Miyama, Eriko & Managi, Shunsuke
(2014)
Global environmental emissions estimate: Application of multiple imputation.
Environmental Economics and Policy Studies, *16*(2), pp. 115-135.

https://doi.org/10.1007/s10018-014-0080-3

Global Environmental Emissions Estimate:
Application of Multiple Imputation

Eriko Miyama[1*] and Shunsuke Managi[1,2]

[1] Graduate School of Environmental Studies, Tohoku University
6-6-20 Aramaki-Aza Aoba, Aoba-Ku, Sendai 980-8579, Japan
*Corresponding author: miyama@mail.kankyo.tohoku.ac.jp
[2] Institute for Global Environmental Strategies

Abstract

A new database called the World Resource Table (WRT) is constructed in this study. Missing values are known to produce complications when constructing global databases. This study provides a solution for applying multiple imputation techniques and estimates the global environmental Kuznets curve (EKC) for CO_2, SO_2, PM10, and BOD. Policy implications for each type of emission are derived based on the results of the EKC. Finally, we predicted the future emissions trend and regional share of CO_2 emissions. We found that East Asia and South Asia will be increasing their emissions share while other major CO_2 emitters will still produce large shares of the total global emissions.

Keywords: Global emissions; Multiple Imputation; Environmental Kuznets Curve; Missing data; Forecasting.

Acknowledgement;

We are grateful to Hajime Katayama for his detailed and helpful comments and constructive criticism. The following financial support is acknowledged: S-11 from the Environment Research and Technology Development Fund and Environmental Economy Program, Ministry of the Environment of Japan and MEXT KAKENHI.

1. Introduction

History tells us that environmental quality may deteriorate as a result of economic development because an increase in economically valuable activities with massive industrial production and transportation often consume natural resources and emit pollutants and greenhouse gases (GHGs). However, we can mitigate the environmental deterioration by adopting environmentally friendly innovations, which may appear after countries achieve a certain level of economic development. The environmental Kuznets curve (EKC) is a variation of the Kuznets Curve (Kuznets 1955) that represents such relationship between economic development and environmental degradation. According to the EKC hypothesis, relationship between economic development and environmental quality can be described as an inverted U-shaped curve, but this is merely a hypothesis that has yet to be empirically tested. For sustainable development, environmental quality must be maintained or improved with the economic growth, which was argued in *Our Common Future* by the World Commission on Environment and Development (1987). By investigating the EKC hypothesis, we can predict whether the economic growth will be sustainable in terms of natural resource management.

However the existence of global EKC has not yet been confirmed even though over a thousand empirical studies have been undertaken. The literature has shown that model specification, econometric methods, and features of the dataset substantially affect the results of the analyses of EKC estimations (Millimet et al. 2003; Stern 2004; Bertinelli and Strobl 2005). In particular, deleting low-income countries from analyses because of data availability issues is common in the field and might bias estimation results. One of the reasons to cause such a problem is data availability. Especially information about natural resource and environmental quality tend to be missing for developing countries, so the necessary information for panel data regression is frequently missing in global datasets. The problem of missing data is widely recognized in the field of applied economics analysis, including EKC estimations, because missing data may cause estimation bias and also affect the projection of future emissions/concentrations. In addition, when the range of analyses on EKCs are limited to certain regions and countries, it may cause misunderstanding of whole picture of relationship between economic development and environmental quality.

In order to cope with the missing data problem on panel data analysis, three types of methods are used: deleting samples or variables with missing values, single imputation, and multiple imputation. The most common method of sample deletion is listwise deletion that samples with missing values are deleted before statistical analysis. Sample deletion used to be the major way to deal with missing data until the computer calculation technology has been developed after 1990s. Deleting sample with missing values may cause estimation bias if the missing pattern is not at random. Single imputation is imputing a value for the missing data unit. Linear/ spline regression and hot deck/ cold deck imputation are major single imputation methods used for statistical analysis. Hot deck imputation is also called “matching” that if a sample has a missing value, other complete samples, which have similar values for other variables are chosen and the value of that samples are substituted for the missing value (Andridge and Little 2010). Single imputation is the most common way of addressing the problem of missing values (e.g., as applied in the World Bank databases), but tends to underestimate the error variance of missing data (see Junninen et al. 2004: 2906-2907).

On the other hand, multiple values are estimated for the missing data imputation by multiple imputation. Multiple imputation mitigates the problem of estimation bias and underestimation of standard errors. However, multiple imputation has not been widely used until recently because it requires large computational power to undertake the calculations required to generate a sufficient number of imputed datasets. The rapid development of computer technology since the turn of the century has made multiple imputation one of the more common methods currently used to address missing data issues, particularly in medical science studies in which some data of respondents are often missing. Multiple imputation enables future levels of emissions for each country—even low-income countries—to be estimated, which allows political goals to be set for such countries.

Therefore, in this article, we apply multiple imputation to EKC estimation to global panel data analysis and emission projection to draw a path to the environmental friendly and low carbon society. By imputing missing values in the dataset, we can include more countries for longer time periods. We first estimate the EKC. Second, we compare the results of the estimation by multiple imputation with listwise deletion. Third, using the results of the EKC estimation, we show the future projections of global emissions of CO_2 by focusing on each region's share of the total emissions of emerging economies.

We implement the analysis described above on four major environmental indices, which have various rate of missingness: CO_2/ SO_2 emissions per capita, PM10[1] concentration per cubic meter, and BOD (biological oxygen demand) per day per worker[2]. CO_2 and SO_2 emissions are the most common environmental indices for EKC analyses and have low rates of missing data. PM10 concentration is one of the major indices to evaluate air quality, and also has a long and wide range of data collection and the lowest missing rate in our dataset. The BOD indicates how much the water microorganisms consume oxygen to resolve organic matters in the water. The higher BOD is detected, the more polluted the water resource is, thus is important to measure water pollution; however, not every country collect BOD data and this index thus has the highest rate of missing data among the four indices.

In the next section, we first describe the abstract of our dataset. Second, we explain the missing mechanisms of the dataset, which is the main reason why we must use multiple imputation. Third, the data imputation and EKC estimation methods are described. Forth, after implementing the multiple imputation, we use the imputed global dataset to estimate the global EKCs for the four environmental indices, which have different rates of data missing. Using the estimated parameters of the EKC, we project the CO_2 emissions until 2018. Finally, we summarize our findings and discuss the implications of this study.

2. Data

We include industry- and environment-related variables in the imputation model, as shown in Table 1[3]. The data sources are the World Development Indicators (WDI) 2012, Pen World Table (PWT) Ver. 7.1 (Nov. 2012), and Environmental Performance Index (EPI) 2012. The data periods used in this study are 1970-2010 for CO_2 and SO_2, 1990-2010 for PM10, and 1990-2007 for BOD. For the CO_2 emission projection, we used the total population data (medium fertility) from the World Population Prospects (the 2012 revision) by UNDESA and GDP growth rate from the World Economic Outlook by the IMF. We construct a new database titled World Resource Table (WRT) using the imputed method described below. Missing rates of the dataset for CO_2, SO_2, PM10 and BOD are 36.3%, 40.6%, 7.2% and 47.2%, respectively.

Countries in the dataset are categorized by geographical region and income level, which are controlled by dummy variables in the regression. In total, 181 countries are included for CO_2, SO_2, and PM10, and 97 countries are included for BOD, based on data availability[4]. The 181 countries are categorized into 12 regional subcategories (Table 2): (1) South Asia, (2) Central Asia, (3) Middle East and North Africa, (4) Sub-Saharan

Africa, (5) Latin America and the Caribbean, (6) North America, (7) Western Europe, (8) Central Asia, (9) Western Asia and Eastern Europe (10) Pacific Oceania, (11) Southeast Asia, and (12) East Asia. Thirty countries are categorized as high-income OECD countries, 19 countries are high-income non-OECD countries, 49 countries are middle-income countries, 49 countries are lower middle-income countries, and 34 countries are low-income countries. The missing data rate is higher in Pacific Oceania and Sub-Saharan Africa for SO_2, CO_2, and BOD. Data for PM10 are missing more often in the regions with developed economies compared with regions that have more emerging economies.

Data imputation enables us to include more countries; therefore, we can categorize countries into smaller geographical units of 12 that capture the detailed regional feature of the EKCs. Our approach is more inclusive and detailed compared to the EKC literatures, in which geographical information is treated either as larger categories, such as Asia, Europe, North America, and Africa, or as individual countries (Lee et al., 2010; Grossman and Krueger 1991; Orubu and Omotor 2011). Especially PM10 has been studied only for a limited number of countries and cities, i.e., US, Mexico, and Italy (Dasgupta et al. 2002; Mazzanti et al. 2007).

3. Methods

Missing mechanisms

We create a large database for the WRT by imputing the missing data. Before choosing an imputation method, we must first identify the “missing mechanism” of the dataset, which is defined by Rubin (1976). The missing mechanism tells us whether we can derive unbiased estimator from the dataset, it thus helps us determine which imputation method to adopt (Cranmerand and Jeff Gill 2013). The missing mechanism is classified into three separate patterns: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) or nonignorable (NI).

Let D denote a matrix of $n \times k$, where n is the sample size and k is the number of variables. D embraces all dependent and independent variables, including some missing values,to be used in subsequent analyses. We denote D_{obs} as the observed values and D_{miss} as the missing values in D; thus, $D = [D_{obs}, D_{miss}]$.

MCAR

The missing mechanism is MCAR if the probability of missingness is identical for all the data units, i.e., $p(M|D) = p(M)$, which indicates that the missing pattern M is independent of D. For example, when the missing pattern is determined by coin flips, the missing mechanism of the dataset is MCAR. If the missing mechanism is MCAR, the listwise/pairwise deletion and single/multiple imputation methods do not cause the estimators to be biased. Sample deletion in this context only affects efficiency.

MAR

The second missing mechanism is called MAR, such that the missing pattern M depends on D_{obs} but is independent of D_{miss}; more formally, $p(M|D) = p(M|D_{obs})$. In this case, missingness is determined by the observed values of variables without the missing values. For example, if the probability of the missingness of a variable is higher for older people, and the “Age” variable, which has no missing data, is included in the dataset, then the missing mechanism of the dataset is MAR. In the case of MAR, listwise/pairwise deletion causes estimation bias, whereas single/multiple imputation methods do not.

MNAR/NI

When the missing pattern is determined by the missing values D_{miss}, the missing mechanism is MNAR or NI. When the missing mechanism is MNAR, the probability of missingness is dependent on the missing data; thus, the missing pattern cannot be predicted because we do not have information about the missing values. The methods discussed above—listwise/pairwise deletion and single/multiple imputation methods—may cause the estimators to be biased when the missing mechanism is MNAR. However, we can convert the MNAR dataset to MAR by adding auxiliary variables (AV), by which we can predict the missing pattern M. The missing mechanism of a dataset cannot be determined with certainty because we do not have information about missing values. To avoid causing bias with the MNAR datasets, we must use priors and expert information to make the dataset MAR when it is highly likely that the given dataset is MNAR.

We verify the missing mechanism of the dataset by using correlations of the indicator matrix and values of the dataset (see Kabacoff 2011: 360- 362). One way to verify a missing mechanism is to examine the correlation of observed values and matrix of a missing pattern[5]. We generate a dataset with indicator variables that are coded 1 for missing and 0 for observed. The resulting 0, 1 matrix is called the “shadow matrix”. All of the correlations between the shadow matrix and observed values are lower than 0.4, which indicates that the missing mechanism of the dataset can be assumed to be either MCAR or MAR[6]. If the missing mechanism of the dataset is MAR, multiple imputation is an appropriate way to deal with the missing unit in the dataset.

Multiple imputation with EMB

The idea of multiple imputation has been introduced by Rubin (1977). While only one value is imputed for a missing value in case of the single imputation, different values are imputed for the missing elements in the multiple imputation process, such that *m* numbers of imputed datasets are generated based on observed values. As with Single imputation, there are various ways to generate the multiple imputed datasets. Among those, two algorithms are widely used for multiple imputation, the imputation-posterior (IP) approach and expectation maximization with importance sampling (EMis) (King et al. 2001). IP is a method based on a Markov chain and Monte Carlo algorithm that requires both expertise and a lengthy computational time. EMis is based on the expectation maximization (EM) algorithm, the iterative estimation method, which requires less expertise and is faster than IP. These two methods have been used as major algorithms for multiple imputation. However, both IP and EMis have the disadvantage that they require considerable computational time and expertise. In addition, a large panel dataset that includes cross-sectional and time series information, may not be properly treated by these methods.

A newly introduced method of multiple imputation that is implemented by *Amelia*, a statistical package for R (Honaker and King 2010) addresses the missing data of a panel dataset. Amelia can handle both cross-sectional and time-series features of panel data. The imputation method used in Amelia is based on the EM algorithm with bootstrapping (EMB), which can efficiently estimate missing values. The process of multiple imputation using the EMB is shown in Figure 1. The EMB is suitable for large datasets because the drawing process of the mean vector and joint covariance matrix is simplified by bootstrapping. Bootstrapping has better lower order assumptions than the parametric approaches implemented by EMis and IP (see Honaker and King 2010: 564-565).

Regarding the number of imputed datasets, many articles in the literature have indicated that 5 to 20 imputations are sufficient for consistent analyses (see King et al. 2001, p.53; Gelman and Hill 2006, p.542). However, most recent studies have found that estimation results with small numbers of imputed datasets can be biased. Generating more than 100 imputations is recommended because of the reduced computational time of today's technology (Graham et al. 2007). Following the recent study, we execute 100 imputations for our analysis.

EKC estimation

After the imputation, we estimate the regression model for the EKC. We calculate the OLS estimators for the parameters in regression model described by equation (1). E_{it} is the emission/concentration level in country *i* at time *t*, X is the GDP per capita, *Year* is a time trend variable, α is the intercept, μ_i is a fixed effect of the region (we have 12 regions here), and ε_{it} is an error term. To consider nonlinearity, we include the second and third power terms of GDP per capita.

$$\mathrm{E}_{it} = \alpha + \beta_1 X_{it} + \beta_2 {X_{it}}^2 + \beta_3 {X_{it}}^3 + \mu_i + Year + \varepsilon_{it} \qquad (1)$$

We recognize that arguments have been made that sing a parametric regression may not produce reliable results because the functional form and distribution are assumed in advance; therefore, semi-parametric methods have become commonly used to estimate the relationship between environmental degradation and income (Bertinelli and Strobl 2005; Azomahou et al. 2006; Tsurumi and Managi 2010). However, the main purpose of this study is not to investigate the existence of the EKC but to compare the regression results from two methods: listwise deletion and multiple imputation. In addition we use the estimated parameters for future emission projection. We therefore implement parametric estimation and focus on a simple form of the EKC.

In our study, we adopt the integration method developed by Rubin and Schenker (1986) to combine multiple imputed datasets into one result because the method can consider variance among m estimations. To combine the multiple imputed datasets using this method, we use estimated coefficients and standard errors. Therefore we select parametric regressions for the EKC estimation. By Rubin and Schenker’s (1986) method, the variance of standard errors among the imputed datasets is used to calculate the integrated standard error, which means that too large a missing rate for small datasets may make the integrated estimated coefficients insignificant.

4. Results

4.1. Results of multiple imputations

Figures 2-5 are scatter plots of imputed mean values and observed values from the WRT.

The open circles are imputed mean values, and the filled circles are observed values. For all the four environmental indicators, the missing data are imputed for various levels of GDP per capita nearly uniformly. The data for CO_2 exhibits a clear increasing relationship with the GDP per capita. The data for PM10 and BOD are more scattered compared with other indices. Overall, the multiple imputation succeeds in reproducing a reasonable distribution of the imputed dataset based on the observed data.

4.2. EKC estimations

Tables 3-6 list the estimated coefficients and standard errors of the panel data regressions. All four environmental indices exhibit a significant relationship with the GDP per capita both before and after the imputation. The signs for the terms of the GDP per capita do not change after imputation, although their magnitudes are slightly different. The significance of the estimated parameters does not differ between the listwise deletion and multiple imputation. This may be because the sample size is large enough for the coefficients to be statistically significant with small standard error even after the multiple imputation.

Figure 6-9 show the estimated EKC for each environmental index. The year trend and regional fixed effects for the 12 regions are averaged to describe the mean relationship between the environmental indices and GDP per capita. The broken line is the fitted value derived from the listwise deletion, and the solid line is the result of multiple imputation.

CO_2

The curve for CO_2 emissions per capita exhibits a monotonic increase (see Figure 6). The results of the listwise deletion and multiple imputation show almost the same trend except the fitted value for the multiple imputation dataset is slightly larger than that for the listwise deletion. At higher than 80,000 dollars GDP per capita, where number of observations is more limited than for the lower GDP countries, the slope is steeper for the listwise deletion. With the imputed dataset, a linear relationship is found between the CO_2 emission per capita and GDP per capita.

This result is different from early studies of EKC, such as those by Holtz-Eakin and Selden (1995), who found an inverted U-shaped relationship between the global CO_2 emissions per capita and GDP per capita. Their estimated emissions peaked at 35,428 US dollars of GDP per capita. However, our imputed dataset does not support this EKC relationship. Many other empirical studies have found a monotonously increasing relationship between CO_2 emissions per capita and GDP per capita within the observed income levels (e.g., Shafik 1994; Heil and Selden 2001). Our results thus confirm the empirical results of those studies and imply that merely GDP per capita growth alone does not contribute to the low-carbon society.

SO_2

SO_2 emissions per capita shows a U (N)-shaped trend (Figure 7). Countries reduce SO_2 emissions more efficiently than CO_2 emissions until a GDP per capita of 90,000 international dollars. The emissions peak for the imputed results at lower GDP levels is at 31,800 international dollars per capita, which is a higher level of GDP per capita than in the results of the listwise deletion. The results with the imputed dataset show more realistic figures compared with those of the listwise results because we could minimize the bias of the estimators with the imputation.

Technologies, which are the source of SO_2 emissions, normally decrease as the GDP per capita increases. Therefore, where most countries remain in the developing stage, the SO_2 emissions per capita are increasing at lower economic levels and begins to decrease at higher GDP levels, which is consistent with recent studies (e.g., Iwami 2004; Yaguchi et al. 2007; Coleman 2009). SO_2 is considered a local pollutants that is easier to control than global emissions. Sulfur, as a byproduct of industrial production in factories, can be reduced by using an end-of-pipe filter, which is a widely applicable technology even in developing countries. These two factors—being a local pollutant and the existence of applicable clean technology—may help SO_2 emissions be lower at the middle and high-middle income countries.

PM10

The relationship between PM10 concentrations and the GDP per capita is an inverted N-shaped curve (Figure 8). The PM 10 concentration peak is at 84,100 international dollars for the imputed dataset with a higher concentration level compared to the result of the listwise deletion. Many articles in the literature that have found an inverted U-shaped relationship between the PM10 concentration and GDP per capita have used city-level data for air pollution (e.g., Grossman and Krueger 1995), which may have a strong relationship with GDP per capita because the population and industry concentration in urban areas is strongly related to the GDP growth. However, our result shows that the PM10 concentration increases at 25,000 dollars GDP per capita and starts to decrease again at 85,000 dollars GDP per capita that such level of income is still too high for many countries because the mean GDP per capita is 4,712 international dollars for our global dataset. Therefore, results from previous studies showing an inverted U-shaped relationship are too optimistic for current global situation, and we must consider how to decrease the PM10 emissions from emerging nations whose GDP per capita is below 85,000 dollars.

BOD

The level of BOD shows the overall decreasing trend relative to increases of the GDP per capita (Figure 9). The BOD has the highest missing rate among the four environmental indices in our analysis; therefore, the results differ the most between the listwise deletion and multiple imputation, and the difference is larger at the higher GDP per capita. Up to approximately 80,000 dollars GDP per capita, the BOD emission does not decrease much. The emissions increase from 20,000 to 70,000 dollars GDP per capita, which means that for most of the countries, controlling the BOD may still be a problem for economic development. Managi et al. (2009) found that the BOD of a country decreases as the trade openness of the country increases. Therefore, encouraging trade to increase income may also improve the water quality of the country.

4.3. Prediction of emission level

Figures 10 and 11 show the predicted annual trend and regional shares of CO_2 emissions from 1992 to 2018. The predicted emission levels are calculated by the estimated parameters. We multiply the 2010 GDP from the PWT (in 2005 international dollars) by the growth rate of each country's GDP (in constant national currency) from WEO2013 to derive the projection of the GDP per capita from 2011 to 2018.

Our estimation results show that the annual CO_2 emission levels in 2018 in Asia will be almost twice the emission levels in 1992 (see Figure 10). North America, East Asia and Europe are the three largest CO_2 emitters throughout the period. The share of Europe and North America will decline gradually until 2018whereas South Asia and East Asia will increase their emissions share.

Increases in the emission share for East Asia and South Asia can be explained by population pressures in the regions. Because of increasing population pressures, the decreasing trend in per capita emissions is insufficient to cause a reduction in total CO_2 emissions. China and India—the two world most populated countries, thus the two largest emitters of CO_2 in Asia— will further increase their population and emission levels per capita. Emissions from Central Asian countries are also growing. With massive population growth, these countries will continue to become a global warming threat.

5. Conclusion

We constructed a new database, the World Resource Table (WRT), in this study. We provide an application of multiple imputation so that coverage increases compared with the existing databases described in previous studies. This study then estimates the global EKC using WRT. First, by choosing the appropriate imputation method and model, we imputed reasonable values into datasets with missing values. Second, we estimated global EKC for four environmental indices—CO_2, SO_2, PM10, and BOD. Finally, we predicted the future emission trend and regional share of CO_2 emissions.

We found that with a large sample such as the global dataset, which we used for our analysis, increase in the standard errors by multiple imputation do not affect the statistical significance of the estimated parameters. Therefore the regression parameters for the GDP remain significant at a 1% significance level after the imputation, such that we can produce an obvious relationship of environmental indices with GDP per capita.

We also calculated the projection of future global CO_2 emission levels and each region's share. As a result of rapid population growth, East Asia and South Asia, where China and India are located, will increase their emission shares by 2018, whereas other major CO_2 emitters such as Europe and North America will still produce a large share of the total global emissions. The total CO_2 emissions will be twice as large in 2018 compared to 1992. Further studies on environmentally friendly technology and implementation are required to mitigate this trend.

The results of this study indicate that multiple imputation can serve to expand the datasets of environmental indices with missing values for panel data analysis. These results will contribute to the prediction of future trends of global environmental quality and help set the goals for constructing a low carbon society. Since East Asia and South Asia are predicted to increase their emission share rapidly in the world, further studies focusing on Asian countries are urgently needed.

---

[1] PM10 stands for Particulate Matter up to 10 micrometers in size.
[2] BOD is an index to evaluate water quality.

[3] Dataset used for this paper is basically the one from the data used in Miyama and Managi (2014). In Miyama and Managi (2014), we implemented multiple imputation and panel data analysis to Asian countries. We expand the panel data analysis to the global dataset in this article.
[4] If the country does not exist during the covered period because it is occupied by other countries, the country is excluded from the analysis until its year of independence. Thus, the dataset is unbalanced.
[5] This method is a simple example to verify the missing mechanisms. More formal tests to determine the MCAR are introduced by Little (1988).
[6] See Miyama and Managi (2014) for detailed result of the missing mechanism test.

Table 1 Variables used for multiple imputation

Variables	Unit	Source	Period
GDP per capita	1,000 constant 2005 international dollar (I$) per person	PWT	1970-2010
Investment Share of GDP Per Capita	% of PPP Converted GDP Per Capita at 2005 constant prices	PWT	1970-2010
Openness of Economy	% at 2005 constant prices	PWT	1970-2010
SO_2 emission per capita	SO_2 emissions kg per person	EPI	1970-2010
CO_2 emission per capita	CO_2 emissions kg per person	EPI	1970-2010
School enrollment, primary	%	WDI	1970-2010
Total Population	person	WDI	1970-2010
Manufacturing, value added	% of GDP	WDI	1970-2010
Organic water pollutant emissions (BOD)*	kg per day per worker (country level)	WDI	1990-2007
PM10 concentration	micrograms per cubic meter	WDI	1990-2010
CO_2 per GDP	CO_2 emissiona kg per GDP	EPI	1970-2010
Renewable electricity	% of electricity production	EPI	1970-2010

* Organic water pollutants are measured by biochemical oxygen demand (BOD),
which refers to the amount of oxygen that bacteria in water will consume in breaking down waste.

Note: PWT indicates Penn World Table Version 7.1 (2012 November), WDI indicates World Development Indicators 2012, and EPI indicates the Environmental Performance Index 2012.

Table 2 List of countries categorized by region

Region	Country
South Asia	Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, Sri Lanka
Middle East & North Africa	Algeria, Bahrain, Djibouti, Egypt-Arab Rep., Iran-Islamic Rep., Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Malta, Morocco, Oman, Qatar, Saudi Arabia, Syrian Arab Republic*, Tunisia, United Arab Emirates
Sub-Saharan Africa	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Dem. Rep.*, Congo-Rep., Côte d'Ivoire, Equatorial Guinea, Eritrea*, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, São Tomé and Principe, Senegal, Seychelles, Sierra Leone, Somalia*, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
Latin America & Caribbean	Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba*, Dominica, Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico*, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela-RB
North America	Bermuda*, Canada, United States
Western Europe	Albania, Austria, Belgium, Bulgaria, Cyprus*, Czech Republic*, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia-FYR, Netherlands, Norway, Poland, Portugal, Romania*, Russian Federation, Slovak Republic, Spain, Sweden, United Kingdom
Central Asia	Kazakhstan, Kyrgyz Republic, Tajikistan, Turkmenistan, Uzbekistan
Western Asia	Armenia, Azerbaijan, Turkey
Eastern Europe	Belarus, Georgia, Hungary, Moldova, Switzerland, Ukraine*
Pacific Oceania	Australia, Fiji, Kiribati, Marshall Islands, Micronesia-Fed. Sts., New Zealand, Palau*, Papua New Guinea, Samoa, Solomon Islands, Tonga, Vanuatu
Southeast Asia	Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Philippines, Singapore, Thailand, Vietnam
East Asia	China, Hong Kong-SAR, China, Japan, Korea-Rep., Macao-SAR, China, Mongolia

Note: All the countries listed above are included for the multiple imputation and EKC estimations, but countries with * are excluded from the CO_2 emission projection due to the data availability.

Table 3 Results of the EKC estimation (CO_2)

Dep. var: CO2 emissions kg per capita	Listwise deletion		Multiple imputation	
	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	458.5	315.2	1262.18***	401.52
GDP	441.3***	21.9	388.37***	41.04
GDP^2	−3.15***	0.68	−1.38	1.69
GDP^3	0.02***	0.01	0.004	0.02
South Asia	−395.1	388.3	−1218.31***	431.49
Middle East & North Africa	2368***	302.6	1835.29***	403.53
Sub-Saharan Africa	−326.2	315.7	−1228.82***	387.48
Latin America & Caribbean	−651.1**	296.3	−1531.14***	371.94
North America	7199***	492.3	10322.99***	1241.35
Western Europe	631.5**	294.8	454.93	395.29
Central Asia	3142***	478.7	2240.1***	531.08
Western Asia	−1.84	498.1	−531.85	562.61
Eastern Europe	−301.7	396.9	−604.59	484.13
Pacific Oceania	1501***	489.1	−153.14	414.57
Southeast Asia	−1165***	346.0	−1668.3***	423.67
Year	−14.06***	5.03	−12.39***	4.73
Number of obs.	4291		6902	

*** Significantly different from zero at the 1% significance level

** Significantly different from zero at the 5% significance level

* Significantly different from zero at the 10% significance level

Table 4 Results of the EKC estimation (SO_2)

Dep. var: SO2 emissions kg per capita	Listwise deletion		Multiple imputation	
	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	13.19***	3.17	16.86***	3.38
GDP	3.19***	0.27	2.49***	0.25
GDP^2	−0.1***	0.01	−0.06***	0.01
GDP^3	0.0009***	0.0001	0.0003***	0.0001
South Asia	−4.35	3.97	−3.36	3.87
Middle East & North Africa	13.94***	3.08	15.96***	3.33
Sub-Saharan Africa	10.12***	3.18	3.82	3.35
Latin America & Caribbean	0.17	2.99	0.69	3.2
North America	80.37***	5.16	123.52***	15.72
Western Europe	25.56***	3.03	21.63***	3.24
Central Asia	25.99***	5.26	22.38***	4.92
Western Asia	3.92	5.32	3.99	5.45
Eastern Europe	22.77***	4.2	16.86***	4.41
Pacific Oceania	31.81***	5.12	14.12***	4.29
Southeast Asia	−0.61	3.48	−2.56	3.73
Year	−0.56***	0.06	−0.62***	0.05
Number of obs.	4291		6902	

*** Significantly different from zero at the 1% significance level

** Significantly different from zero at the 5% significance level

* Significantly different from zero at the 10% significance level

Table 5 Results of the EKC estimation (PM10)

Dep. var: PM10 concentration micrograms per cubic meter	Listwise deletion		Multiple imputation	
	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	107.3***	4.28	108.95***	4.2
GDP	−2.56***	0.22	−2.65***	0.21
GDP^2	0.07***	0.01	0.07***	0.01
GDP^3	−0.0004***	0.00004	−0.0004***	0.00004
South Asia	−1.31	4.93	−3.01	4.85
Middle East & North Africa	16.0***	4.34	13.83***	4.29
Sub-Saharan Africa	−14.7***	4.25	−17.09***	4.15
Latin America & Caribbean	−26.39***	4.16	−27.32***	4.09
North America	−40.38***	6.91	−41.33***	6.09
Western Europe	−36.76***	4.23	−38.05***	4.17
Central Asia	−26.88***	5.49	−27.45***	5.36
Western Asia	−1.24	6.18	2.85	6.04
Eastern Europe	−35.44***	5.16	−36.47***	5.06
Pacific Oceania	−49.47***	5.06	−50.97***	4.49
Southeast Asia	−20.67***	4.74	−22.25***	4.67
Year	−1.89***	0.10	−1.86***	0.10
Number of obs.	3908		6902	

*** Significantly different from zero at the 1% significance level

** Significantly different from zero at the 5% significance level

* Significantly different from zero at the 10% significance level

Table 6 Results of the EKC estimation (BOD)

Dep. var: BOD g per day per worker	Listwise deletion		Multiple imputation	
	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	173.5***	8.95	178.02***	8.05
GDP	−5.1***	0.69	−4.04***	0.61
GDP^2	0.16***	0.03	0.11***	0.02
GDP^3	−0.001***	0.0003	−0.001***	0.0002
South Asia	−7.89	16.91	0.46	0.3
Middle East & North Africa	16.53**	8.21	10.51	7.1
Sub-Saharan Africa	55.61***	8.55	51.18***	7.83
Latin America & Caribbean	100.2***	8.43	87.01***	8.2
North America	19.21*	11.43	17.28	10.79
Western Europe	25.61***	7.16	19.71***	6.67
Central Asia	38.77***	10.07	33.83***	9.25
Western Asia	6.93	10.69	−2.06	10.29
Eastern Europe	118.8***	10.18	106.79***	10.63
Pacific Oceania	105.5***	10.29	104.6***	9.73
Southeast Asia	−18.92**	8.98	−8.68	8.5
Year	0.66**	0.33	−5.92	14.93
Number of obs.	905		1725	

*** Significantly different from zero at the 1% significance level

** Significantly different from zero at the 5% significance level

* Significantly different from zero at the 10% significance level

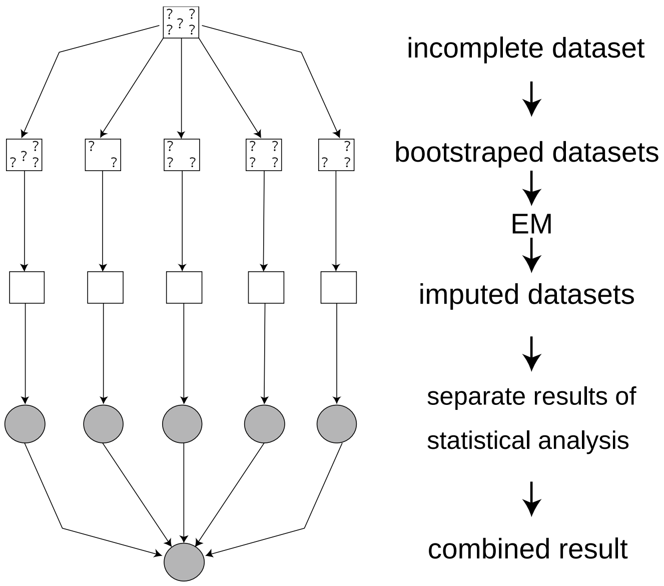

Fig. 1 Imputation process by EMB algorithm

Source: Honaker et al. 2011, p.4

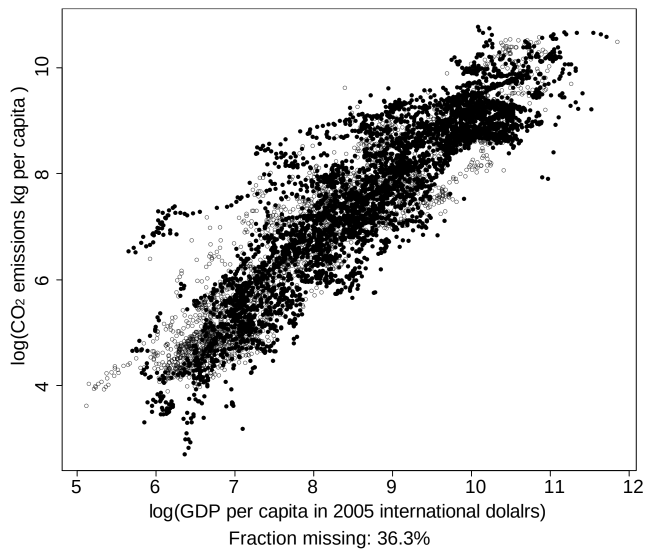

Fig. 2 Scatter plot of observed values and imputed values（CO_2）

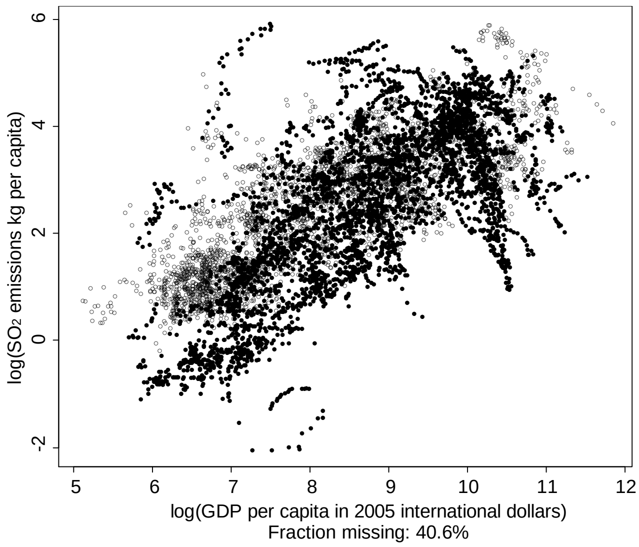

Fig. 3 Scatter plot of observed values and imputed values（SO_2）

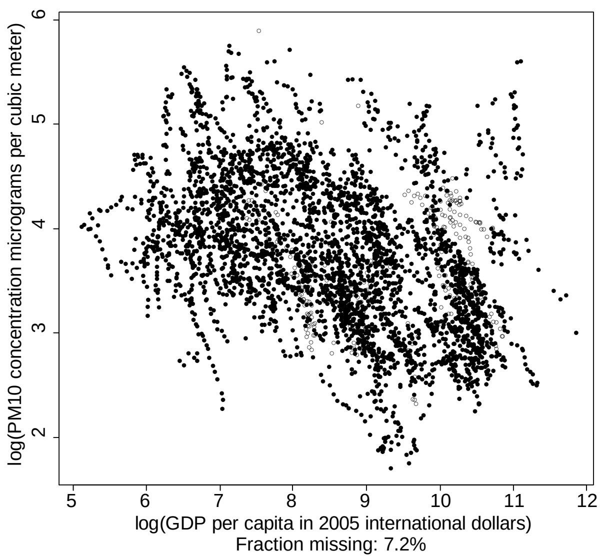

Fig. 4 Scatter plot of observed values and imputed values（PM10）

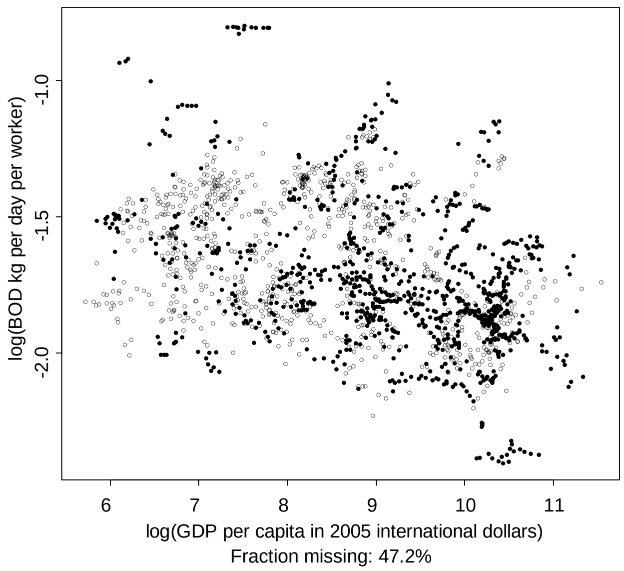

Fig. 5 Scatter plot of observed values and imputed values（BOD）

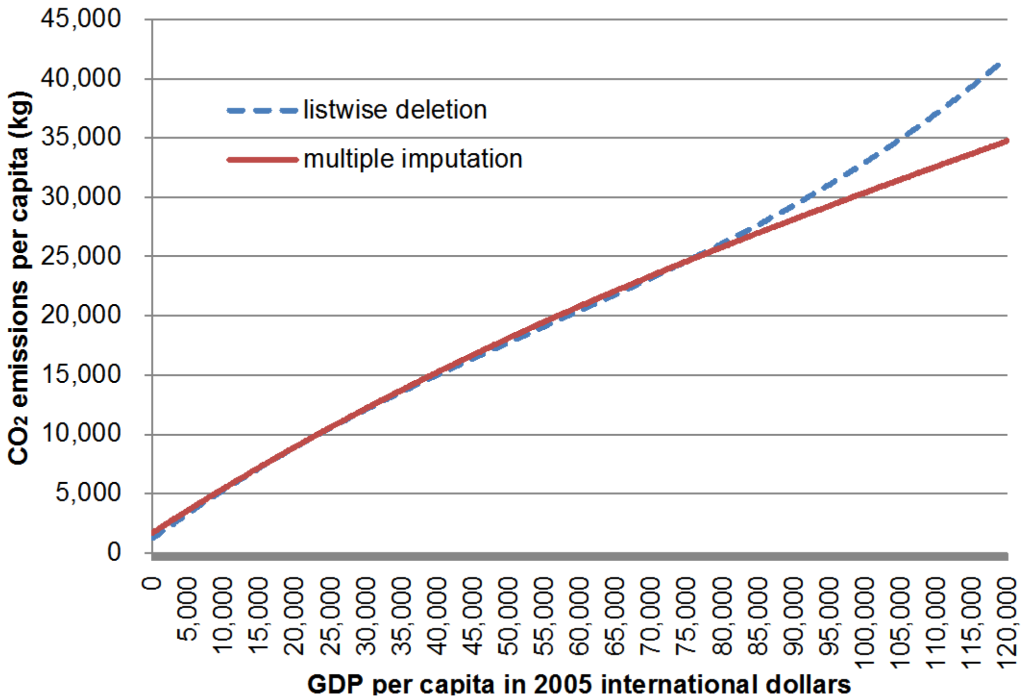

Fig. 6 The estimated EKC for CO_2 emissions

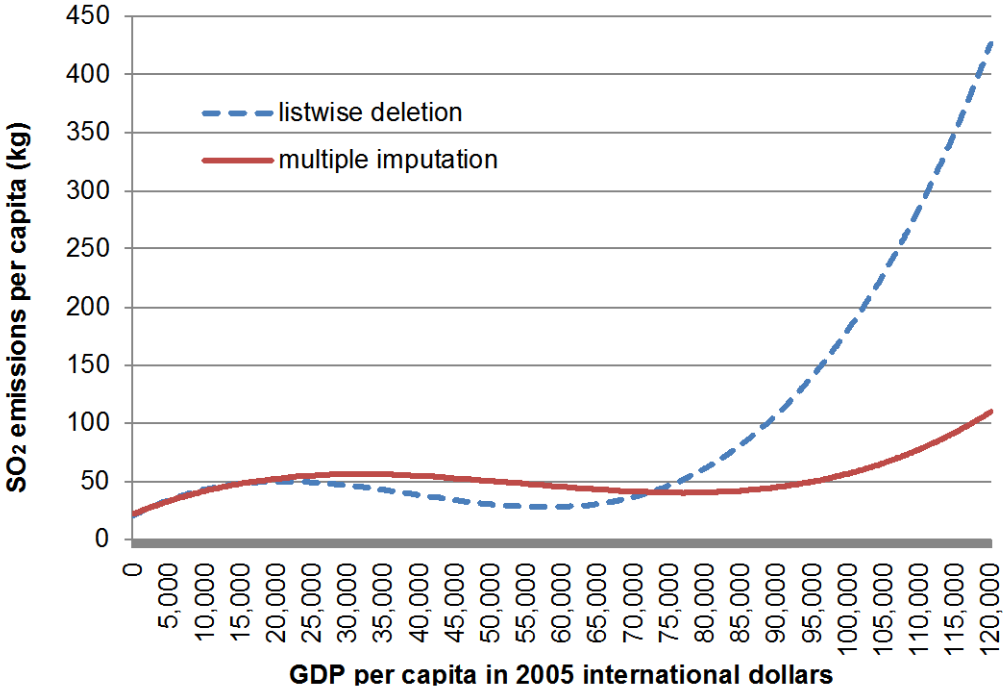

Fig. 7 The estimated EKC for SO_2 emissions

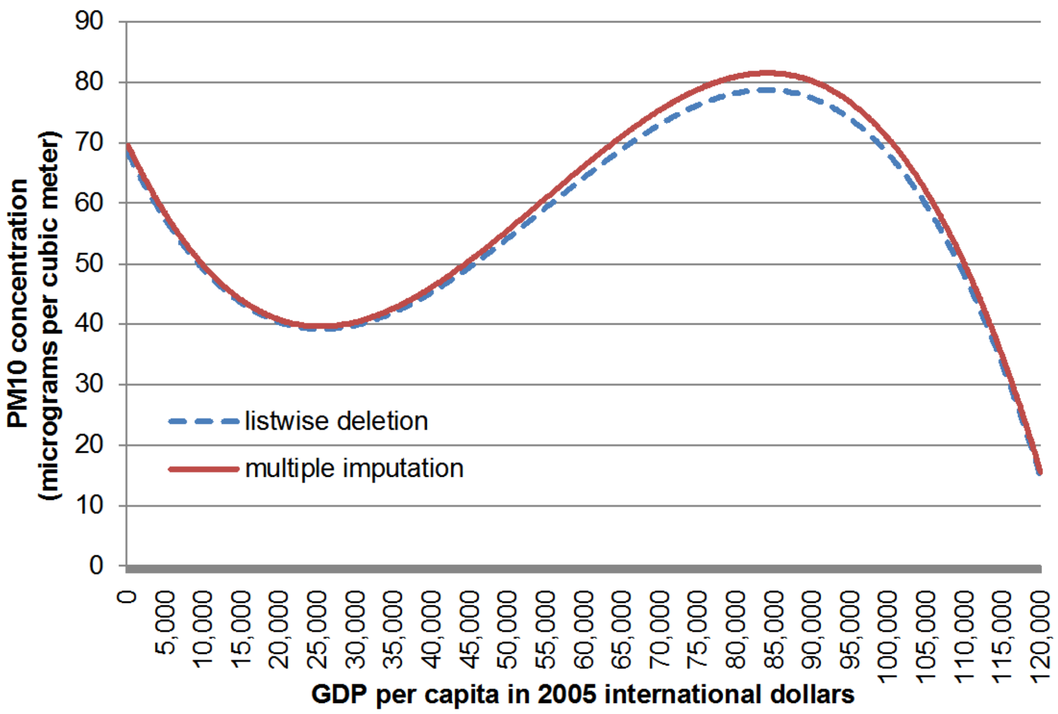

Fig. 8 The estimated EKC for PM10 concentration

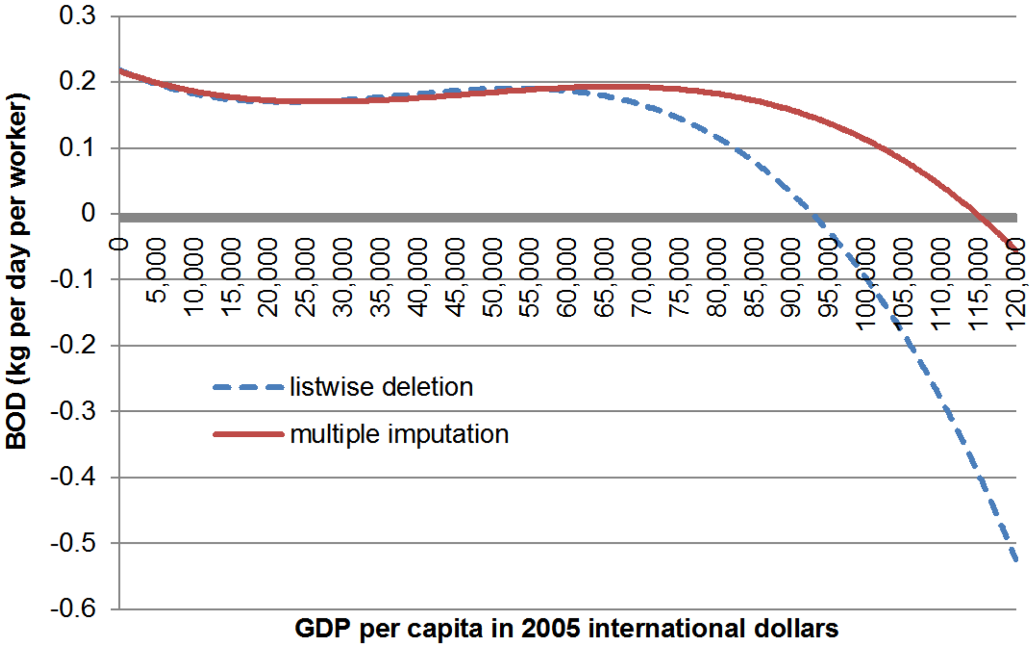

Fig. 9 The estimated EKC for BOD

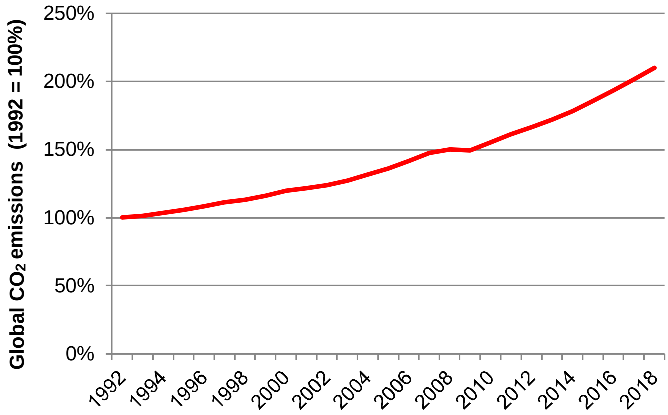

Fig. 10 Annual trend of predicted global CO_2 emissions

Note: The emission level in 1992 is set to be 100%.

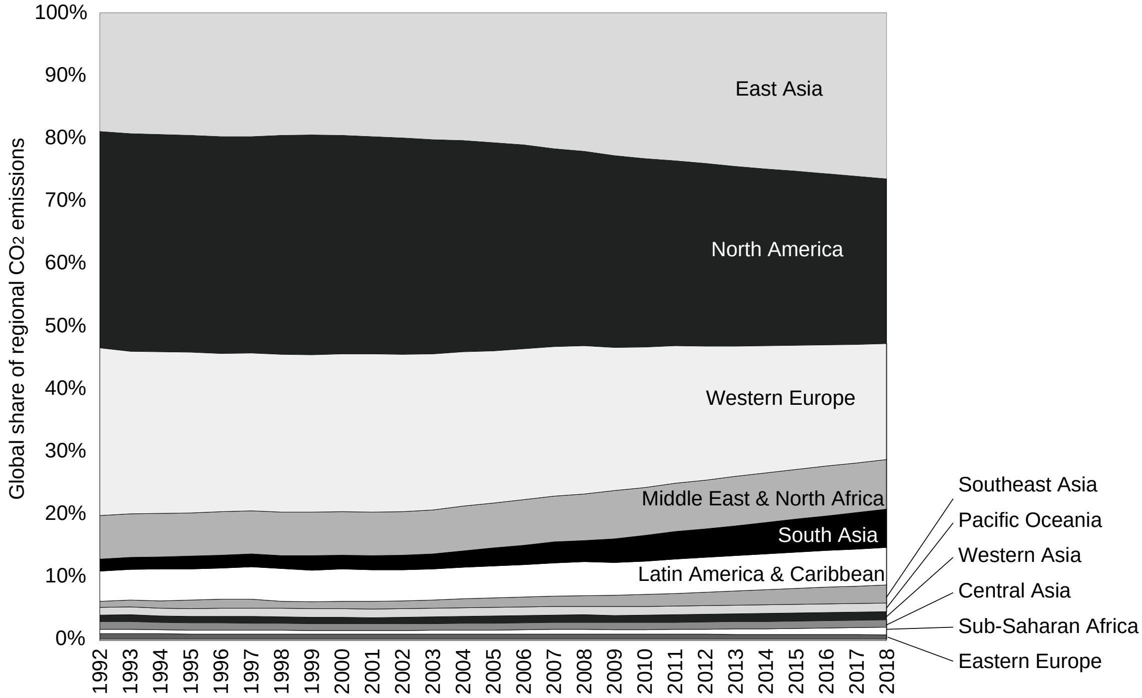

Fig. 11 Annual trend of predicted CO_2 emissions share

References

Andridge R. R, Little R. J. A (2010) A review of hot deck imputation for survey non-response. Int Stat Rev 78:40-64

Azomahou T, Laisney F, Nguyen Van P (2006) Economic development and CO_2 emissions: a nonparametric panel approach. J Public Econ 90:1347-1363

Bertinelli L, Strobl E (2005) The environmental Kuznets curve semi- parametrically revisited. Econ Lett 88:350-357

Coleman A (2009) A model of spatial arbitrage with transport capacity constraints and endogenous transport prices. Am J Agric Econ 91:42-56

Cranmerand S. J, Gill J (2013) We have to be discrete about this: a non-parametric imputation technique for missing categorical data. Br J Polit Sci 43:425-449

Dasgupta S, Laplante B, Wang H, Wheeler D (2002) Confronting the environmental Kuznets curve. The J Econ Perspectives 16:147-168

Gelman A, Hill J (2006) Data Analysis Using Regression and Multilevel/ Hierarchical Models. Cambridge University Press, Cambridge

Graham J, Olchowski A, Gilreath T (2007) How many imputations are really needed? some practical clarifications of multiple imputation theory. Prev Sci 8:206-213

Grossman GM, Krueger AB (1991) Environmental impacts of a North American free trade agreement. NBER Working Papers 3914, National Bureau of Economic Research, Inc

Grossman GM, Krueger AB (1995) Economic growth and the environment. The Q J Econ 110:353-377

Heil MT, Selden TM (2001) Carbon emissions and economic development: future trajectories based on historical experience. Environ Dev Econ 6:63-83

Holtz-Eakin D, Selden TM (1995) Stoking the fires? CO_2 emissions and economic growth. J Public Econ 57:85-101

Honaker J, King G (2010) What to do about missing values in time series cross-section data. Am J Polit Sci 54:561-581

Honaker J, King G, Blackwell M (2011) Amelia II: a program for missing data. J Stat Softw 45:1-47

Iwami T (2004) Economic development and/or environmental quality: emissions of CO_2 and SO_2 in East Asia. Discuss Pap F series March 2004

Junninen H, Niska H, Tuppurainen K, Ruuskanen J, Kolehmainen M (2004) Methods for imputation of missing values in air quality data sets. Atmospheric Environ 38:2895-2907

Kabacoff R (2011) R in Action: Data Analysis and Graphics with R. Manning Publications Co

King G, Honaker J, Joseph A, Scheve K (2001) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. The Am Polit Sci Rev 95:49-69

Kuznets S (1955) Economic growth and income inequality. The Am Econ Rev 45:1-28

Lee CC, Chiu YB, Sun CH (2010) The environmental Kuznets curve hypothesis for water pollution: do regions matter? Energy Policy 38:12-23

Little RJA (1988) A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 83:1198-1202

Managi S, Hibiki A, Tsurumi T (2009) Does trade openness improve environmental quality? J Environ Econ Manag 58:346-363

Mazzanti M, Montini A, Zoboli R (2007) Environmental Kuznets curves for GHGs and air pollutants in Italy: evidence from sector environmental accounts and provincial data. Econ Polit 24:369-406

Millimet D, List J, Stengos T (2003) The environmental Kuznets curve: real progress or misspecified models? Rev Econ Stat 85:1038-1047

Miyama E, Managi S (2014) The environmental Kuznets curve in Asia. In: Managi S (Eds) Handbook of Environmental Economics in Asia. Routledge, New York, USA (forthcoming)

Orubu CO, Omotor DG (2011) Environmental quality and economic growth: searching for environmental Kuznets curves for air and water pollutants in Africa. Energy Policy 39:4178-4188

Rubin DB (1976) Inference and missing data. Biom 63:581-592

Rubin DB (1977) The design of a general and flexible system for handling non-response in sample surveys. Working document prepared for the U.S. Social Security Administration.

Rubin DB, Schenker N (1986) Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. J Am Stat Assoc 81:366-374

Shafik N (1994) Economic development and environmental quality: an econometric analysis. Oxf Econ Pap 46:757-773

Stern D (2004) The rise and fall of the environmental Kuznets curve. World Dev 32:1419-1439

Tsurumi T, Managi S (2010a) Decomposition of the environmental Kuznets curve: scale, technique, and composition effects. Environ Econ Policy Stud 11:19-36

Tsurumi T, Managi S (2010b) Does energy substitution affect carbon dioxide emissions-income relationship? J Jpn Int Econ 24:540-551

World Commission on Environment and Development (1987) Our Common Future. Oxford University Press, Oxford

Yaguchi Y, Sonobe T, Otsuka K (2007) Beyond the environmental Kuznets curve: a comparative study of SO_2 and CO_2 emissions between Japan and China. Environ Dev Econ 12:445-470