emissions are continuously in relation to the GDP and petrol prices, and lowering of emissions has not been possible except for the economic growth rates, being almost equal to zero. Further, even for the negative growth rates, positive impact of production on emissions cannot be rejected. The results suggest strong positive effects of economic growth on emissions. Further, after eliminating the effects of the crisis by isolating the recession and crisis periods, economic growth has a continuous positive impact on CO_2 emissions. As discussed in "Introduction" and "Literature review," without isolating the effects of crises and without differentiating between the business cycle regimes, the generalized results could be misleading in terms of deriving a conclusion in terms of the shape of the EKC. If the GDP -> CO_2 line is evaluated, the causal impact of GDP on CO_2 follows a decreasing path that never falls below the zero line. If taken as the shape of the EKC curve, instead of an inverted U, the overlook of the shape reveals a U-shape that accelerates at highly positive and moderate negative economic growth rates. The results for the petrol prices are also in the expected direction in the spirit of MS and MS-VAR literature (Krolzig 1998; Hamilton 1990) underlining the influence of petrol prices on economic growth cycles. In our results, increases in petrol prices are expected to have a negative impact on production on the supply side in addition to a negative impact on fossil fuel consumption on the demand side. Accordingly, the

overlook of the causal impact of increases in petrol prices on CO_2 emissions is negative. Though not directly analyzed in the study, the usage of hydroelectric energy and various renewable energy sources after 1960s, in addition to the decline in the dependence on petrol after 1970s are important factors on emissions.

Thirdly, the causal impacts of the variables on petrol prices are given in Fig. 5a, b for the UK. The causal impact of GDP growth on petrol price is positive up to the sixth decile and negative afterwards; however, the dashed-dotted line follows a path very close to zero, suggesting very little impact of GDP of the UK on petrol prices. One cannot conclude that economic growth rates in the UK have no impact on petrol prices, but this impact is limited. As a result, compared to the previous graphs (Fig. 4a, b), figures show the major direction of causality flows from petrol to GDP and to CO_2. This relation is also maintained in the expansions. In addition, the petrol prices are most significantly under the influence of its past values, the previous values of the petrol prices. This type of path dependence is evident, since the largest fluctuations occur in the dotted line compared to the solid- and dashed-dotted lines. In the recessions, the path followed by the petrol is positive up to the fifth decile, following a negative and sharp decline towards to tenth, and a negative effect of petrol prices on itself is maintained up to the third decile during expansions after which the impact becomes positive. The combined analysis of both regimes shows a U-shaped path followed by the petrol, reaching a trough at zero economic growth, following a negative but increasing pattern afterwards in the expansions.

The overall results confirm the impact of petrol prices on business cycles, which in return has strong effects on CO_2 emissions. By evaluating Figs. 3, 4, and 5, the results indicate that GDP growth rates, CO_2 emissions and petrol prices show different dynamics in terms of magnitude and direction in the UK, and these complex relations cannot be summarized by an EKC-type-inverted U finding. Based on the MS-VAR-MLP approach, there is clear evidence of the causal effects among the GDP and petrol prices, which affects CO_2 emissions nonlinearly at various levels. In particular, by focusing on Fig. 4, the positive causal effect of GDP growth rates on CO2 emissions does not only occur during expansions, but also during recessions: the impact approaches zero only when the economy approaches the last periods of recession. The emissions follow a steady, S-shaped, positive path during expansion and the positive impact settles at a horizontal path during moderate and high growth in expansions. However, as economic growth approaches its peak around the eighth to tenth decile, CO_2 emissions start to follow another accelerating path. The results of the paper clearly show that the impact on CO_2 emissions are highly nonlinear, and they differ significantly from the findings obtained in the literature, which utilized the polynomial regressions and various nonlinear approaches, including the TAR, MS, and MS-VAR

Fig. 5 The causal effects of the input variables on petrol prices in the UK

recessionary regime   expansionary regime
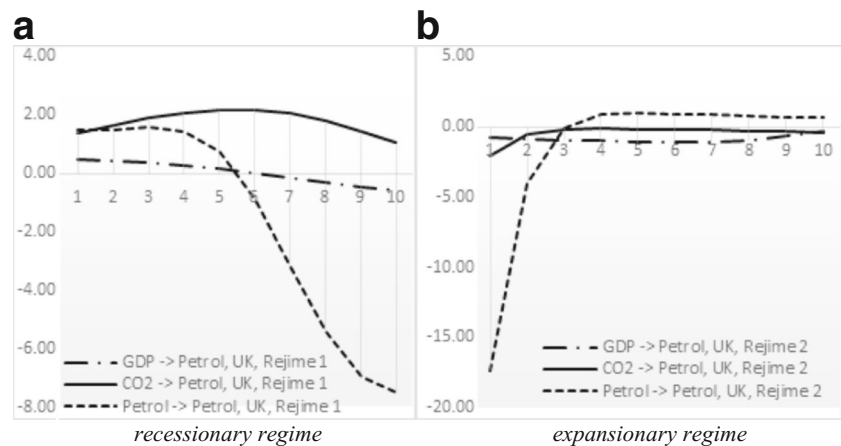
models. Further, compared to the MS-VAR model, the MS-VAR-MLP approach reveals more complex nonlinear relations, since the relations are allowed to follow NN processes within distinct regimes. The analysis determines that the relationship between the variables are not only depends on which regime the economy is in, but also on the magnitude (or size) of GDP growth rates, CO_2 emissions and petrol prices. Compared to the literature given in "Literature review," these findings provide significant deviations especially from those obtained with the linear polynomial regression models. Linear models assume the parameters to be constant over the sample period, which also assumes the relationship between GDP and CO_2 emissions being rather stable implicitly. In addition to observed regime dependency and magnitude dependency in our models, many important factors could result in trajectory changes in the environment-economic growth relationship. As mentioned in Fig. 1, the important historical factors leading to trajectory changes include the first, second, and third industrial revolutions, trade liberalization, and financial liberalization policies, various economic crises, including the 1929 Great Depression, Oil Shocks in 1970s, the Asian Crisis in mid 1990s, the ERM crisis in late 1990s, the Great Recession of 2008, and World Wars I and II. The results suggest that the shape of the EKC is altered historically depending on the phase of the business cycle in addition to magnitude dependency of the responses of environment and economic growth variables within a nonlinear setting. As a result, the econometric analyses with samples covering a long historical period should consider the incorporation of more advanced methodologies, including machine learning and neural networks.

MS-VAR-MLP and sensitivity results: the USA The results for the USA are given in Table 2. The model selected for the USA is a two-regime MS(2)-VAR(3)-MLP(9-3-6) model with nine input variables in the input layer and three neurons in each layer that are linked to the output layer that produces three predicted values for each regime consisting of six predictions (output vectors) in total.

The MSE error criteria and rho coefficients point to satisfactory training and test results. The MSE errors of the first vector, dly_t, were calculated as 0.0049 and 0.0005 for the training and test samples of the first regime and 0.000376 and 0.000165 for the training and test samples of the second regime. The correlations between the target and output variables are significantly large. The highest training performance was obtained for the first regime of the dly_t vector, and the lowest was obtained for the second regime of the dlp_t vector. The results suggest good generalization and goodness of fit. Similar to the UK's model, the interpretations will be conducted through the MS-VAR-MLP based sensitivity analysis though the parametric interpretation is also possible.

Similar to the analysis conducted for UK, the horizontal axis represents the decile range for the GDP growth rates in the recessionary and in the expansionary regimes. In Fig. 6a, the impact of petrol prices on GDP growth rates is positive and largest for the most negative values of GDP growth rates, and the positive impact decays towards the fifth decile, staying positive and horizontal afterwards during recessions. In the expansionary regime (Fig. 6b), this negative impact is maintained for the whole decile ranges during expansions. As for the UK, the causal impact of petrol price varies for different magnitudes of GDP growth rates. In terms of the impact of the petrol prices in the UK, the majority of decile ranges suggests a negative impact of petrol prices on GDP with various degrees in both regimes. The economic interpretation is obtained similar to the results for the UK. Further, the results should be evaluated with caution, considering the petrol price shocks that also coincide with the recessionary regimes. These negative and large supply side shocks are coupled with drastic declines in the GDP growth rates (Hamilton 1990), increases in petrol prices are coupled with negative economic growth rates especially in the deep recessions.

The impact of CO_2 emissions on GDP growth rates is represented by the solid-dotted line. The line follows a path close to the horizontal line suggesting almost no effect of emission growth increases on GDP growth during the

Fig. 6 The causal effects of the input variables on GDP growth in the USA

recessionary regime   expansionary regime
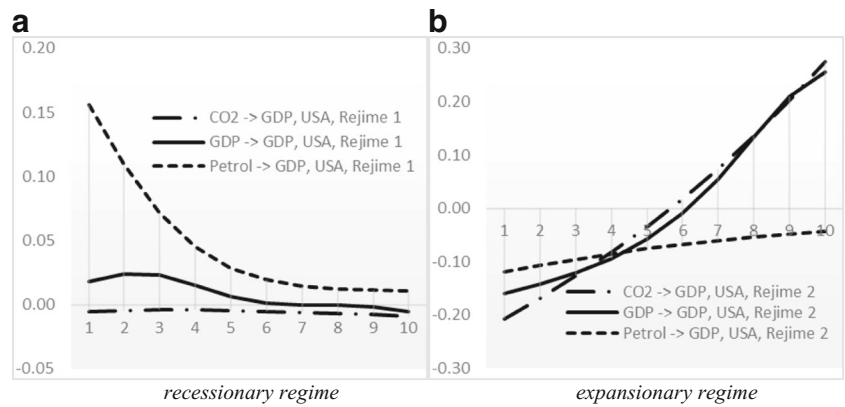
recessions. In contrast, the results suggest the negative impact of emissions on GDP growth rates up to the sixth decile after which the impact becomes positive in the expansionary regime. The result is in favor of a simultaneous occurrence between the GDP and CO_2 emissions. Similar to the results obtained for the UK, the co-movement of CO_2 on GDP also occurs for the USA. However, the direction of the effect should be evaluated after the investigation of the impact of GDP on CO_2 for USA in Fig. 7 coupled with Fig. 6, which deserves special importance in the analysis of EKC in terms of the motivation of the study.

The impact of GDP and petrol prices on CO_2 emission growth rates is depicted in Fig. 7a, b for the USA. In Fig. 7a, the impact of GDP growth rates on CO_2 emissions (solid-dashed) shows a positive impact for the lowest GDP growth rate deciles, follows a decreasing path that reaches zero at the sixth decile, the path becomes increasing after reaching the lowest point at the eighth decile. The impact of GDP growth rates on emissions accelerates and crosses zero at the tenth decile. In Fig. 7b, the impact becomes positive after the second decile and follows an increasing path, suggesting an accelerating effect of economic growth on emissions in the expansionary regime. The overall evaluation suggests a U-shape relation rather than an inverted U relation between GDP and CO_2 growth rates. The results also expose that

different dynamics exist for the recessions compared to the expansions and the impact of GDP growth rates on CO_2 emissions is highly asymmetric at different magnitudes of economic growth, and the aggregate evaluation of two regimes suggest that the effect is positive for the majority of the economic growth decile ranges in the USA.

If the impact of petrol price growth rates on CO_2 growth rates is evaluated, an overlook suggests a U-shaped relation, which hint the negative impact of petrol prices on GDP. In the recessions, the causal impact of petrol prices on CO_2 emissions follows a positive path only for the lowest deciles. After the fourth, the impact becomes negative. In the expansionary regime, the impact of petrol prices at the first decile is highest; however, as the higher deciles are reached, the negative impact is maintained. The results are in the expected direction, the increases in petrol prices are expected to have a negative impact on both productions on the supply side and on fossil fuel consumption on the demand side in the 1871–2016 period.

Similar to the results for the UK, the dynamics modeled by the MS-VAR-MLP for the USA deviate from the EKC examinations in the existing literature. Our results determined a U-shaped curve, where the response of the CO_2 emissions varies nonlinearly depending on the magnitude of the GDP growth in distinct regimes. Accordingly, the environmental

Fig. 7 The causal effects of the input variables on CO2 emissions in the USA

recessionary regime   expansionary regime
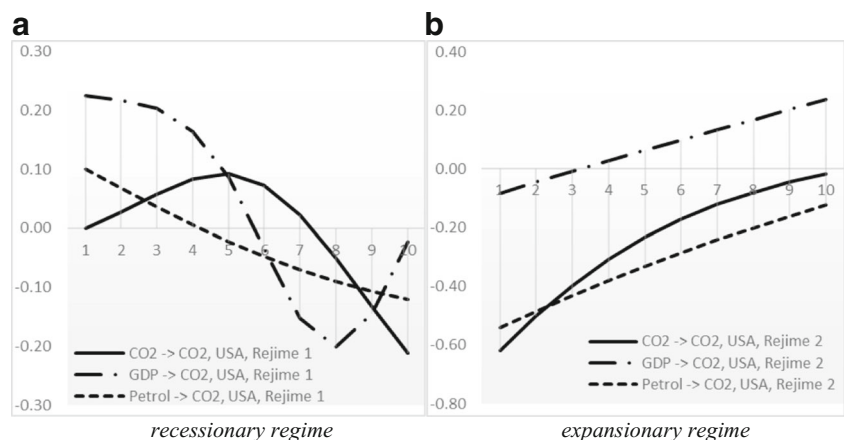
deterioration increases as per capita income growth changes, and the response of environmental deterioration is asymmetric depending on the phase of the business cycle. The results also show that at higher levels of economic growth, the elasticity of demand for environmental quality becomes highly positive. The finding is in line with the findings of Atasoy (2017) and Apergis 2016 who suggest that the environmental quality is sacrificed to achieve higher economic growth at the majority of states of the USA.

Lastly, the causal impacts of CO_2 emission growth rate on itself are indicated by a solid line, which follows a U-shape in Fig. 8a, b. The pattern mimics to that found for the GDP growth rates (dashed-dotted line). The shape is in favor of path dependence of CO_2 emissions, dependence to its past values similar to the results for the UK. The resemblance of the GDP growth and CO_2 emissions curves also confirm the co-occurrence of CO_2 emission and GDP growth rates for the USA.

The causal impact of the variables on petrol prices is shown in Fig. 8a, b. The causal impact of GDP growth rates on petrol prices (dashed-dotted) is positive for the entire decile range and decays towards zero only at the tenth decile, for zero economic growth rates in the recession. During the expansions, the negative effect of GDP on petrol prices continues; however, the effect decelerates and reverses after the sixth decile, where it achieves a positive and increasing path. Accordingly, GDP growth of the USA has a positive impact on petrol prices in the recessions and except for low economic growth rates in the expansions. The path followed by the impact of CO_2 emissions (solid line) is negative and fluctuating up to the last deciles of the recession. After the ninth decile, the effect of the CO_2 emissions becomes positive and increasing. In the overall expansions, the effect is positive, but decays towards zero at the eighth decile. The results for the USA differentiated complex causal dynamics among the analyzed variables compared to the UK. However, there is a clear positive and accelerating effect of GDP growth rates on CO_2 emissions during the expansions in addition to positive effects of GDP during the deep recessions such as crises, whereas the lowering of emissions could only occur during slight recessions in the neighborhood of zero economic growth.

Comparison of generalization performances of MS-VAR and MS-VAR-MLP models The overall results suggest significant dynamics in terms of direction, size, and magnitude, all of which may be evaluated using the recommended MS-VAR-MLP methodology. The results also indicate that although MS-VAR models provide important improvements over the linear VAR model, especially in terms of evaluating asymmetric and nonlinear causal relationships between different regimes, the MS-VAR-MLP model provides a richer analysis within each regime, since the processes in each regime are modeled

with VAR-MLP processes.[1] The MS-VAR model accounts for asymmetry and nonlinearity for causal relationships among regimes and the MS-VAR-MLP further improves the causality analysis by nonlinearity within each distinct regime.

To investigate the discussion above, the models are compared in terms of MSE, MAE, and RMSE error criteria for their forecast accuracy. The results are provided in Table 3. For the MS-VAR models, single MSE, MAE, and RMSE statistic is reported. For the MS-VAR-MLP models, these statistics are reported separately for each regime. The table is conducted to provide one-step-ahead forecast evaluations; nevertheless, out-of-sample forecast practices are also possible. Due to the aim of the study, the models are investigated solely in terms of in-sample generalization. As noted in the empirical section, the training of the MS-VAR-MLP models was conducted with early stopping, to avoid over-fitting.[2] For the MS-VAR model, a general statistic that represents the whole regression space is reported. For the MS-VAR-MLP model, each regime is simulated to produce predictions to be compared to the results of the MS-VAR to ease the interpretation. The model with the lowest error criteria statistic is taken as having better one-step-ahead forecast accuracy.

The MS-VAR-MLP results for the UK suggest that in terms of MSE, the MS-VAR-MLP model performs better for the dly, for the dlco2 and for the dlp vectors in both of the regimes over the MS-VAR model. For the MS-VAR model, the MSE statistics for the dly, dlco2, and dlp vectors are calculated as 0.0078, 0.2291, and 0.2549, which are drastically larger than those obtained for the MS-VAR-MLP. Same finding also holds for the MAE and RMSE statistics. The overall results suggest significant improvement of MS-VAR-MLP over the MS-VAR for UK. The results obtained for the USA suggest that in two out of three vectors, the dly and dlp vectors, the MS-VAR-MLP model provides better forecast accuracy in terms of MSE, MAE, and RMSE criteria. However, the MS-VAR model performed better for both dlco2 vectors in both regimes, though the second regime of the MS-VAR-MLP could be taken to perform close in terms of the MAE statistic (0.0381) compared the MS-VAR (0.0339). This finding also holds for the MSE calculated for the second regime of the MS-VAR-MLP model having a very close performance with MSE = 0.0033. Compared to the significant improvement achieved by the MS-VAR-MLP model over the MS-VAR for the UK, this finding holds two out of three vectors of the

---

[1] It should be noted that in the MS-VAR analysis, the MS-VAR model was accepted against the linear VAR model in terms of certain portmanteau tests. The results are available from the authors upon request and are not included in the paper to save space.

[2] Early stopping algorithm continuously checks the MSE in the test and validation samples during estimation. If not conducted, lower values for in-sample MSE and various error criteria are also possible. However, without early stopping, the models could suffer from over-fitting and their out-of-sample performances could deteriorate significantly (for details, see Bishop 1995).

Fig. 8 The causal effects of the input variables on petrol prices in the USA

recessionary regime   expansionary regime
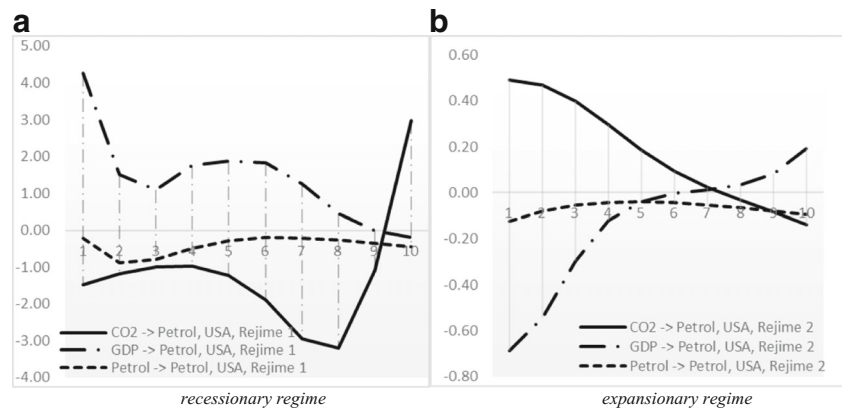
MS-VAR-MLP for the USA and in the GDP growth rate and petrol price growth rate vectors, the MSE, MAE, and RMSE error criteria are significantly lower than the error criteria calculated for the MS-VAR model. The overall evaluation of the statistics suggests that the MS-VAR-MLP performs better in forecast accuracy in both countries; however, due to the focus of the study on CO_2 emissions, the two models estimated for the USA are assumed to provide almost equal forecast accuracy. It should be noted that the result for forecasting accuracy holds only for the data set in consideration. Consequently, the improvement of the results for the neural network models in terms of forecasting is highly sensitive to the number of observations. Hence, though the data set covers the period from 1871 to 2016, the length of data is restrictive insofar as the data were collected annually. As a result, the selection of the optimum number of neurons is highly sensitive when trying to maintain parsimony and loss caused by degrees of freedom. Thus, the results obtained for forecasting cannot be generalized and should be evaluated for different and larger datasets. The overall findings of the study suggest that MLP augmented MS-VAR-MLP models provide significant improvement over the MS-VAR models in terms of the interpretability of causality and capturing the complex nonlinear relations between the analyzed variables. The MS-VAR-MLP model makes important contributions by providing differentiated dynamics that are asymmetric and nonlinear in relation to size, magnitude, and direction for the investigated variables. For future studies, the investigation of the EKC relation with the MS-VAR-MLP model with monthly data could provide a comparatively larger sample size. This analysis could hinder important insights regarding the emissions, income, and petrol relation. However, due to the long-run characteristic of the EKC, which requires a sample that covers various stages of economic development historically, the analysis is conducted with the use of annual data and the results suggest strong nonlinear effects of the variables under investigation.

Discussion and policy recommendations The MS-VAR-MLP model allows asymmetric and size-dependent causality; therefore, there are additional differences when compared to the

MS-VAR models. The proposed MS-VAR-MLP model also aims at the derivation of regime-specific sensitivity analysis, which allows the researcher to investigate regime-dependent asymmetric and nonlinear causality between the analyzed variables. Additionally, the proposed regime-dependent sensitivity analysis extended MS-VAR-MLP approach also has the capability to provide visual interpretations, i.e., graphical mappings of causal relations, which show the magnitude and regime-dependent causality.

The MS-VAR-MLP model overcomes the necessity of filter selection for the MS-VAR models. The MS-VAR models are generally estimated with either the EM or ML algorithms, and the selection of filter has been debated. As discussed, the Hamilton and the Kalman filters had been suggested and the majority of empirical studies could be considered as utilizing the Hamilton filter. In addition, trajectory changes resulting from oil shocks, economic crises, policy changes, and wars in such a historical data could result in achieving MS-VAR regimes that capture only these abrupt shifts in the time series. This could result in non-rejection of certain regimes, which capture a small number of observations including the outliers. Since the hidden Markov chains are inherent in the MLP models and the NN learning algorithms, no filtering is necessary in the estimation of the MS-VAR-MLP model. With the use of the learning algorithms and inherent nature of the hidden Markov chains in the MLP models, the MS-VAR-MLP model is capable of modeling such large span of data without the application of any filters. The MS-VAR-MLP models proposed are fully parametric and therefore allow ease in interpretability though the functional representations are highly complex compared to the linear regression models and nonlinear MS-VAR models. By the use of the sensitivity analysis, the causal interpretation of the neural network is possible considering the neurons, the weights of each variable (and its lagged variants as in the time series approach) and the relevant connections between the neurons, input layer, and the output layer. The sensitivity approach allows evaluation of asymmetry with visual representations for different episodes, i.e., recessions and expansions.

Table 1 MS(2)-VAR(4)-MLP(12-2-6) Model, UK

a. Training results

Vectors:	Architecture	Train. perf.	Test perf.	Train. error	Test error	Epoch
$dly(t)$	Regime 1: MLP 12-2-1*	0.846	0.530	0.008	0.019	20
$dly(t)$	Regime 2: MLP 12-2-1	0.544	0.295	0.012	0.012	7
$dlco2(t)$	Regime 1: MLP 12-2-1	0.861	0.988	0.004	0.008	25
$dlco2(t)$	Regime 2: MLP 12-2-1	0.717	0.126	0.004	0.012	14
$dlp(t)$	Regime 1: MLP 12-2-1	0.836	0.894	0.005	0.015	58
$dlp(t)$	Regime 2: MLP 12-2-1	0.947	0.505	0.001	0.002	60

b. Parameter estimates

Connections:	$dly(t)$ vector		$dlco2(t)$ vector		$dlp(t)$ vector	
	Reg. 1	Reg. 2	Reg. 1	Reg. 2	Reg. 1	Reg. 2
$dly(t\text{-}1)$ --> hn 1	2.80	3.14	0.54	−3.52	0.17	1.11
$dly(t\text{-}1)$ --> hn 2	−0.20	−2.06	1.46	−4.37	−1.31	−0.81
$dlco2(t\text{-}1)$ --> hn 1	−1.16	1.77	−0.14	−0.37	4.80	−12.67
$dlco2(t\text{-}1)$ --> hn 2	0.10	−6.04	2.47	−7.98	0.00	0.67
$dlp(t\text{-}1)$ --> hn 1	0.51	0.52	−0.82	−1.26	0.15	−5.65
$dlp(t\text{-}1)$ --> hn 2	1.29	−0.67	−0.28	3.92	−4.22	−4.53
$dly(t\text{-}2)$ --> hn 1	−0.53	−4.88	0.77	−6.91	0.87	0.17
$dly(t\text{-}2)$ --> hn 2	−0.12	−1.44	1.05	−1.76	2.70	1.32
$dlco2(t\text{-}2)$ --> hn 1	0.81	−2.07	1.98	−0.39	−2.28	−2.07
$dlco2(t\text{-}2)$ --> hn 2	0.34	0.63	−0.29	4.58	−1.57	3.34
$dlp(t\text{-}2)$ --> hn 1	1.42	0.09	−0.02	−1.64	−0.66	−0.54
$dlp(t\text{-}2)$ --> hn 2	2.31	−1.39	−2.57	−3.02	−2.66	−5.95
$dly(t\text{-}3)$ --> hn 1	5.20	−0.46	−1.40	−1.74	−0.11	−1.28
$dly(t\text{-}3)$ --> hn 2	−1.88	0.50	1.44	1.11	−0.64	−0.50
$dlco2(t\text{-}3)$ --> hn 1	−5.18	−0.56	−0.66	0.32	2.72	0.98
$dlco2(t\text{-}3)$ --> hn 2	−1.50	0.78	−0.72	−0.66	0.49	0.36
$dlp(t\text{-}3)$ --> hn 1	1.23	−0.46	−1.06	−0.71	−0.11	0.71
$dlp(t\text{-}3)$ --> hn 2	1.50	0.25	−1.23	1.74	−2.21	0.76
$dly(t\text{-}4)$ --> hn 1	0.15	0.43	−0.42	0.91	0.33	−0.77
$dly(t\text{-}4)$ --> hn 2	−1.06	0.25	0.48	0.90	0.84	−0.77
$dlco2(t\text{-}4)$ --> hn 1	0.23	0.47	0.44	−0.67	−0.85	0.81
$dlco2(t\text{-}4)$ --> hn 2	0.47	−0.15	0.50	−0.19	−0.76	−1.64
$dlp(t\text{-}4)$ --> hn 1	4.42	0.03	−0.51	0.90	−0.30	0.74
$dlp(t\text{-}4)$ --> hn 2	1.85	0.13	−1.28	0.72	−0.70	2.87
ib --> hn 1	−0.10	0.06	0.61	−2.00	2.87	−0.22
ib --> hn 2	−1.04	−0.41	−0.39	−3.82	0.72	2.88
hn 1 --> output	1.14	0.30	3.00	3.03	2.58	3.41
hn 2 --> output	1.40	−1.14	2.57	−2.82	−3.21	3.00
hb --> output	−1.66	0.77	−2.23	0.26	0.51	−2.79

Trans. probabilities:			Prob.	Duration		
p_{11}	0.53	Reg. 1	0.2771	2.14		
p_{22}	0.82	Reg. 2	0.7229	5.57		

Regime-specific correlations:	Regime 1		Regime 2			
	dly	dlco2	dlp	dly	dlco2	dlp
dly	1.00	−0.04	−0.03	1.00	−0.03	−0.04
dlco2	−0.04	1.00	−0.50	−0.03	1.00	0.56
dlp	−0.03	−0.50	1.00	−0.04	0.56	1.00

The estimated model is denoted as MS(r)-VAR(p)-MLP(i-h-o) where r stands for the number of regimes, p is the optimum lag length selected for the MS-VAR specification of the model. In the neural networks architecture, the i, h and o specifies the number of inputs, the number of hidden units and the number of outputs, respecttively. Accordingly, MS(2)-VAR(4)-MLP(12-2-6) model is a 2 regime model with 3 input variables having 4 lags each; therefore, the number of inputs total to 12 inputs. The input layer that consists of 12 inputs linked to the hidden layer consisting of 2 nonlinear sigmoid type activation functions or hidden units. The hidden layer is linked to the output layer that produces 3 target variables for each 2 regime, totaling to 6 output variables in the output layer defined as an identity function in the output layer

Further abbreviations are as follows. *hn* hidden neuron, *ib* input bias, *hb* hidden bias. For the training and test performances, rho statistics and MSE error criteria are reported. The epoch of convergence is reported under *epoch*

*Though the overall model structure is determined as MS(2)-VAR(4)-MLP(12-2-6), only one predicted variable exists in each vector of the model; therefore, the architecture in each vector is in the form of MLP (12-2-1)

As discussed in the literature section, a relatively small body of research focused on the MS-VAR models to investigate the environmental degradation and pollution relationship compared to the good amount of the polynomial regressions. Further, various papers also underlined the importance of more advanced methods, including the BP-NN and SVM. The MS-VAR-MLP further contributes by dividing the regression space into two or more regimes to investigate regime-dependent behavior. The MS-VAR-MLP model is also capable in overcoming the difficulties achieved with the MS-VAR

Table 2 MS(2)-VAR(3)-MLP(9-3-6) Model, USA

a. Training results

Vectors:	Architecture	Train. perf.	Test perf.	Train. error	Test error	Epoch
dly(t)	Regime 1: MLP 9-3-1*	0.891	0.977	0.005	0.001	61
	Regime 2: MLP 9-3-1	0.681	0.692	0.000	0.000	22
dlco2(t)	Regime 1: MLP 9-3-1	0.882	0.924	0.006	0.023	11
	Regime 2: MLP 9-3-1	0.613	0.519	0.001	0.001	26
dlp(t)	Regime 1: MLP 9-3-1	0.859	0.513	0.002	0.007	30
	Regime 2: MLP 9-3-1	0.554	0.500	0.002	0.002	17

b. Parameters estimates

Connections:	dly(t) vector		dlco2(t) vector		dlp(t) vector	
	Reg. 1	Reg. 2	Reg. 1	Reg. 2	Reg. 1	Reg. 2
dlco2(t-1) --> hn 1	− 1.47	0.68	2.70	0.44	− 5.22	5.49
dlco2(t-1) --> hn 2	2.09	0.47	− 1.85	1.13	− 10.83	1.38
dlco2(t-1) --> hn 3	1.17	0.39	1.62	− 0.14	− 10.64	1.69
dly(t-1) --> hn 1	0.25	− 0.03	0.42	− 2.88	1.27	− 0.53
dly(t-1) --> hn 2	− 1.03	0.08	0.17	− 0.15	4.42	− 4.82
dly(t-1) --> hn 3	1.05	0.69	1.68	− 1.82	− 2.95	− 2.08
dlp(t-1) --> hn 1	1.89	0.30	− 1.11	0.47	− 3.70	− 0.41
dlp(t-1) --> hn 2	− 2.78	0.27	− 1.55	− 1.27	2.17	4.62
dlp(t-1) --> hn 3	1.23	0.65	1.59	1.11	4.46	− 1.59
dlco2(t-2) --> hn 1	0.70	0.60	− 0.54	− 0.61	10.35	1.71
dlco2(t-2) --> hn 2	5.81	− 2.36	2.52	− 0.28	− 10.15	− 3.49
dlco2(t-2) --> hn 3	4.57	0.09	0.58	− 0.45	2.26	− 3.42
dly(t-2) --> hn 1	− 0.30	− 0.27	− 1.34	0.50	3.42	− 1.59
dly(t-2) --> hn 2	− 2.32	0.53	1.19	0.12	− 11.54	− 9.17
dly(t-2) --> hn 3	1.82	− 0.44	− 0.86	0.58	2.69	0.12
dlp(t-2) --> hn 1	1.14	− 1.18	0.37	− 0.03	8.22	− 0.68
dlp(t-2) --> hn 2	− 7.25	0.88	0.44	0.57	1.93	8.87
dlp(t-2) --> hn 3	2.04	− 0.14	− 0.90	− 1.25	− 4.64	− 2.90
dlco2(t-3) --> hn 1	1.60	− 1.73	0.25	− 0.21	− 3.23	0.99
dlco2(t-3) --> hn 2	4.94	4.87	− 4.34	0.20	− 5.33	0.45
dlco2(t-3) --> hn 3	3.39	− 1.54	− 1.59	− 0.44	− 4.76	0.10
dly(t-3) --> hn 1	− 0.57	0.62	2.78	0.04	− 3.48	− 4.10
dly(t-3) --> hn 2	− 2.50	− 1.89	− 2.80	0.14	0.67	− 4.66
dly(t-3) --> hn 3	1.31	0.12	1.16	0.12	0.77	− 1.42
dlp(t-3) --> hn 1	1.28	2.46	− 0.29	− 0.22	8.45	4.08
dlp(t-3) --> hn 2	− 5.65	− 3.18	− 1.08	0.18	− 4.91	− 1.18
dlp(t-3) --> hn 3	1.45	− 0.07	1.55	− 1.12	− 4.64	− 3.73
ib --> hn 1	− 0.22	1.43	−0.18	− 1.34	5.68	1.21
ib --> hn 2	− 0.06	0.15	− 0.34	− 0.77	− 2.57	0.66
ib --> hn 3	− 1.04	− 1.27	1.87	− 0.74	6.50	0.41
hn 1 --> output	0.42	2.97	1.61	1.70	− 2.96	0.27
hn 2 --> output	1.84	1.52	8.71	0.79	− 0.51	3.01
hn 3 --> output	− 0.83	2.93	4.85	0.46	2.48	− 1.72
hb --> output	− 0.48	− 3.43	− 8.00	0.07	0.55	0.40

Trans probabilities			Prob.	Duration		
p_{11}	0.60	Reg. 1	0.3791	2.51		
p_{22}	0.76	Reg. 2	0.6209	4.11		

Regime specific correlations:	Regime 1		Regime 2			
	dly	dlco2	dlp	dly	dlco2	dlp
dly	1.00	0.57	0.19	1.00	0.66	− 0.09
dlco2	0.57	1.00	0.32	0.66	1.00	− 0.36
dlp	0.19	0.32	1.00	− 0.09	−0.36	1.00

The estimated model is denoted as MS(r)-VAR(p)-MLP(i-h-o), where r stands for the number of regimes, p is the optimum lag length selected for the VAR specification in each regime. In the neural networks architecture, the i, h, and o specifies the number of inputs, the number of hidden units and the number of outputs, respectively. Accordingly, an MS(2)-VAR(3)-MLP(9-3-6) model is a two-regime model with each variable enters the model with three lags; therefore, the total no. of input variables is 9. The input layer is linked to the hidden layer that consists of three nonlinear sigmoid type activation functions. The hidden layer is linked to the output layer that produces three target variables for each two regimes, totaling to six output variables in the output layer with a linear activation function in the output layer

*Though the overall model structure is determined as MS(2)-VAR(3)-MLP(9-3-6), only one predicted variable exists in each vector of the model, therefore architecture in each vector is in the form of MLP (9-3-1)

model in terms of modeling the magnitude dependent, in addition to regime-dependent characteristics between the causal relations among the analyzed variables. The neural networks in each regime provide a more detailed analysis of the relationships between the variables: instead of the ceteris-paribus approach, as done in the linear econometric models, the

Table 3 Comparisons of MS-VAR and MS-VAR-MLP Models

MS-VAR-MLP Model of UK

	dly vector		dlco2 vector		dlp vector	
Criteria	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
MSE	0.0003	0.0002	0.0012	0.00054	0.00079	0.00631
MAE	0.0117	0.0112	0.0254	0.01495	0.02123	0.05471
RMSE	0.0190	0.0143	0.0356	0.02325	0.02814	0.07943

MS-VAR Model of UK

Criteria	dly vector	dlco2 vector	dlp vector
MSE	0.0078	0.2291	0.2549
MAE	0.0345	0.3701	0.3362
RMSE	0.0883	0.4787	0.5048

MS-VAR-MLP Model of USA

	dly vector		dlco2 vector		dlp vector	
Criteria	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
MSE	0.0003	0.0009	0.0088	0.0033	0.0064	0.0064
MAE	0.0122	0.0221	0.0728	0.0381	0.0471	0.0548
RMSE	0.0173	0.0313	0.0939	0.0576	0.0803	0.0806

MS-VAR Model of USA

Criteria	dly vector	dlco2 vector		dlp vector
MSE	0.0017	0.0023		0.0123
MAE	0.0314	0.0339		0.0762
RMSE	0.0407	0.0475		0.1110

MSE, MAE, and RMSE are mean squared error, mean absolute error, and root mean squared error respectively

variables in the analysis are nonlinear functions of themselves and are not held constant. Therefore, as an economist's perspective, the model assumes mutatis-mundi, i.e., variables cannot be assumed constant, and the changes in the relations between the variables are not only specific to the regime. The economy is at but also the stages in each regime result in differentiated causal relations depending on the magnitude. In the MS-VAR-MLP specification, complex nonlinear relations could be effectively captured by the use of neural networks.

The main advantages of the MS-VAR-MLP over its MS-VAR counterpart could be summarized as (i) its flexibility in terms of modeling regime-dependent asymmetry, while allowing neural networks specifications within each regime; therefore, having no a-priori functional form between the analyzed series, (ii) the model benefits from the improved generalization capabilities of NN learning algorithms; hence, augments the estimation of the MS-VAR models with the neural networks algorithms, overcoming problems, such as convergence to local minima or mistakenly defining outliers as distinct regimes. (iii) The model relaxes the ceteris-paribus assumption, since the variables are not assumed to be constant at different size and magnitudes. (iv) The model could be extended to the regime-dependent sensitivity analysis to investigate the impact of each variable on the other variables analyzed, (v) in terms of the investigated EKC relation, the model provides important insights, the relations between emissions,

economic growth, and petrol prices are more complex than assumed and the model is capable of contributing by capturing such relations.

One should also note that the MS-VAR-MLP model has disadvantages over the MS-VAR model: (i) the estimation of the MS-VAR-MLP model requires comparatively larger sample sizes given the number of parameters to be estimated, the loss in degrees of freedom is larger and this could be controlled only for a certain degree with SIC information criteria, (ii) if the researcher aims at estimating the model with more than two regimes, with each additional regime, the number of parameters to be estimated is multiplied, the requirement of large sample becomes more serious. (iii) Though the MS-VAR-MLP model proposed in the study is parametric but the model assumes no ceteris-paribus, the parametric interpretation is harder than its MS-VAR variant. However, one could overcome the last disadvantage given above with the suggested regime-dependent sensitivity analysis. In summary, the overall evaluation suggests a trade-off between complexity and easiness in interpretability. Depending on the complexity of the analyzed phenomenon, the researcher could choose between MS-VAR and its MLP variant. As in the case for EKC, one hint of such complexity is the derivation of significantly differentiated findings in the literature depending on the econometric methodology and the selection of variables. As a result, if one selects polynomial regressions over more complex analysis techniques, the over generalization of the

phenomenon could result in wrong policy implications and linearity in parameters could be mostly misleading.

In addition to abovementioned advantages and disadvantages, similar to the discussion given in the literature section, another limitation of our paper is the selection of CO_2 as the variable of pollution. Our selection is based on two pillars, first one is the availability of CO2 data starting from the eighteenth century and the second is to reach comparable results with the majority of literature that follow early studies in the investigation of the EKC. By doing so, the paper aimed at underlining the capability of the MS-VAR-MLP model in capturing the complex relations between emissions, income and petrol prices, and providing comparable results to papers utilizing polynomial regressions and their panel variants. Recent studies propose the differentiation of the data representing types of ecological footprint, including land and water resources. Al-Mulali et al. (2016) and Bello et al. 2018) show that the impact of the explanatory variables could be significantly altered on the type of ecological footprint selected including, water, land, and air. Gil-Alana and Solarin (2018) investigate the effectiveness of the environmental policies of the USA by global and per capita NOx and VOC emissions for data covering 50 years, and show that the environmental policies implemented in 1965, 1967, 1970, 1977, and 1990 are effective in reducing emissions. They also note different estimates for fractionality parameters with wide confidence intervals after the policy implementations. Another important paper is Solarin and Al-Mulali (2018), which investigate the impact of the foreign direct investment (FDI) and economic factors, including GDP in 20 countries by differentiating between types of environmental degradation as CO_2, carbon footprint and ecological footprint. Their results reveal the necessity of investigation of the EKC with ecological footprint indicators, in addition to pollution indicators generally accepted by the scholars. Due to availability of data and to obtain comparable results to the mainstream papers following long-run analysis of the EKC, the paper is limited to the investigation of CO_2 series. The application of the MS-VAR-MLP model to other ecological footprint indicators is left for future studies if larger datasets are available.

Our findings based on the suggested MS-VAR-MLP model and the proposed regime-dependent sensitivity analyses lead to important policy suggestions. If the impact of economic growth on CO2 emissions growth is evaluated for the UK and for the USA, the negative effect of economic growth on emissions is restricted to the cases, where economic growth rates are very close to 0%. Even for the majority of first decile ranges that represent very negative growth rates, the impact is positive. This result confirms the previous results obtained that suggest the selection of economic prosperity over environment for the majority of cases. Further, the petrol prices in the history had strong impact on the business cycles in both countries. The policies towards the environment should take the complex

relations between the analyzed variables into consideration and even for the analyzed developed countries, the environmental degradation could not be reversed unless various measures are taken. One important point to be made regarding the Rostow-type analysis is that, at the last stages of development, the society is transformed into a state of high mass consumption, in which the share of industry is lowered and the economy is dominated with the service sectors. Even though this is the case, in the stage of high mass consumption, the domestic demand for various manufactured goods is satisfied with the supply of goods originated from the industries in the developing world including China. As a result, the results for the EKC curve are restricted to be domestic. The role of high mass consumption in the developed world on the pollution created through the production in the East should not be overseen. The environmental degradation is a public good with negative externalities to the world. Though such effects are not in the focus of our study, as CDIAC data showed, the CO_2 emissions in China increased by three times in the last two decades. Further, as shown by Liu et al. (2018), with a slowly decreasing trend, more than 70% of the energy demand is satisfied with non-renewable energy sources, such as coal. Solarin and Al-Mulali (2018) point at the finding that the lack of environmental policies in the developing countries, including China create incentives for attracting FDI to their home country under the strict and contrasting environmental policies of the developed countries. Shen et al. (2017) investigate the environmental policies, FDI, and manufacturing relations, and their findings are in favor of the validity of the pollution haven hypothesis for China as a location for various pollutant industries.

China is not in the focus of the study and nor is the only responsible country for the global emissions and global warming. As noted in the introduction section, the shift of production from the more developed to less developed countries is not only a transfer of emissions to the developing world but also is a result of the increased GDP and consumption levels of the more developed countries. Even an inverted U-shaped EKC would have been observed for the developed countries, the role in their high mass consumption of goods produced in the developing countries translate into a derived demand of emissions in the world. In addition to the carbon taxes both at the consumption and production level in the developed countries, the developed countries should take role in investing on non-polluting energy production in the developing countries. These policies could include indirect policies, such as carbon tariffs; however, direct policies, such as investing in renewable energies, such as solar and wind in developing nations should be taken into agendas. Nevertheless, emissions and global warning are global public goods and the negative externality created by the developed countries should be internalized by the developed nations themselves, in addition to developing countries to achieve global environmental sustainability.

Conclusion

This study aimed to analyze the dynamics between CO_2 emissions, GDP growth rates, and petrol prices for the UK and the USA during the period 1871–2016. For this purpose, the study contributed to the literature by providing a neural network approach to nonlinear causality analysis based on the newly introduced regime-switching MS-VAR-MLP models, which allow causality to follow nonlinearity and asymmetry in terms of direction, size, and magnitude among different regimes and within each regime under analysis. The MS-VAR-MLP model provided visual sensitivity analysis that eases the interpretation of complex relations between the analyzed variables. The proposed MS-VAR-MLP model introduced the regime-switching to the MLP (and VAR-MLP) processes governed by Markov chains. With MS-VAR-MLP-based regime-dependent sensitivity analysis, the general dependency of causal effects of each variable to another could be easily modeled and interpreted. As a result, the proposed methodologies allow the researcher to investigate dependency of the causality between the analyzed variables to the magnitude (size) of each variable within each distinct regime of the business cycle.

The empirical results suggested that the MS-VAR-MLP model and the regime-dependent sensitivities confirmed the causal relations between the GDP growth rates, petrol price growth rates, and CO2 emissions during both of the regimes, expansions and recessions, with varying magnitudes within each distinct regime. The overlook of the results showed that the causality from GDP to CO_2 growth rates could not be rejected and the positive impact of economic growth on emissions failed to be reversed at various magnitudes of economic growth. The findings should be taken as a rejection of an inverse U-shaped EKC relation for the analyzed countries. The results are listed in detail as follows. The determination of the direction of causality deserved special importance in the study. The empirical results obtained for the MS-VAR-MLP and the regime-dependent sensitivity analyses confirmed that the direction of relationship ran not only from GDP to CO_2, but also in the opposite direction. By including the impact of petrol prices to the environment-economic development relation, important information had been gathered not only in terms of its impact on the business cycles only, but also in terms of the two-directional relation between the emissions and the petrol prices. In contrast to the general expectation that economic development and CO_2 emissions were closely linked, GDP and CO_2 variables also had significant effects on the petrol prices in both the UK and USA. As discussed in the "Literature review" section, in addition to the USA's influence in the petrol market, the UK's production and emissions had strong influence on the petrol prices through its production and refinery operations in addition to its influential role in a global scale.

The results favored that the relationship between the economic and environmental variables determined as being far from an inverted U-shape. Instead, the relation showed a J- or an S-shaped curve depending on the regime the economy is in. If two regimes are analyzed together to cover the whole decile ranges of economic growth, the shape of the relation is close to a U-shape. The J and S curves in two regimes suggested that as economic growth was accelerated, CO_2 emissions accelerated by following a sigmoidal S curve, and the acceleration rates differed drastically at various magnitudes of economic growth. Further, the dependence on magnitudes of variables within each regime had shown significant deviations from the findings, which could not be achieved with the polynomial regressions that assumed a stable relation for the whole period, in addition to the MS-VAR models that assumed constancy of parameters and the stable relation for the determined regimes.

The impact of petrol price growth rates had significant nonlinear impacts at various stages of economic growth. In the recessions, the causal impact of petrol prices on CO_2 emissions followed a decreasing but positive path so that the impact became negative as growth approached 0 %. In the expansionary regimes, the positive impact of petrol prices in the lowest deciles was the largest. As the economy moved towards the deciles that represented higher petrol price growth rates, the impact was comparatively lower. Similar complex nonlinear relations were captured for both countries with differences. Moreover, the causal impact of CO_2 emissions and GDP showed similarities in terms of the path followed. This co-movement suggested a significant lead-lag relation between GDP growth rates and CO_2 emissions, pointing to a co-occurrence of CO_2 emissions and GDP growth rates for the overall results.

The results obtained for the MS-VAR-MLP model further enhanced the nonlinear structure of the causality between variables. Using the MS-VAR-MLP model, this paper aimed at nonlinear causality analysis in light of the provided pointwise sensitivity methodology. According to the sensitivity results, the causal links between petrol price growth rates, GDP growth rates, and CO_2 emissions had different impacts depending on the size and magnitude of GDP growth rates even within the expansionary and recessionary regimes. According to the results, positive causal effects of GDP growth rates on CO_2 emissions could not be rejected during recessions and expansions, while the GDP growth rates clearly had a positive impact on CO_2 emissions. One interesting finding was the impact decayed towards zero at the last episodes of recessions only. Further, during the last decile ranges in the expansions where the economic growth rate was moderate and high, the emissions followed a positive and accelerating path. Hence, the positive path followed by emissions proved that the impact of economic growth on CO_2 emission growth rates possessed important departure from the literature. Therefore, the proposed hybrid approach that benefited from the MS-VAR,

neural networks and sensitivity analysis offered flexibility in terms of analyzing the complex nonlinear causal relations among variables without assuming linear and stable relations.

Due to number of parameters to be estimated for the MS-VAR-MLP similar to other NN models, the large sample size is a necessity and is recommended for researchers. In addition to the required large samples, the findings of the study are also limited to the investigation of the CO_2 data instead of evaluating different ecological footprint indicators. As discussed in the discussions section, the justification of CO_2 is based on aiming at providing comparable results to the literature to highlight the complexity of the EKC relation compared to the findings in the literature that follow long-run data approaches. In addition, the availability of the CO_2 data starting from the eighteenth century also limited the study to the investigation of CO_2 emissions. As of our knowledge, the differentiation between various forms of footprint, such as land and water are not available for this long series. However, if the frequency of the data could be increased, such as monthly data, the application of the MS-VAR-MLP to other ecological footprint indicators could provide important insights; this research is left for future studies.

References

Alam, Begum IA, Buysse J, Huylenbroeck GV (2012) Energy consumption, carbon emissions and economic growth nexus in Bangladesh: cointegration and dynamic causality analysis. Energy Policy 45: 217–225

Al-Mulali U, Solarin SA, Sheau-Ting L, Ozturk I (2016) Does moving towards renewable energy cause water and land inefficiency? An empirical investigation. Energy Policy 93:303–314

Ang JB (2007) CO_2, emission, energy consumption and output in France. Energy Policy 35:4772–4778

Apergis N (2016) Environmental Kuznets curves: new evidence on both panel and country-level CO2 emissions. Energy Econ 54:263–271

Arouri MEH, Jawadi F, Nguyen DK (2012a) Nonlinearities in carbon spot-futures price relationships during Phase II of the EU ETS. *Econ Model 29*(3):884–892

Arouri ME, Youssef A, M'henni H, Rault C (2012b) Energy consumption, economic growth and CO2 emissions in Middle East and North African countries. Energy Policy 45:342–349

Aslanidis N, Iranzo S (2009) Environment and development: is there a Kuznets curve for CO2 emissions? Appl Econ *41*(6):803–810

Atasoy BS (2017) Testing the environmental Kuznets curve hypothesis across the U.S.: evidence from panel mean group estimators. Renew Sus. Energ Rev 77:731–747

Bello MO, Solarin SA, Yen YY (2018) The impact of electricity consumption on CO2 emission, carbon footprint, water footprint and ecological footprint: the role of hydropower in an emerging economy. J Environ Manag 219:218–230

Bergin T (2008) Oil majors' output growth hinges on strategy shift. Reuters. https://www.reuters.com/article/us-oilmajors-production/oil-majors-output-growth-hinges-on-strategy-shift-idUSL169721220080801. Accessed 19.01.2018

Bildirici M (2013) Economic growth and electricity consumption: MS-VAR and MS-Granger causality analysis. OPEC Ener Rev 37(4): 447–476

Bildirici M, Ersin Ö (2009) Improving forecasts of GARCH family models with the artificial neural networks: an application to the daily returns in Istanbul stock exchange. Exp Sys with App 36:7355–7362

Bildirici M, Ersin Ö (2013) Forecasting oil prices: smooth transition and neural network augmented GARCH family models. J Pet Sci Eng 109:230–240

Bildirici M, Ersin Ö (2014) Modeling Markov switching ARMA-GARCH neural networks models and an application to forecasting stock returns. Sci World J 2014:1–22

Bildirici M, Ersin Ö (2018) Economic growth and CO2 emissions: an investigation with smooth transition autoregressive distributed lag models for the 1800–2014 period in the USA. Environ Sci Pollut Res 25(1):200–219

Bildirici M, Gökmenoğlu S (2017) Environmental pollution, hydropower energy consumption and economic growth: evidence from G7 countries. Renew Sust Energ Rev 75:68–85

Bishop C (1995) Neural networks for pattern recognition, 1st edn. Oxford, New York

Bloomberg Businessweek (2018) Company Overview of BP Exploration & Production Inc. https://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=2414356. Accessed 19 Jan 2018

BP (2018) History of BP. https://www.bp.com/en/global/corporate/who-we-are/our-history.html. Accessed 18 Jan 2018

CDIAC, 2016. Carbon dioxide information analysis center database. http://cdiac.ornl.gov. Accessed 11 Nov 2016

Charfeddine L (2017) The impact of energy consumption and economic development on ecological footprint and CO2 emissions: evidence from a Markov switching equilibrium correction model. Energy Econ 65:355–374

Cheng B, Titterington DM (1994) Neural networks: a review from statistical perspective. Stat Sci 9(1):49–54

Chevallier J (2011a) Macroeconomics, finance, commodities: interactions with carbon markets in a data-rich model. Econ Model 28(1–2):557–567

Chevallier J (2011b) A model of carbon price interactions with macro-economic and energy dynamics. Energy Econ 33:1295–1312

Cole MA, Rayner AJ, Bates JM (1997) The environmental Kuznets curve: an empirical analysis. Environ Dev Econ 2:401–416

Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Math Control Signals Syst 2:303–314

Dimopoulos Y, Bourret P, Lek S (1995) Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process Lett 2(6):1–4

Dimopoulos I, Chronopoulos J, Chronopoulou-Sereli A, Lek S (1999) Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). Ecol Mod 120(2–3):157–165

Engelbrecht AP, Cloete I, Zurada JM (1995) Determining the significance of input parameters using sensitivity analysis. In: Mira J, Sandoval F (eds) From natural to artificial neural computation. IWANN 1995. Lecture Notes in Computer Science, vol 930. Springer, Berlin, Heidelberg

Engelbrecht AP, Flectcher L, Cloete I (1999) Variance analysis of sensitivity information for pruning neural networks. In: IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339). https://doi.org/10.1109/IJCNN.1999.832657

Ersin Ö (2009) Türkiye'de Fiyatlar Genel Düzeyine İlişkin Maliye Teorisinin Doğrusal Olmayan Zaman Serisi Modelleri Bakımından İncelenmesi. PhD. Thesis, Yıldız Tech. Uni., Inst. of Soc. Sci., Dept. of Econ., Istanbul

Ersin Ö (2016) The nonlinear relationship of environmental degradation and income for the 1870–2011 period in selected developed countries: the dynamic panel-STAR approach. Procedia Econ Fin 38: 318–339

Esteve V, Tamarit C (2012) Threshold cointegration and nonlinear adjustment between CO 2 and income: the environmental Kuznets curve in Spain, 1857–2007. Energy Econ 34(6):2148–2156

Fezzi C, Bunn DW (2009) Structural interactions of European carbon trading and energy prices. The Journal of Energy Markets 2(4):53

Fodha M, Zaghdoud O (2010) Economic growth and pollutant emissions in Tunisia: an empirical analysis of the environmental Kuznets curve. Energy Policy 38(2):1150–1156

Fosten J, Morley B, Taylor T (2012) Dynamic misspecification in the environmental Kuznets curve: evidence from CO2 and SO2 emissions in the United Kingdom. Ecol Econ 76:25–33

Gevrey M, Dimopoulos I, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol Model 160:249–264

Ghosh S (2010) Examining carbon emissions economic growth nexus for India: a multivariate cointegration approach. Energy Policy 38: 3008–3014

Gil-Alana LA, Solarin SA (2018) Have U.S. environmental policies been effective in the reduction of U.S. emissions? A new approach using fractional integration. Atmos Pollut Res 9:53–60

Granger CWJ, Terasvirta T (1993) Modelling dynamic nonlinear economic relationships, first edn. Oxford Uni. Press, Oxford

Grossman G, Krueger A (1991) Environmental impacts of a North American free trade agreement. NBER Working Papers 3914, pp 1–57. http://www.nber.org/papers/w3914. Accessed 7 Sept 2018

Guo Z, Ward M, Rundensteiner E, Ruiz C (2011) Pointwise local pattern exploration for sensitivity analysis. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp 131–140. https://doi.org/10.1109/VAST.2011.6102450

Hadzima-Nyarko M, Nyarko E, Moric D (2011) A neural network based modelling and sensitivity analysis of damage ratio coefficient. Expert Syst Appl 38:13405–13413

Halkos GE, Tsionas EG (2001) Environmental Kuznets curves: Bayesian evidence from switching regime models. Energy Econ 23:191–210

Hamilton JD (1990) Analysis of time series subject to regime changes. J Econ 45:39–70

Hamilton JD (2011) Historical oil shocks. NBER Working Paper 16790, pp 1–52. http://www.nber.org/papers/w16790.pdf. Accessed 9 Sept 2018

Jalil A, Mahmud SF (2009) Environment Kuznets curve for CO2 emissions: a cointegration analysis for China. Energy Policy 37(12): 5167–5172

Kim S, Lee K, Nam K (2010) The relationship between CO2 emissions and economic growth: the case of Korea with nonlinear evidence. Energy Policy 38(10):5938–5946

Krolzig HM (1998) Econometric modelling of Markov-switching vector autoregressions using MSVAR for Ox. http://fmwww.bc.edu/ec-p/software/ox/Msvardoc.pdf. Accessed 21 Jan. 2018

Krolzig HM (2000) Predicting Markov-switching vector autoregressive processes. Department of Economics and Nuffield College, Oxford. https://pdfs.semanticscholar.org/2b99/ebe2736ea800384370db30325ec27f6d5347.pdf. Accessed 7 Sept 2018

Krolzig HM, Clements MP (2002) Can oil shocks explain asymmetries in the US business Cycle? Empir Econ 27(2):185–204

Krolzig HM, Toro J (2005) Classical and modern business cycle measurement: the European case. Span Econ Rev 7(1):1–21

Kuan C-M, White H (1994) Artificial neural networks: an econometric perspective (with discussions). Econometric Rev 13:1–91 139–143

Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S (1996) Application of neural networks to modelling nonlinear relationships in ecology. Ecol Model 90(1):39–52

Lindmark M (2002) An EKC-pattern in historical perspective: carbon dioxide emissions, technology, fuel prices and growth in Sweden 1870–1997. Ecol Econ 42(1–2):333–347

Liu J, Zhang X, Song X (2018) Regional carbon emission evolution mechanism and its prediction approach driven by carbon trading-a case study of Beijing. J Clean Prod 172:2793–2810

Martinez-Zarzoso I, Maruotti A (2013) The environmental Kuznets curve: functional form, time-varying heterogeneity and outliers in a panel setting. Environmetrics 24:461–475

Mensah JT (2014) Carbon emissions, energy consumption and output: a threshold analysis on the causal dynamics in emerging African economies. Energy Policy 70:172–182

Menyah K, Wolde-Rufael Y (2010) Energy consumption. pollutant emissions and economic growth in South Africa. Energy Econ 32(6): 1374–1382

Molas G, Yamazaki F (1995) Neural networks for quick earthquake damage estimation. Earthq Eng Struct Dyn 24:505–516

Olivier J, Janssens-Maenhout G, Muntean M, Peters J (2015) Trends in the global CO2 emissions: 2015 report. PBL Netherlands Environmental Assessment Agency Publication 1803, pp 1–98. http://www.pbl.nl/en/publications/trends-in-global-co2-emissions-2015-report. Accessed 18 Sept 2016

Olteanu M, Rynkiewicz J, Maillet B (2004) Nonlinear analysis of shocks when financial markets are subject to changes in regime. ESANN 2004 proceedings, pp 87–92

Özokçu S, Özdemir Ö (2017) Economic growth, energy, and environmental Kuznets curve. Renew Sustain Energy Rev 72:639–647

Pan Z, Wang Y, Wu C, Yin L (2017) Oil price volatility and macroeconomic fundamentals: a regime switching GARCH-MIDAS model. J Empir Financ 43:130–142

Pao HT, Tsai CM (2010) CO2 emissions, energy consumption and economic growth in BRIC countries. Energy Policy 38(12):7850–7860

Park J, Hong T (2013) Analysis of South Korea's economic growth, carbon dioxide emission, and energy consumption using the Markov switching model. Renew Sustain Energy Rev 18:543–551

Plassmann F, Khanna N (2006) Household Income and Pollution: Implications for the Debate About the Environmental Kuznets Curve Hypothesis. J Environ Dev 15(1):22–41

Pesaran MH, Shin Y, Smith RJ (2001) Bounds testing approaches to the analysis of level relationships. J Appl Econ 16(3):289–326

Rasli A, Qureshi M, Isah-Chikaji A, Zaman K, Ahmad M (2018) New toxics, race to the bottom and revised environmental Kuznets curve: the case of local and global pollutants. Renew Sustain Energy Rev 81(2):3120–3130

Rehman MU (2018) Do oil shocks predict economic policy uncertainty?. Physica A Stat Mech Appl 498:123–136

Richmond AK, Kaufmann RK (2006) Is there a turning point in the relationship between income and energy use and/or carbon emissions? Ecol Econ 56:176–189

Roach T (2015) Hidden regimes and the demand for carbon dioxide from motor-gasoline. Energy Econ 52:306–315

Rostow WW (1960) The stages of economic growth: a non-communist manifesto, third edn. Cambridge University Press, Cambridge

Sanjari F, Delangizan S (2010) Carbon emissions and economic growth: the Iranian experience. SSRN, 1–7. http://ssrn.com/abstract=1635233. Accessed 21 Jan 2018

Selden TM, Song D (1994) Environmental quality and development: is there a Kuznets curve for air pollution? J Environ Econ Environ Mgmt 27:147–162

Shafik N, Bandyopadhyay S (1992) Economic growth and environmental quality: time series and cross-country evidence. World Bank Working Paper Series 904, pp 1–55

Shen J, Wei YD, Yang Z (2017) The impact of environmental regulations on the location of pollution-intensive industries in China. J Clean Prod 148:785–794

Sims CA (1980) Macroeconomics and reality. Econometrica 48(1):1–48

Sinha A, Shahbaz M (2018) Estimation of environmental Kuznets curve for CO2 emission: role of renewable energy generation in India. Renew Energy 119:703–711

Solarin SA, Al-Mulali U (2018) Influence of foreign direct investment on indicators of environmental degredation. Environ Sci Pollut Res. https://doi.org/10.1007/s11356-018-2562-5

Stern DI, Enflo K (2013) Causality between energy and output in the long-run. Energy Econ 39:135–146

Stern DI, Common MS, Barbier EB (1996) Economic growth and environmental degradation: the environmental Kuznets curve and sustainable development. World Dev 24(7):1151–1160

Sun W, Liu M (2016) Prediction and analysis of the three major industries and residential consumption CO2 emissions based on least squares support vector machine in China. J Clean Prod 112:144–153

Swanson N, White H (1997) A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. Rev Econ Stat 79:540–550

Terasvirta T (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. J Am Stat Assoc 89:208–218

Tong H (1990) Non-linear time series: a dynamical system approach, first edn. Oxford University Press, Oxford

Tucker M (1995) Carbon dioxide emissions and global GDP. Ecol Econ 15(3):215–223

U.K. Trade and Investment Report (2012) http://www.ukti.gov.uk/export/countries/asiapacific/middleeast/iraq/businessopportunity/345360.html. Accessed 19 Jan 2018

U.S. Energy Information Administration EIA (2018) https://www.eia.gov/todayinenergy/detail.php?id=16971. Accessed 19 Jan 2018

Unruh GC, Moomaw WR (1998) An alternative analysis of apparent EKC-type transitions. Ecol Econ 25:221–229

Wang SS, Zhou DQ, Zhou P, Wang QW (2011) CO2 emissions, energy consumption and economic growth in China: a panel data analysis. Energy Policy 39:4870–4875

White H (1992) Artificial neural networks: approximation and learning theory, first edn. Blackwell, Oxford

Winchester N, Ledvina K (2017) The impact of oil prices on bioenergy, emissions and land use. Energy Econ 65:219–227

Wutsqa DU, Guritno S, Guritno Z (2006) Forecasting performance of VAR-NN and VARMA models. In: Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications Universiti Sains Malaysia. http://staff.uny.ac.id/sites/default/files/132048772/penang%20paperbaru.pdf. Accessed 7 Sept 2018

Xu T (2018) Investigating environmental Kuznets curve in China–aggregation bias and policy implications. Energy Policy 114:315–322

Yavuz NC, Yilanci V (2013) Convergence in per capita carbon dioxide emissions among G7 countries: a TAR panel unit root approach. Environ Resour Econ 54(2):283–291

Zambrano-Monserrate M, Silva-Zambrano C, Davalos-Penafiel J, Zambrano-Monserrate A, Ruano M (2018) Testing environmental Kuznets curve hypothesis in Peru: the role of renewable electricity, petroleum and dry natural gas. Renew Sustain Energy Rev 82:4170–4178

Zhang XP, Cheng XM (2009) Energy consumption, carbon emissions, and economic growth in China. Ecol Econ 68:2706–2712

Zi C, Jie W, Hong-Bo C (2016) CO2 emissions and urbanization correlation in China based on threshold analysis. Ecol Indic 61:193–201