ISSN 2004-2965 Energy Proceedings, Vol. 20, 2021

International Conference on Applied Energy 2021
Nov. 29 - Dec. 5, 2021, Thailand/Virtual
Paper ID: 523

Compositional Linear Model Clarifies Country-Level Energy Mix Drivers

Shaofang Xue[1], Ying Fan[2*]

1 School of Economics and Management, Beihang University

2 School of Economics and Management, Beihang University (Corresponding Author)

ABSTRACT

The energy mixes, which describe energy consumption structure by fuel type, are complex compositional data that cannot directly apply the multivariate statistical methods. Consequently, we apply a special compositional linear regression model to study how the country-level energy mix is influenced. The main findings of 45 countries by income group over 1990-2019 are: 1) The dependence of energy use on energy resource for each fuel type is confirmed in upper-middle-income countries. 2) Non-fossil use is driven by hydroelectricity resources and economic level in high-come and upper-middle-income countries, and by oil resources in lower-middle-income and low-income countries; 3) The energy transition as economy grows presents two typical patterns: high-income and upper-middle-income countries shift to non-fossils and natural gas, while the lower-middle-income and low-income shift to coal and oil.

Keywords: energy mix, compositional linear model, energy resource endowment, income group

NONMENCLATURE

Abbreviations	
EM	Energy Mix
ERE	Energy Resource Endowment
PS	Population Size
EDL	Economic Development Level
ilr	Isometric Log-Ratio
mtoe	Million Tonnes of Oil Equivalent

1. INTRODUCTION

At present, the energy transition towards a low-carbon and clean direction has become the prevailing trend of the energy systems of all countries in the world. And the low-carbon performance is directly reflected in energy mix (EM). Therefore, analysing the influencing factors of energy mix could help suggest personalized energy transition paths to different countries.

From the perspective of control theory, Fig. 1 depicts the energy system with EM as the output, energy resource endowment (ERE) as the input. EM is compositional energy consumption structure by fuel type. For instance, the global share of oil, natural gas, coal, hydroelectricity, nuclear energy and other renewable energy in 2020 is 29%, 24%, 27%, 3%, 5%, and 12% respectively (Enerdata, 2021). ERE refers to the amount of energy resources possessed by a country and can be measured in various ways. Besides, among the numerous controlled conditions of the energy system, population, economy (i.e. affluence) and technology are generally selected as vital factors in models that study environmental impacts, such as IPAT and STIRPAT (York et al, 2003).

Existed research on the country-level (or national-level) energy mix focuses on the numerical proportion of each fuel and handle it separately in the model (Burke, 2013; Csereklyei et al., 2017), ignoring the relative information contained in. And there is little literature that involves compositional data analysis in studying how the country-level EM is influenced. By adopting a compositional linear regression model, we aim at analysing the impacts of ERE, population size (PS) and

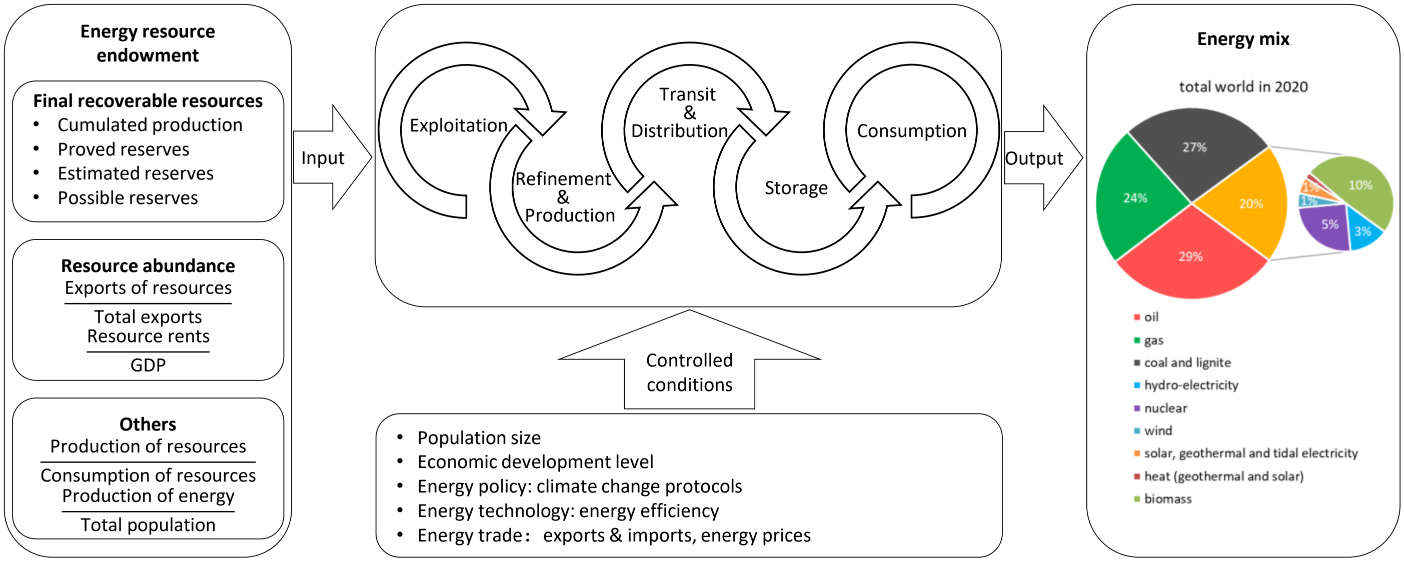

Fig. 1. Schematic diagram of the energy system

economic development level (EDL) on the country-level EM. Our findings provide compositional results and contribute to the energy transition rule by income group, thus favouring the achievement in country-level carbon targets.

2. METHODS & DATA

2.1 Theory

2.1.1 Concepts of composition data

A D-part composition x is defined in a D-dimensional simplex space $S^D=\{x=(x_1,x_2,\dots,x_D):x_i\geq 0,\ \sum_{i=1}^{D}x_i=1\}$. And the constraints are that all components of a composition are non-negative and add up to one. Our **EM**=(*oilcon*, *gascon*, *coalcon*, *nonfossil*) consists of four parts, namely primary energy consumption of oil, natural gas, coal and non-fossil energy. And the sum of *oilcon*, *gascon*, *coalcon*, *nonfossil* within each **EM** equals 100%.

Aitchison (1986) introduced several compositional operations that replaced the conventional geometric operations in Euclidean space. The *Closure* operation $C(x)=\left[\frac{x_1}{\sum_{i=1}^{D}x_i},\frac{x_2}{\sum_{i=1}^{D}x_i},\dots,\frac{x_D}{\sum_{i=1}^{D}x_i}\right]$ ensures that all components within a composition x sum up to 100% and preserves the ratios between the components. The *Perturbation* operation $x\oplus y=C\left[x_1\cdot y_1,x_2\cdot y_2,\dots,x_D\cdot y_D\right]$ functions as compositional sum of two compositions x and y. The *Powering* operation $\lambda\odot x=C\left[x_1^{\lambda},x_2^{\lambda},\dots,x_D^{\lambda}\right]$ is a compositional scalar multiplication of composition x and constant λ.

In order to apply multivariate statistical methods, we must eliminate the troublesome constraints of non-negativity and sum-of-one. Here we use *Isometric Log-Ratio* (ilr) transformation (see Eq. (1)) to convert compositional data from Simplex space S^D to (D-1)-dimensional Euclidean space R^{D-1} (Van den Boogaart, Tolosana-Delgado, 2013).

$$ilr(x)=ln(x)\cdot V^t \qquad (1)$$

The D×(D-1)-element matrix V^t given by Eq. (2) is a quasi-orthonormal basis that includes (D-1) linearly independent columns.

$$V^t=\begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \cdots & -\frac{1}{\sqrt{(D-1)\times(D-2)}} & -\frac{1}{\sqrt{D\times(D-1)}} \\ \sqrt{\frac{1}{2}} & -\frac{1}{\sqrt{6}} & \cdots & -\frac{1}{\sqrt{(D-1)\times(D-2)}} & -\frac{1}{\sqrt{D\times(D-1)}} \\ 0 & \sqrt{\frac{2}{3}} & \cdots & -\frac{1}{\sqrt{(D-1)\times(D-2)}} & -\frac{1}{\sqrt{D\times(D-1)}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\frac{D-2}{D-1}} & -\frac{1}{\sqrt{D\times(D-1)}} \\ 0 & 0 & \cdots & 0 & \sqrt{\frac{D-1}{D}} \end{pmatrix} \qquad (2)$$

Similarly, the inverse ilr transformation given by Eq. (3) calculates the original composition by the aforementioned matrix V.

$$x=C(e^{ilr(x)\cdot V}) \qquad (3)$$

2.1.2 Compositional linear regression model

In our compositional linear regression model (see Eq. (4)), *oilres*, *gasres*, *coalres*, *hydres* are EREs of oil, natural gas, coal and hydroelectricity respectively; PS is total population; EDL is measured by GDP per capita; $\boldsymbol{\alpha}$, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$, $\boldsymbol{\beta_4}$, $\boldsymbol{\beta_5}$, $\boldsymbol{\beta_6}$ are unknown compositional parameters; and $\boldsymbol{\varepsilon}$ are compositional residuals that have a normal distribution on the simplex with null compositional expectation and constant variance.

$$\mathbf{EM}=\boldsymbol{\alpha}\oplus oilres\odot\boldsymbol{\beta}_1\oplus gasres\odot\boldsymbol{\beta}_2\oplus coalres\odot\boldsymbol{\beta}_3 \oplus hydres\odot\boldsymbol{\beta}_4\oplus \mathrm{PS}\odot\boldsymbol{\beta}_5\oplus \mathrm{EDL}\odot\boldsymbol{\beta}_6\oplus\boldsymbol{\varepsilon} \quad (4)$$

To estimate the parameters in Eq. (4), we rewrite it as a multivariate linear regression model in Eq. (5).

$$\mathrm{ilr}(\mathbf{EM})=\mathrm{ilr}(\boldsymbol{\alpha})+oilres\cdot\mathrm{ilr}(\boldsymbol{\beta}_1)+gasres\cdot\mathrm{ilr}(\boldsymbol{\beta}_2)+coalres\cdot\mathrm{ilr}(\boldsymbol{\beta}_3) +hydres\cdot\mathrm{ilr}(\boldsymbol{\beta}_4)+\mathrm{PS}\cdot\mathrm{ilr}(\boldsymbol{\beta}_5)+\mathrm{EDL}\cdot\mathrm{ilr}(\boldsymbol{\beta}_6)+\mathrm{ilr}(\boldsymbol{\varepsilon}) \quad (5)$$

Generally, the intercept $\boldsymbol{\alpha}$ is interpreted as the expected composition when all independent variables are zero; and the slope $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$, $\boldsymbol{\beta}_4$, $\boldsymbol{\beta}_5$, $\boldsymbol{\beta}_6$ is interpreted as the perturbation when the independent variables increase one unit (Van den Boogaart, Tolosana-Delgado, 2013). A component of the compositional coefficient less than 1/D decreases the corresponding component of **EM**, and vice versa.

2.2 Data

Though the total amount of energy consumption in a country is irrelevant to the analysis of its EM, we still carefully choose those countries that rank top in the world in 2018 for total primary energy consumption. The justification for doing so lies in that those countries cover more than 90% of global primary energy consumption in recent years. Furthermore, the sample selected is comparable to the whole world in terms of income group, geological region and international organization, thus being quite representative.

The annual energy and economic data of 45 countries from 1990 to 2019 are downloaded from Enerdata - Global Energy & CO2 Data (Enerdata, 2021). We refer to World Bank Open Data and divide sample countries into three income groups (World Bank, 2021). The units of the components of **EM**, with their primary consumption measured in mtoe (million tonnes of oil equivalent), is percent (%); the units of *oilres*, *gasres*, *coalres*, *hydres* are thousand barrels, thousand cubic meters, thousand tonnes and thousand kWh respectively, measured by proved energy reserves per capita; the unit of PS is thousand people; and the unit of EDL is thousand US dollars at constant price and exchange rate (2015) per capita.

The data are processed using R (Kabacoff, 2015) by three steps: 1) Missing value analysis and imputation with global linear trend; 2) ilr transformation for compositional **EM**; 3) Z-score Standardization for all the independent variables.

3. RESULTS & DISCUSSION

The brief descriptive statistics are given in Table 1. EM varies apparently among different income groups. Oil is the main energy source (≥35%) in both high-income and upper-middle-income groups, followed by natural gas (≥25%), coal (≥15%). Meanwhile, lower-middle-income and low-income group use non-fossil energy (36.18%) most, and primarily consume natural gas (29.12%) and oil (23.83%) among the fossil energy. EREs of all fuel types are the highest in high-income group, and so is EDL. The total population sees the highest in lower-

Table 1 Summary statistics by income group

Group	High-income		Upper-middle-income		Lower-middle-income & Low-income	
Country	United States, Japan, Germany, Canada, South Korea, France, Saudi Arabia, United Kingdom, Italy, Australia, Spain, Poland, United Arab Emirates, Netherlands, Belgium, Sweden, Qatar, Czechia, Chile		China, Russia, Brazil, Iran, Indonesia, Mexico, Turkey, Thailand, South Africa, Malaysia, Argentina, Kazakhstan, Iraq, Venezuela, Colombia		India, Nigeria, Pakistan, Ukraine, Egypt, Vietnam, Algeria, Philippines, Uzbekistan, Ethiopia, Bangladesh	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
oilcon	36.29	12.64	39.90	18.93	23.83	13.89
gascon	27.89	21.10	25.50	17.32	29.12	26.42
coalcon	17.76	16.50	19.98	23.75	10.87	13.47
nonfossil	18.05	16.33	14.62	12.68	36.18	29.61
oilres	3.05	7.99	0.97	1.95	0.06	0.11
gasres	1068.49	4611.09	82.84	120.90	25.42	38.16
coalres	0.40	1.35	0.26	0.45	0.08	0.20
hydres	1.82	4.13	1.68	1.84	0.69	0.92
ps	50488.70	65056.17	158129.75	309012.42	181637.02	307573.07
edl	33.00	14.47	7.40	5.17	1.73	0.99

middle-income and low-income countries.

We implement the compositional data analysis by package "compositions" in RStudio. The estimated ilr transformed coefficients of Eq. (5) are given in Table 2, and the inverse compositional coefficients are illustrated in Fig. 2.

3.1 High-income group

The resource dependency is a close relationship between ERE and EM of the same fuel type, where each ERE type encourages the same type of energy consumption. In high-income countries, the resource dependency is found in oil, coal and hydroelectricity, but not in natural gas. The growth in oil resources increases the consumption of itself (32.9%) and natural gas (41.8%), and decreases that of coal (20.6%) and non-fossil energy (4.7%). Oppositely, the growth in natural gas resources increases the consumption of coal (28.6%) and non-fossil energy (36.8%), and decreases that of oil (15.7%) and itself (18.8%), indicating that the natural gas use might not rely on domestic resources, but relate to international trade flow. A boost in coal resources promotes its consumption (33.2%), and simultaneously prevents the consumption of other fuels. Especially, the abundance of coal resources could inhibit the use of non-fossil energy (19.3%). Similar to the advantage that coal resources bring to coal use, a boost in hydroelectricity resources promotes the corresponding non-fossil energy use (36.7%) as well. However, hydroelectricity resources have the strongest restraint on natural gas use (21.9%), which seems to be a rival of natural gas.

Besides, population growth improves the consumption of oil (25.4%) and coal (28.8%), and lower that of natural gas (23.4%) and non-fossil energy (22.4%); by contrast, economic growth enhances the consumption of natural gas (27.9%) and non-fossil energy (36.2%), and reduces that of oil (20.2%) and coal (15.7%).

On the whole, high-income countries with more oil (coal) resources and population tend to consume more oil (coal), which indicates a dependence of energy use on its EREs. As for non-fossil drivers, relatively high natural gas and hydroelectricity resources help develop non-fossils. This is because natural gas can solve the intermittency of renewable energy and hydroelectricity is the main force of non-fossil energy. On the contrary,

Table 2 ilr transformed coefficients by income group

Group	High-income			Upper-middle-income			Lower-middle-income & Low income		
	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3
oilres	0.171***	−0.481***	−1.620***	−0.404***	0.084*	−0.211***	−0.453***	−1.285***	1.916***
	(0.029)	(0.044)	(0.082)	(0.027)	(0.046)	(0.050)	(0.110)	(0.138)	(0.093)
gasres	0.127***	0.417***	0.512***	0.415***	−0.850***	−0.614***	0.885***	0.409***	−2.502***
	(0.028)	(0.042)	(0.079)	(0.025)	(0.043)	(0.047)	(0.108)	(0.136)	(0.091)
coalres	−0.006	0.273***	−0.277***	0.090***	0.987***	−0.334***	0.255***	0.488***	0.019
	(0.023)	(0.035)	(0.065)	(0.023)	(0.039)	(0.043)	(0.064)	(0.080)	(0.054)
hydres	−0.184***	0.055	0.485***	−0.101***	0.027	0.558***	0.214***	0.631***	−0.161***
	(0.023)	(0.035)	(0.066)	(0.024)	(0.042)	(0.045)	(0.071)	(0.090)	(0.060)
PS	−0.056**	0.136***	−0.121*	−0.130***	0.536***	0.135***	−0.148**	0.692***	−0.326***
	(0.025)	(0.038)	(0.070)	(0.022)	(0.038)	(0.041)	(0.065)	(0.082)	(0.055)
EDL	0.231***	−0.338***	0.485***	0.380***	−0.698***	0.176***	−0.158*	0.326***	−0.945***
	(0.028)	(0.043)	(0.079)	(0.026)	(0.044)	(0.048)	(0.086)	(0.108)	(0.073)
Constant	−0.352***	−0.585***	−0.832***	−0.499***	−0.857***	−0.628***	−0.124**	−1.349***	0.376***
	(0.022)	(0.034)	(0.063)	(0.021)	(0.036)	(0.039)	(0.060)	(0.075)	(0.050)
Observations	570	570	570	450	450	450	330	330	330
R^2	0.427	0.378	0.475	0.579	0.812	0.536	0.277	0.535	0.801
Adjusted R^2	0.421	0.371	0.470	0.573	0.810	0.529	0.264	0.526	0.798
Residual Std. Error	0.531	0.801	1.493	0.445	0.764	0.832	1.083	1.361	0.916
	(df=563)	(df=563)	(df=563)	(df=443)	(df=443)	(df=443)	(df=323)	(df=323)	(df=323)
F Statistic	70.043***	57.024***	85.066***	101.338***	319.514***	85.190***	20.676***	61.950***	217.116***
	(df=6; 563)	(df=6; 563)	(df=6; 563)	(df=6; 443)	(df=6; 443)	(df=6; 443)	(df=6; 323)	(df=6; 323)	(df=6; 323)

Note: 1. * p<0.1; ** p<0.05; *** p<0.01.

2. Standard errors are in parentheses.

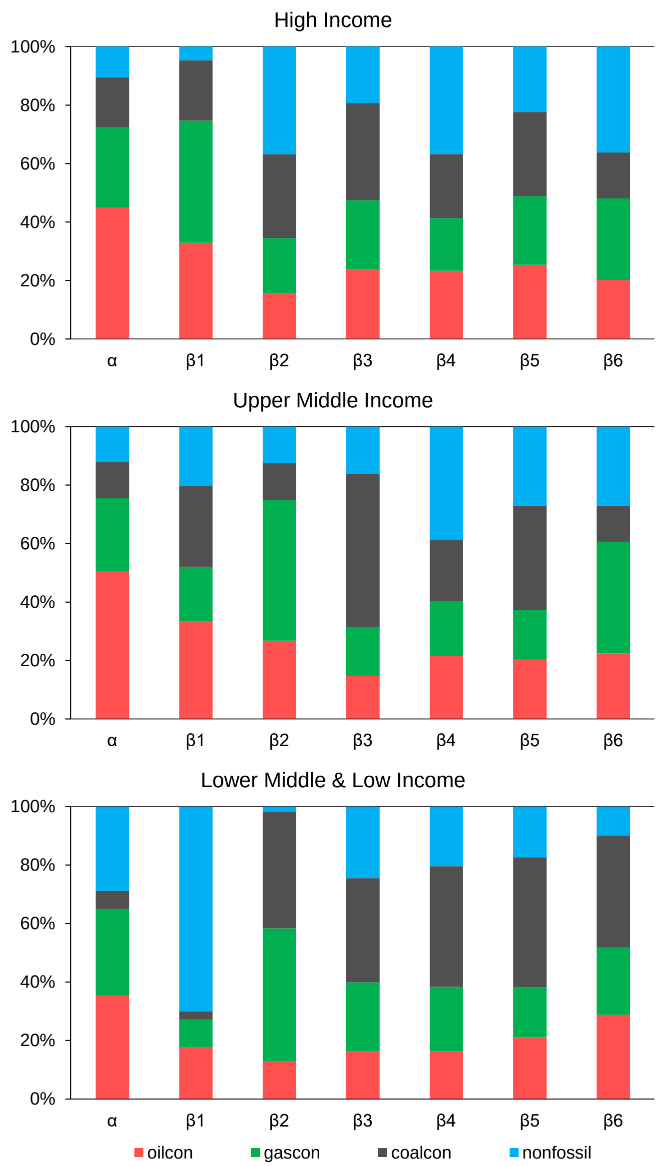

Fig. 2. Estimated results of inverse ilr transformed compositional coefficients by income group. The horizontal axis is the compositional coefficient; the vertical axis is the estimated impact measured in per cent; and each bar is the estimation of one component in a compositional coefficient. By definition, a bar less than a quarter decreases its proportion in EM, and vice versa.

abundant oil and coal resources hinder the EM transition towards non-fossil energy. Particularly, high-income countries that have smaller populations and more developed economies are more likely to use less coal and more non-fossils.

3.2 Upper-middle-income group

There is a resource dependency among upper-middle-income countries for oil, natural gas, coal and non-fossil energy (33.3%, 48.1%, 52.6%, and 38.8% respectively). Oil use is positively affected by the resources of itself and natural gas (26.8%), and negatively affected by coal resources (14.7%), hydroelectricity resources (21.6%), PS (20.3%) and EDL (22.4%). This suggests a linkage relation between oil use and natural gas resources, as well as a competitive relation between coal and oil in upper-middle-income countries. Natural gas use is positively affected by the resources of itself and EDL (38.3%), negatively affected by other EREs and PS (16.9%). This intimates that upper-middle-income countries with ample natural gas reserves, developed economy, and small populations are apt to use more natural gas. Coal use is positively affected by oil resources (27.7%) and PS (35.8%), and negatively affected by natural gas resources (12.7%), hydroelectricity resources (20.8%) and EDL (12.4%). Hence, population growth causes upper-middle-income countries consume a higher proportion of coal, yet economic growth leads to a minimization in coal utilization; natural gas and hydroelectricity stand out against coal as a cleaner alternative energy.

As for fossil fuel promoters, those upper-middle-income countries with more natural gas resources use more oil; those with ample natural gas reserves, small populations and developed economy use more natural gas; those with more coal resources and larger population tend to use more coal.

In addition, non-fossil energy use is positively affected by PS (27.0%) and EDL (27.0%), and negatively affected by the resources of oil (20.3%), natural gas (12.5%) and coal (16.0%). On that account, population and economic growth in upper-middle-income countries could result in expansion of non-fossil energy use, while richer fossil resources draw forth smaller non-fossil energy use.

3.3 Lower-middle-income & low-income group

In lower-middle-income and low-income countries, we can only observe the resource dependency in natural gas and coal. Oil resources in lower-middle-income and low-income countries have a promoting effect on non-fossil energy (70.1%), implying their role in balancing the intermittent renewable energy as necessary as natural gas resources to non-fossil energy in high-income countries. Natural gas resources promote the consumption of itself (45.4%) and coal (40.1%), and reduce that of oil (13.0%) and non-fossil energy (1.6%). Except for oil resources, coal resources, hydroelectricity resources, PS and EDL all increase coal consumption (35.7%, 41.2%, 44.4%, 38.3% respectively), and decrease other energy use. Unexpectedly, hydroelectricity resources negatively affect non-fossil energy use (20.4%), resulting by the slightly ambiguous data in our model which misses other non-fossil energy resources such as nuclear, wind, and solar energy. Apart from coal consumption, EDL also stimulates oil consumption (28.8%). This means lower-middle-income and low-income countries are still pursuing an EM that relies on high proportion of traditional fossil energy.

4. CONCLUSIONS

To examine the driving factors of the country-level EM, we allow for ERE, PS and EDL in a compositional linear model and conclude the characterized EM evolution patterns by income group. During this process, we identify EM analysis gaps and recognise the need to apply compositional data methods. Our method treats EM as a multivariate composition rather than an ordinary univariate data. Such a consideration not only dissect the country-level EM in detail, but also produce sufficiently structured results. Moreover, we verify the resource dependency, summarize the driving factors of non-fossil and fossil utilization, provide several energy transition directions with economic growth by income group.

We draw the main conclusions as follow. First, the exact relationship between ERE and EM of the same fuel type is merely found in upper-middle-income countries. Second, with respect to the driving factors of non-fossil energy utilization, hydroelectricity resources and EDL matter both in high-income and upper-middle-income countries. Specifically, natural gas resources and PS also function in high-income and upper-middle-income countries respectively. Instead, oil resources have a dominant position in developing lower-middle-income and low-income countries' non-fossils. Third, population expansion increases coal consumption of all income groups. Last, in terms of energy transition with economic growth, high-income and upper-middle-income countries incline towards non-fossils and natural gas, whereas lower-middle-income and low-income countries incline towards coal and oil, which are generally recognised as comparatively dirty fuels.

There are a few suggestions from the standpoint of policymakers. First, it might be harder for a fossil-rich country to reduce fossil proportion in energy mix. On the one hand, a country could benefit from a plenteous domestic fossil resources and enjoy a period of prosperity; on the other hand, the same fossil-rich country could also suffer from an excessively high fossil proportion and carry a burden of pollution. When confronted with sustainable energy challenges, policymakers who are too satisfied with resource abundance might fall behind those fossil-scarce competitors. Second, it is the economic level that classify a country's energy transition rule. Thus, policymakers should take measures appropriate to their particular situation and national circumstances.

Notwithstanding, we accept the existing limitations that should be modified and handled with in future works: 1) Our model neglects the dynamic evolution of EM and the influencing factors from a time-varying perspective. 2) Our independent variables exclusively consider ERE, PS and EDL, which is too uncomplicated to make the model closer to reality.

ACKNOWLEDGEMENT

We acknowledge financial support from the National Natural Science Foundation of China (71690245).

REFERENCE

[1] Enerdata. Global Energy & CO2 Data. https://www.enerdata.net/research/energy-market-data-co2-emissions-database.html, Accessed on 2021-04-08.

[2] York R, Rosa EA, Dietz T. STIRPAT, IPAT and ImPACT: analytic tools for unpacking the driving forces of environmental impacts. Ecological Economics 2003;46:351-365.

[3] Burke PJ. The national-level energy ladder and its carbon implications. Environmental and Developement Economics 2013;18:484-503.

[4] Csereklyei Z, Thurner PW, Langer J, Küchenhoff H. Energy paths in the European Union: A model-based clustering approach. Energy Economics 2017;65:442–457.

[5] Aitchison J. The Statistical Analysis of Compositional Data. London, New York: Chapman and Hall; 1986.

[6] Van den Boogaart KG, Tolosana-Delgado R. Analyzing Compositional Data with R. Berlin Heidelberg: Springer; 2013.

[7] World Bank. World Bank Open Data. https://data.worldbank.org/, Accessed on 2021-04-08.

[8] Kabacoff RI. R in Action: Data Analysis and Graphics with R. 2nd ed. Shelter Island: Manning; 2015.