

Green Complexity, Economic Fitness, and Environmental Degradation: Evidence from U.S. State-Level Data

İbrahim Tuğrul Çınar

Anadolu University: Anadolu Universitesi

İlhan Korkmaz

Dumlupınar University: Kutahya Dumlupınar Universitesi

Muhammet Yunus Şişman (✉ myunus.sisman@dpu.edu.tr)

Dumlupınar Üniversitesi: Kutahya Dumlupınar Universitesi <https://orcid.org/0000-0002-9791-1382>

Research Article

Keywords: Green Complexity, Economic Complexity, Sub-national level, SO₂, PM₁₀, CO₂ emissions

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1035080/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Environmental Science and Pollution Research on March 29th, 2022. See the published version at <https://doi.org/10.1007/s11356-022-19859-8>.

Abstract

Green production is one of the major debates as environmental degradation poses threats globally. The paper attempts to explore the relationship between green economy and environmental quality by using Economic Fitness approach. We develop a Green Complexity Index (GCI) dataset consists of 290 traded green-labeled products for the US States between 2002 and 2018. We analyze the environmental performance of green production using the GCI data at the sub-national level. Findings indicate that exporting more complex green products has insignificant effects on local (i.e., Sulfur dioxide, Particulate Matter 10) and global pollutants such as Carbon dioxide (CO₂), even accounting for per capita income. Yet, overall economic complexity has a significant negative impact on the emission levels implying that sophisticated production significantly improves environmental quality in the US. The insignificant impact of GCI on environmental degradation suggests that green product classifications should incorporate the production and end-use stages of goods to limit the adverse environmental effects of green-labeled products. The study, therefore, provides policy implications for green industrial policies.

JEL codes: O18, Q56, R11

1. Introduction

Global ecological degradation has been raising environmental awareness in modern societies. Local governments and international institutions are promoting environmentally friendly products and services to limit the adverse effects of industrialization. Policymakers increasingly adopt environmental strategies to stimulate the production of climate-neutral and sustainable products and to prevent environmentally hazardous goods in the markets. Scholarly works accompany these efforts and provide extensive theoretical background for the "green economy."

The existing literature has evolved into a fairly large area known as the Environmental Kuznet Curve (EKC) hypothesis, which mainly focuses on the relationship between income level and environmental degradation (i.e., Meadows et al. 1972; Grossman and Krueger, 1991; Shafik and Bandyopadhyay, 1992; Panayotou 1993, and Selden and Song, 1994). More recently, environmental impacts of economic activities are outlined by using Economic Complexity Index (ECI) developed by Hidalgo and Hausmann (2009). The economic complexity approach introduced by Hidalgo and Hausmann (2009) emerged as an empirical innovation using computational complex network techniques within the structuralist approach of development economics and has significantly impacted the growth and development literature since its inception (Gala et al., 2018). The ECI, considered a robust indicator of economic growth, is calculated based on the ubiquity and diversity of products and measures the sophisticated manufacturing capabilities of a country's production structure. In this context, several studies (e.g., Can and Gozgor, 2017; Doğan et al., 2019; Yilanci and Pata, 2020) examine the relationship between economic complexity and environmental degradation. Can and Gozgor (2017) finds a negative relationship between air pollutants (CO₂) and economic complexity in developed countries, while, Doğan et al. (2019) and Yilanci

and Pata (2020) suggest that the relationship between CO₂ and economic complexity is positive for developing countries.

Among others, Neagu (2019), Chu (2020), and Pata (2020) show that EKC type inverted U shape is valid between ECI and CO₂. However, Dinda (2004) states that empirical evidence for the EKC hypothesis is mostly valid for local pollutants such as SO₂ and Suspended Particulate Matters (SPM). Between global pollutants such as CO₂ and income level shows a monotonic relationship rather than an inverted U shape. The ECI is questioned due to the linear computation approach by a number of studies in the literature (Tacchella et al., 2012; Caldarelli et al., 2012; Tacchella et al., 2013; and Cristelli et al., 2013). In this context, Tacchella et al. (2012) developed Economic Fitness Index (EFI) based on non-linear fixed-point iteration. They claim to eliminate the conceptual and application-related defects they identified in the Hidalgo and Hausmann (2009) method. Boleti et al. (2021), in their study for 88 countries, examined the relationship between economic complexity and environmental performance by using the ECI approach as well as the ECI+, which is equivalent to the economic fitness algorithm. They concluded that economic complexity improves environmental performance, but negatively affects air pollution such as CO₂ and PM 2.5. In their prominent study, Mealy and Teytelboym (2020) introduces a green complexity index (GCI) using the environmental product lists reported by WTO, OECD, and APEC. They develop GCI based on economic complexity (Hidalgo and Hausmann, 2009) and economic fitness (Tacchella et al., 2012) approaches and examined the relationship between environmental degradation and green product complexity for 122 countries. Their findings indicate that countries having higher GCI experience lesser ecological degradation, i.e., lower CO₂ emissions. However, cross-country data inconsistency is a major issue, particularly in environmental research, due to the significant differences between emission measurement methodologies across countries (Stern et al., 1996; List and Gallet, 1999; De Groot et al., 2004; Carson, 2010; and Awaworyi Churchill et al., 2020).

This study aims to develop a green product complexity index to explore the link between environmental degradation and green production in the United States at the sub-national level. Empirical analysis at the sub-national level allows us to minimize the previously mentioned data inconsistency problem, particularly in environmental research. We employ economic fitness approach (Tacchella et al., 2012) to 290 products at HS-6 level listed as green products by OECD, APEC, and WTO. Then, EFI, GCI, and environmental data are estimated by the fractional polynomial regression method, which has several desirable features such as providing more flexible functional forms and allowing powers to be logarithmic, non-integer, or to be repeated. Findings reveal that GCI has an insignificant impact on environmental quality in the US, implying that exporting more complex green products does not affect emission levels. On the other hand, we find that EFI has statistically significant coefficients indicating an inverted U-shape, particularly for SO₂ in the US.

The remainder of the study is organized as follows. Section 2 introduces the theoretical background for our economic fitness and green product complexity index database for fifty-one US States. The next section overviews the data used for the analysis and empirical framework. Following the results and discussion, the paper concludes with key remarks.

2. Theoretical Background And Methodology

2.1. Computing Economic Fitness Index of States

Economic fitness approach developed by Tacchella et al. (2012) is based on the complex network structure as in Hidalgo and Hausmann (2009). This network structure is represented by an adjacency matrix that enables to do numerical measurement. While countries are at the rows of the adjacency matrix, exported products are at the columns. In this respect, an adjacency matrix with c countries, where p products are exported, will be a country-product matrix of cxp dimensions and consisting of 1s and 0s in each element denoted by M_{cp} . Both economic complexity and economic fitness approaches use Revealed Comparative Advantage (RCA) index developed by Balassa (1965) to obtain 1s and 0s. In this way, it is possible to obtain information whether or not a country is an important exporter of a product. Accordingly, RCA index of a product p exported by a country c can be defined as Equation (1) below.

$$RCA_{cp} = \frac{\sum_{c'} q_{c'p}}{\sum_{p'} q_{cp'}} \frac{\sum_{c'p'} q_{c'p'}}{\sum_{c'p} q_{c'p}} \quad (1)$$

If the value obtained from the Equation (1) is greater than or equal to 1, M_{cp} will take the value 1, otherwise 0. Formally, this can be represented as in Equation (2).

$$\begin{aligned} RCA_{cp} \geq 1 &\Rightarrow M_{cp} = 1 \\ RCA_{cp} < 1 &\Rightarrow M_{cp} = 0 \end{aligned} \quad (2)$$

Tacchella et al. (2012) showed that the products exported by less diversified countries are generally ordinary products, while highly diversified countries export both ordinary and sophisticated products. Thus, the products exported by diversified countries give us almost no information about the level of sophistication of these products. Therefore, it is meaningless to use the average diversity levels of the exporting countries to determine the level of sophistication of products as claimed by Hidalgo and Hausmann (2009).

According to Tacchella et al. (2012), the way to mathematically represent the structure described above is to define a nonlinear relationship. From this point of view, they proposed an iteration process that obtains fixed points of the system by defining the fitness (F_c) and product complexity (Q_p) in a nonlinear coupled equation system as given in Equation (3).

$$\left. \begin{aligned} \tilde{F}_c^{(n)} &= \sum_p M_{cp} Q_p^{(n-1)} \\ \tilde{Q}_p^{(n)} &= \frac{1}{\sum_c M_{cp} \tilde{F}_c^{(n-1)}} \end{aligned} \right\} \rightarrow \begin{aligned} F_c^{(n)} &= \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c} \\ Q_p^{(n)} &= \frac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle_p} \end{aligned} \quad (3)$$

Referring to the theory of evolution, Tachella et al. (2012) coined their approach as economic fitness. In economic fitness algorithm given in Equation (3), F_c is proportional to the sum of a country's exports weighted by product complexity values, while Q_p is inversely proportional to the number of countries exporting the product.

Equation (3) shows a two-stage iteration process. First, $\tilde{F}_c^{(n)}$ and $\tilde{Q}_p^{(n)}$ intermediate variables are calculated using the relevant formulas, and then these intermediate variables are normalized at each iteration stage and $F_c^{(n)}$ and $Q_p^{(n)}$ values are reached. For the solution of the coupled equation system given in Equation (3), $F_c^{(0)} = 1$ and $Q_p^{(0)} = 1$ are given as the initial condition. Although Tacchella et al. (2012) stated that the fixed point solutions of the coupled equation system calculated with Equation (3) are stable and independent of the initial condition, Morrison et al. (2017) brought up the issue of the instability of the system. Servedio et al. (2018) state that the original economic fitness algorithm successfully reveals the ranking of countries for different years, yet they also state that there are some parts of the system that are open to improvement.

The new economic fitness algorithm developed by Servedio et al (2018) is shown in Equation (4).

$$\begin{aligned} F_c^{(n)} &= \phi_c + \sum_{p'} M_{cp'} / P_{p'}^{(n-1)} \\ P_p^{(n)} &= \pi_p + \sum_{c'} M_{c'p} / F_{c'}^{(n-1)} \end{aligned} \quad (4)$$

In Equation (4), the product complexity is now given by P_p^{-1} . In the system where the initial condition is given as $F_c^{(0)} = P_p^{(0)} = 1$, as in Equation (3), by adding two values greater than zero, such as ϕ_c and π_p to each equation, the system is provided to have a structure that is not defined by a multiplicative constant. In this way, the system does not need to be normalized at every stage as it was done earlier (Servedio et al., 2018). In Equation (4), ϕ_c represents the self-fitness value of a country. Accordingly, even if a country does not export at all, it will have a fitness value of ϕ_c . On the other hand, π_p expresses the minimum value that P_p will take in the case that a product is not exported by any country ($M_{cp} = 0, \forall c$), thus the maximum value that the product complexity expressed as $P_p = Q_p^{-1}$ can take for any product. Servedio et al. (2018) stated that this situation can only be valid for innovative products that have not

been produced yet, while they stated that the complexity value of products that have not yet been invented will be at the maximum level.

To make Equation (4) parameter free and facilitate the algorithm, a common value is assigned such that $\phi_c = \pi_p = \theta$; $\forall c, p$ and rescaled quantities $\tilde{P}_p = P_p / \theta$ and $\tilde{F}_c = F_c \theta$ are introduced. In this way, after rearranging equation (4) following equation can be obtained.

$$\begin{aligned} \tilde{F}_c^{(n)} &= \theta^2 + \sum_p M_{cp} / \tilde{P}_p^{(n-1)} \\ \tilde{P}_p^{(n)} &= 1 + \sum_{c'} M_{c'p} / \tilde{F}_{c'}^{(n-1)} \end{aligned} \tag{5}$$

As soon as the θ parameter is much smaller than the typical value of M_{cp} matrix elements, i.e., much smaller than unity, the fixed point in terms of \tilde{F}_c and \tilde{P}_p almost does not depend on θ (Servedio et al. 2018).

Operti et al. (2018) developed two criteria, exogenous fitness and endogenous fitness, in order to apply the economic fitness approach in a robust way at the regional level. Endogenous fitness; It is calculated using the state-product matrix based on the RCA index, which is the ratio of a product p exported by states to the total exports of states, and the ratio of that product p to the total export of the whole country. Except for the state-product matrix, the standard economic fitness value algorithm is applied exactly. Besides, exogenous fitness; Based on the assumption that product complexity values should not change all over the world, it is based on calculating the economic fitness value for the countries of the world with the standard method and using the product complexity data obtained on a global basis to calculate the fitness value for the states. Hereby, it is prevented from obtaining deviant product complexity values for products that are not produced locally or produced by very few states but widely produced in the world. Accordingly, for the EFI that will be calculated at the state level, the product complexity ($Q_p = (P_p - 1)^{-1}$) vector calculated according to the new EFI algorithm shown in Equation (5) and M state-product matrix are multiplied as follows:

$$F_{p \text{ exogenous}} = \sum_p M_{sp} Q_p \tag{6}$$

Here, M_{sp} denotes the elements of the binary RCA matrix, M , in which 51 U.S. states and exported products classified according to the 6 digit Harmonized System (HS6) are arranged in the rows and columns, respectively. Accordingly, if a state has a comparative advantage in the export of a product, the relevant M_{sp} element will take the value 1, otherwise it will take the value 0. Q_p is the global product complexity vector calculated on the basis of 206 countries.

The M_{sp} matrix to be used for the calculation of the state-level exogenous fitness index and the M_{cp} country-product matrices prepared to calculate the global product complexity vector should be created

separately for each year and the calculation given in Equation (5) and Equation (6) should be repeated respectively for each year. The analyzes were carried out over 500 iterations and the value of $\theta=10^{-6}$.

2.2. Computing Green Complexity Index of States

The Green Complexity Index (GCI) developed by Mealy and Teytelboym (2020) is constructed for US states in addition to EFI in this study. Following Mealy and Teytelboym (2020), the environmentally friendly products are obtained from WTO Core List, OECD customized product list of environmental goods, OECD illustrative product list of environmental goods and in the APEC list. There are 295 environmental products in total. Hence, M matrix, is expected to be 51x295. However, the size of the matrix may decrease to around 51x290 in some years due to zero export values. Using equation (6), this Msp matrix is multiplied by the global product complexity vector to obtain the state level GCI values. Here, Qp is the sub-vector of the product complexity vector calculated on the basis of 206 countries, which includes environmental products. This process is repeated for each year separately for the period between 2002-2018.

2.3. Estimation Methodology

Panel data analysis provides a useful analysis tool in terms of providing more variability and less collinearity. Baltagi (2005) states that the fixed effect estimation is more appropriate for a sample of states, companies, or countries with similar conditions. In addition, availability of state-level export data restricts the time dimension of our study. Therefore, the analysis methods applied for large T may cause biased results in this study. For this reason, we used fixed effect estimation method to examine the relationship between GCI, EFI, and local air pollutants for US states that display a more homogeneous structure in terms of environmental policies and laws.

On the other hand, when the Environmental Kuznets Curve (EKC) literature is examined, it is seen that the estimated baseline model is polynomial models that include the quadratic or cubic forms of GDP. These quadratic and cubic forms added to the models can save researchers from falling into functional form bias to some extent. However, there are criticisms that these quadratic or cubic terms, which are added to the models to take into account the nonlinear relationship, are quite restrictive (Aslanidis, 2009). In fact, without knowing the exact form of this nonlinear relationship, quadratic or cubic terms are added to models by researchers and then predictions are made over these forms. At this point, the fractional polynomial regression approach developed by Royston and Altman (1994) emerges as an alternative model that provides more flexibility compared to the regular polynomial models used to test the EKC hypothesis. Fractional polynomial regression is a method that allows logarithmic, non-integer, or repeated powers, allowing us to choose the most appropriate functional form among a much more considerable range of functional forms and to determine whether the independent variable is important to our model (Royston, 2017; Royston et al., 1999). From this point of view, it was also aimed to examine the possible non-linear relationship between explanatory variables and explained variables without falling into functional form bias by making fractional polynomial fixed effect regression estimation in the study.

In the fractional polynomial regression approach, the Function Selection Procedure (FSP) based on the closed test procedure is applied to determine the most convenient functional form. FSP starts with a highest degree fractional polynomial regression and statistically tests the ability to reduce that model to a first degree or a linear model.

According to FSP, the highest degree fractional polynomial allowed by the researcher is first tested for the case where the variable is omitted from the model. If this test statistic is significant, testing procedure continues for linear, second order, or third order fractional polynomials. Otherwise, the variable should be omitted from the model and the testing procedure stops (Royston, 2017).

3. Data

Following Mealy and Teytelboym (2020), environmental product list was obtained by combining the products in OECD customized product list of environmental goods (244 products), OECD illustrative list (120 products), WTO Core list (26 products) and APEC List (52 products). The environmental product list consists of 295 products. It should be noted that products are classified according to HS6 in all lists. However, the versions of the HS codes may differ. For example, the OECD customized list is organized according to HS2007, and the WTO Core List is organized according to HS 2002. The APEC list is available separately for HS2002, HS2007 and HS2012. In order to collate these coding into a single dataset, all environmental product lists and export data on the basis of countries and states are converted to HS1992 by using the Eurostat RAMON classification converter. In order to calculate the green complexity index at the country level, the export data of the countries classified according to HS6t level can be easily obtained from the BACII and COMTRADE data sets. However, in order to be able to conduct a subnational analysis, export data are published at subnational level over the years is needed. As the United States provides export data at HS6 level for each state we prefer to examine the relationship between GCI, EFI and environmental degradation for US. On the other hand, since CO₂, real GDP and industrial energy consumption data is reported for 50 states and 1 federal district we limited our sample with these 51 cross sections.

Data for real GDP and industrial energy consumption available until 2018, CO₂ data up to 2017 and state-based export data according to HS6 starting from 2002. In this context, our analysis covers 2002-2018 for the dependent variables PM₁₀ and SO₂ and 2002-2017 for CO₂.

BACI export data set in HS6 is used to calculate product complexity data on a global scale. The BACI data set is freely accessible and quite useful because of the data it provides by overlapping the declarations of the exporting and importing countries on the trade of goods through the United Nations COMTRADE data (Gaulier ve Zignago, 2010; Cristelli vd. 2013). Among the countries included in the BACI data set, the countries with the status of micro-states that have no exports are excluded and as a result, global product complexity calculations are made over 206 countries.

Six-digit export data for US states were acquired from the US Census Bureau's USA Trade Online database. Unadjusted energy related CO₂ emission data for CO₂ data were obtained from the US Energy

Information Administration (EIA). SO₂ and PM₁₀ data are taken from the Environmental Protection Agency. As recommended by Dinda (2004), these CO₂, PM₁₀, and SO₂ data are proportioned to mid-year population estimate data from the Bureau of Economic Analysis for US states to allow the use of relative data instead of absolute. In this context, CO₂, SO₂ and PM₁₀ data are in tons per capita. To calculate the state level real GDP per capita values, Real GDP (2012 million dollars) data were proportioned to the mid-year population estimates of the states.

Natural logarithmic transformations of all variables are used in our analyses. The Green Complexity value for Alaska in 2010 is zero, its natural logarithmic transformation becomes undefined. For this reason, analyzes involving the Green Complexity variable has a missing value.

We used two additional control variables, electricity consumption per capita and population density, in our estimations. Population density, is thought to be effective on environmental degradation since the earliest researches such as Grossman and Krueger (1991) and Selden and Song (1994). On the other hand, according to Dinda (2004) most of the air pollutants such as SO₂, PM₁₀ and CO₂ are energy related. Thus, we used electric energy consumption per capita as a proxy for energy use.

4. Results

The graphs of the calculated state-level GCI and EFI values by years are given below. In Figure 1, the left panel shows the five states with the highest average EFI values between 2002 and 2018; the right panel gives the five states with the lowest average fitness for this period.

The left panel of Figure (1) shows us that New Jersey stands out as the state with the highest EFI value. In addition, it is noteworthy that while the EFI values of states such as Florida and California have increased over the years, the state of Pennsylvania has experienced a slight decrease. In Figure 1, we can also observe that states with high EFI values show a more stable development path over the years. Alaska stands out as the state that lags behind in terms of EFI, i.e. productive capabilities, due to its economy based largely on unsophisticated products such as fishing and forestry. One interesting point in Figure (1) is that, as one of the states with the highest per capita income, DC, is ranked among the lowest EFI states. The main reason for this is that the DC's economic structure is business service oriented rather than production.

In addition to EFI, state level visual inspection can be made in terms of GCI from the Figure (2) below. Figure 2 demonstrates the states with the highest (left panel) and the lowest (right panel) average GCI values for the period of 2002-2018.

In Figure (2), three states with the highest EFI, namely Illinois, Florida, and Pennsylvania, are also among the states with the highest GCI. In addition, Alaska, Hawaii, North Dakota, and DC, which are among the lowest EFI values, are also among the states with the lowest GCI values. It is noteworthy that states with high EFI such as New Jersey and California are not among states with high GCI. This shows us that while

New Jersey and California have a high accumulation of productive capabilities, this accumulation seems not to be concentrated on green products.

Correlations and descriptive statistics of the variables we used for our analysis are given in Table 1 and Table 2, respectively.

Table 1
Correlations

	GCI	EFI	GDP pc	Electric Cons. pc	Population Density
GCI	1.000				
EFI	0.883	1.000			
GDP pc	-0.172	-0.213	1.000		
Electric Cons. pc	-0.141	-0.247	-0.087	1.000	
Population Density	0.442	0.566	0.387	-0.269	1.000
<i>N</i>	866				

In Table 1, the high correlation between GCI and EFI is remarkable. For this reason, we preferred not to include the GCI and EFI variables together in an estimated model. Apart from this, the correlations between other variables are considered to be reasonable. In Table 2, descriptive statistics for logarithmic transformations of the variables are given. Alaska, is the only state with a negative logarithmic value due to its low population density.

Table 2
Summary Statistics

	Mean	Standard Deviation	Min	Max
GCI	7.308	0.822	1.975	8.701
EFI	9.938	0.778	6.970	11.273
GDP pc	10.824	0.256	10.336	12.123
Electric Cons. pc	9.442	0.305	8.775	10.332
Population Density	3.631	1.527	-0.833	8.398

The models we run for SO₂, PM₁₀ and CO₂ dependent variables are given in Table 3. First, we questioned the significance of the GCI and EFI variables separately. Both variables are not significant for three different air pollution variables.

Table 3
Univariate Fixed Effect Estimations for GCI and EFI

	S02pc		PM10pc		CO2pc	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
GCI	-0.009		0.001		0.007	
	(0.068)		(0.037)		(0.018)	
EFI		-0.166		0.025		-0.033
		(0.222)		(0.102)		(0.051)
Constant	-3.046 ^{***}	-1.459	-2.559 ^{***}	-2.799 ^{***}	2.994 ^{***}	3.372 ^{***}
	(0.495)	(2.180)	(0.274)	(0.991)	(0.127)	(0.502)
R-squared within	0.771	0.770	0.282	0.276	0.686	0.687
F stat.	49.295	45.017	20.328	24.104	24.151	31.866
prob.	0.000	0.000	0.000	0.000	0.000	0.000
Number of obs.	866	867	866	867	815	816
Heteroskedasticity and autocorrelation robust standard errors are in parentheses.						
* p < 0.10, ** p < 0.05, *** p < 0.01. Year dummies are not reported.						

We expand the results in Table 3 with our control variables and that are reported in Table 4. Accordingly, there is no linear relationship between GCI and EFI variables. However, considering the fact that the analysis results using linear fixed effect estimation may contain functional form bias, alternative estimations based on fractional polynomial recession have been made.

From Table 3 and Table 4, we can see that there is no statistical relationship between GCI, EFI, and local air pollutants. However, it should be kept in mind that the fixed effect estimation is based on the assumption that the relationship between the variables is linear. Considering that polynomial models are used to avoid functional form bias in the EKC literature, estimation over alternative functional forms will yield more reliable estimation results. In this context, the test results called fractional polynomial selection procedure or function selection procedure proposed by Royston (2017), separately for both our GCI and EFI variables, are given in Table 5 and Table 6, respectively.

Table 4
Multivariate Fixed Effect Estimations for GCI and EFI

	SO2pc		PM10pc		CO2pc	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
GCI	0.029		0.003		0.018	
	(0.057)		(0.035)		(0.013)	
EFI		-0.057		0.006		0.002
		(0.216)		(0.104)		(0.035)
GDP pc	0.317	0.243	-0.417	-0.465	0.061	0.061
	(0.914)	(0.923)	(0.389)	(0.408)	(0.104)	(0.099)
Electric.Cons pc.	2.223**	2.229**	0.382	0.404	0.646***	0.640***
	(1.018)	(1.026)	(0.375)	(0.378)	(0.111)	(0.113)
Population Density	0.698	0.695	0.727	0.709	-0.399**	-0.399**
	(1.069)	(1.093)	(0.594)	(0.598)	(0.192)	(0.197)
Constant	-30.204***	-28.678**	-4.294	-3.960	-2.407	-2.250
	(9.302)	(10.767)	(4.229)	(4.581)	(1.454)	(1.365)
R-squared within	0.793	0.791	0.292	0.287	0.776	0.775
F stat.	32.569	35.282	15.249	14.063	43.650	39.047
prob.	0.000	0.000	0.000	0.000	0.000	0.000
Number of obs.	866	867	866	867	815	816

The first column of Table 5 shows the null hypotheses of FSP. Accordingly, the null hypothesis of omitting the GCI variable from the model cannot be rejected for both the second order (M=2) and third order (M=3) polynomial functional form specifications for SO2pc and PM10pc. This shows that, as reported in our fixed effect estimations, GCI is statistically unrelated with the SO2 and PM10 variables for both linear and nonlinear functional forms.

In the FSP, if the null hypothesis for omitting the variable in question is significant, we can say that there is an association between the variables, and moreover, statistically significant null hypothesis for linearity (second row), refers to a nonlinear relationship (Royston, 2017). According to Table 5, the functional form of the relationship between the GCI and CO2 should be linear. However, the linear functional specification in our fixed effect estimation shows us that the relationship between GCI and CO2 is not statistically significant.

Table 5
Function Selection Procedure for GCI

	SO2pc		PM10pc		CO2pc	
	Second order fractional polynomial form	Third order fractional polynomial form	Second order fractional polynomial form	Third order fractional polynomial form	Second order fractional polynomial form	Third order fractional polynomial form
Omitted	0.897	0.714	0.859	0.937	0.047**	0.030**
Linear	0.863	0.617	0.765	0.848	0.158	0.116
M=1	0.777	0.435	0.637	0.679	0.179	0.121
M=2	0.946	–	0.413	–	0.357	–
M=3	–	–	–	–	–	–
Number of models tested	164	44	164	44	164	44
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$						

When the functional selection procedure is applied for second (M=2) and third (M=3) order polynomial forms for the EFI variable, the results are as in Table 6.

Table 6
Function Selection Procedure for EFI

	SO2pc		PM10pc		CO2pc	
	Second order fractional polynomial form	Third order fractional polynomial form	Second order fractional polynomial form	Third order fractional polynomial form	Second order fractional polynomial form	Third order fractional polynomial form
Omitted	0.044**	0.036**	0.000***	0.002***	0.005***	0.014**
Linear	0.027**	0.019**	0.000***	0.001***	0.003***	0.006***
M=1	0.019**	0.010***	0.000***	0.000***	0.002***	0.004***
M=2	0.263	–	0.003***	–	0.057*	–
M=3	–	–	–	–	–	–
Number of models tested	164	44	164	44	164	44
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$						

In Table 6, the null hypothesis for omitting the EFI, along with the linear and first order fractional specifications of EFI, are all statistically significant and thus, should be rejected. Table 6 shows us that the EFI variable should not be linearly estimated, as in our fixed effect models, but rather second order fractional polynomial specification of EFI should be used to test the relationship between SO2, PM10, and CO2 variables.

While the best specification for SO2 and CO2 variables for EFI is the second order fractional polynomial form, third order fractional polynomial form is the best specification for PM10.

However, as stated in Royston (2017), the probability of falling into type II error increases as the degrees of variables tested in the functional selection procedure are increased. For this reason, it would be appropriate to choose the most parsimonious model. From this point of view, for the PM10pc variable, the second-order fractional polynomial of the EFI variable is preferred and included in the model.

According to these test results, it can be said that fixed effect estimations, may yield biased results due to functional form misspecification. Therefore, the fractional polynomial fixed effect estimation results we estimated based on the test results above are given in Table 7 below.

Table 7
Second Order Fractional Polynomial Estimation

	SO2	PM10	CO2
EFI-1	0.036*	-1474.680***	-303.435**
	(0.018)	(447.972)	(126.274)
EFI-2	-0.014*	868.063***	178.526**
	(0.007)	(279.143)	(80.125)
GDP pc	0.251	-0.456	0.061
	(0.919)	(0.406)	(0.098)
Electric Cons. pc	2.199**	0.387	0.636***
	(1.017)	(0.370)	(0.112)
Population Density	0.889	0.869	-0.362*
	(1.138)	(0.603)	(0.199)
Constant	-33.283***	-9.623**	-3.396*
	(9.872)	(4.190)	(1.747)
R-squared within	0.793	0.301	0.778
F stat.	44.420	14.700	71.167
prob.	0.000	0.000	0.000
Number of obs.	867	867	816
Heteroskedasticity and autocorrelation robust standart errors are in parentheses.			
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Year dummies are not reported.			

Since the EFI is added to the model as a second order fractional polynomial, two variables included to the model additionally as EFI-1 and EFI-2. Both variables are significant for SO2pc, PM10pc and CO2pc. Graphs showing predicted values and observations for estimated fractional polynomial models for EFI are given below.

According to these graphs, it is seen that the EFI variable is not only significant, but also exhibits a similar structure to the inverted U-shaped for US. This is especially evident for the SO2 variable.

Overall, the results indicate no evidence of a relationship between GCI and air pollution at the regional level. This finding seems not compatible with the findings of Mealy and Teytelboym (2020). In addition, a non-linear relationship found between EFI and air pollution. Moreover, this relationship is in the form of an

inverted U shape for SO₂. Dinda (2004) states that inverted U shape is significant between GDP and especially SO₂ and PM₁₀, but controversial for CO₂ data. He makes this assertion with reference to Holtz-Eakin and Selden (1995), Roberts and Grimes (1997) and Dinda (2001). Our results contain similar findings for the EFI variable as Dinda (2004) put forward for GDP. In the same vein, Pata (2020), in his study for US, found that there is an inverted U shape between ECI and CO₂.

Concluding Remarks

Institutional and public environmental awareness stimulates global demand for environmental products and services. Governments and corporations are getting more sensitive against ecologically hazardous production. For instance, recently, the European Union agreed on the European Green Deal (EGD) which promotes environmentally friendly product markets and sets new product standards to eliminate the adverse effects of environmentally hazardous production.

We attempt to analyze the nexus between green production and environmental quality by exploiting sub-national data for US States. The analysis consists of two stages. First, we developed a green product complexity index dataset for each state. Later, environmental data and green and overall product complexity indices are estimated by fixed effect and the fractional polynomial regression method, which allows more flexible functional forms.

We find that higher green complexity index levels have an insignificant effect on emission levels in the US States. Contrary to Mealy and Teytelboym (2020), our findings indicate that exporting more sophisticated green products does not yield a better air quality. This may be due to the current green product classifications which fail to incorporate the production and end-use stages of goods or services of green-labeled products. Thus, green-labeled products may have adverse environmental effects from their production to final consumption.

In contrast to the GCI, findings suggest that economic complexity index which includes all production regardless of green or non-green classifications has significantly reduces the Sulfur dioxide, Particulate Matter 10, and CO₂ levels. In line with the existing literature (e.g., Neagu, 2019; Chu, 2020; and Pata, 2020), we find an inverted U-shape relationship between EFI and emission levels particularly for SO₂. This paper extends the literature in many folds. First, we provide a new dataset (i.e., green product complexity index) for each US states. The GCI data can be used for future research on green production. Furthermore, we outline the link between green production and environmental quality at the sub-national level. Sub-national analysis provides more robust estimation in environmental studies as the significant differences between emission measurement methods across countries create cross-country data inconsistency.

Although subnational analysis offers a more homogeneous environment for researchers compared to cross-country studies, unfortunately they do not allow for a completely homogeneous research sample. In addition, it should be kept in mind that subnational studies are less generalizable. For this reason, more subnational studies for different countries are needed for a more reliable results.

Declarations

Author contributions: All authors have made significant contributions to this study.

Çınar, İbrahim Tuğrul: Conceptualization, supervision, data curation, investigation, writing – review & editing

Korkmaz, İlhan: Conceptualization, methodology, analysis, software, writing – review & editing

Şişman, Muhammet Yunus: Conceptualization, literature review, writing – review & editing

Data availability: This study used the secondary data. Thus, all the data information with details is available in the “Data” section.

Ethics Declarations

Ethics approval: Not applicable

Consent to participate: Not applicable

Consent to publish: Not applicable

Conflict of interest: All authors declare that they have no conflict of interest

Funding: Not applicable

References

1. Albeaik S, Kaltenberg M, Alsaleh M, Hidalgo CA (2017) Improving the economic complexity index. arXiv preprint arXiv:1707.05826.
2. Andrew D, Thompson R, Morris R, Pellegrino C (2001) Environmental Goods and Services: The Benefits of Further Global Trade Liberalisation. OECD.
3. Aslanidis N (2009) Environmental Kuznets curves for carbon emissions: A critical survey.
4. Awaworyi Churchill S, Inekwe J, Ivanovski K, Smyth R (2020) The Environmental Kuznets Curve across Australian states and territories. *Energy Economics* 90. doi:10.1016/j.eneco.2020.104869
5. Balassa B (1965) Trade liberalisation and “revealed” comparative advantage 1. *The manchester school* 33(2) 99-123.
6. Boleti E, Garas A, Kyriakou A, Lapatinas A (2021) Economic Complexity and Environmental Performance: Evidence from a World Sample. *Environmental Modeling Assessment* 26(3) 251-270.
7. Caldarelli G, Cristelli M, Gabrielli A, Pietronero L, Scala A, Tacchella A (2012) A network analysis of countries’ export flows: firm grounds for the building blocks of the economy. *PloS one* 7(10) e47278.
8. Can M, Gozgor G (2017) The impact of economic complexity on carbon emissions: evidence from France. *Environ Sci Pollut Res Int* 24(19) 16364-16370. doi:10.1007/s11356-017-9219-7

9. Carson RT (2010) The Environmental Kuznets Curve: Seeking Empirical Regularity and Theoretical Structure. *Review of Environmental Economics and Policy* 4(1) 3-23. doi:10.1093/reep/rep021
10. Chu LK (2020) Economic structure and environmental Kuznets curve hypothesis: new evidence from economic complexity. *Applied Economics Letters* 1-5. doi:10.1080/13504851.2020.1767280
11. Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, Pietronero L (2013) Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one* 8(8) e70726.
12. De Groot HLF, Withagen CA, Minliang Z (2004) Dynamics of China's regional development and pollution: an investigation into the Environmental Kuznets Curve. *Environment and Development Economics* 9(4) 507-537. doi:10.1017/s1355770x0300113x
13. Dinda S (2004) Environmental Kuznets curve hypothesis: a survey. *Ecological economics* 49(4) 431-455.
14. Dogan B, Saboori B, Can M (2019) Does economic complexity matter for environmental degradation? An empirical analysis for different stages of development. *Environ Sci Pollut Res Int* 26(31) 31900-31912. doi:10.1007/s11356-019-06333-1
15. Gabrielli A, Cristelli M, Mazzilli D, Tacchella A, Zaccaria A, Pietronero L (2017) Why we like the ECI+ algorithm. arXiv preprint arXiv:1708.01161.
16. Gala P, Rocha I, Magacho G (2018) The structuralist revenge: economic complexity as an important dimension to evaluate growth and development. *Brazilian Journal of Political Economy* 38(2) 219-236.
17. Gaulier G, Zignago S (2010) Baci: international trade database at the product-level (the 1994-2007 version) Retrieved from: https://mpra.ub.uni-muenchen.de/31398/1/MPRA_paper_31398.pdf
18. Grossman GM, Krueger AB (1991) Environmental impacts of a North American free trade agreement (0898-2937) Retrieved from https://www.nber.org/system/files/working_papers/w3914/w3914.pdf
19. Hidalgo CA, Hausmann R (2009) The building blocks of economic complexity. *Proceedings of the national academy of sciences* 106(26) 10570-10575.
20. Lapatinas A, Garas A, Boleti E, Kyriakou A (2019) Economic complexity and environmental performance: Evidence from a world sample. Retrieved from https://mpra.ub.uni-muenchen.de/92833/1/MPRA_paper_92833.pdf
21. List JA, Gallet CA (1999) The environmental Kuznets curve: does one size fit all? *Ecological economics* 31(3) 409-423.
22. Meadows DH, Meadows DL, Randers J, Behrens WW (1972) *The limits to growth*. New York 102(1972) 27.
23. Mealy P, Teytelboym A (2020) Economic complexity and the green economy. *Research Policy*. doi:10.1016/j.respol.2020.103948
24. Morrison G, Buldyrev SV, Imbruno M, Arrieta OAD, Rungi A, Riccaboni M, Pammolli F (2017) On economic complexity and the fitness of nations. *Scientific Reports* 7(1) 1-11.

25. Neagu O (2019) The Link between Economic Complexity and Carbon Emissions in the European Union Countries: A Model Based on the Environmental Kuznets Curve (EKC) Approach. *Sustainability* 11(17) doi:10.3390/su11174753
26. Neagu O, Teodoru M (2019) The Relationship between Economic Complexity Energy Consumption Structure and Greenhouse Gas Emission: Heterogeneous Panel Evidence from the EU Countries. *Sustainability* 11(2) doi:10.3390/su11020497
27. Operti FG, Pugliese E, Andrade Jr JS, Pietronero L, Gabrielli A (2018) Dynamics in the Fitness-Income plane: Brazilian states vs World countries. *PloS one* 13(6) e0197616.
28. Panayotou T (1993) Empirical tests and policy analysis of environmental degradation at different stages of economic development. ILO Working Papers 992927783402676 International Labour Organization.
29. Pata UK (2021) Renewable and non-renewable energy consumption economic complexity CO2 emissions and ecological footprint in the USA: testing the EKC hypothesis with a structural break. *Environ Sci Pollut Res Int* 28(1) 846-861. doi:10.1007/s11356-020-10446-3
30. Royston P, Altman DG (1994) Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Applied Statistics* 43:429-67
31. Royston P, Ambler G, Sauerbrei W (1999) The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*: 28:964-74
32. Royston P (2017) Model selection for univariable fractional polynomials. *The Stata Journal* 17(3) 619-629.
33. Sauvage J (2014) The Stringency of Environmental Regulations and Trade in Environmental Goods. OECD Trade and Environment Working Papers 2014/03. OECD.
34. Selden TM, Song D (1994) Environmental quality and development: is there a Kuznets curve for air pollution emissions? *Journal of Environmental Economics and management* 27(2) 147-162.
35. Servedio VD, Buttà P, Mazzilli D, Tacchella A, Pietronero L (2018) A new and stable estimation method of country economic fitness and product complexity. *Entropy* 20(10) 783.
36. Shafik N, Bandyopadhyay S (1992) Economic growth and environmental quality: time series and cross section evidence. Policy research working paper N° WPS904 World Bank.
37. Steenblik R (2005) Environmental Goods: A Comparison of the APEC and OECD Lists. OECD Trade and Environment Working Paper No. 2005-04. OECD.
38. Stern DI, Common MS, Barbier EB (1996) Economic growth and environmental degradation: the environmental Kuznets curve and sustainable development. *World development* 24(7) 1151-1160.
39. Swart J, Brinkmann L (2020) Economic Complexity and the Environment: Evidence from Brazil. In: Leal Filho W, Tortato U, Frankenberger F. (eds) *Universities and Sustainable Communities: Meeting the Goals of the Agenda 2030*. World Sustainability Series. Springer Cham. https://doi.org/10.1007/978-3-030-30306-8_1.

40. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L (2012) A new metrics for countries' fitness and products' complexity. *Scientific Reports* 2 723.
41. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L (2013) Economic complexity: conceptual grounding of a new metrics for global competitiveness. *Journal of Economic Dynamics and Control* 37(8) 1683-1691.
42. Yilanci V, Pata UK (2020) Investigating the EKC hypothesis for China: the role of economic complexity on ecological footprint. *Environ Sci Pollut Res Int* 27(26) 32683-32694. doi:10.1007/s11356-020-09434-4

Figures

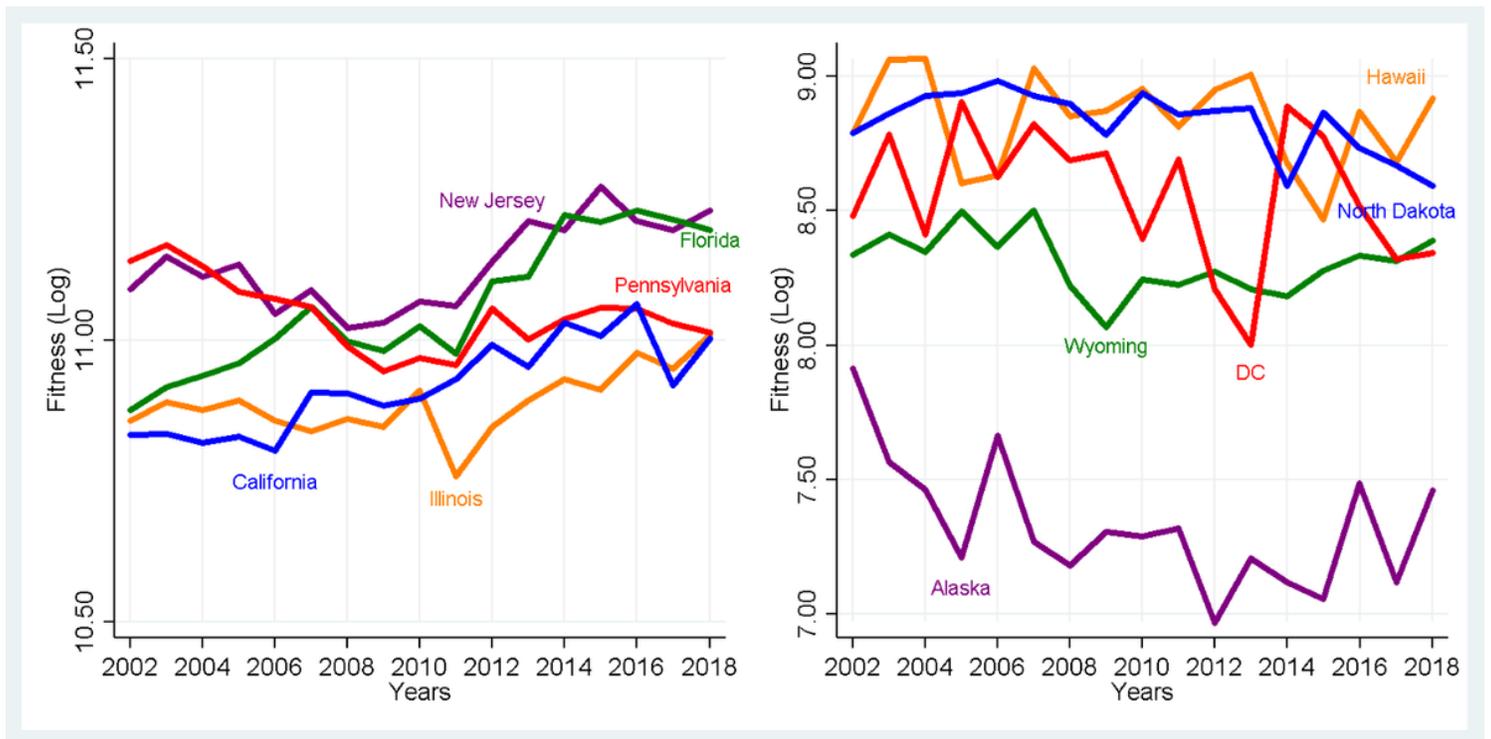


Figure 1

States with the highest (left panel) and lowest (right panel) average EFI values (2002-2018)

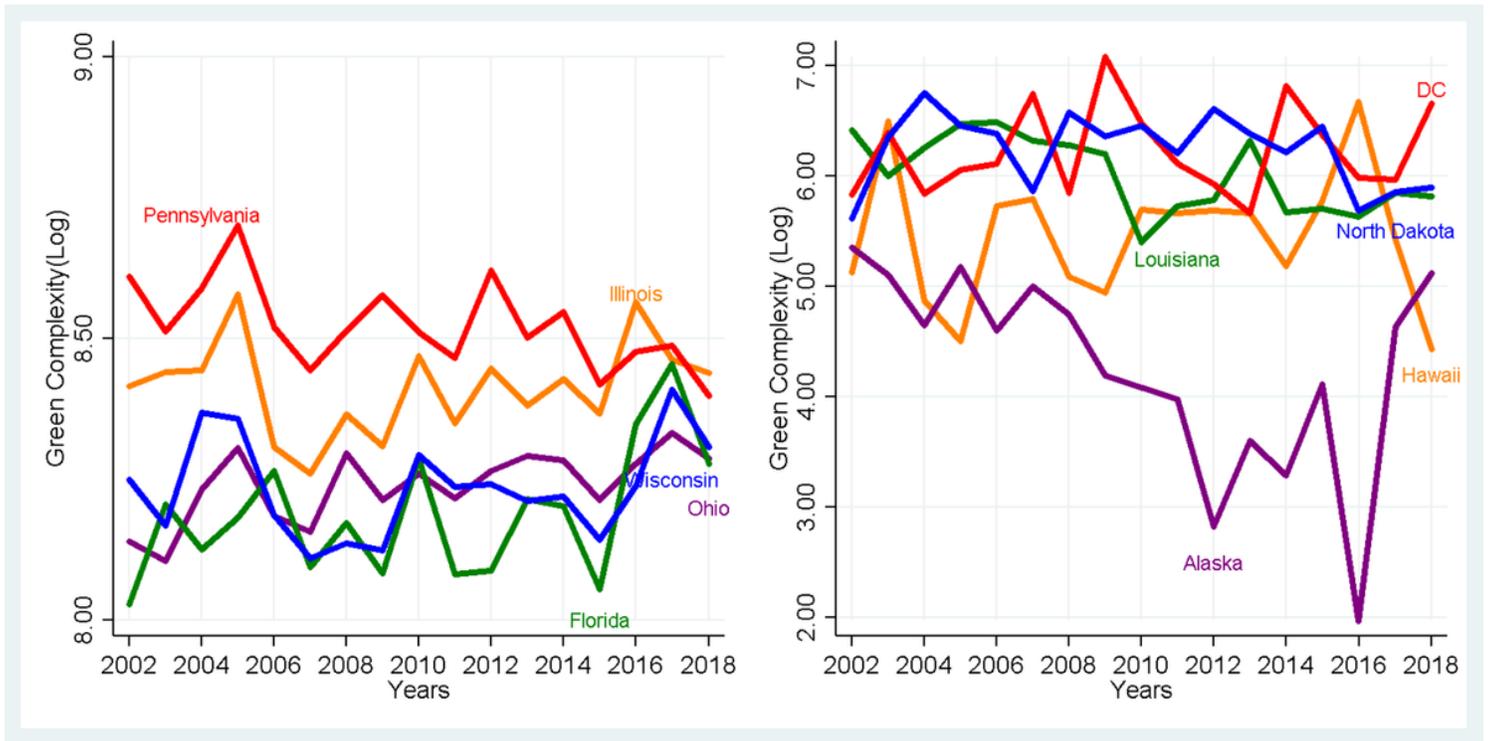


Figure 2

States with the highest (left panel) and lowest (right panel) average GCI values (2002-2018)

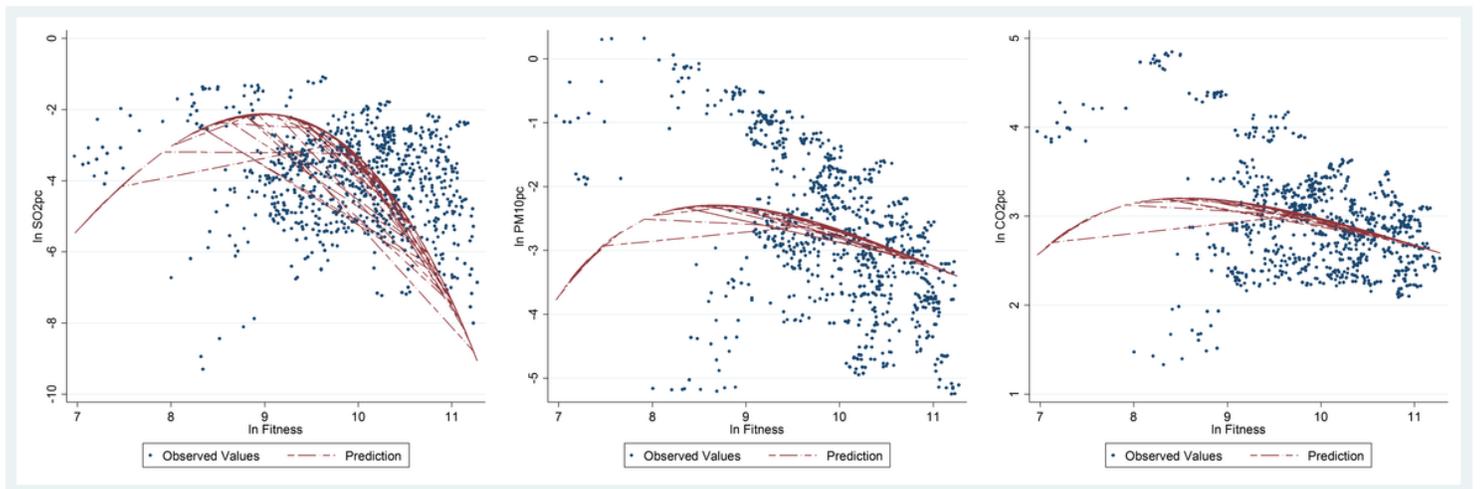


Figure 3

Observed and Predicted Values for EFI