This is a repository copy of *An environmental Kuznets curve for global forests: an application of the mi-lasso estimator*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/218851/

Version: Published Version

Article:

Cherodian, R. and Fraser, I. (2024) An environmental Kuznets curve for global forests: an application of the mi-lasso estimator. Forest Policy and Economics, 168. 103304. ISSN 1389-9341

https://doi.org/10.1016/j.forpol.2024.103304

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Forest Policy and Economics 168 (2024) 103304

Contents lists available at ScienceDirect

Forest Policy and Economics

journal homepage: www.elsevier.com/locate/forpol

An environmental Kuznets curve for global forests: An application of the mi-lasso estimator

Rowan Cherodian [a], Iain Fraser [b,*]

[a] Population Health, School of Medicine and Population Health, University of Sheffield, Barber House, 387 Glossop Road, Sheffield S10 2HQ, UK
[b] School of Economics and DICE, University of Kent, Canterbury, CT2 7NP, UK

ARTICLE INFO

JEL codes:
C21
Q23
Q56

Keywords:
Forestry
Spatial dependence
Turning points
Mi-lasso estimator

ABSTRACT

In this study, we employ a Moran's *i* based Lasso (Mi-Lasso) methodology to address the spatial dependence of an unspecified functional form, investigating the association between a country's economic growth and the rate of deforestation. Our aim is to explore the existence of a forestry environmental Kuznets curve (EKC). Our approach to handling spatial dependence overcomes limitations identified in existing EKC literature. We estimate a series of cross-sectional data models spanning the period from 1990 to 2020 for 146 countries. Our findings indicate a non-linear relationship, revealing a change peak rate of deforestation over time. Additionally, we observe that the income threshold at which the deforestation rate begins to decrease changes over time with differences observed between model specifications. Crucially, our results highlight that failing to account for spatial dependence leads to a significant absolute upward bias in ordinary least squares (OLS) estimates of income and worse model fit.

1. Introduction

Understanding the relationship between environmental degradation and economic development is an important question, especially in the context of accelerating climate change. Grossman and Krueger (1991), Shafik and Bandyopadhyay (1992) and Grossman and Krueger (1995) are among the first to examine the relationship between economic growth and environmental quality using the environmental Kuznets curve (EKC). Subsequently, a huge literature has developed examining empirically and theoretically the EKC. Insightful reviews and surveys of the literature are provided by Stern (2017), Shahbaz and Sinha (2019), and Purcel (2020).

One natural resource that has been the subject of significant empirical EKC research is forestry (EKCf). Forests, are an important resource at both global and local levels. At the global level, forests play a pivotal role in carbon storage (Seymour and Busch, 2016) whereas at the local level, forests play a key role in supporting biodiversity. The extent of biodiversity loss of forest specialist species is reported by Almond et al. (2022) to have been 53% between 1970 and 2018. However, the economic value of forest ecosystem services is still globally significant (Taye et al., 2021).

The main reason why forestry has been examined in relation to the EKC hypothesis is because of the well documented extent of deforestation that has occurred (Williams, 2003, 2008). Historically, human activity has played a crucial role in reshaping the forest landscape for at least 6000 years, with timber being used as a coforforpolre input for economic development (Williams, 2008). Recently, the FAO (2021) reports that the world is still experiencing net forest loss, although at a declining rate. In contrast, Song et al. (2018) using satellite data found since 1982 that total forest cover has increased, as the decline in tropical forest cover has been outweighed by the gain in forest cover in boreal, subtropical, and temperate regions. This growth in forest cover is described in the literature as the forest transition (Mather, 1992; Wolfersberger et al., 2015; Barbier et al., 2017; Benedek and Ferto, 2020). It can occur at the country or region level and indicates a change from net forest area loss to net gain. Over the same period economic growth has been relatively constant and Caravaggio (2020*b*) argues this is potentially evidence for the presence of the EKCf. In addition, over the last 20 years, carbon dioxide emissions from land-use, land-use change, and forestry have shown a slight decrease (Friedlingstein et al., 2021).

Importantly, when investigating the EKCf, only a few studies have tried to account for the spatial dimension of deforestation that is

* Corresponding author.
E-mail addresses: r.cherodian@sheffield.ac.uk (R. Cherodian), i.m.fraser@kent.ac.uk (I. Fraser).

https://doi.org/10.1016/j.forpol.2024.103304
Received 12 March 2024; Received in revised form 1 August 2024; Accepted 5 August 2024
Available online 24 August 2024
1389-9341/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

inherent in the data (McPherson and Nieswiadomy, 2005; Mills and Waite, 2009; Mills Busa, 2013).[1] The need to consider the spatial aspects of the data stem from the simple observation that many countries in close proximity to each other may well have similar forest resources management strategies, uses based on common social norms and types of forest. As such there is good reason to assume that we need to at minimum take account of spatial dependence in country level data sets. This observation is supported by the findings reported in Mills and Waite (2009) and Mills Busa (2013).

Our paper, makes two contributions to the literature. First, we examine the EKCf by robustly accounting for spatial correlation using Eigenvector Spatial Filtering (ESF) (Griffith, 2000, 2003). This approach means that we can avoid possible mis-specification of the spatial parts of the model as we no longer need to generate a weighting matrix prior to model estimation. To implement ESF we select the relevant subset of eigenvectors using Morans' *i* Lasso (Mi-Lasso) that is proposed by Barde et al. (2023). The Mi-Lasso spatial filtering procedure and importantly, the statistical properties of the model estimator and the assumptions necessary for consistent eigenvector selection are known.

Second, due to the availability of forest data from the FAO (2021) global forest resource assessment, we consider four time periods: 1990–2000; 2000–2010; 2010–2015; and 2015–2020. GDP and the controls are all taken from the starting year, i.e., 1900, 2000, 2010 and 2015. This allows us to see the effect that initial income has on the annual deforestation rate over each of the time periods. Importantly, it means that we estimate the EKCf at different points in time. This is a departure from much of the existing literature that typically reports key model results such as turning points for the EKC/EKCf over the entire time period of a sample using time series or panel data methods. There are exceptions in the literature including Chow and Li (2014) and Bernard et al. (2015). As explained by Chow and Li (2014) they estimate a series of cross-section model specifications to avoid the econometric issues that emerge when employing time series data and/or panel data. For example, issues with regard to spurious correlation among the regressors can exist as a result of unit roots in the data. Furthermore, given that many of the data series examined in the literature are at least 30 or 40 years long there is good reason to assume that the shape and behaviour of the EKCf may change over time. Within the wider literature it has been noted by Apergis (2016) that the EKC relationship is not time invariant. Similar results have been reported by Mikayilov et al. (2018) who employ time varying coefficient cointegration to take account of this issue. They also note that if model parameters are time varying but are assumed to be fixed that regression models will yield spurious results. This data issue also relates to the possibility of there being structural breaks in data series which if ignored can mean that many of the time series tests employed in model development to establish if data are stationary will be biased.

Our approach effectively treats each period as structurally different (in parameters) and not just evolving with some time-dependence which is the underlying rationale for dynamic panels that employ time lags. In addition, we also avoid the problem that a one period lag may be different at different points in time (i.e., non-stationarity). For the econometric approach we have implemented our results indicate that the likely turning point of the EKCf is not stable such that the level of GDP required before a country experiences an increase in forest cover changes over time which supports the findings reported by Apergis (2016) and Mikayilov et al. (2018).

The rest of the paper is outlined as follows, Section 2 briefly reviews the antecedent literature, Section 3 then reviews the existing econometric approaches to dealing with spatial dependence and details the method we employ in this paper. Section 4 then describes the data used and Section 5 presents our results. Finally, Section 7 offers our concluding remarks.

[1] A larger number of papers have estimated spatial models when looking at other environmental indicators (Jeetoo and Chinyanga, 2023).

2. Literature review

2.1. EKCf antecedent literature

In terms of empirical research on the EKCf, Bandyopadhyay (1992) were the first to examine forests as an environmental indicator for 77 countries over the period 1961 to 1988, finding minimal support for the EKCf hypothesis. Subsequently the EKCf was examined by Cropper and Griffiths (1994), using panel data for 64 tropical countries identifying for Latin American and African countries evidence of the EKCf, but not for Asian countries. Many other studies have followed including Koop and Tole (1999); Bhattarai and Hammig (2001a); Ehrhardt-Martinez et al. (2002); Araujo et al. (2009); Oliveira and Almeida (2011); Damette and Delacote (2012); Polomé and Trotignon (2016); Leblois et al. (2017); Wang et al. (2019); Caravaggio (2020b); Murshed (2020); Ajanaku and Collins (2021); Farooq and Dar (2022); Pablo-Romero et al. (2023). These studies have employed various types of data including country level panel data sets as well as within country data. There are also a varied collection of econometric methods that have been employed to take account of data structure and associated econometric and modelling requirements (Ajanaku and Collins, 2021).

Another issue that has attracted a great deal of interest in the literature is the selection of control variables to employ in model specifications. Pablo-Romero et al. (2023) provides a useful summary of control variables used in studies examining the EKCf. Typical variables employed include: demographic variables such as population variables; political institutions; measures of institutional quality; energy use; direct foreign investment; trade openness; and agricultural land use. The rational for why specific variables are included or excluded is frequently ad hoc in that the choices do not have no theoretical justification. In addition, how a specific variable is measured has been examined in detail. There are discussion regarding the appropriate choice of the dependent variable as well as examples of which way to measure a control variable (e.g., trade openness (Tameko, 2024)).

In terms of findings reported in the literature, Choumert et al. (2013) provides an excellent meta-analysis of the EKCf. They consider 69 cross-country and country-specific studies between 1992 and 2012, and find that more recent studies with better data, and applying newer econometric techniques tend not to find the hypothesised inverted U-shape. However, and in contrast, Caravaggio (2020*b*), reviews more recent literature and conducts cross-country analysis, finding evidence supporting the EKCf hypothesis. Thus, much of the EKCf literature provides mixed results when looking for the presence of the EKCf, such that the existence of the EKCf remains an open research question (Ajanaku and Collins, 2021). These mixed findings reflect the wider EKC literature as noted by Bernard et al. (2015).

2.2. EKCf and spatial correlation

A standard core assumption placed on cross-sectional units in econometric analysis is independence. If this assumption is violated, i.e., the cross-sectional units are dependent, then the estimates will be biased and/or inconsistent. In cross-country or regional analysis such as that being conducted for the EKCf, it is hard to justify that the cross-sectional units (countries or regions) are independent.

In general, the standard approach to spatial modelling in economics requires the researcher to define a spatial weighting matrix (SWM) that explicitly defines the the pair-wise spatial interactions and which parts of the model are spatially correlated.[2] In this paper, we employ a different approach to dealing with the spatial dependence, given, we view the spatial parameters in the underlying model as nuisance parameters, as we are only interested in the direct effects, i.e., the coefficients on income per capita, controlling for any possible spatial effects. Therefore, rather than trying to explicitly specify a structural spatial model and then estimating the corresponding spatial parameters, we instead employ ESF (Griffith, 2000, 2003). The ESF methodology approximates the spatially correlated parts of the underlying model using a subset of eigenvectors from the SWM as regressors in a linear regression framework. ESF has an additional advantage, as demonstrated in simulations by Cherodian (2023), that the procedure is robust to mis-specification of the SWM. To select the relevant subset of eigenvectors, we use the Morans' i Lasso (Mi-Lasso) proposed in Barde et al. (2023).

In terms of the ECKf Mills and Waite (2009) and Mills Busa (2013) account for the spatial correlation using an approach called principal coordinate of neighbour matrices (PCNM) spatial filtering (Borcard and Legendre, 2002). This approach accounts for the spatial correlation in a model in an agnostic way. PCNM uses principal coordinates (i.e., eigenvectors) from a singular value decomposition of a truncated Euclidean distance matrix. The truncation is implemented through a distance threshold, whereby distances exceeding this threshold are assigned an arbitrarily large value, they propose $4(\alpha)$ where α representing the threshold. The number of candidate eigenvectors is sensitive to the threshold and coordinates with negative eigenvalues cannot be used as they are complex. Both Mills and Waite (2009) and Mills Busa (2013) only use two eigenvectors without discussing how the eigenvectors were selected or what threshold was used for truncation in their analysis. Mills and Waite (2009) estimate a panel quantile regression model with and without the two eigenvectors. They find the eigenvectors are significant but do not qualitatively change the results. Mills Busa (2013), also estimate a panel quantile regression model, finding both eigenvectors are significant but do not discuss or present results without the eigenvector. How these eigenvectors are selected and under what conditions consistent selection is achieved in the procedure used is unclear.

Another strand of the EKC and EKCf literatures also considers spatial dependence but refers to it as cross-sectional dependence (e.g., Apergis (2016); Erdogan (2024); Tameko (2024)). This part of the literature generally employs panel data specifications and undertakes various statistical tests of time series properties of the data as well as tests for cross-sectional dependence such as the Pesaran CD test proposed by Pesaran (2021). Both Apergis (2016) and Tameko (2024) report evidence of cross-sectional dependence in their data but it is then less clear how exactly they deal with this in the resulting models that are estimated. For example, Tameko (2024) reports a large set of results generated by an array of different model specifications that yield significant variation with regard to the turning point of the EKCf. However, the fact that cross-sectional dependence is being identified supports our use of econometric model estimators that explicitly take account of this data property.

3. Method

3.1. Model overview

In general, initial empirical investigations into the EKCf typically estimate the following reduced-form regression specification:

$$\boldsymbol{f} = \beta_1\boldsymbol{y} + \beta_2\boldsymbol{y}^2 + \beta_3\boldsymbol{y}^3 + \boldsymbol{X}\boldsymbol{\zeta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{f}$ is an $n\times 1$ vector of forest cover, $\boldsymbol{y}$ is an $n\times 1$ vector of GDP, $\boldsymbol{X}$ is an $n\times q$ matrix of controls and $\boldsymbol{\varepsilon}$ is an $n\times 1$ error vector. Importantly, this is a reduced form equation where the control variables included vary significantly in the literature. A useful summary of recent examples in the EKCf literature is provided by Pablo-Romero et al. (2023).

The classic EKCf inverted u-shape requires a positive β_1 coefficient, a negative β_2 coefficient, and a zero β_3 coefficient. This would indicate that the forest cover falls (deforestation rate increases) initially with $\boldsymbol{y}$ until a point and then rises (falls) as $\boldsymbol{y}$ increases further. The majority of studies testing ECKf hypotheses adopt this type of model specification and estimate the model using conventional statistical techniques. However, we note, that Caravaggio (2020*a*) has argued that a second turning point can occur as a result of forest recovery at the point of maximum reforestation.

Only a limited number of papers have endeavored to consider spatial (cross-sectional) dependence in the examination of either the EKC or the ECKf. As noted, spatial modelling in economics typically requires the researcher to pre-define two key components:

1. An $n\times n$ SWM where the elements define the pair-wise spatial interactions, i.e., the spatial structure; and
2. Which parts of the structural model are spatially correlated, i.e. a spatial economic model.

Some commonly used spatial economic models are, the first-order spatial autoregressive (SAR(1)) model and the spatial error model (SEM).[3] A problem with pre-specifying spatial models like the SAR(1) or SEM is that there is no guarantee the chosen model is the correct specification. A standard robustness check that is applied in spatial economics is assessing the sensitivity of the parameter estimates to different SWMs. If model estimates prove sensitive to the choice of SWM, researchers often attribute this to the choice of SWM. However, as demonstrated by LeSage and Pace (2014), as long as the SWMs are reasonably well correlated, results should not be overly sensitive to the exact SWM employed. Instead the observed sensitivity may stem from the misspecification of the spatial economic model rather than the choice of SWM. As such LeSage and Pace (2014) argue that researchers should prioritise specifying the spatial model over finding the 'ideal' SWM.

Earlier EKC research incorporating spatial dependence such as Maddison (2006), and Wang and Ye (2017), estimated SAR(1) model specifications. In contrast, Hao et al. (2016) employed a model a spatial lag of the dependent variable and exogenous variables, to examine coal consumption, and Chang et al. (2021) estimated a dynamic panel with fixed effects and a spatial lag of the dependent variable using Generalised Method of Moments.

3.2. Underlying model specification

Our underlying model is:

$$\Delta\mathbf{f} = \sum_{i=1}^{3}\beta_i\mathbf{y}^i + \mathbf{X}\boldsymbol{\zeta} + g(\mathbf{W}, \Delta\mathbf{f}, \mathbf{X}, \mathbf{y}) + \mathbf{u} \tag{2}$$

where $\Delta\boldsymbol{f}$ is the change in forest cover, and $g(\boldsymbol{W}, \Delta\boldsymbol{f}, \boldsymbol{X}, \boldsymbol{y})$ is some linear in parameter function of the spatial weights matrix $\boldsymbol{W}$ and possibly $\Delta\boldsymbol{f}$, $\boldsymbol{X}$, and $\boldsymbol{y}$ which can include not just first-order spatial lags but also higher-order spatial lags (powers). This type of model specification has

[2] An alternative name used to describe spatial correlation is cross-sectional dependence which is widely used in econometrics and within the global vector autoregressive (GVAR) literature the weighting matrix is referred to as the connectivity matrix (Elhorst et al., 2021).

[3] A SAR(1)is model with a first order spatial lag of the dependent variable and a SEM is a model where the errors follow a pure spatially autoregressive process.

generally been employed in much of the earlier EKCf literature.

Turing to the SWM $\boldsymbol{W}$ we employ a specification similar to Csereklyei and Stern (2015). We use a binary connectivity matrix where pairs of countries get a one if they are neighbours and zero otherwise. Our definition of neighbour is if they share a land border (this includes lakes), and for countries with no land borders, we judge their nearest neighbour. For example, New Zealand's nearest neighbour is Australia and Australia closest neighbour is Indonesia, thus New Zealand and Australia has one and two neighbour, respectively. The matrix is symmetric, and the diagonal elements are set equal to zero.[4]

3.3. Moran's i based lasso

With our underlying model shown in eq. (2), there is substantial uncertainty over which parts of the model exhibit spatial correlation, i.e. the functional form of $g(\boldsymbol{W}, \Delta \boldsymbol{f}, \boldsymbol{X}, \boldsymbol{y})$. However, as we have already noted, we are not explicitly interested in estimating any corresponding spatial parameters, the coefficients of interest are the β_i's. Thus, rather than explicitly specifying which parts of the models are spatially correlated and estimating a standard spatial economic model like the SAR(1) or SEM, we approximate $g(\mathbf{W}, \Delta \mathbf{f}, \mathbf{X}, \mathbf{y})$ using a subset of eigenvectors from the SWM rather than eigenvectors from a lower triangular distance matrix like the filtering procedure of (Borcard and Legendre, 2002). See Appendix C for a technical explanation of how eigenvectors of the SWM can be used to approximate the spatial terms in a general spatial economics model.

With Eigenvector Spatial Filtering (ESF), the idea of using eigenvectors from the SWM as explanatory variables to control/proxy for any spatially correlated omitted variables i.e. $g(\mathbf{W}, \Delta \mathbf{f}, \mathbf{X}, \mathbf{y})$, was proposed by Griffith (2000, 2003). A spectral decomposition of $\boldsymbol{W}$ is defined as $\mathbf{W} = \mathbf{EDE}'$ where $\boldsymbol{E}$ is an $n \times n$ matrix of eigenvectors and $\boldsymbol{D}$ is a diagonal matrix of eigenvectors. If the full set of eigenvectors are used this approach then yields the high-dimensional ESF reduced form model:

$$\Delta \mathbf{f} = \sum_{i=1}^{3} \beta_i \mathbf{y}^i + \mathbf{X}\zeta + \mathbf{E}\gamma + \mathbf{u} \tag{3}$$

where γ is a vector of unknown constants. We view $\boldsymbol{E}\gamma$ as a first order approximation of $g(\mathbf{W}, \Delta \mathbf{f}, \mathbf{X}, \mathbf{y})$. It is important to note that eq. (3) is a high-dimensional linear equation as there are more parameters $(3 + k + n)$ than observations (n). Thus, estimation by OLS is infeasible. However, Griffith (2000, 2003) argue each of the eigenvectors can be viewed as a distinct spatial pattern and only a specific subset of these patterns (eigenvectors) will be related to the dependent variable Δf and will thus have non-zero coefficients, i.e., the parameter vector γ is sparse. OLS estimation is possible if just the subset of relevant eigenvectors is used. However, as the relevant subset is unknown, a selection procedure is required. To solve this selection problem, we use the Lasso-based procedure proposed in Barde et al. (2023).

Selection via Lasso was first proposed by Seya et al. (2015), with the objective function:

$$[\widehat{\boldsymbol{\beta}}, \widehat{\zeta}, \widehat{\gamma}] \in min\left\{ \left\| \Delta \mathbf{f} - \sum_{i=1}^{3} \beta_i \mathbf{y}^i - \mathbf{X}\zeta - \mathbf{E}\gamma \right\|_2^2 + \theta \|\gamma\|_1 \right\} \tag{4}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]'$ and θ is the Lasso tuning parameter. It is important to note here that the choice of θ determines the number of selected eigenvectors. With this model specification the selection problem can now be considered a tuning parameter calibration problem. Seya et al. (2015) proposed using k-fold cross-validation prediction accuracy to estimate θ. However, as the aim of ESF is to eliminate patterns of spatial correlation, there is no guarantee the target of prediction accuracy will achieve this. Additionally, when the Lasso tuning parameter is derived by cross-validation, the existing results on theoretical error bounds require the assumptions of random/independent observations to hold (Chetverikov et al., 2021). Given the eigenvectors are derived from a matrix that describes the dependence relationship between the observations, these bounds are unlikely to hold within the ESF framework.

Instead, we propose using the more intuitive Moran's i based Lasso of (Mi-Lasso) proposed by Barde et al. (2023), where the Lasso tuning parameter is calibrated from the standardised Moran's i (z) of the OLS residuals of (1).Specifically they replace θ in (4) with $\theta = z^{-2}$. The intuition behind using the squared inverse transformation is based in assumption that when the level of spatial correlation (z) is low/small only a small set of eigenvectors necessary, which requires a large value of θ and vice versa high levels of spatial correlation, the squared inverse transformation achieves this Additionally, squaring z insures θ is positive which is necessary for Lasso to have a unique solution (Hastie et al., 2015). Barde et al. (2023) also shows the conditions necessary for consistent eigenvector selection and derives a finite sample performance bound for the Lasso-based procedure. We estimate the model by post Lasso i.e. OLS estimation with the selected eigenvectors of Mi-Lasso included. We also take account of heteroscedasticity by employing robust standard errors.

4. Data

In this paper, we are working with several cross-sections of country level data between 1990 and 2020. The cross-sections we employ vary in size from 124 countries up to 146. Our forest data is drawn from the FAO (2021) global forest resource assessment. This data was first published at 10-year intervals, and subsequently at 5-year intervals. In our analysis we use data for the years 1990, 2000, 2010, 2015 and 2020.

Table 1 presents the variables, along with descriptive statistics, we employ in our analysis. Our preferred dependent variable following Hyde (2012) is the annual rate of deforestation. In addition, we have employed a measure of the annual change in forest cover as an alternative dependent variable as a means to assess robustness of our model results. The formula used to derive these variables are shown in appendix section B. We also note, that the annual change in forest cover over the four periods is small but negative.

The variable, we use to measure income is real Gross Domestic Product per capita (GDPpc) at chained purchasing power parity in 2017 dollars ($1000) taken from the Penn World Tables (PWT) Version 10.01 (Feenstra et al., 2015) for each of the sampling years..[5] In our model, GDPpc serves as a proxy for economic development. We observe that the distribution of GDPpc has become more skewed over time.

Turning to our control variables, the existing EKCf literature has a somewhat eclectic set of reduced form model specifications that employ a varied set of control variables. We employ several controls in our model specification that are taken from the FAO (2021) which is in keeping with the antecedent literature (Pablo-Romero et al., 2023). First, we have a measure of forest area (Fora). This variable has been used in previous research by Culas (2012). This measure appears to stay relatively constant over the time periods considered in our study at the aggregate level but this does not reveal the changes occurring at the country level. Second, we have a measure of the annual ratio of agricultural land area divided by total land area (ALg). This measure is frequently employed within the EKCf literature (Pablo-Romero et al., 2023). Third, We employ a measure of population density (Popd). Finally, following Bhattarai and Hammig (2001b), Wehkamp et al. (2018) and Murshed (2022), we additively combine measures of civil liberties (CL) and political freedoms (PR) variation as a proxy for institutions and democracy (PRCL). It has been reported in the EKCf

[4] The SWM used is provided in the data appendix.

[5] The PWT data is available here: https://www.rug.nl/ggdc/productivity/pwt/?lang=en

Table 1
Descriptive statistics and definitions.

Variable	Definition	Countries	Mean	St. Dev.	Min	Max
		Period: 1990–2000				
Δf	Annual change in forest cover	146	−0.07	1.06	−4.23	5.55
Df	Annual rate of deforestation	146	−0.01	1.12	−3.51	7.16
y	Real GDP pc ($1000)	146	10.77	15.36	0.71	144.84
Fora	Forest area (1000 ha)	124	33.08	99.27	0.03	808.95
ALg	Annual ratio of agricultural land to total land area	124	0.21	1.45	−6.1	9.19
Popd	Populations density	124	0.86	1.13	0.01	7.93
PRCL	Political freedoms + Civil liberties	124	7.85	4.08	2.00	14.00
		Period: 2000–2010				
Δf	Annual change in forest cover	146	−0.06	0.76	−2.51	2.82
Df	Annual rate of deforestation	146	−0.04	0.76	−2.24	3.21
y	Real GDP pc ($1000)	146	12.51	16.28	0.54	111.00
Fora	Forest area (1000 ha)	142	28.68	91.74	0.03	809.27
ALg	Annual ratio of agricultural land to total land area	142	0.1	1.05	−2.83	3.99
Popd	Populations density	142	0.95	1.23	0.02	9.81
PRCL	Political freedoms + Civil liberties	142	7.26	3.87	2.00	14.00
		Period: 2010–2015				
Δf	Annual change in forest cover	146	−0.11	1.11	−5.90	7.59
Df	Annual rate of deforestation	146	0.08	1.14	−5.25	8.84
y	Real GDP pc ($1000)	146	17.20	17.73	0.71	83.52
Fora	Forest area (1000 ha)	143	28.12	90.53	0.03	815.14
ALg	Annual ratio of agricultural land to total land area	143	0.11	1.4	−10.96	3.2
Popd	Populations density	143	1.08	1.39	0.02	11.34
PRCL	Political freedoms + Civil liberties	143	6.93	3.85	2.00	14.00
		Period: 2015–2020				
Δf	Annual change in forest cover	146	−0.23	0.74	−3.56	1.13
Df	Annual rate of deforestation	146	−0.21	0.72	−3.32	1.16
y	Real GDP pc ($1000)	146	18.17	18.16	0.85	82.38
Fora	Forest area (1000 ha)	143	27.97	90.34	0.03	814.93
ALg	Annual ratio of agricultural land to total land area	143	0.06	0.81	−2.87	4.21
Popd	Populations density	143	1.15	1.50	0.02	12.00
PRCL	Political freedoms + Civil liberties	143	7.03	3.97	2.00	14.00

literature that differences in institutional and governance between countries do have different effects on forest cover. These data are obtained from Freedom House (Freedom-House, 2024) with each index measured on a declining scale of 1–7.[67]

A visual examination of the raw data is presented in Fig. 1, which shows a scatter plot of the average deforestation rate against initial real GDP per capita for the four periods considered. The curvature of the solid line is obtained through the application of a non-parametric Locally Weighted Scatterplot Smoothing (LOWESS) (Cleveland and Devlin, 1988),[8] estimator, offering preliminary evidence of a non-linear relationship between the the deforestation rate and country level GDPpc.

5. Results

We now present our first set of model results in Table 2 where we compare the Mi-Lasso estimated (filtered - including the selected eigenvectors - columns 2, 4, 6, and 8) results with the (unfiltered - columns 1, 3, 5, and 7) OLS results for the four periods considered. Columns 1 and 2 are for 1990–2000, columns 3 and 4 are for 2000–2010, columns 5 and 6 are for 2010–2015 and columns 7 and 8 show the years 2015–2020.

Table 2 shows the filtered and unfiltered results for the four time periods considered without any controls included.[9] The unfiltered OLS results (columns 1, 3, 5, and 7) indicate a complex cubic relationship between the rate of deforestation and income for three out of four time periods examined. Similarly, when we examine the filtered Mi-Lasso estimates, we also find that for three out of four time periods that we observe a cubic relationship. The time period for which the EKCf relationships appears to disappear is 2010–2015.

Importantly, when the selected eigenvectors are included, the magnitude of the estimated coefficients and associated standard errors shrink, and the EKCf relationship changes somewhat. Also, the inclusion of the eigenvectors is further justified by the improvement in the fit of the filtered models (adjusted R^2) and given that the partial F-test on the Mi-Lasso selected eigenvectors is statistically significant at the 1 % level for three periods and at the 10 % level for the fourth. This shows that the underlying spatial process explains a substantial part of a countries deforestation rate, and by correctly accounting for this process, we can get a very good model using just a simple model set-up.

Next, we consider model results once the controls are included.

The results in Table 3 show the unfiltered (OLS - columns 9, 11, 13 and 15) and the filtered (Mi-Lasso - columns 10, 12, 14 and 16) model specifications. The controls we have employed are frequently found to be statistically insignificant except for the ratio of agricultural land area divided by total land area (ALg) for the earlier time period 2000–2010 and in 2015–2020 for several of the controls. The impact of the inclusion of the controls is seen by the adjusted R^2 being lower than in the corresponding columns in Table 2. However, the impact of the inclusion of the eigenvectors (filtering) remains significant. The adjusted R^2 for the filtered models is significantly higher and the partial F-test on the selected eigenvectors is always significant at the one or 10 % level for all four periods considered. We can also see that the income coefficients are similar to the model specifications without the inclusion of the controls (i.e., Table 2) in that the absolute magnitude of the coefficients as well as their standard errors are generally equal or smaller compared to the

[6] The data from Freedom House that we use is available at `https://freedomhouse.org/report/freedom-world`

[7] We did consider trade openness as a control variable (Gräbner et al., 2021). We used the trade share ratio and found that it seriously reduced model performance and was therefore not included in our final model specification.

[8] LOWESS function splits the data into subsets and then fits a low degree polynomial to each of the subsets using weights least squares.

[9] The estimates of the eigenvector coefficients for all model specifications are available upon request.

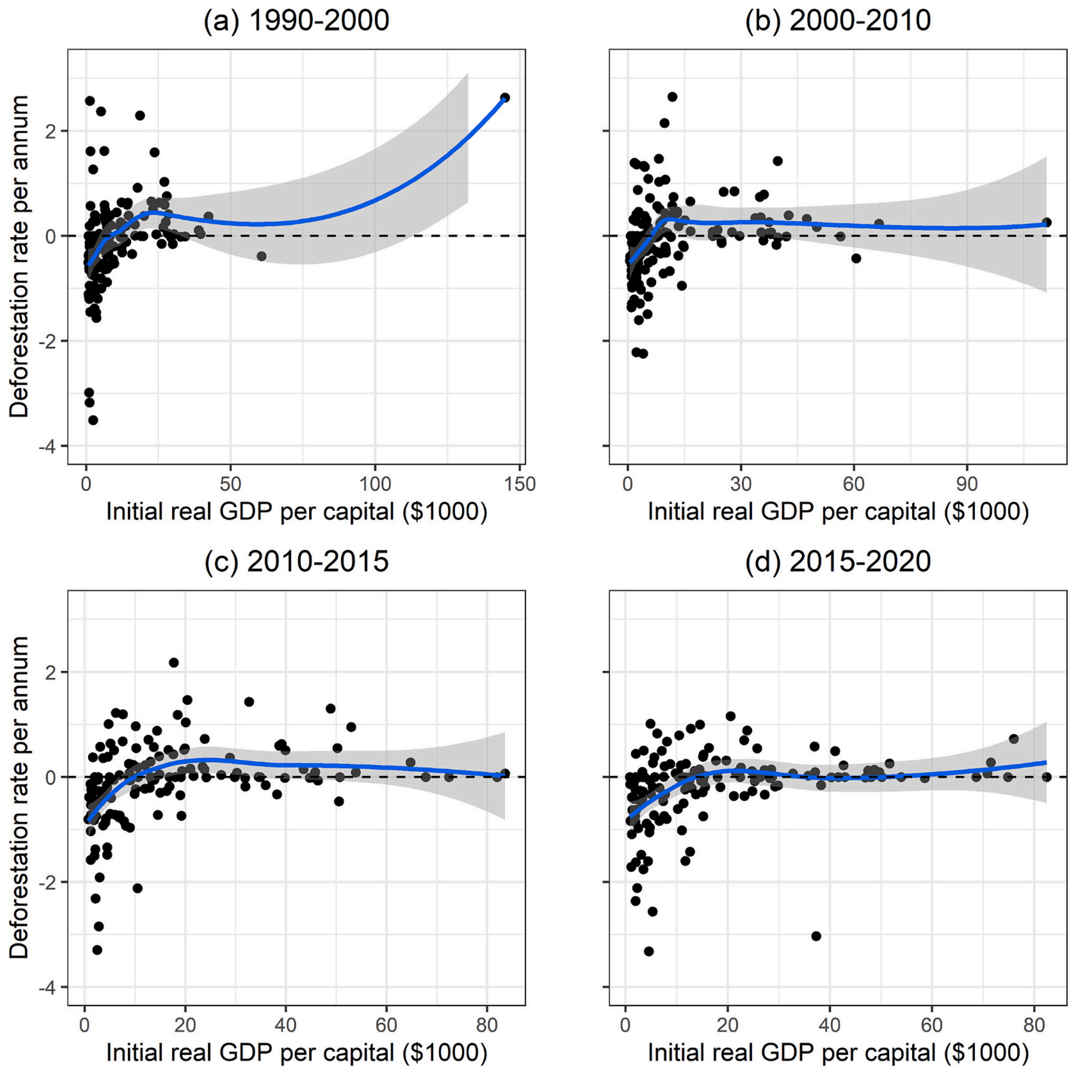

Fig. 1. Annual deforestation rate and GDP per capita.
The solid line is a LOWESS (locally weighted scatterplot smoothing) function. Each plot includes 146 countries.

Table 2
Results excluding controls - dependent variable annual deforestation rate.

	Dependent variable:							
	Df							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
y	0.103^{***}	0.034^{**}	0.078^{***}	0.027^{***}	0.049	0.059^{*}	0.077^{***}	0.051^{***}
	(0.024)	(0.016)	(0.016)	(0.008)	(0.041)	(0.030)	(0.017)	(0.010)
$\boldsymbol{y}^2$	-0.003^{***}	-0.001^{**}	-0.002^{***}	-0.001^{***}	−0.001	−0.001	-0.002^{***}	-0.001^{***}
	(0.001)	(0.0004)	(0.0005)	(0.0002)	(0.001)	(0.001)	(0.001)	(0.0003)
$\boldsymbol{y}^3$	0.00001^{***}	0.00001^{**}	0.00001^{***}	0.00001^{***}	0.00001	0.00001	0.00002^{***}	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00001)	(0.00001)	(0.00000)	(0.00000)
Constant	-0.581^{***}	−0.059	-0.465^{***}	-0.162^{**}	−0.507	-0.581^{**}	-0.787^{***}	-0.667^{***}
	(0.149)	(0.123)	(0.097)	(0.065)	(0.325)	(0.245)	(0.141)	(0.081)
Period	1990–2000	1990–2000	2000–2010	2000–2010	2010–2015	2010–2015	2015–2020	2015–2020
Estimator	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso
Adjusted R^2	0.129	0.748	0.116	0.842	0.029	0.237	0.141	0.791
Partial F-statistic	–	19.05^{***}	–	30.28^{***}	–	3.36^{*}	–	15.74^{***}

*p<0.1; **p<0.05; ***p<0.01. Figures in parenthesis are robust standard errors. OLS is a regression with no eigenvectors included and Mi-Lasso is a regression which includes the selected eigenvectors from Mi-Lasso (Barde et al., 2023). The 'Partial F-statistic' is for an F-test on the included eigenvectors. All regressions include 146 countries.

Table 3
Results including controls - dependent variable annual deforestation rate.

	Dependent variable:							
	Df							
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
y	0.106***	0.091***	0.075***	0.048***	0.061*	0.063**	0.073***	0.073***
	(0.027)	(0.029)	(0.018)	(0.011)	(0.034)	(0.030)	(0.018)	(0.017)
y^2	−0.003***	−0.002***	−0.002***	−0.001***	−0.001	−0.002*	−0.002***	−0.002***
	(0.001)	(0.001)	(0.001)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.001)
y^3	0.00001***	0.00001***	0.00001***	0.00001***	0.00001	0.00001	0.00002***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00001)	(0.00001)	(0.00000)	(0.00000)
For a	−0.001	−0.001	−0.0004	−0.001***	−0.0002	−0.0002	0.0001	0.0001
	(0.001)	(0.001)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
ALg	0.036	0.040	−0.107*	−0.111***	0.001	0.013	−0.230*	−0.192**
	(0.100)	(0.102)	(0.059)	(0.042)	(0.030)	(0.034)	(0.123)	(0.090)
Popd	0.019	−0.012	0.029	−0.003	0.134	0.089	0.048	0.061**
	(0.080)	(0.066)	(0.033)	(0.022)	(0.088)	(0.061)	(0.030)	(0.029)
PRCL	−0.008	0.003	0.035*	0.007	−0.002	−0.026	0.013	0.029*
	(0.040)	(0.034)	(0.021)	(0.010)	(0.023)	(0.024)	(0.014)	(0.015)
Period	1990–2000	1990–2000	2000–2010	2000–2010	2010–2015	2010–2015	2015–2020	2015–2020
Estimator	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso
Adjusted R^2	0.104	0.306	0.151	0.652	0.029	0.239	0.191	0.329
Partial F-statistic	–	2.14	–	28.853***	–	3.38*	–	5.67***
Countries	124	124	142	142	143	143	143	143

*p<0.1; **p<0.05; ***p<0.01. Figures in parenthesis are robust standard errors. OLS is a regression with no eigenvectors included and Mi-Lasso is a regression which includes the selected eigenvectors from Mi-Lasso (Barde et al. (2023)). The 'Partial F-statistic' is for an F-test on the included eigenvectors. The number of countries varies due to missing covariate data, this is especially acute for 1990, as many of the countries were part of the USSR then.

unfiltered OLS estimates for three of the time periods. The exception is 2010–2015, models 13 and 14 for which we now see that the quadratic term on income is marginally statistically significant for the filtered model specification (14).

Turning to Table 4 this shows the number of selected eigenvectors and their significance for the filtered models estimated. This table reports the number of eigenvectors that are statistically significant at the 1, 5 and 10% levels for all models. What we observe is that once we include controls in our model specifications far fewer eigenvectors are selected. This result varies by time period and generally the reduction in eigenvectors selected positively correlates with a reduction in the adjusted R^2. The likely reason for this occurring is that controls not only do not improve model fit but are likely introducing noise that impairs the performance of the Mi-Lasso estimator.

The meaning of our model results in regard to the EKCf, Figs. 2 and 3 show model predictions for non-parametric LOWESS functions of the fitted values from the regressions presented in Tables 2 and 3 against initial levels of GDPpc in the four periods examined. For each of the time periods, we identify the initial turning points that allow us to compare differences that occur between model estimates and across the four time periods.

The first thing to note, is that there are marginal differences between the Mi-Lasso and OLS results in terms of the turning points when we include and exclude the controls. Concentrating on the predicted model results that include the controls (Fig. 3) for the period 1990–2000 (Fig. 3(a)), the turning point for GDPpc is the quite similar for the Mi-Lasso and OLS estimates. However, in stark contrast, for the period 2000–2010 (Fig. 3(b)) the turning point is now significantly higher for the OLS specification. Also, the Mi-Lasso turning point is much lower for 2000–2010 compared to 1990–2000 indicating that the turning point can change location over time. We also note that for these two periods that the shape of the LOWESS estimator results are significantly different for high values of GDPpc. This result likely occurs because of some outliers in the data which can be seen in Fig. 1.

Next, if we examine Fig. 3(c) and (d), the turning points are much higher for the period 2010–2015 than any other periods and again the OLS turning point is somewhat larger than the Mi-Lasso equivalent. Interestingly, when we examine the turning point for 2015–2020, we see a significant decrease in the point estimates which again suggests that the EKCf is not stationary. Also, like the earlier periods the shape of the LOWESS estimator functions varies significantly across the time periods.

Finally, we graphically summarise the predicted average in deforestation and GDPpc in Fig. 4. This figure illustrates the revealed level of variation in the estimated quadratic turning point for the four time periods examined. The variation that is revealed is striking and it suggests that the EKCf relationship varies across time and that the variation is not monotonic.

6. Discussion

Overall, our results confirm that after accounting for spatial correlation, there is a non-linear relationship between the average

Table 4
Number and significance of selected eigenvectors - dependent variable annual deforestation rate.

	(2)	(10)	(4)	(12)	(6)	(14)	(8)	(16)
No of Eigenvectors	32	3	55	22	1	1	51	3
Significant at 1% level	25	0	50	18	0	0	43	2
Significant at 5% level	5	3	3	4	0	0	4	1
Significant at 10% level	1	0	1	0	1	1	2	0
Not significant	1	0	1	0	0	0	2	0
Period	1990–2000	1990–2000	2000–2010	2000–2010	2010–2015	2010–2015	2015–2020	2015–2020
Controls	excluded	included	excluded	included	excluded	included	excluded	included

Column number correspond to the equivalent columns in Table 2 and 3.

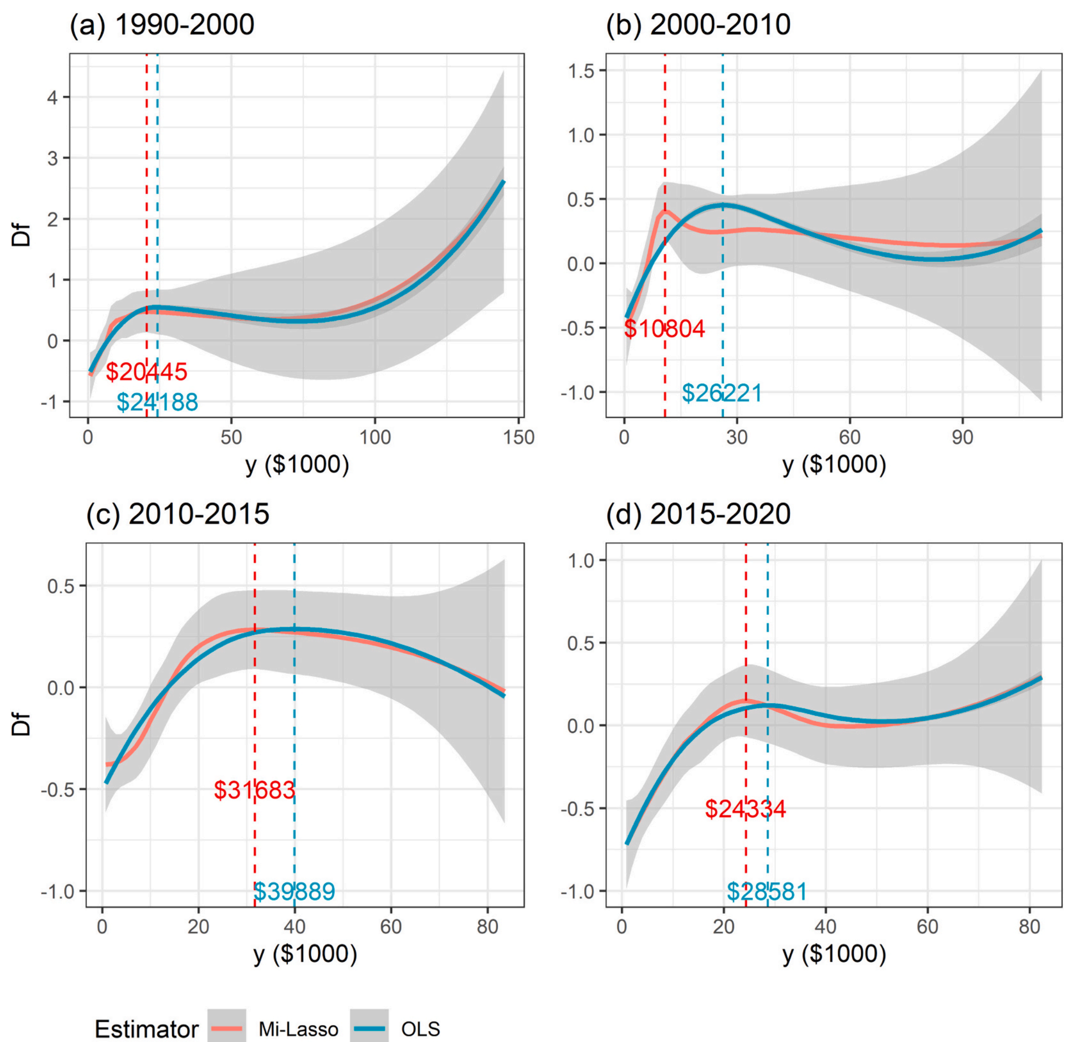

Fig. 2. Predicted deforestation rate and GDP per capita - without controls.
$\widehat{Df}$ are the fitted values from the regression results presented in Table D.1 and the blue and red lines are a LOWESS (locally weighted scatterplot smoothing) function, and the dotted lines show the turning points.

deforestation rate and GDPpc. However, this relationship appears to be changing with time and there is no obvious increase or decrease of the key turning point which varies significantly across the four periods examined. Like Mills and Waite (2009); Mills Busa (2013), we find the inclusion of eigenvectors significant, but unlike Mills and Waite (2009); Mills Busa (2013), we find their inclusion changes the parameter estimates of GDPpc which in turn has an influence on the predicted turning point of the relationship.

For many of the model specifications examined, we find evidence of a non-linear relationship that is changing with over time. We have identified this feature of the EKCf because we have estimated a series of cross-sectional models. This approach to model estimation is in keeping with that employed by Chow and Li (2014) and Bernard et al. (2015) who argue that econometric limitations associated with panel data estimation are frequently overlooked in much of the empirical literature. Importantly, by taking this approach, we observe a more complicated relationship in some periods than others and we also find evidence that the inverse-U shape postulated by the EKCf is potentially too simplistic. Our results also indicate that the turning point for the average change in forest cover appears to be unstable over time.

There are also differences in our estimates between the Mi-Lasso and OLS model specifications as well when we included or exclude control variables. The impact of different estimators on key model results has previously been observed in the literature (e.g., Tameko (2024)). In particular, our results indicate that the identification of a unique and stable turning point for the EKCf is unlikely. This may not be surprising given that the institutional and policy environment that shapes the use of forest resources is changing and as such the incentives facing economic agents are changing. In this context trying to empirically model and capture all of the potential influences on forests is infeasible. Researchers may well be better advised, if they wish to continue estimating EKC model specifications, to indirectly capture the many possible spatial influences using the Mi-Lasso estimator employed here.

In terms of the Mi-Lasso estimator employed in this analysis it is clear that its performance is significantly influenced by the inclusion of additional explanatory variables. Indeed, the introduction of several standard and typical control variables used in the literature had a detrimental impact on model performance. This finding raises an interesting question about how best to model a reduced form EKCf relationship. In principle, as opposed to employing a set of ad hoc controls that are not theoretically justified, researchers may well be better served to extract information from the latent spatial dependence in the data that is implicitly taking account of many of the relationships that the control variables are attempting to proxy for. Indeed, with such an atheoretical model there could be a philosophical justification for employing the Mi-Lasso estimator as opposed to control variables. This

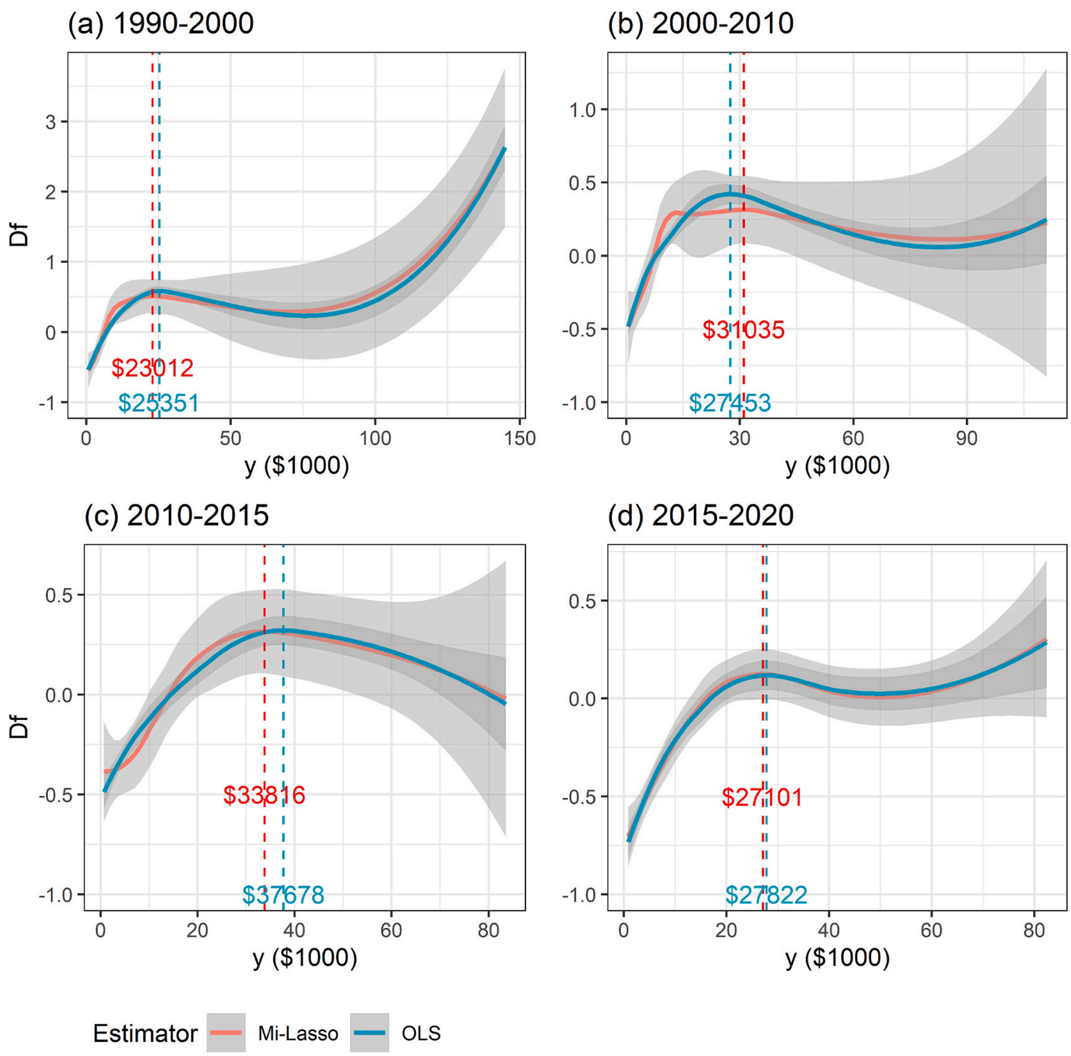

Fig. 3. Predicted deforestation rate and GDP per capita - with controls.

$\widehat{Df}$ are the fitted values from the regression results presented in Table 3 and the blue and red lines are a LOWESS (locally weighted scatterplot smoothing) function, and the dotted lines show the turning points.

specific issue warrants further investigation. At the same, we also note that Stern (2017) is somewhat critical of the empirical literature examining the EKC and that new econometric models specification are required that are derived from economic theory that examines the conditions required for the EKC to exist (Shibayama and Fraser, 2014; Alonso-Carrera et al., 2019).

Another econometric issue that is raised by this analysis relates to the structure and form of the SWM. As noted, in the antecedent literature researchers attempt to specify this part of the model and frequently discuss issues with model fit in terms of poor SWM form. With the Mi-Lasso estimator the SWM is recovered indirectly and from the results presented here it does appear to change over time. Thus, assuming that spatial dependence or at least the important parts of the spatial structure is fixed and not random is another limitation of the standard approach taken to dealing with spatial dependence.

Finally, there are some limitations with the research presented as well as scope for further developments. First, the ESF and the use of a Mi-Lasso estimator to select the relevant subset of eigenvectors is currently limited to cross-sectional data sets. Clearly being able to extend this methodology to panel data would enable a far richer set of model comparisons. This modelling extensions remains an option for future research. Second, the data used in the current study has made significant use of the FAO (2021) global forest resource assessment. There are extensive discussions within the literature regarding limitations of this data. This has resulted in researchers considering alternative data sources. For example, Tameko (2024) considers several alternative data sources that are available and that offer the potential to improve our understanding of the EKCf (e.g., Hansen et al. (2013); CCI (2017); Liu et al. (2020). Examining these data with the econometric methodology we have presented here would be an interesting extension.

7. Conclusions

In this paper, we have employed the Mi-Lasso estimator to take account of spatial dependence for the EKCf. The approach we have employed is new to the EKCf literature and it overcomes issues with how spatial dependence has previously been accounted for using the SWM. In particular, by employing the Mi-Lasso estimator procedure, we take account of the spatial dependence of an unobserved functional form plus avoiding issues of model mis-specification that can arise from the chosen form of the SWM.

To illustrate the utility of this method, we have employed data drawn from the (FAO, 2021) global forest resource assessment between 1990 and 2020 for a sample of 146 countries. This data has been used to re-investigate the relationship between a country's GDP per capita and the rate of deforestation. Overall, our results confirm that after

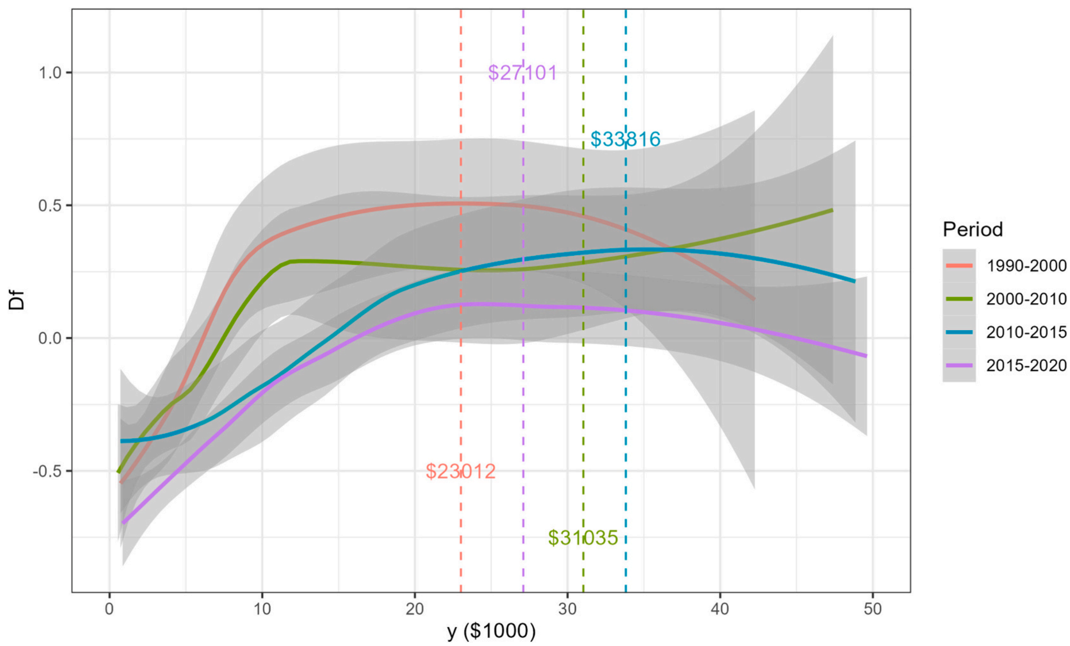

Fig. 4. Predicted deforestation rate and GDP per capita - Mi-Lasso comparison.
$\widehat{Df}$ are the fitted values from the Mi-Lasso regression results presented in Table 3 and the lines are a LOWESS (locally weighted scatterplot smoothing) function, and the dotted lines show the turning points.

accounting for spatial correlation, there is a non-linear relationship between the deforestation rate and GDPpc. This results is keeping with many of the earlier studies that have examined this issue. Importantly, if researchers ignore spatial correlation and report OLS estimates of the GDPpc turning point these have been shown to exhibit a tendency to be larger whilst the Mi-Lasso yields higher levels of model fit. Another important finding of our research, given that we have estimated a series of cross-sectional models, is that EKCf turning point appears to be changing over time. Importantly, there is no obvious trend increase or decrease of the turning point which varies significantly across the four periods examined. We also observe that the turning points differ between the econometric specifications employed across the four time periods. Taken together, our results indicate that identification of the EKCf at the country level must take account of spatial dependence and that the turning point is likely to be time varying.

CRediT authorship contribution statement

Rowan Cherodian: Writing – review & editing, Software, Methodology, Formal analysis, Conceptualization. **Iain Fraser:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Will make data and code available on publication

Appendix A. Countries used in analysis

Albania, Algeria, Angola, Argentina, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia (Plurinational State of), Brazil, Bulgaria, Cameroon, Canada, Chile, Colombia, Congo, Costa Rica, Côte d'Ivoire, Cuba, Cyprus, Democratic Republic of the Congo, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Finland, France, Gabon, Ghana, Greece, Guatemala, Haiti, Honduras, Hungary, India, Indonesia, Iraq, Ireland, Iran (Islamic Republic of), Israel, Italy, Jamaica, Japan, Jordan, Kenya, Republic of Korea, Lebanon, Luxembourg, Malaysia, Mexico, Morocco, Mozambique, Nepal, Netherlands, New Zealand, Nicaragua, Nigeria, Norway, Pakistan, Panama, Paraguay, China, Peru, Philippines, Poland, Portugal, Romania, Senegal, South Africa, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syrian Arab Republic, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, United Kingdom, tanzania, United States of America, Uruguay, Venezuela (Bolivarian Republic of), Viet Nam, Zambia, Zimbabwe, Afghanistan, Armenia, Azerbaijan, Belarus, Belize, Bhutan, Bosnia and Herzegovina, Botswana, Brunei Darussalam, Burkina Faso, Burundi, Cambodia, Central African Republic, Chad, Croatia, Czech Republic, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Gambia, Georgia, Germany, Guinea, Guinea-Bissau, Guyana, Kazakhstan, Kyrgyzstan, Lao People's Democratic Republic, Latvia, Lesotho, Liberia, Libyan Arab Jamahiriya, Lithuania, Macedonia, Madagascar, Malawi, Mali, Mauritania, Moldova/republic of, Mongolia, Montenegro, Namibia, Niger, Papua New Guinea, Russian Federation, Rwanda, Saudi Arabia, Serbia, Sierra Leone, Slovakia, Slovenia, Somalia, Suriname, Tajikistan, Timor-Leste, Turkmenistan, Uganda, Ukraine, United Arab Emirates, Uzbekistan, Yemen, French Guyana, Myanmar, Eswatini.

Appendix B. Data construction formula

This table provides details of the formula used to construct several variables employed in model estimation.

Table B.1: Variable formulas.

Annual change in forest cover (Δf)

$$\Delta f = \left(\left(\frac{\text{forest area}_{y=2000}}{\text{forest area}_{y=1990}}\right)^{1/10} - 1\right) \times 100$$

Annual rate of deforestation (Df)

$$Df = \left(\left(\frac{\text{forest area}_{2000} - \text{forest area}_{1990}}{\text{forest area}_{1990}}\right) \times 100\right) \div 10$$

Annual ratio of agricultural land to total land area (ALg)

$$agland = \left(\frac{\text{agri land}_{year}}{\text{land area}_{year}}\right)$$

$$ALg = \left(\left(\frac{\text{agland}_{2000} - \text{agland}_{1990}}{\text{agland}_{1990}}\right) \times 100\right) \div 10$$

Appendix C. ESF technical example

This appendix provides a technical example of how eigenvectors for a SWM can be used in a linear regression framework to approximate a general spatial economic model. Consider a model with a pth order spatial lags of the dependent variable and first order spatial lags of the exogenous variables;

$$\begin{aligned}\mathbf{y} &= \sum_{i=1}^{p} \mathbf{W}^i \mathbf{y} \rho_i + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\psi + \mathbf{v} \\ &= \mathbf{W}\left(\sum_{i=2}^{p} \mathbf{W}^i \mathbf{y} \rho_i + \mathbf{X}\psi\right) + \mathbf{X}\boldsymbol{\beta} + \mathbf{v}\end{aligned} \tag{5}$$

Where $\boldsymbol{W}$ is a symmetric weights matrix and $\boldsymbol{v}$ is a disturbance term fulfilling the usual assumptions of the linear regression model. A spectral decomposition of $\boldsymbol{W}$ gives

$$\mathbf{W} = \mathbf{E}\mathbf{D}\mathbf{E}' \tag{6}$$

where $\boldsymbol{E}$ is a $n \times n$ matrix of mutually orthogonal eigenvectors and $\boldsymbol{D}$ is a corresponding $n \times n$ diagonal matrix of eigenvalues.

Let $\mathbf{M_E} = \mathbf{I} - \mathbf{E}(\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'$ be an orthogonal projection matrix where $\boldsymbol{I}$ is an $n \times n$ identity matrix.

Now substituting (6) into (5) and pre-multiplying by M_E,

$$\begin{aligned}\mathbf{M_E}\mathbf{y} &= \mathbf{M_E}\mathbf{E}\Delta\mathbf{E}'\left(\sum_{i=2}^{p} \mathbf{W}^i \mathbf{y} \rho_i + \mathbf{X}\psi\right) + \mathbf{M_E}\mathbf{X}\boldsymbol{\beta} + \mathbf{M_E}\mathbf{v} \\ \mathbf{M_E}\mathbf{y} &= \mathbf{M_E}\mathbf{X}\boldsymbol{\beta} + \mathbf{M_E}\mathbf{v}\end{aligned} \tag{7}$$

By the Frisch–Waugh–Lovell (partial regression) theorem (7) is equivalent to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \mathbf{v} \tag{8}$$

where $\boldsymbol{\gamma}$ is a $n \times 1$ vector of unknown constants.

It is important to note here that estimation of (8) by OLS is infeasible as it is a high-dimensional linear model. Griffith (2000, 2003) argued that only a subset eigenvectors will have non-zero coefficients, i.e. $\boldsymbol{\gamma}$ is sparse. Under this sparsity assumption (8) can be reduced to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}^*\boldsymbol{\gamma}^* + \mathbf{v} \tag{9}$$

where $\boldsymbol{E}^*$ is an $n \times s$ matrix with $s < n - k$ and $\boldsymbol{\gamma}^*$ is the corresponding $s \times 1$ parameter vector of non-zero coefficients.

To estimate the relevant s eigenvectors $\boldsymbol{E}^*$ we use the Moran's i based Lasso procedure described in Section 3.3.

Appendix D. Model results using change in forest cover

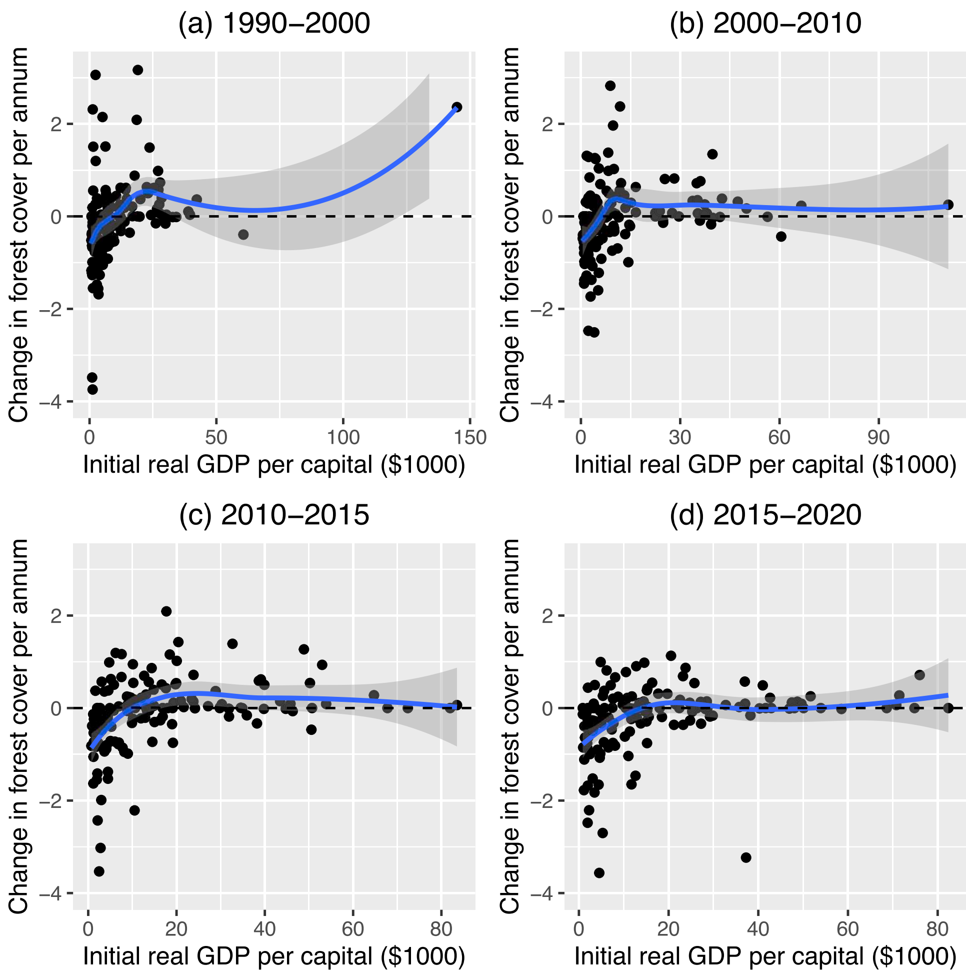

Fig. 5: Average change in forest cover and GDP per capita.
The solid line is a LOWESS (locally weighted scatterplot smoothing) function. Each plot includes 146 countries.

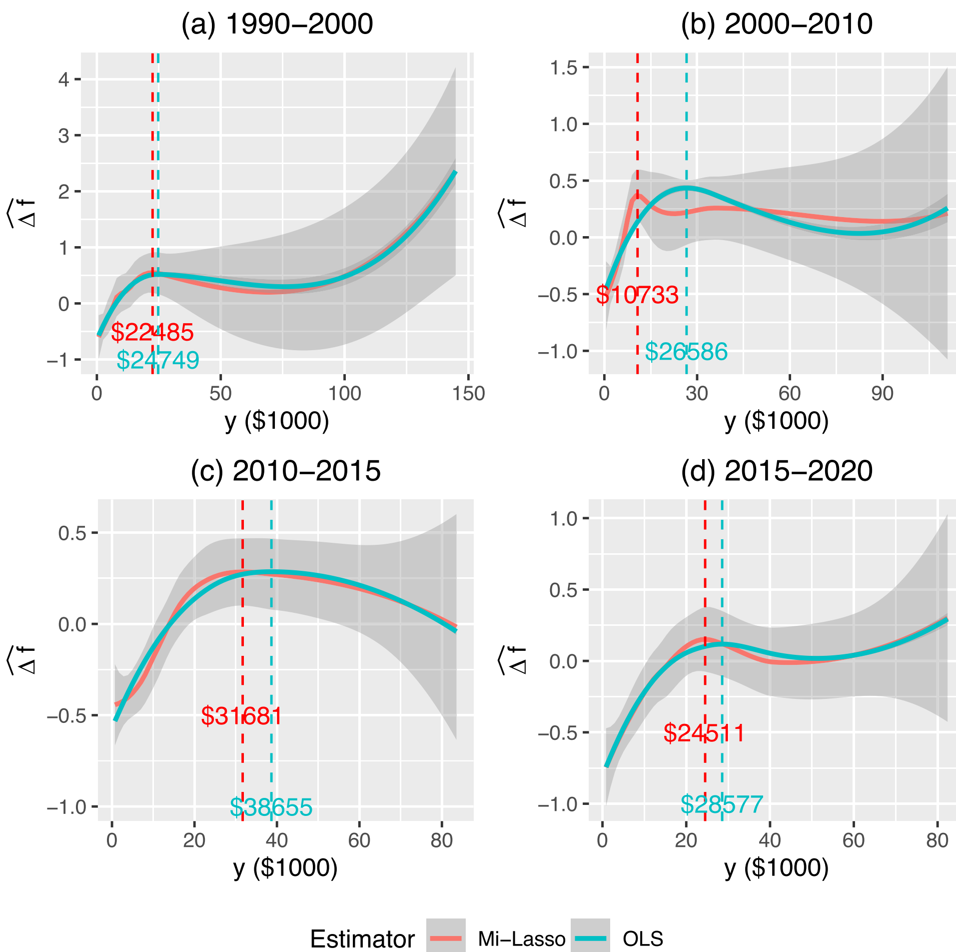

Fig. 6: Predicted average change in forestry and GDP per capita without controls.
$\widehat{\Delta f}$ are the fitted values from the regression results presented in Table D.1 and the blue and red lines are a LOWESS (locally weighted scatterplot smoothing) function, and the dotted lines show the turning points.

Table D.1: Results excluding controls - dependent variable change in forest cover.

	Dependent variable:							
	Δf							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\boldsymbol{y}$	0.104***	0.049***	0.079***	0.026***	0.054	0.064**	0.079***	0.055***
	(0.022)	(0.011)	(0.016)	(0.007)	(0.037)	(0.028)	(0.017)	(0.011)
$\boldsymbol{y}^2$	−0.003***	−0.001***	−0.002***	−0.001***	−0.001	−0.001	−0.002***	−0.001***
	(0.001)	(0.0003)	(0.0005)	(0.0002)	(0.001)	(0.001)	(0.001)	(0.0003)
$\boldsymbol{y}^3$	0.00001***	0.00001***	0.00001***	0.00001***	0.00001	0.00001	0.00002***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00001)	(0.00001)	(0.00001)	(0.00000)
Constant	−0.644***	−0.186**	−0.497***	−0.175***	−0.571*	−0.642***	−0.813***	−0.695***
	(0.155)	(0.085)	(0.100)	(0.065)	(0.294)	(0.228)	(0.146)	(0.084)
Period	1990–2000	1990–2000	2000–2010	2000–2010	2010–2015	2010–2015	2015–2020	2015–2020
Estimator	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso
Adjusted R^2	0.163	0.916	0.141	0.920	0.060	0.260	0.155	0.855
Partial F-statistic	–	65.10***	–	33.54***	–	3.74*	–	12.30***

*p<0.1; **p<0.05; ***p<0.01. Figures in parenthesis are robust standard errors. OLS is a regression with no eigenvectors included and Mi-Lasso is a regression which includes the selected eigenvectors from Mi-Lasso (Barde et al., 2023). The 'Partial F-statistic' is for an F-test on the included eigenvectors. All regressions include 146 countries.

Table D.2: Results including controls - dependent variable change in forest cover.

	Dependent variable:							
	Δf							
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
y	0.110***	0.106***	0.076***	0.041***	0.064**	0.066**	0.074***	0.076***
	(0.026)	(0.026)	(0.018)	(0.010)	(0.032)	(0.029)	(0.019)	(0.018)
y^2	−0.003***	−0.003***	−0.002***	−0.001***	−0.001	−0.002*	−0.002***	−0.002***
	(0.001)	(0.001)	(0.0005)	(0.0003)	(0.001)	(0.001)	(0.001)	(0.001)
y^3	0.00001***	0.00001***	0.00001***	0.00000***	0.00001	0.00001*	0.00002***	0.00001***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00001)	(0.00001)	(0.00000)	(0.00000)
Fora	−0.001	−0.002***	−0.0003	−0.001***	−0.0001	−0.0002	0.0002	0.0001
	(0.001)	(0.001)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
ALg	0.028	0.006	−0.109*	−0.120***	−0.004	0.008	−0.242*	−0.227**
	(0.096)	(0.081)	(0.059)	(0.041)	(0.030)	(0.033)	(0.130)	(0.114)
Popd	0.027	−0.051	0.032	−0.003	0.126	0.083	0.049	0.045
	(0.074)	(0.060)	(0.033)	(0.021)	(0.079)	(0.056)	(0.032)	(0.031)
PRCL	−0.001	0.004	0.035*	0.004	−0.005	−0.028	0.013	0.024
	(0.035)	(0.028)	(0.020)	(0.010)	(0.023)	(0.025)	(0.015)	(0.015)
Period	1990–2000	1990–2000	2000–2010	2000–2010	2010–2015	2010–2015	2015–2020	2015–2020
Estimator	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso	OLS	Mi-Lasso
Adjusted R^2	0.116	0.347	0.156	0.713	0.040	0.241	0.189	0.287
Partial F-statistic	–	2.78**	–	57.02***	–	3.77*	–	4.45**
Countries	124	124	142	142	143	143	143	143

*p<0.1; **p<0.05; ***p<0.01. Figures in parenthesis are robust standard errors. OLS is a regression with no eigenvectors included and Mi-Lasso is a regression which includes the selected eigenvectors from Mi-Lasso (Barde et al. (2023)). The 'Partial F-statistic' is for an F-test on the included eigenvectors. The number of countries varies due to missing covariate data, this is especially acute for 1990, as many of the countries were part of the USSR then.

Table D.3: Number and significance of selected eigenvectors - dependent variable change in forest cover.

	(2)	(10)	(4)	(12)	(6)	(14)	(8)	(16)
No of Eigenvectors	57	6	59	28	1	1	48	2
Significant at 1% level	48	0	52	27	0	0	35	0
Significant at 5% level	7	5	4	1	0	1	10	2
Significant at 10% level	2	1	2	0	1	0	2	0
Not significant	0	0	1	0	0	0	1	0
Period	1990–2000	1990–2000	2000–2010	2000–2010	2010–2015	2010–2015	2015–2020	2015–2020
Controls	excluded	included	excluded	included	excluded	included	excluded	included

Column number correspond to the equivalent columns in Table D.1 and D.2.

References

Ajanaku, B., Collins, A., 2021. Economic growth and deforestation in african countries: is the environmental kuznets curve hypothesis applicable? Forest Policy Econ. 129, 102488.

Almond, R., Grooten, M., Juffe Bignoli, D., Petersen, T., 2022. Living Planet Report 2022 - Building a Nature-Positive Society, Annual report. World Wildlife Fund.

Alonso-Carrera, J., de Miguel, C., Manzano, B., 2019. Economic growth and environmental degradation when preferences are non-homothetic. Environ. Resour. Econ. 74, 1011–1036.

Apergis, N., 2016. Environmental kuznets curves: new evidence on both panel and countrylevel co2 emissions. Energy Econ. 54, 263–271.

Araujo, C., Bonjean, C.A., Combes, J.-L., Motel, P.C., Reis, E.J., 2009. Property rights and deforestation in the brazilian amazon. Ecol. Econ. 68 (8–9), 2461–2468.

Barbier, E.B., Delacote, P., Wolfersberger, J., 2017. The economic analysis of the forest transition: a review. J. For. Econ. 27, 10–17.

Barde, S., Cherodian, R., Tchuente, G.. Moran's I Lasso for Models with Spatially Correlated Data, Working Papers. https://arxiv.org/pdf/2310.02773.pdf.

Benedek, Z., Ferto, I., 2020. Does economic growth influence forestry trends? An environmental kuznets curve approach based on a composite forest recovery index. Ecol. Indic. 112, 106067.

Bernard, J.-T., Gavin, M., Khalaf, L., Voia, M., 2015. Environmental kuznets curve: tipping points, uncertainty and weak identification. Environ. Resour. Econ. 60 (2), 285–315.

Bhattarai, M., Hammig, M., 2001a. Institutions and the environmental kuznets curve for deforestation: a crosscountry analysis for latin america, africa and asia. World Dev. 29 (6), 995–1010.

Bhattarai, M., Hammig, M., 2001b. Institutions and the environmental kuznets curve for deforestation: a crosscountry analysis for latin america, africa and asia. World Dev. 29 (6), 995–1010.

Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model. 153 (1–2), 51–68.

Caravaggio, N., 2020a. Economic growth and the forest development path: a theoretical re-assessment of the environmental kuznets curve for deforestation. Forest Policy Econ. 118, 102259.

Caravaggio, N., 2020b. A global empirical re-assessment of the environmental kuznets curve for deforestation. Forest Policy Econ. 119, 102282.

CCI, E.L.C., 2017. 'Product User Guide Version 2.0', UCL-Geomatics: London, UK.

Chang, H.-Y., Wang, W., Yu, J., 2021. Revisiting the environmental kuznets curve in China: a spatial dynamic panel data approach. Energy Econ. 104, 105600.

Cherodian, R., 2023. Eigenvector Spatial Filtering and the Lasso Estimator: Theory and Applications. PhD thesis. University of Kent.

Chetverikov, D., Liao, Z., Chernozhukov, V., 2021. On cross-validated lasso in high dimensions. Ann. Stat. 49 (3), 1300–1317.

Choumert, J., Motel, P.C., Dakpo, H.K., 2013. Is the environmental kuznets curve for deforestation a threatened theory? A meta-analysis of the literature. Ecol. Econ. 90, 19–28.

Chow, G.C., Li, J., 2014. Environmental kuznets curve: conclusive econometric evidence for co2. Pac. Econ. Rev. 19 (1), 1–7.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83 (403), 596–610.

Cropper, M., Griffiths, C., 1994. The interaction of population growth and environmental quality. Am. Econ. Rev. 84 (2), 250–254.

Csereklyei, Z., Stern, D.I., 2015. Global energy use: decoupling or convergence? Energy Econ. 51, 633–641.

Culas, R.J., 2012. Redd and forest transition: tunneling through the environmental kuznets curve. Ecol. Econ. 79, 44–51.

Damette, O., Delacote, P., 2012. On the economic factors of deforestation: what can we learn from quantile analysis? Econ. Model. 29 (6), 2427–2434.

Ehrhardt-Martinez, K., Crenshaw, E.M., Jenkins, J.C., 2002. Deforestation and the environmental kuznets curve: a cross-national investigation of intervening mechanisms. Soc. Sci. Q. 83 (1), 226–243.

Elhorst, J.P., Gross, M., Tereanu, E., 2021. Cross-sectional dependence and spillovers in space and time: where spatial econometrics and global var models meet. J. Econ. Surv. 35 (1), 192–226.

Erdogan, S., 2024. On the impact of natural resources on environmental sustainability in african countries: a comparative approach based on the ekc and lcc hypotheses. Res. Policy 88, 104492.

FAO, 2021. Global Forest Resources Assessment 2020, Main Report. FAO UN, Rome, Italy.

Farooq, U., Dar, A.B., 2022. Is there a kuznets curve for forest product footprint?– empirical evidence from India. Forest Policy Econ. 144, 102850.
Feenstra, R.C., Inklaar, R., Timmer, M.P., 2015. The next generation of the penn world table. Am. Econ. Rev. 105 (10), 3150–3182.
Freedom-House, 2024. Freedom in the World 2024: The Mounting Damage of Flawed Elections and Armed Conflict. Freedom House.
Friedlingstein, P., Jones, M.W., O'Sullivan, M., et al., 2021. Global carbon budget 2021. Earth Syst. Sci. Data Discuss. 2021, 1–191.
Gräbner, C., Heimberger, P., Kapeller, J., Springholz, F., 2021. Understanding economic openness: a review of existing measures. Rev. World Econ. 157, 87–120.
Griffith, D.A., 2000. A linear regression solution to the spatial autocorrelation problem. J. Geogr. Syst. 2 (2), 141–156.
Griffith, D.A., 2003. Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization. Springer Science & Business Media.
Grossman, G.M., Krueger, A.B., 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914. National Bureau of Economic Research.
Grossman, G.M., Krueger, A.B., 1995. Economic growth and the environment. Q. J. Econ. 110 (2), 353–377.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., et al., 2013. High-resolution global maps of 21st-century forest cover change. science 342 (6160), 850–853.
Hao, Y., Liu, Y., Weng, J.-H., Gao, Y., 2016. Does the environmental kuznets curve for coal consumption in China exist? New evidence from spatial econometric analysis. Energy 114, 1214–1223.
Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman & Hall/CRC.
Hyde, W.F., 2012. The Global Economics of Forestry. Routledge.
Jeetoo, J., Chinyanga, E.R., 2023. A spatial econometric analysis of the environment kuznets curve and pollution haven hypothesis in sub-saharan africa. Environ. Sci. Pollut. Res. 30 (20), 58169–58188.
Koop, G., Tole, L., 1999. Is there an environmental kuznets curve for deforestation? J. Dev. Econ. 58 (1), 231–244.
Leblois, A., Damette, O., Wolfersberger, J., 2017. What has driven deforestation in developing countries since the 2000s? Evidence from new remote-sensing data. World Dev. 92, 82–102.
LeSage, J.P., Pace, R.K., 2014. The biggest myth in spatial econometrics. Econometrics 2 (4), 217–249.
Liu, H., Gong, P., Wang, J., Clinton, N., Bai, Y., Liang, S., 2020. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. Earth Syst. Sci. Data 12 (2), 1217–1243.
Maddison, D., 2006. Environmental kuznets curves: a spatial econometric approach. J. Environ. Econ. Manag. 51 (2), 218–230.
Mather, A.S., 1992. The forest transition. Area 24 (4), 367–379. URL: http://www.jstor.org/stable/20003181.
McPherson, M.A., Nieswiadomy, M.L., 2005. Environmental kuznets curve: threatened species and spatial effects. Ecol. Econ. 55 (3), 395–407.
Mikayilov, J.I., Hasanov, F.J., Galeotti, M., 2018. Decoupling of co2 emissions and gdp: a time-varying cointegration approach. Ecol. Indic. 95, 615–628.
Mills Busa, J.H., 2013. Dynamite in the ekc tunnel? Inconsistencies in resource stock analysis under the environmental kuznets curve hypothesis. Ecol. Econ. 94, 116–126.
Mills, J.H., Waite, T.A., 2009. Economic prosperity, biodiversity conservation, and the environmental kuznets curve. Ecol. Econ. 68 (7), 2087–2095.
Murshed, M., 2020. Revisiting the deforestation-induced ekc hypothesis: the role of democracy in Bangladesh. GeoJournal 1–22.
Murshed, M., 2022. Revisiting the deforestation-induced ekc hypothesis: the role of democracy in Bangladesh. GeoJournal 87 (1), 53–74.
Oliveira, R., Almeida, E., 2011. Deforestation in the brazilian amazonia and spatial heterogeneity: a local environmental kuznets curve approach. In: 57th Annual North American Meetings of the Regional Science Association International.
Pablo-Romero, M.P., Sánchez-Braza, A., Gil-Pérez, J., 2023. Is deforestation needed for growth? Testing the ekc hypothesis for latin america. Forest Policy Econ. 148, 102915.
Pesaran, M.H., 2021. General diagnostic tests for cross-sectional dependence in panels. Empir. Econ. 60 (1), 13–50.
Polomé, P., Trotignon, J., 2016. Amazonian deforestation, environmental kuznets curve and deforestation policy: a cointegration approach. In: Environmental Kuznets Curve and Deforestation Policy: A Cointegration Approach (February 16, 2016).
Purcel, A.-A., 2020. New insights into the environmental kuznets curve hypothesis in developing and transition economies: a literature survey. Environ. Econ. Policy Stud. 22, 585–631.
Seya, H., Murakami, D., Tsutsumi, M., Yamagata, Y., 2015. Application of lasso to the eigenvector selection problem in eigenvector-based spatial filtering. Geogr. Anal. 47 (3), 284–299.
Seymour, F., Busch, J., 2016. Why Forests? Why Now?: The Science, Economics, and Politics of Tropical Forests and Climate Change. Brookings Institution Press.
Shafik, N., Bandyopadhyay, S., 1992. Economic Growth and Environmental Quality: Timeseries and Cross-Country Evidence, vol. 904. World Bank Publications.
Shahbaz, M., Sinha, A., 2019. Environmental kuznets curve for co2 emissions: a literature survey. J. Econ. Stud. 46 (1), 106–168.
Shibayama, K., Fraser, I., 2014. Nonhomothetic growth models for the environmental kuznets curve. Int. Econ. Rev. 55 (3), 919–942.
Song, X.-P., Hansen, M.C., Stehman, S.V., Potapov, P.V., Tyukavina, A., Vermote, E.F., Townshend, J.R., 2018. Global land change from 1982 to 2016. Nature 560 (7720), 639–643.
Stern, D.I., 2017. The environmental kuznets curve after 25 years. J. Bioecon. 19, 7–28.
Tameko, A.M., 2024. The effects of trade openness on deforestation in the Congo basin countries. Forest Policy Econ. 162, 103189.
Taye, F.A., Folkersen, M.V., Fleming, C.M., Buckwell, A., Mackey, B., Diwakar, K., Le, D., Hasan, S., Saint Ange, C., 2021. The economic values of global forest ecosystem services: a meta-analysis. Ecol. Econ. 189, 107145.
Wang, J., Xin, L., Wang, Y., 2019. Economic growth, government policies, and forest transition in China. Reg. Environ. Chang. 19 (4), 1023–1033.
Wang, Z., Ye, X., 2017. Re-examining environmental kuznets curve for china's city-level carbon dioxide (co2) emissions. Spatial Stat. 21, 377–389.
Wehkamp, J., Koch, N., Lübbers, S., Fuss, S., 2018. Governance and deforestation—a meta-analysis in economics. Ecol. Econ. 144, 214–227.
Williams, M., 2003. Deforesting the Earth: From Prehistory to Global Crisis. University of Chicago Press.
Williams, M., 2008. A new look at global forest histories of land clearing. Annu. Rev. Environ. Resour. 33 (1), 345–367.
Wolfersberger, J., Delacote, P., Garcia, S., 2015. An empirical analysis of forest transition and land-use change in developing countries. Ecol. Econ. 119, 241–251.