Journal of Physics: Complexity

PAPER • OPEN ACCESS

The degree of economic development pattern of economy

To cite this article: Yuan-Yuan Guo and Xiao-Pu Han 2024 *J. Phys. Complex.* **5** 025002

Journal of Physics: Complexity

RECEIVED
2 July 2023

REVISED
13 February 2024

ACCEPTED FOR PUBLICATION
11 March 2024

PUBLISHED
16 April 2024

PAPER

The degree of economic development pattern of economy

Yuan-Yuan Guo and Xiao-Pu Han*

Alibaba Research Center for Complexity Sciences, Hangzhou Normal University, Hangzhou 311121, People's Republic of China

* Author to whom any correspondence should be addressed.

E-mail: xp@hznu.edu.cn

Keywords: the degree of economic development pattern, similarity analysis, correlation analysis, the relative indicators

Abstract

In this article, we explore the concept and measurement of the degree of economic development pattern (DEDP) of economy, which refers to the extent to which the development of an economy can serve as a reference for other economies. Utilizing 76 macroeconomic indicators across 217 economies, the economic development paths in a standardized space of economy is compared to identify variations in DEDP through the regression analysis on the relationship between the similarity of development paths and the growth rate on gross domestic product (GDP) per capita. To measure DEDP of economy from different perspective, two types of metrics are constructed. One is the determination coefficient of regression analysis, which exhibits significant positive correlations with population size of economy, uncovering differences of development paths among economies of varying population sizes. The other type of metrics is based on the consistency on regression coefficients and effectively explains disparities among economies in the growth rate on GDP per capita, economic complexity index and economic fitness. These findings reveal the differences in development paths among different countries from the perspective of referentiality for development patterns, suggesting the potential existence of the paths with more universal meaning to economic development.

1. Introduction

'Economic development pattern' refers to the overall approach or strategy that a country or region uses to achieve economic growth, social development, and sustainable progress [1]. Every country strives to find a development path that fits its unique national conditions and can sustain long-period growth momentum. Based on their environment, resources, culture, social and economic systems, and other factors, countries have gradually explored various economic development patterns in their long-term development practices. Among them, some countries that achieved rapid growth for a long period, such as Japan after World War II, South Korea in the late twentieth century, and China in the last four decades, have developed several remarkable economic development patterns that have attracted extensive attention from scholars and been partly used for reference by other countries [2–4]. Although different countries must fully consider their own national conditions when choosing their economic development paths and patterns, and there are often failures when learning from the development patterns of other countries [5], it is noteworthy to identify a set of common features in these successful economic development patterns, such as promoting market integration [6], giving full play to government's active role [7], emphasizing infrastructure and public services [8], adopting an open economic system [9], strengthening investment in education and technology [10], and promoting innovation and scientific and technological development [11]. These shared characteristics are indicative of fundamental principles underlying economic development and constitute focal points for cross-country learning, thereby indicating the potential existence of a generalizable pattern in economic development.

At the same time, the research in development economics and growth economics also put forward a series of development theories based on the general sense, such as the neoclassical growth model [12], institutional change theory [13], endogenous technological change theory [14], market growth theory [15], trying to give the interpretation of economic development driving from a universal sense. However, there are still serious differences between these theories on several key issues [16], such as the perception of the role of the market, whether the government needs to intervene in the market, and the understanding on the source and role of economic inequality, which weakens the universality of their theories.

In recent years, a novel research approach that integrates meso- or micro-economic information into macroeconomic analysis has achieved remarkable success in exploring economic development patterns. This approach is typically driven by a substantial amount of meso- or micro-economic data, including trade indicators at the product level, patent-level technology and innovation information, and combines structural-analysis-based methods that are widely utilized in complexity sciences. This new methodology has led to the development of a series of novel macroeconomic indicators that effectively interpret economic development patterns from a meso–macro perspective, such as the economic complexity index (ECI) [17] and economic fitness [18, 19]. It has also been extended to analyze various macroeconomic issues [20–23], providing valuable insights for formulating economic policies.

This article tries to explore economic development pattern from a perspective that differs from both traditional methods of constructing macroeconomic indicators and the structural analysis-driven approach that integrates meso- or micro-economic information, and examines the concept of 'Degree of Economic Development Pattern' (DEDP). It approaches this from the perspective of quantifying the extent to which an economy's development pattern can serve as a reference for others. By establishing an economy as the benchmark, the DEDP of that benchmark economy is determined by how well the similarity economic development pattern between other economies and the benchmark can explain their respective economic development. An economy with a higher level of DEDP indicates that other economies with similar development patterns tend to have a higher level in economic development, indicating that the development pattern of this economy can serve as a more suitable reference for the development of other economies. Based on this perspective, we aggregate numerous macroeconomic indicators to construct economic development paths to explore economic development pattern, and obtain metrics of DEDP through similarity analysis on the economic development paths. The DEDP metrics constructed using this approach can capture a series of underlying characteristics of economic development patterns while maintaining a relatively low level of data dependency.

This main part of this article is organized as follows: in section 2, the actual macroeconomic indicators are first transformed to establish a comparable standardized space, and then the similarity between different economies in each indicator is defined, and the determination coefficient is computed through regression analysis to serve as a metric for DEDP. Subsequently, in section 3, the relationship between the above DEDP metric and population size is examined. Finally, in section 4, a set of DEDP metrics based on the consistency of regression coefficients is constructed, and their correlations with macroeconomic indicators pertinent to economic development are observed.

2. Similarity analysis of development paths

2.1. Data and construction of the development paths in a standardized space

The calculation of the similarity between the benchmark economy and other economies in terms of their economic development patterns is a crucial step in determining an economy's DEDP. The economic development patterns of an economy usually are hidden within its development paths. To establish metrics for an economy's DEDP based on the development paths, the process involves three fundamental steps: firstly, the development paths of each economy are constructed in a standardized space based on multiple macroeconomic indicators of economies. Subsequently, the similarities of development paths between different economies are calculated. Finally, the metrics of DEDP are determined through regression analysis that examines the relationship between economic growth and the calculated similarities, as well as the further analysis based on the results of regression analysis.

The development paths of each economy are obtained from a comprehensive transformation of multiple macroeconomic indicators. We employed a total of 76 macroeconomic indicators across 11 indicator categories such as transportation, market integration, foreign trade, military, innovation, social assets, labor force, sub-industry, human development level, urbanization, and finance, as shown in table A1 in appendix. In addition to these indicators, the data also includes per capita gross domestic product (GDP per capita) and population data of each economy. In the 76 macroeconomic indicators used in this study, the human development index was sourced from the United Nations Development Programme website (www.undp.org), while the remaining indicators were obtained from the World Bank database (www.worldbank.org). This dataset covered a total of 264 countries, regions, and organizations, with 217 being national entities or regions (they are called 'economies' in this article), and spanning a period of 61 years (from 1960 to 2020). We also collected two new emerging product-level-based macroeconomic indexes of economies: ECI and economic fitness. ECI is a measure of an economy's capacity [17]. The ECI data is sourced from the

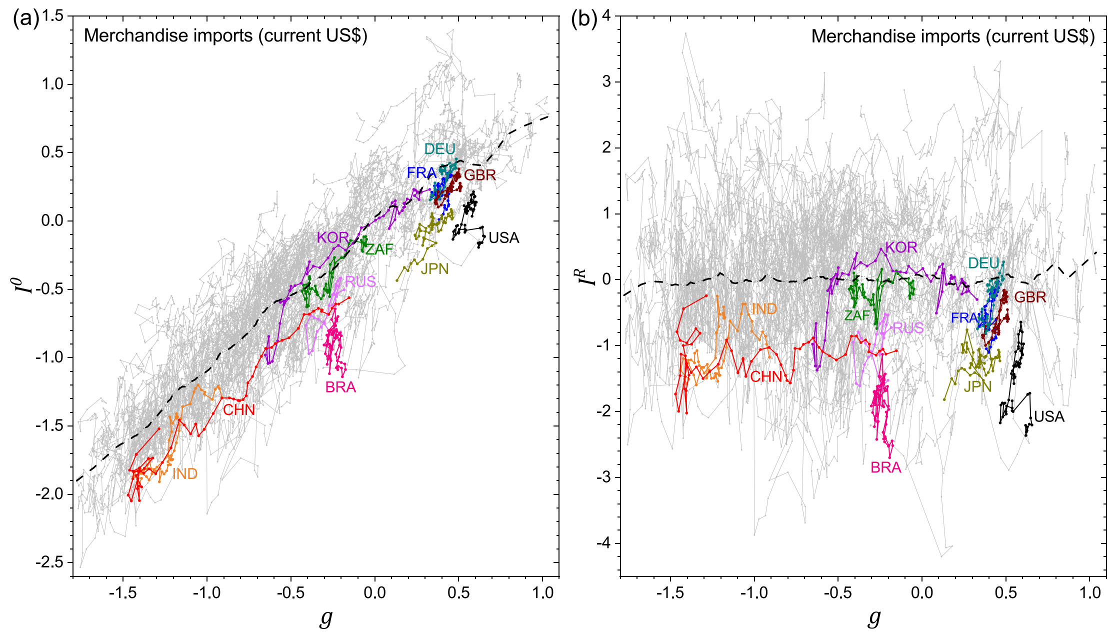

Figure 1. The trajectories of each economy in the space of economic indicator (merchandise imports (current US$) as an example) and the logarithmic relative GDP per capita g. The indicator shown in panels (a) and (b) is the original indicator and the correspond relative indicator, respectively. The color curves and the gray curves respectively correspond to representative economies and other economies, and the black dashed lines show the sliding-window-averaged curves (window width $w = 0.2$).

'Economic Complexity and International Development' project by the Harvard Growth Lab (https://atlas.cid.harvard.edu). It covers 133 countries and spans a time period of 26 years, from 1995 to 2020. Economic fitness is a metric for product complexity that is calculated by the data of product-level exports of economy and shows well predictive ability on economic growth [18, 19]. The data of economic fitness of each economy is collected from Databank of the World Bank (https://databank.worldbank.org), covering 149 economies from 1995 to 2015.

For the 76 macroeconomic indicators, we first performed the following steps as preprocessing. Firstly, among these macroeconomic indicators, if an indicator is derived from the sum of per capita indicators, it will be restored as a per capita indicator. Subsequently, we replaced each indicator with its ratio to the corresponding world average to eliminate the impact of changes in statistical scope and measurement capabilities across different years for each indicator. Given the heterogeneity of these indicators, we further use the logarithmic value of the ratio as the processed indicator value I^0 for each indicator. Similarly, for the GDP per capita of each economy, due to its heterogeneity, we also convert it into the logarithm of the ratio to the world average value, which is called the logarithmic relative GDP per capita (denoted by g), as the indicator of the economic development level of each economy.

We construct a standardized space that enables the comparison of each indicator across different economies and years within it. We plot the trajectories of each economy in the space of I^0 vs. g for every indicator. As the example shown in figure 1, numerous macroeconomic indicators exhibit a strong dependence on g, showing the strong impact of the economic development level of economy. Under such circumstances, identical values on the indicator in economies with distinctly different levels of development can conceal underlying systemic disparities. Consequently, the validity of direct comparisons based solely on the absolute value of I^0 is greatly reduced. So the first step of the construction is to eliminate the impact of economic development levels on these indicators, the method is as follows:

For an indicator (e.g. the kth indicator I_k^0), plot the trajectories of every economy in the space of I_k^0 vs. g. And then, we set a narrow sliding window with each data point as the center of the horizontal coordinate. For example, the left border and right border of the sliding window for the ith data point with coordinate $(g_i, I_k^0(g_i))$ is $g_i - w/2$ and $g_i + w/2$, where w is the width of the sliding window. The average value $\langle I_k^0 \rangle (g_i)$ of I_k^0 for all data points in the sliding window can be considered as the global expected value of I_k^0 at $g = g_i$. By using this method, we obtained the global expected value of all data points. Compared with the original indicator value $I_k^0(g_i)$ (the ith data point, say), the deviation of each data point from its global expected value $\delta I_k^0(g_i) = I_k^0(g_i) - \langle I_k^0 \rangle (g_i)$ expresses the difference to the average behavior on the relationship I_k^0 vs. g and can better reflect the systemic differences across economies with different economic development level, because the hidden characteristics in the I_k^0 of each economy covered by the impact of economic development level

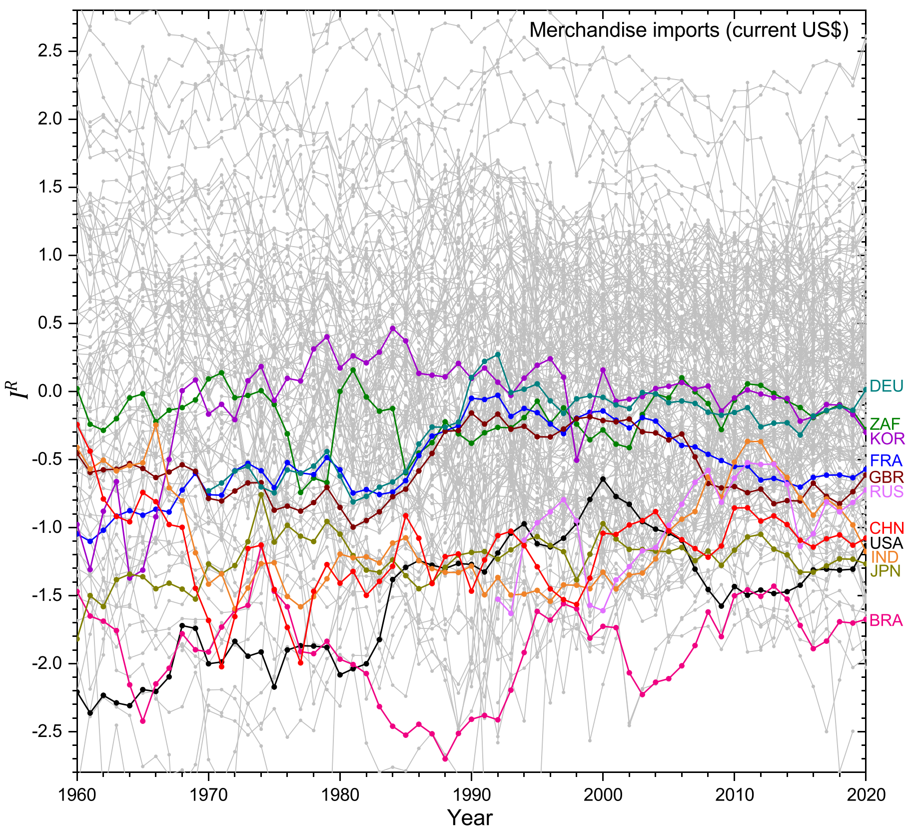

Figure 2. The curves of the relative indicator (merchandise imports (current US$) as an example) of each economy over the years (1960–2020). The color curves and the gray curves correspond to representative economies and other countries respectively.

are revealed. Several previous studies in mining urban characteristics and innovation ability have confirmed the effectiveness of the relative indicators constructed based on this type of deviation [24–27].

We further standardize this deviation for each data point within its corresponding sliding window:

$$I_k^{\mathrm{D}}(g_i) = \frac{\delta I_k^0(g_i)}{\sigma\left(I_k^0(g)\right)}, \tag{1}$$

where $\sigma(I_k^0(g))$ is the standard deviation of the value of $I_k^0(g)$ of all data points in the sliding window. The curve of I_k^{D} of each economy still contains the impact of the fluctuation in the indicator's global level over the years. To exclude this impact, for each economy (economy i, say), the difference between I_k^{D} and the global average value at each year (year Y, say) is calculated:

$$I_k^{\mathrm{R}}(Y)|_i = I_k^{\mathrm{D}}(Y)|_i - \langle I_k^{\mathrm{D}}(Y)\rangle, \tag{2}$$

where $\langle I_k^{\mathrm{D}}(Y)\rangle$ is the average of $I_k^{\mathrm{D}}(Y)$ of all economies at year Y. Through this calculation, all indicators are converted into a set of standardized new sequences I^{R} that excludes both the impact of economic development level and world fluctuations.

These I^{R} exhibit the development paths of every economy in a 76 dimensional standardized space and will be the base of the following similarity analysis. An example of the trajectories of the indicator I^0 and the corresponding relative indicator I^{R} for economies in the space I^0 vs. g and I^{R} vs. g are respectively plotted in figures 1(a) and (b), and their corresponding curves of I^{R} over year are shown in figure 2. Since these I^{R} mainly are constructed by the relative level between I^0 and the corresponding global expected level, they are called 'the relative indicators' in the following discussion.

2.2. The similarity and the regression analysis

In the calculation of the similarity in the development paths between two economies, we use the following three measures to describe the features in the path of a relative indicator: the level, the trend, and the variability. For the kth relative indicator $I_k^{\mathrm{R}}(Y)$, the level value is directly set as $I_k^{\mathrm{R}}(Y)$. The trend $\kappa(Y)$ refers to the regression coefficients obtained from linear regression on the relationship between $I_k^{\mathrm{R}}(y)$ and year y in a sliding year window with m years width and Y as the middle year (the regression model is $I_k^{\mathrm{R}}(y) = \kappa(Y)y + c$), while the mean of regression residuals $\zeta(Y)$ is used as a measure of variability at year Y.

The similarity in the path of the relative indicator between two economies is quantified by the Euclidean distance. For the relative indicator I_k^{R}, the distance between economy i and economy j at year Y is:

$$D_k^{ij}(Y) = \sqrt{d_I(Y)^2 + d_\kappa(Y)^2 + d_\zeta(Y)^2}, \tag{3}$$

where d_I, d_κ and d_ζ respectively are the difference on the standardized values of the level I_k^{R}, the trend κ and the variability ζ: $d_I(Y) = I_j^{\mathrm{R}*}(Y) - I_i^{\mathrm{R}*}(Y)$, $d_\kappa(Y) = \kappa_j^*(Y) - \kappa_i^*(Y)$, $d_\zeta(Y) = \zeta_j^*(Y) - \zeta_i^*(Y)$, where $I^{\mathrm{R}*}$, κ^* and ζ^* denote the z-score standardized value of I_k^{R}, κ and ζ, respectively. The mean of D_k^{ij} for all valid years (that is, for which both economies have data) is defined as the similarity S_k^{ij} in the path of the relative indicator I_k^{R} between economy i and economy j. Since S_k is distance-based, a larger value indicates a lower level of similarity.

By setting an economy (e.g. economy i) as the benchmark economy and employing the above approach, we can obtain the similarity of other economies relative to the benchmark economy on each relative indicator. And then, for each relative indicator (e.g. the kth relative indicator I_k^{R}), we calculate the z-score standardize value of the similarity of other economies (e.g. economy j) relative to the benchmark economy (economy i) on I_k^{R}:

$$S_k^{ij,*} = \frac{S_k^{ij} - \langle S_k^i \rangle}{\sigma\left(S_k^i\right)}\Big|_{j \neq i}, \tag{4}$$

where $\langle S_k^i \rangle$ and $\sigma(S_k^i)$ are the average value and the standard deviation of the set of the similarity relative to the benchmark economy (economy i) on I_k^{R}, respectively. Noticed that, for economy's missing indicator, its standardize value of the similarity is just set to 0.

Finally, by using the z-score standardized value of the similarity of other economies relative to the benchmark economy (economy i) in every relative indicators as the independent variable and setting their average growth δg per year on the relative GDP per capita g as the dependent variable, we perform multivariate regression analysis. The regression model is:

$$\delta g_j = \sum_k q_k^i S_k^{ij,*} + \epsilon^i, \tag{5}$$

where the subscript j denotes economy j ($j \neq i$), the subscript k denotes the kth relative indicator, q_k^i is the regression coefficient, and ϵ^i is the error term of benchmark economy (economy i). Given the correlation among the similarities in different relative indicators, the ridge regression method is employed to mitigate multicollinearity concerns. The determination coefficient R_i^2 of this multivariate regression analysis is adopted to be the metric of DEDP of the economy i.

3. Correlation analysis

Setting the width of sliding year window $m = 11$, we calculated the value R^2 of each economy in the term from 1960 to 2020, and obtained effective values of R^2 of 207 economies. The global average of R^2 is 0.326, with some economies with higher DEDP reaching a level over 0.56, suggesting that the degree of similarity in development paths between other economies and these high DEDP economies can largely explain the differences in their speed of economic development, also indicating the higher prediction accuracy on the predicted value $E(\delta g)$ of δ using the regression model of equation (5). The relationships between the predicted value $E(\delta g)$ and actual δg of various economies when using several high-DEDP representative economies as benchmark economies are plotted in the figure 3, where the Pearson correlation coefficients r and the Spearman correlation coefficients ρ between $E(\delta g)$ and δg are over 0.7 for some benchmark economies. The values of R^2, r and ρ of 30 economies with the highest R^2 are listed in table 1.

We examined the relationships between R^2 and population size of economy. The logarithmic average $\langle \log_{10} p \rangle$ of population over years in this term (from 1960 to 2020) as a measure of an economy's population size in this analysis. As shown in figure 4, R^2 exhibits a significant positive correlation with population size $\langle \log_{10} p \rangle$ of economy, with a Pearson correlation coefficient of 0.474 (the significance level $P < 0.001$) and a Spearman correlation coefficient of 0.482 ($P < 0.001$), and the partial correlation coefficient controlling the effect of the average g over years ($\langle g \rangle$) is 0.477 ($P < 0.001$), indicating that the populous economies usually have higher level of economic development pattern.

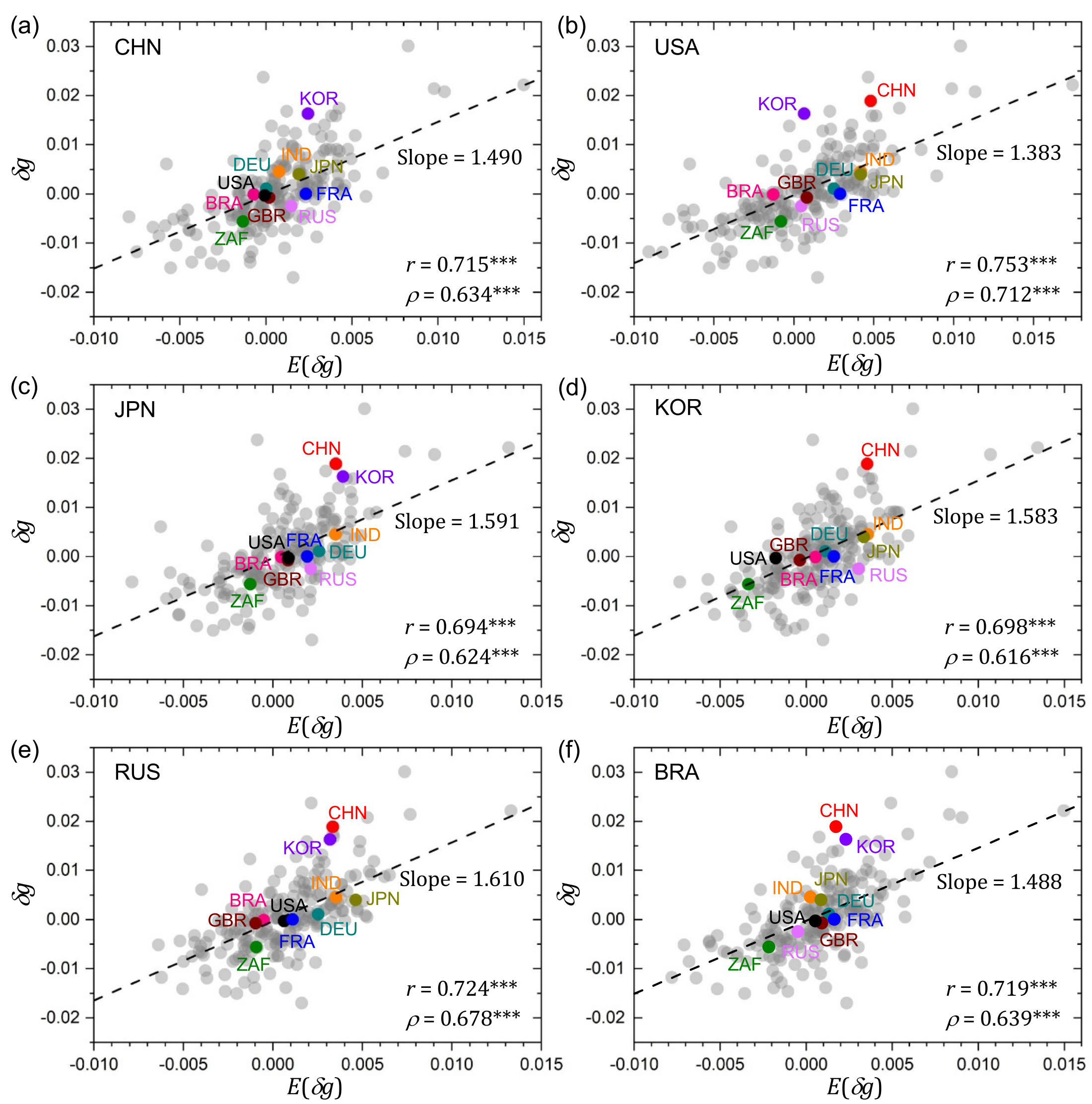

Figure 3. The relationships between the actual δg and the predicted value $E(\delta g)$ of δg for each economy in different benchmark economies. The benchmark economies of panels (a)–(f) are China (the mainland), the United States of America, Japan, Republic of Korea, Russia and Brazil, respectively. The color data points and the gray data points correspond to representative economies and other countries respectively. The dashed lines show the linear-fitting curves. One, two, and three asterisks represent significance levels of $P < 0.001$.

Table 1. The list of 30 economies with the highest R^2. All of correlation coefficients have the significance $P < 0.001$.

Rank	Econ.	R^2	r	ρ	Rank	Econ.	R^2	r	ρ
1	ECU	0.563	0.772	0.691	16	TZA	0.472	0.726	0.646
2	MEX	0.538	0.759	0.682	17	RWA	0.471	0.722	0.588
3	TTO	0.533	0.763	0.722	18	LAO	0.468	0.713	0.632
4	ROU	0.530	0.752	0.695	19	GIN	0.467	0.724	0.629
5	AZE	0.524	0.744	0.656	20	PRY	0.464	0.731	0.666
6	USA	0.524	0.753	0.712	21	AUS	0.463	0.716	0.653
7	MDG	0.514	0.747	0.706	22	CZE	0.463	0.719	0.664
8	BOL	0.512	0.745	0.664	23	KWT	0.463	0.714	0.627
9	GEO	0.512	0.749	0.708	24	BRA	0.462	0.719	0.639
10	PER	0.490	0.730	0.636	25	NPL	0.462	0.711	0.632
11	GBR	0.486	0.732	0.649	26	PHL	0.460	0.711	0.637
12	BEL	0.479	0.727	0.680	27	CRI	0.459	0.721	0.639
13	SLV	0.477	0.730	0.667	28	PAK	0.459	0.715	0.648
14	BRN	0.476	0.725	0.626	29	CHN	0.456	0.715	0.634
15	COG	0.476	0.727	0.644	30	IDN	0.453	0.725	0.646

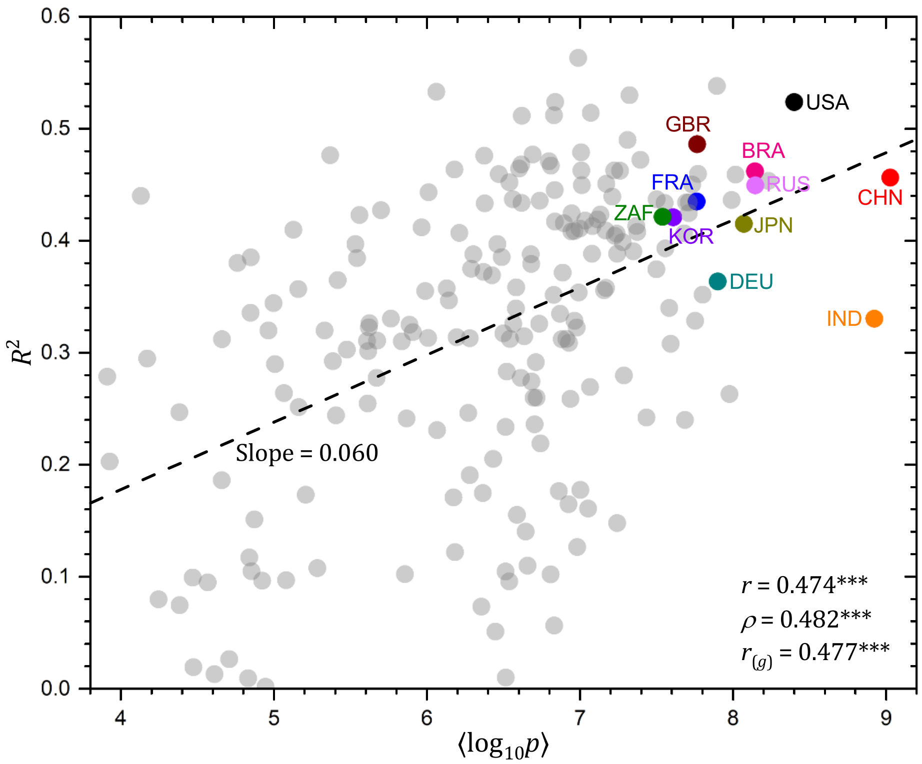

Figure 4. The correlation between each economy's DEDP R^2 and the logarithmic population size $\langle \log_{10} p \rangle$, where the color points and the gray points respectively correspond to representative economies and other economies, and the dashed line is the linear fitting curve. Three asterisks represent significance levels of $P < 0.001$.

As shown in figure 5, certain economies with low R^2 values exhibit significant missing indicator issues. Consequently, we repeat the above correlation analysis after excluding a subset of economies with substantial missing indicator. For a given economy, the economy's missing indicator rate μ is defined as the proportion of indicators that do not include that economy out of the total number of indicators. The correlation coefficients (Pearson correlation coefficient r and Spearman correlation coefficient ρ) and the partial correlation coefficients $r_{(g)}$ (controlling the impact of $\langle g \rangle$) in the cases after excluding the economies with higher μ are listed in table 2 (the threshold of μ is 50% and 30%). Although the correlation between R^2 and $\langle \log_{10} p \rangle$ is slightly weakened after excluding high-μ economies, it is still significant. These results further confirm the significant positive correlation between R^2 and population size of economy.

The correlations between each economy's R^2 and other macroeconomic indicators that are relevant to long-term economic development are also observed. These indicators include δg, the average value $\langle g \rangle$ of g over years in the period from 1960 to 2020, the average value $\langle K_{\mathrm{E}} \rangle$ of ECI over years in the period from 1995 to 2020, and the logarithmic average value $\langle \log_{10} F \rangle$ of economic fitness over years in the period from 1995 to 2015. However, R^2 does not show any significant correlations with these indicators.

4. The consistency analysis based on regression coefficients

4.1. The consistency analysis based on the global average regression coefficients

Equation (5) as a regression model that all the independent variables are standardized, the absolute value of the regression coefficient is correlated to the effect of the factor on the dependent variable. The regression coefficients of each indicator for each benchmark economy are plotted in figure 5. Due to the lower S means higher similarity, the negative regression coefficient indicates that economies with smaller differences in this indicator compared to the benchmark economy are more likely to have faster economic growth. Conversely, for positive regression coefficients, the opposite is true. It can be observed that there is high-level consistency in the regression coefficients obtained from different benchmark economies. We therefore calculate the average $\langle q \rangle$ of regression coefficient q of each indicator, and the indicators with high absolute value of $\langle q \rangle$ are listed in table 3, most of which are the trade indicators, showing the strong impact of international trade on economic development.

Figure 5 also implies that there is a degree of consistency between the regression coefficients for the different benchmark economies. For the sequence $\{q\}$ of regression coefficients of each benchmark economy

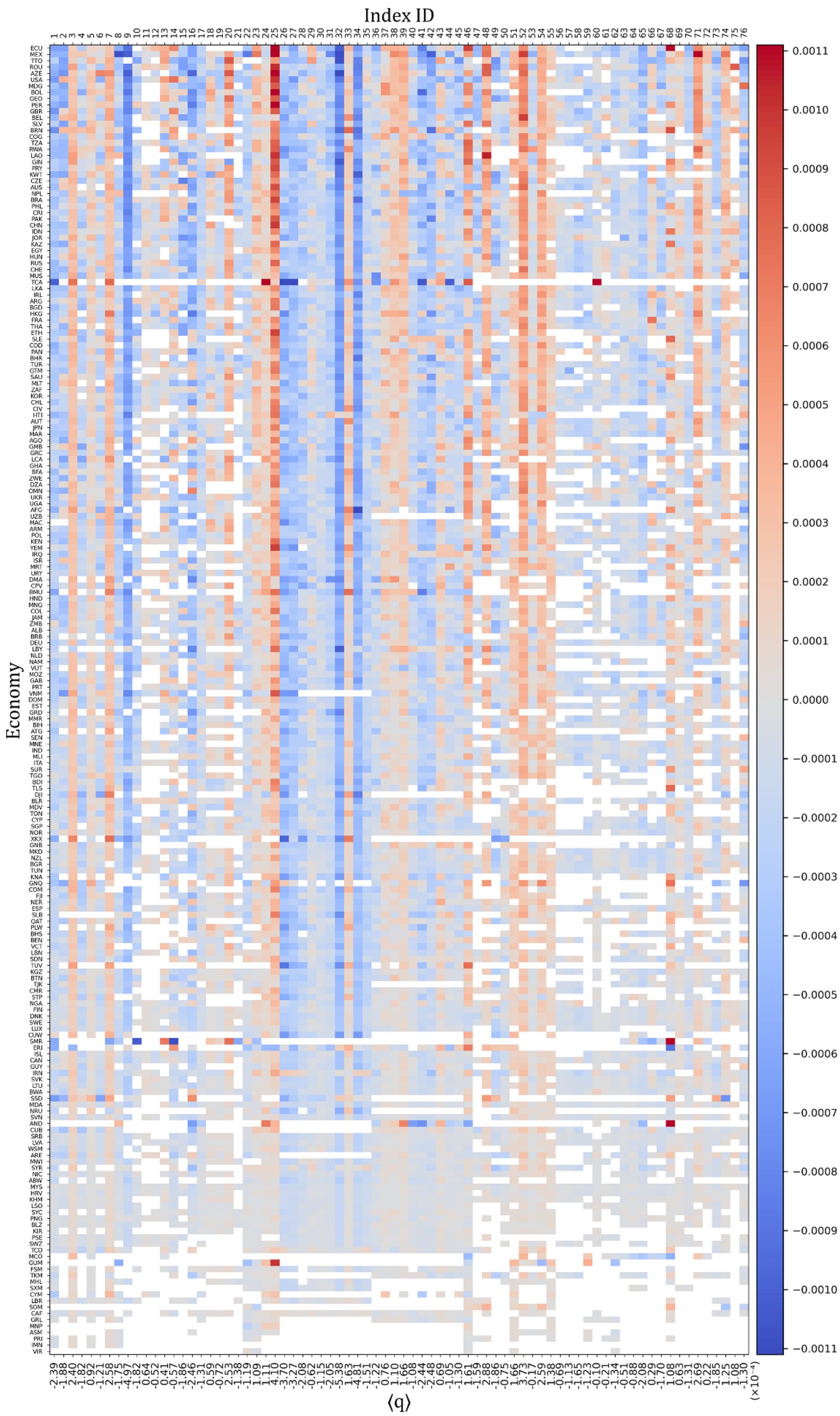

Figure 5. The regression coefficients of each indicator for each benchmark economy. Economies are listed in R^2 value from highest to lowest. The values listed in the bottom are the average $\langle q\rangle$ of the regression coefficients of each indicator.

Table 2. Correlations between R^2 and $\langle \log_{10} p\rangle$.

	r	ρ	$r_{(g)}$
All economies (total 207)	0.474 ($P < 0.001$)	0.482 ($P < 0.001$)	0.477 ($P < 0.001$)
Economies with $\mu \leqslant 50\%$ (total 185)	0.361 ($P < 0.001$)	0.403 ($P < 0.001$)	0.380 ($P < 0.001$)
Economies with $\mu \leqslant 30\%$ (total 157)	0.354 ($P < 0.001$)	0.355 ($P < 0.001$)	0.365 ($P < 0.001$)

Table 3. List of indicators with $|\langle q\rangle| > 3.0 \times 10^{-4}$.

Category	ID	Name of indicator	$\langle q\rangle(\times 10^{-4})$
Financial sector	9	Adjusted savings net national savings (current US$)	−4.97
Trade	25	Merchandise imports (current US$)	4.10
Trade	26	Commercial service exports (current US$)	−3.70
Trade	27	Commercial service imports (current US$)	−3.27
Trade	32	Transport services (% of commercial service exports)	−5.38
Trade	34	Travel services (% of commercial service exports)	−4.81
Infrastructure	52	Air transport, passengers carried	3.73

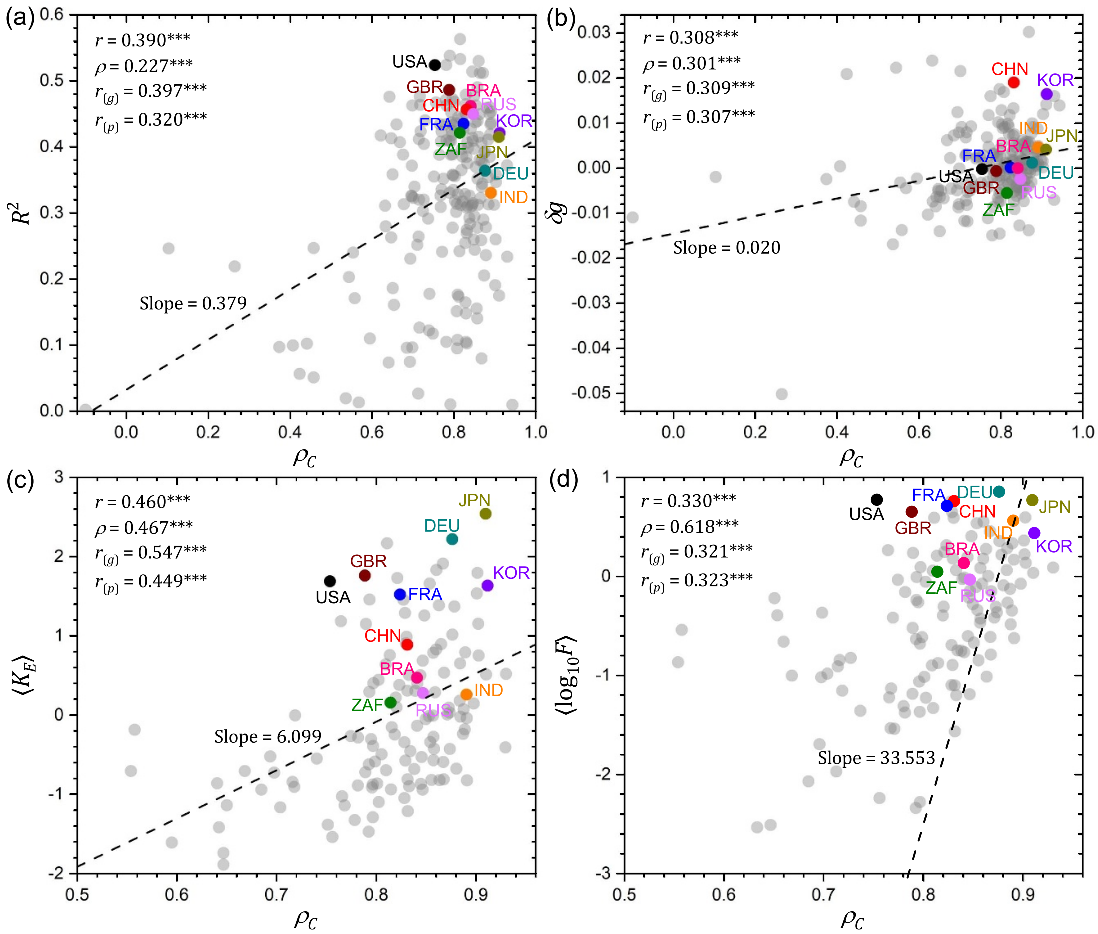

Figure 6. The correlations between benchmark economy's consistency (ρ_C) based on global average regression coefficients and their $R2$ (panel (a)), δg (panel (b)), $\langle K_E\rangle$ (panel (c)), and $\langle \log_{10} F\rangle$ (panel (d)). The color points and the gray points in each panel respectively correspond to representative economies and other economies, and the dashed lines show the linear fitting. Three asterisks represent significance levels of $P < 0.001$.

and the sequence $\{\langle q\rangle\}$ of average regression coefficients, we use the correlation coefficient (Spearman correlation coefficient ρ_C) between them to be the measure of their consistency. In the calculation of correlation coefficients, for the economy with missing indicators, the corresponding indicators are removed from the sequence $\{\langle q\rangle\}$ of average regression coefficients, and then we calculate the correlation coefficients between the remaining part of the two sequences. This consistency is an indicator for another aspect of DEDP of the benchmark economy, which measures the difference in direction of its explanation with the global average explanation direction when using the similarity S to explain economic growth.

This consistency (measured by ρ_C) is positively correlated with R^2 (figure 6(a)), and the Pearson correlation coefficient r and the Spearman correlation coefficient ρ are 0.390 and 0.227, respectively. ρ_C also exhibits a weak correlation with the population size $\langle \log_{10} p\rangle$, as indicated by the correlation coefficients of $r = 0.247$ and $\rho = 0.240$. Moreover, we find that a significant correlation between ρ_C and δg, with Pearson correlation coefficient and Spearman correlation coefficient of 0.308 and 0.301, as shown in figure 6(b). Although the correlation is not strong, the partial correlation coefficients $r_{(g)}$ and $r_{(p)}$ remain significant ($r_{(g)} = 0.309$ and $r_{(p)} = 0.307$, both with $P < 0.001$) after excluding the effects of $\langle g\rangle$ and population size

Table 4. The list of top 20 economies on the five indicators ($\rho_{\mathrm{P}}^{\delta g}$, ρ_{P}^{K}, ρ_{P}^{F}, $\rho_{\mathrm{P}}^{\delta K}$, and $\rho_{\mathrm{P}}^{\delta F}$) of explanatory power of coefficient benchmark economy. One, two, and three asterisks represent significance levels of $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

Rank	Econ.	$\rho_{\mathrm{P}}^{\delta g}$	Econ.	ρ_{P}^{K}	Econ.	ρ_{P}^{F}	Econ.	$\rho_{\mathrm{P}}^{\delta K}$	Econ.	$\rho_{\mathrm{P}}^{\delta F}$
1	POL	0.417***	AUT	0.730***	DEU	0.786***	VNM	0.420***	OMN	0.278***
2	HUN	0.408***	DEU	0.726***	AUT	0.770***	IDN	0.339***	VNM	0.275***
3	HKG	0.374***	SWE	0.715***	CHE	0.747***	THA	0.299***	BWA	0.269***
4	JPN	0.369***	CZE	0.661***	CZE	0.741***	JOR	0.262**	BTN	0.262**
5	CZE	0.367***	BEL	0.659***	LKA	0.737***	AZE	0.255**	RWA	0.260**
6	NOR	0.363***	HUN	0.653***	SWE	0.731***	LAO	0.254**	MOZ	0.235**
7	BEL	0.362***	FRA	0.644***	GBR	0.730***	BHR	0.253**	KWT	0.232**
8	CHE	0.362***	CHE	0.640***	ISR	0.715***	SMR	0.252**	TCA	0.211*
9	ROU	0.359***	AUS	0.635***	FRA	0.712***	LBY	0.251**	QAT	0.207*
10	SVN	0.356***	ISR	0.634***	HUN	0.709***	HKG	0.249**	ZMB	0.206*
11	IND	0.353***	GBR	0.631***	GRC	0.706***	KHM	0.248**	BHR	0.202*
12	CHN	0.348***	POL	0.631***	ESP	0.701***	CHN	0.238**	AZE	0.197*
13	THA	0.347***	SVK	0.618***	JPN	0.700***	GHA	0.236**	IRN	0.194*
14	LVA	0.346***	ESP	0.615***	ARG	0.700***	BTN	0.229**	LBY	0.193*
15	MLT	0.346***	BRB	0.612***	ITA	0.695***	QAT	0.228**	CAF	0.190*
16	EGY	0.343***	ITA	0.611***	ROU	0.694***	TUR	0.221*	LAO	0.189*
17	AUT	0.343***	GRD	0.607***	KOR	0.690***	TTO	0.207*	MNP	0.185*
18	SWE	0.342***	CAN	0.607***	POL	0.689***	SOM	0.207*	ARE	0.178*
19	DEU	0.341***	NOR	0.603***	FIN	0.688***	LBR	0.204*	SLB	0.175*
20	DNK	0.340***	USA	0.602***	USA	0.684***	ALB	0.203*	TTO	0.170*

$\langle \log_{10} p\rangle$ respectively, suggesting that this consistency possesses a certain degree of explanatory power in terms of economic growth.

More remarkable, the consistency is also significantly correlated with ECI and economic fitness of economy. As illustrated in figures 6(c) and (d), ρ_{C} exhibits positive correlations with both $\langle K_{\mathrm{E}}\rangle$ and $\langle \log_{10} F\rangle$ of the benchmark economy, with Spearman correlation coefficients of 0.467 and 0.618, respectively. Excluding the impact of $\langle g\rangle$ and population size ($\langle \log_{10} p\rangle$), the partial correlation coefficients $r_{(g)}$ and $r_{(p)}$ between ρ_{C} and $\langle K_{\mathrm{E}}\rangle$ are 0.547 and 0.449, respectively (as shown in figure 6(c)). Additionally, the partial correlation between ρ_{C} and $\langle \log_{10} F\rangle$ is also significant ($r_{(g)} = 0.321$, $r_{(p)} = 0.323$, see figure 6(d)). These results indicate that the economy with a high ECI or economic fitness usually has a high level of consistency in the direction of its explanation when using similarity S to explain long-term economic growth.

4.2. The consistency analysis based on regression coefficients of benchmark economy

The consistency mentioned above is defined on the benchmark of the global average of regression coefficients. However, using the global average of regression coefficients as the benchmark may not necessarily lead to the best explanatory power. We therefore try to extend the benchmark to the regression coefficients of each economy. That is, by designating an economy (e.g. economy i) as the coefficient benchmark economy, the consistency of the other economy (e.g. economy j) based on the regression coefficients of economy i is measured by the Spearman correlation coefficient $\rho_{\mathrm{C}}^{\mathrm{E}}|_i^j$ between the sequence of regression coefficients $\{q\}_j$ of economy j and $\{q\}_i$ of the coefficient benchmark economy (economy i). In the calculation of similarity, for the case with missing indicators, a method that is similar to the above calculation is adopted: in each pair of economies, if one or two economies have missing indicators, the missing indicators in both sequences of the two economies are removed, and the remaining part of the two sequences is used in the calculation of correlation coefficients.

By successively setting each economy as the coefficient benchmark economy, we calculate the consistency $\rho_{\mathrm{C}}^{\mathrm{E}}$ of every economies' sequence of regression coefficients relative to that of the coefficient benchmark economy, resulting in the consistency set $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ for the coefficient benchmark economy. Subsequently, the correlations between the consistency set $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of each coefficient benchmark economy and several macroeconomic indicators associated with economic growth, including δg, $\langle K_{\mathrm{E}}\rangle$, $\langle \log_{10} F\rangle$, and the average annual changes δK_{E} of $\langle K_{\mathrm{E}}\rangle$ and average annual changes $\delta \log_{10} F$ of $\langle \log_{10} F\rangle$, are observed. Using $\rho_{\mathrm{P}}^{\delta g}$, ρ_{P}^{K}, ρ_{P}^{F}, $\rho_{\mathrm{P}}^{\delta K}$, and $\rho_{\mathrm{P}}^{\delta F}$, to represent the Spearman correlation coefficient between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ with δg, $\langle K_{\mathrm{E}}\rangle$, $\langle \log_{10} F\rangle$, δK_{E}, and $\delta \log_{10} F$, respectively, they assess the explanatory power on these macroeconomic indicators through the regression coefficients of the coefficient benchmark economy, which also serves as a comprehensive set of metrics for evaluating the DEDP of the coefficient benchmark economy from the perspective of explaining the corresponding macroeconomic indicators.

The top 20 economies with the highest $\rho_{\mathrm{P}}^{\delta g}$, ρ_{P}^{K}, ρ_{P}^{F}, $\rho_{\mathrm{P}}^{\delta K}$, and $\rho_{\mathrm{P}}^{\delta F}$, respectively are listed in table 4. The scatter plots of the examples with the highest $\rho_{\mathrm{P}}^{\delta g}$, ρ_{P}^{K}, ρ_{P}^{F}, $\rho_{\mathrm{P}}^{\delta K}$, and $\rho_{\mathrm{P}}^{\delta F}$, are shown in figure 7. For certain

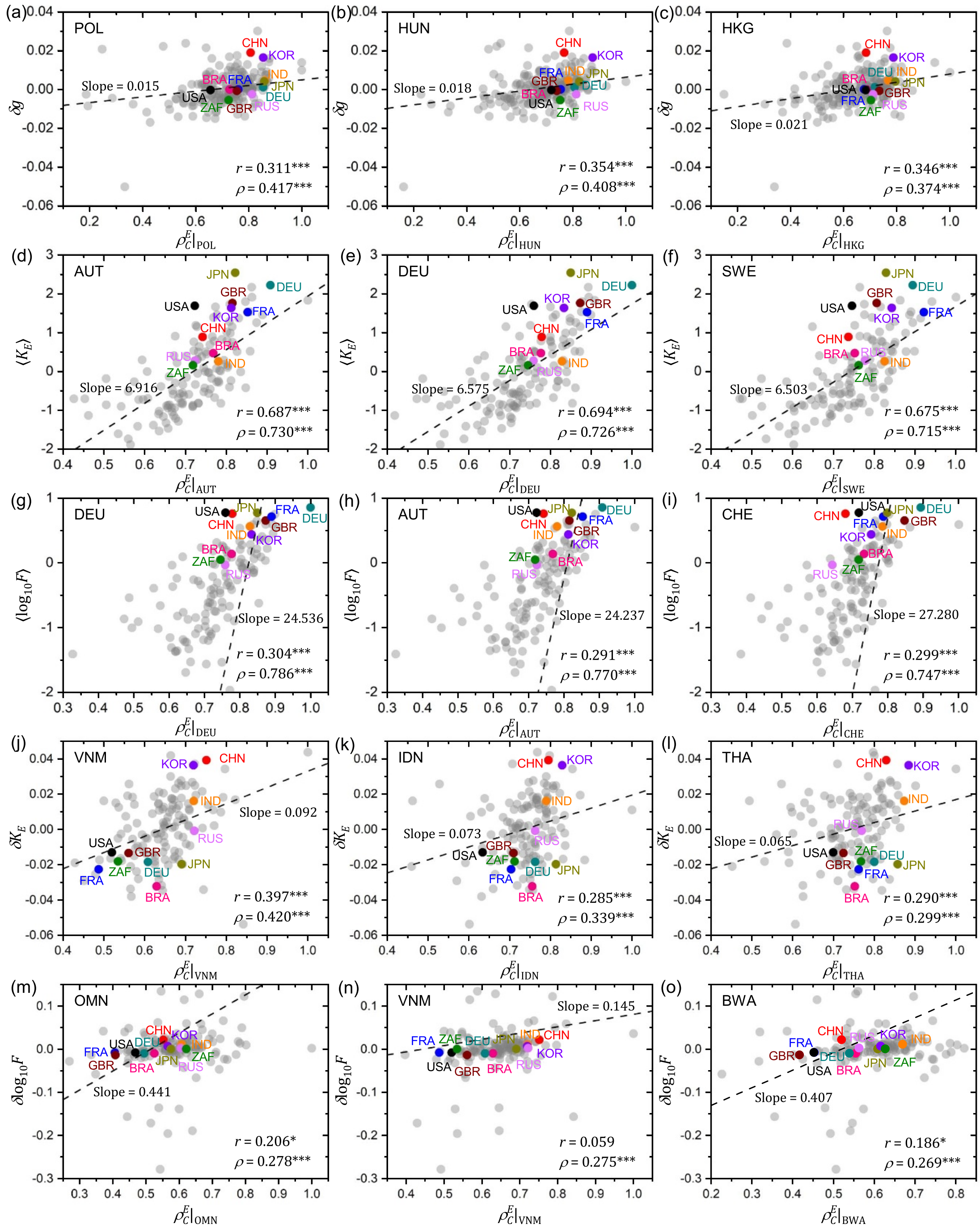

Figure 7. The examples with the strongest correlations (the highest Spearman correlation coefficient) between the consistency set $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of the coefficient benchmark economy and the macroeconomic indicators (δg, $\langle K_{\mathrm{E}}\rangle$, $\langle \log_{10} F\rangle$, δK_{E}, and $\delta \log_{10} F$) of economies. The panels in the first row (panels (a)–(c)), the second row (panels (d)–(f)), the third row (panels (g)–(i)), the fourth row (panels (j)–(l)), and the fifth row (panels (m)–(o)) show the examples of the correlations between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ and δg, $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ and $\langle K_{\mathrm{E}}\rangle$, $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ and $\langle \log_{10} F\rangle$, $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ and δK_{E}, and $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ and $\delta \log_{10} F$, respectively. The color points and the gray points respectively correspond to representative economies and other economies, and the dashed lines show the linear fitting. One, two, and three asterisks represent significance levels of $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

coefficient benchmark economies, their consistency set $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ exhibits remarkable strong correlations with these macroeconomic indicators, particularly $\langle K_{\mathrm{E}}\rangle$ and $\langle \log_{10} F\rangle$. For example, ρ_{P}^{K} of Austria and ρ_{P}^{F} of Germany are as high as 0.730 ($P < 0.001$) and 0.786 ($P < 0.001$), respectively, as shown in table 4, indicating a strong explanatory power for both ECI and economic fitness serving the regression coefficients of these economies as benchmarks. The highest values of $\rho_{\mathrm{P}}^{\delta g}$ and $\rho_{\mathrm{P}}^{\delta K}$ are also over 0.4, indicating that Poland and Vietnam, serving as benchmarks, can also explain the growth of GDP per capita and ECI growth, respectively, to a certain extent. The highest $\rho_{\mathrm{P}}^{\delta F}$ is 0.278 (Oman) and remains significant ($P < 0.001$). In addition, from table 4, we observe that the economies with higher-level metrics of explanatory power for

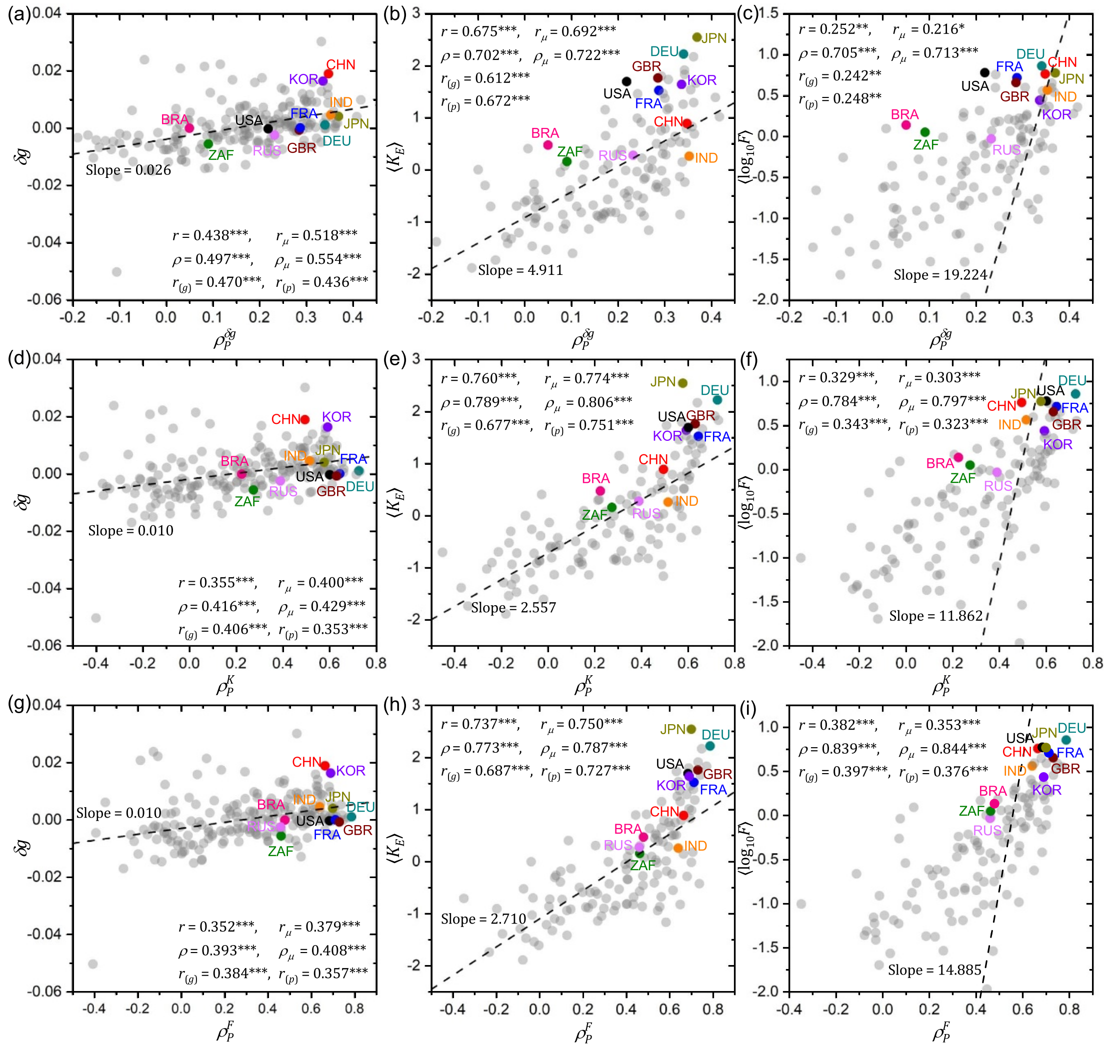

Figure 8. The correlations between three macroeconomic indicators (δg, $\langle K_\mathrm{E} \rangle$ and $\langle \log_{10} F \rangle$) and three metrics ($\rho_\mathrm{P}^{\delta g}$, ρ_P^{K} and ρ_P^{F}) of explanatory power of coefficient benchmark economy. The color points and the gray points respectively correspond to representative economies and other economies, and the dashed lines show the linear fitting. One, two, and three asterisks represent significance levels of $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

different metrics exhibits remarkable differences. For instance, the majority of economies with highest ρ_P^{K} and ρ_P^{F} values are the developed economies in Europe, while those with highest $\rho_\mathrm{P}^{\delta K}$ and $\rho_\mathrm{P}^{\delta F}$ values are the developing economies in Asia and Africa.

Moreover, stronger correlations are observed in the relationships between these metrics of explanatory power ($\rho_\mathrm{P}^{\delta g}$, ρ_P^{K}, ρ_P^{F}) of economy and the macroeconomic indicators of economy. The metrics of explanatory power most strongly related to δg, $\langle K_\mathrm{E} \rangle$, and $\langle \log_{10} F \rangle$ are $\rho_\mathrm{P}^{\delta g}$, ρ_P^{K} and ρ_P^{F}, respectively. As shown in figure 8(a) The Spearman correlation coefficients between δg and $\rho_\mathrm{P}^{\delta g}$ is 0.497 ($P < 0.001$). Meanwhile, the Spearman correlation coefficients between $\langle K_\mathrm{E} \rangle$ and ρ_P^{K}, and between $\langle \log_{10} F \rangle$ and ρ_P^{F} are as high as 0.789 ($P < 0.001$) and 0.839 ($P < 0.001$) (see figures 8(e) and (i)), respectively. The other correlations involving δg, such as δg vs. ρ_P^{K}, and δg vs. ρ_P^{F}, have a similar level close to the correlation between δg and $\rho_\mathrm{P}^{\delta g}$ (see figures 8(d) and (g)), and similar property can be found in the correlations involving $\langle K_\mathrm{E} \rangle$ and $\langle \log_{10} F \rangle$ (see figures 8(b), (c), (f) and (h)). When considering the case of indicator missing, by removing the economies with a missing indicators rate $\mu > 30\%$ from the correlation analysis (the numbers of remaining economies are 157 for the analysis related to δg, 127 for the analysis involving $\langle K_\mathrm{E} \rangle$, and 135 for the analysis involving $\langle \log_{10} F \rangle$), the above Spearman correlation coefficients increase to 0.554 for the correlation between δg and $\rho_\mathrm{P}^{\delta g}$, 0.806 for the correlation between $\langle K_\mathrm{E} \rangle$ and ρ_P^{K}, and 0.844 for the correlation between $\langle \log_{10} F \rangle$ and ρ_P^{F} (see the ρ_μ values in figures 8(a), (e) and (i)), and similar increases on the Spearman correlation coefficients also can be found in other correlations shown in figure 8. By controlling the effect of $\langle g \rangle$ and $\langle \log_{10} p \rangle$, respectively, the partial correlation coefficients $r_{(g)}$ and $r_{(p)}$ for the above correlations still are close to the level of the case

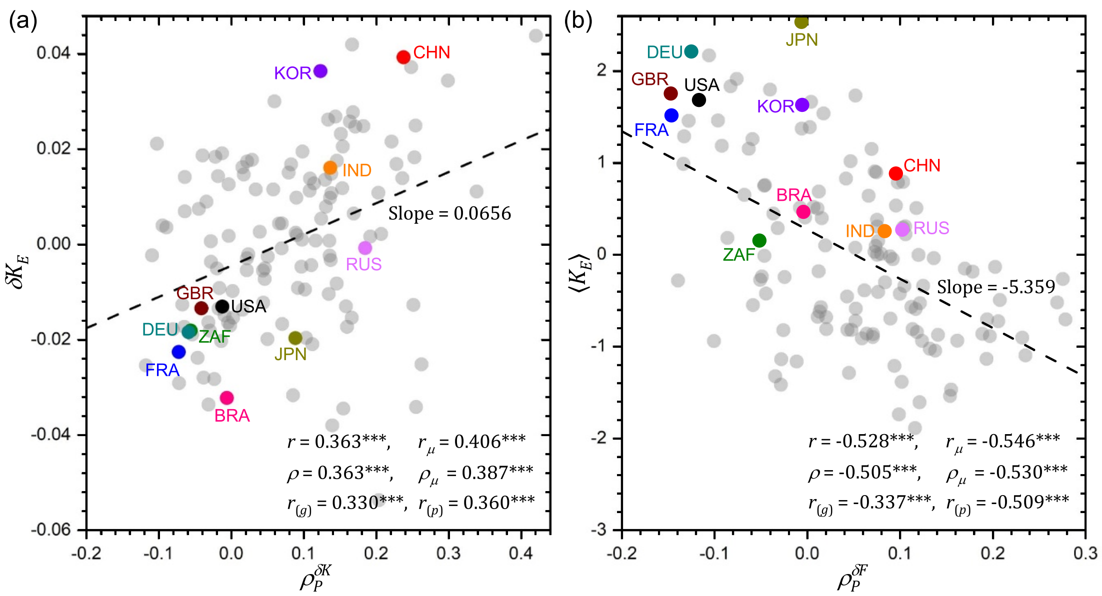

Figure 9. The correlations between $\rho_P^{\delta K}$ and δK_E, $\rho_P^{\delta F}$ and $\langle K_E \rangle$. The color points and the gray points respectively correspond to representative economies and other economies, and the dashed lines show the linear fitting. Three asterisks represent significance levels of $P < 0.001$, respectively.

without factor control, as shown in each panel of figure 8, indicating that these correlations are not a result from other relationships that are relevant to variations in GDP per capita or population size of economy.

The other two metrics $\rho_P^{\delta K}$ and $\rho_P^{\delta F}$ also show significant correlations with δK_E and $\langle K_E \rangle$, respectively, with Spearman correlation coefficients of 0.363 ($P < 0.001$) and -0.505 ($P < 0.001$), as shown in figure 9. Notably, the relationship between $\langle K_E \rangle$ and $\rho_P^{\delta F}$ is the only instance of a significant negative correlation (see figure 9(b)). This finding is pertinent as the majority of economies with a higher $\rho_P^{\delta F}$ belong to the developing economies, which typically exhibit a lower level of ECI.

At last, the notations for main indicators and metrics used in sections 3 and 4 are summarized in table 5.

5. Conclusion

The method for constructing a metric of DEDP can be summarized as follows: in a standardized multiple indicator-year space, the similarity between the development paths of economies in each indicator is defined and calculated. And then, we utilize different economies as benchmark economy to assess the explanatory power (measured by the determination coefficient R^2) of the similarity relative to the development paths of the benchmark economy on the economic development speed of each economy using regression analysis, constructing DEDP metrics based on this explanatory power and the consistency of the sequence of the regression coefficients. The data basis in the construction of DEDP integrates 76 economy-level macroeconomic indicators across 11 categories, providing comprehensive coverage of the external representation of macro-economy in different fields. Since all indicators have undergone relative transformation, that is, the standardized deviation of each indicator relative to the global expectation level is used as the benchmark, and the fluctuations of the world's overall level with each year are excluded, all the indicator data among countries with different levels of economic development in different years are incorporated into a comparable standardized framework [24–27], ensuring the consistency and comparability of the similarity of macroeconomic indicators between economies defined on this deviation. It is important to note that, this approach in the construction of DEDP metrics can be easily generalized to the analysis of other socio-economic indicators. For example, by incorporating more socio-economic indicators into the current framework, a high level of explanatory power and prediction accuracy for economic growth and social development would be achieved.

One of the main findings of this study is the significant positive correlation between R^2 and the population size of an economy. This result suggests that there may be significant disparities in the development paths of economies with different population sizes. In comparison to economies with a smaller population size, the development paths of populous economies are more suitable for serving as a comparative benchmark to explain variations in economic development speed across economies. This implies that the development paths of economies with a larger population size may be closer to certain

Table 5. Notations and descriptions of main indicators and metrics used in the paper.

Indicator type	Notation	Description
The macroeconomic indicator	g	Economy's logarithmic relative GDP per capita of economy
The macroeconomic indicator	$\langle g\rangle$	Economy's average g over years in the period from 1960 to 2020
The macroeconomic indicator	δg	Economy's average growth per year of g in the period from 1960 to 2020
The macroeconomic indicator	$\langle \log_{10} p\rangle$	Economy's logarithmic average of population over years in the period from 1960 to 2020
The macroeconomic indicator	$\langle K_{\mathrm{E}}\rangle$	The mean of ECI of economy in the period from 1995 to 2020
The macroeconomic indicator	δK_{E}	The average annual change of $\langle K_{\mathrm{E}}\rangle$ of economy in the period from 1995 to 2020
The macroeconomic indicator	$\langle \log_{10} F\rangle$	The logarithmic average of economic fitness of economy in the period from 1995 to 2015
The macroeconomic indicator	$\delta \log_{10} F$	The average annual changes of $\langle \log_{10} F\rangle$ of economy in the period from 1995 to 2015
Metric of economy's DEDP	R^2	The determination coefficient of the multivariate regression analysis based on equation (5)
Metric of economy's DEDP	ρ_{C}	The consistency between economy's regression coefficients and the global average of regression coefficients, measured by Spearman correlation coefficient
Metric of economy's DEDP	$\rho_{\mathrm{C}}^{\mathrm{E}}$	The consistency between economy's regression coefficients and the regression coefficients of the coefficient benchmark economy, measured by Spearman correlation coefficient
Metric of economy's DEDP	$\rho_{\mathrm{P}}^{\delta g}$	The explanatory power on δg through the regression coefficients of the coefficient benchmark economy, measured by the Spearman correlation coefficient between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of the coefficient benchmark economy and δg of economy
Metric of economy's DEDP	ρ_{P}^{K}	The explanatory power on $\langle K_{\mathrm{E}}\rangle$ through the regression coefficients of the coefficient benchmark economy, measured by the Spearman correlation coefficient between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of the coefficient benchmark economy and $\langle K_{\mathrm{E}}\rangle$ of economy
Metric of economy's DEDP	ρ_{P}^{F}	The explanatory power on $\langle \log_{10} F\rangle$ through the regression coefficients of the coefficient benchmark economy, measured by the Spearman correlation coefficient between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of the coefficient benchmark economy and $\langle \log_{10} F\rangle$ of economy
Metric of economy's DEDP	$\rho_{\mathrm{P}}^{\delta K}$	The explanatory power on δK_{E} through the regression coefficients of the coefficient benchmark economy, measured by the Spearman correlation coefficient between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of the coefficient benchmark economy and δK_{E} of economy
Metric of economy's DEDP	$\rho_{\mathrm{P}}^{\delta F}$	The explanatory power on $\delta \log_{10} F$ through the regression coefficients of the coefficient benchmark economy, measured by the Spearman correlation coefficient between $\{\rho_{\mathrm{C}}^{\mathrm{E}}\}$ of the coefficient benchmark economy and $\delta \log_{10} F$ of economy
General correlation	r	The Pearson correlation coefficient
General correlation	ρ	The Spearman correlation coefficient
General correlation	$r_{(g)}$	The partial correlation coefficient controlling the impact of the average GDP per capita over years ($\langle g\rangle$)
General correlation	$r_{(p)}$	The partial correlation coefficient controlling the impact of the logarithmic average population size over years ($\langle \log_{10} p\rangle$)
General correlation	r_{μ}	The Pearson correlation coefficient for the economies where the missing indicator rate μ is no more than certain threshold
General correlation	ρ_{μ}	The Spearman correlation coefficient for the economies where the missing indicator rate μ is no more than certain threshold

economic development patterns that hold universal significance. In previous studies, it has been observed that numerous factors that are crucial for economic development often require the support of a large population size. For instance, the establishment of large internal markets [28, 29], the availability of human resources [30], cross-regional industrial division and economic diversification [31], as well as technological development and innovation activities [27, 32, 33] often depend on a sufficient population size. This suggests that populous economies are more likely to leverage the support of their large population to coordinate various factors comprehensively and promote economic development in a multifaceted and comprehensive manner. This could be the underlying reason why the development paths of populous economies are more comparable.

The analysis of the consistency of the sequence of the regression coefficients reveals another perspective for exploring the DEDP of economies. A set of consistency-based metrics, especially the metric $\rho_{\mathrm{C}}^{\mathrm{E}}$ based on the consistency of the sequence of regression coefficients of the coefficient benchmark economy, as well as explanatory power metrics ($\rho_{\mathrm{P}}^{\delta g}$, ρ_{P}^{K} and ρ_{P}^{F}) derived from the correlation between $\rho_{\mathrm{C}}^{\mathrm{E}}$ and macroeconomic indicators, exhibit strong correlations with the growth rate of GDP per capita, ECI, and economic fitness, offering robust explanations for the differences in these macroeconomic indicators across economies. It is worth noting that these consistency-based metrics and R^2 effectively decouple the impact of economic factors and that of demographic factors in economic development patterns: the consistency-based metrics focus on the effective explanations for differences in macroeconomic indicators, while R^2 uncovers the differences in development paths among countries with varying population sizes. This decoupling is unexpected, since we have not directly integrated the separation of these two types of factors into the construction algorithm of the DEDP metric.

More importantly, our findings provide rather significant and strong implications in terms of catch-up policies for the developing countries. Because the economy with higher R^2 reflect that, using their development paths as the benchmark, the differences between other economies' development paths and the benchmark are strongly positively correlated with the speed of economic growth, it implies a potential development strategy for other economies is to try to approach these high-R^2 economies in their development paths. Of course, since development paths are constructed in a standardized space, this 'approach' does not mean directly attempting to replicate the macroeconomic indicators of high-R^2 economy, but rather requires a combination of the domestic economic development stage and current situation.

These results also imply that there may exist one or more virtual development paths in the macroeconomic indicator space that leads to the maximum DEDP metrics. In this case, the difference between the actual development paths of an economy and the DEDP-maximized paths can explain the development speed of each economy to the greatest extent. In other words, the DEDP-maximized paths represent a development model with significant reference value for all economies' development and are likely corresponding to a universal pattern in economic development. The DEDP-maximized paths are more suitable as the absolute benchmark for further analysis and as a guide for the economic development direction of individual economies. The construction and analysis of the DEDP-maximized paths remains to be further studied.

At last, it should be noted that, in the construction of DEDP metrics, the similarity S between economy's development paths used in the regression analysis is defined independently for each macroeconomic indicator, rather than combining them into a single similarity metric. This approach results in difficulties in visually associating the changes in DEDP metrics with those in macroeconomic indicators, thereby posing challenges in the interpretability of the DEDP metrics. The consistency-based metrics, in particular, exhibit weaker interpretability due to their reliance on the similarity of regression coefficients from different coefficient benchmarks. In subsequent studies, further examination is essential to elucidate the distinct manifestations of these DEDP metrics along economic development paths, with the ultimate objective of shaping strategic guidance for the economic development direction of diverse economies.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

Project supported by the National Natural Science Foundation of China (Grant No. 62073112).

Appendix. List of macroeconomic indicators

Table A1. List of macroeconomic indicators.

Category	ID	Name of indicator	Year range	Number of economies	Per capita treatment
Industry-specific	1	Agriculture, forestry, and fishing, value added (% of GDP)	1960–2020	199	
Industry-specific	2	Agriculture, forestry, and fishing, value added per worker (constant 2015 US$)	1991–2019	170	
Industry-specific	3	Industry (including construction), value added (% of GDP)	1960–2020	199	
Industry-specific	4	Industry (including construction), value added per worker (constant 2015 US$)	1991–2019	165	
Industry-specific	5	Services, value added (% of GDP)	1960-2020	196	
Industry-specific	6	Services, value added per worker (constant 2015 US$)	1991–2019	168	
Industry-specific	7	Manufacturing, value added (% of GDP)	1960–2020	197	
Labor	8	Labor force participation rate, total (% of total population ages 15+) (national estimate)	1990–2020	180	
Financial sector	9	Adjusted savings net national savings (current US$)	1970–2019	176	TRUE
Financial sector	10	Lending interest rate (%)	1960–2020	146	
Financial sector	11	Stocks traded, total value (% of GDP)	1975–2020	99	
Financial sector	12	Listed domestic companies, total	1975–2020	100	TRUE
Financial sector	13	Bank capital to assets ratio (%)	2000–2020	133	
Financial sector	14	Total reserves (includes gold, current US$)	1960–2020	178	TRUE
Poverty	15	Gini index (World Bank estimate)	1971–2019	124	
Poverty	16	Human development index (HDI)	1990–2019	179	
Market concentration	17	HH market concentration index	1988–2019	182	
Public sector	18	Revenue, excluding grants (% of GDP)	1972–2020	153	
Public sector	19	Tax revenue (% of GDP)	1972–2020	153	
Public sector	20	Expense (% of GDP)	1972–2020	150	
Public sector	21	Central government debt, total (% of GDP)	1990–2016	93	
Trade	22	Exports of goods and services (% of GDP)	1960–2020	191	
Trade	23	Imports of goods and services (% of GDP)	1960–2020	191	
Trade	24	Merchandise exports (current US$)	1960–2020	197	TRUE
Trade	25	Merchandise imports (current US$)	1960–2020	198	TRUE
Trade	26	Commercial service exports (current US$)	1960–2020	192	TRUE

(Continued.)

Table A1. (Continued.)

Category	ID	Name of indicator	Year range	Number of economies	Per capita treatment
Trade	27	Commercial service imports (current US$)	1960–2020	192	TRUE
Trade	28	Insurance and financial services (% of commercial service exports)	1960–2020	183	
Trade	29	Insurance and financial services (% of commercial service imports)	1960–2020	191	
Trade	30	Computer, communications and other services (% of commercial service exports)	1960–2020	190	
Trade	31	Computer, communications and other services (% of commercial service imports)	1960–2020	191	
Trade	32	Transport services (% of commercial service exports)	1960–2020	190	
Trade	33	Transport services (% of commercial service imports)	1960–2020	191	
Trade	34	Travel services (% of commercial service exports)	1960–2020	190	
Trade	35	Travel services (% of commercial service imports)	1960–2020	190	
Trade	36	Ores and metals exports (% of merchandise exports)	1962–2020	186	
Trade	37	Ores and metals imports (% of merchandise imports)	1962–2020	188	
Trade	38	Agricultural raw materials exports (% of merchandise exports)	1962–2020	186	
Trade	39	Agricultural raw materials imports (% of merchandise imports)	1962–2020	187	
Trade	40	Fuel exports (% of merchandise exports)	1962–2020	183	
Trade	41	Fuel imports (% of merchandise imports)	1962–2020	188	
Trade	42	Food exports (% of merchandise exports)	1962–2020	187	
Trade	43	Food imports (% of merchandise imports)	1962–2020	188	
Trade	44	Manufactures exports (% of merchandise exports)	1962–2020	187	
Trade	45	Manufactures imports (% of merchandise imports)	1962–2020	188	
Urban development	46	Urban population (% of total population)	1960–2020	206	
Urban development	47	Population in urban agglomerations of more than 1 million (% of total population)	1960–2020	119	
Urban development	48	Population in the largest city (% of urban population)	1960–2020	151	

(Continued.)

Table A1. (Continued.)

Category	ID	Name of indicator	Year range	Number of economies	Per capita treatment
Urban development	49	Electric power consumption (kWh per capita)	1960–2014	139	
Urban development	50	Energy use (kg of oil equivalent per capita)	1960–2015	168	
Infrastructure	51	Liner shipping connectivity index (maximum value in 2004 = 100)	2006–2020	161	
Infrastructure	52	Air transport, passengers carried	1970–2020	179	TRUE
Infrastructure	53	Air transport, freight (million ton-km)	1970–2020	176	TRUE
Infrastructure	54	Air transport, registered carrier departures worldwide	1970–2020	179	TRUE
Infrastructure	55	Container port traffic (TEU 20 foot equivalent units)	2000–2020	161	TRUE
Infrastructure	56	Rail lines (total route-km)	1995–2019	109	TRUE
Infrastructure	57	Railways, goods transported (million ton-km)	1995–2019	106	TRUE
Infrastructure	58	Railways, passengers carried (million passenger-km)	1995–2018	103	TRUE
Science & technology	59	Researchers in R&D (per million people)	1996–2018	113	
Science & technology	60	High-technology exports (% of manufactured exports)	2007–2020	165	
Science & technology	61	Technicians in R&D (per million people)	1996–2018	97	
Science & technology	62	Research and development expenditure (% of GDP)	1996–2018	134	
Science & technology	63	Trademark applications, direct resident	1980–2019	156	TRUE
Science & technology	64	Trademark applications, direct nonresident	1980–2019	157	TRUE
Science & technology	65	Trademark applications, total	1980-2019	162	TRUE
Science & technology	66	Industrial design applications, nonresident, by count	1980-2019	119	TRUE
Science & technology	67	Industrial design applications, resident, by count	1980–2019	120	TRUE
Science & technology	68	Scientific and technical journal articles	2000–2018	194	TRUE
Science & technology	69	Charges for the use of intellectual property, receipts (BoP, current US$)	1967–2020	147	TRUE
Science & technology	70	Charges for the use of intellectual property, payments (BoP, current US$)	1960–2020	164	TRUE
Science & technology	71	Patent applications, nonresidents	1980–2019	151	TRUE
Science & technology	72	Patent applications, residents	1980–2019	137	TRUE
Military	73	Military expenditure (% of GDP)	1960–2020	161	

(Continued.)

Table A1. (Continued.)

Category	ID	Name of indicator	Year range	Number of economies	Per capita treatment
Military	74	Armed forces personnel (% of total labor force)	1990–2019	169	
Military	75	Arms exports (SIPRI trend indicator values)	1960–2020	68	TRUE
Military	76	Arms imports (SIPRI trend indicator values)	1960–2020	171	TRUE

ORCID iD

Xiao-Pu Han https://orcid.org/0000-0002-3413-6665

References

[1] Ouyang Y 2016 *The Development of Bric and the Large Country Advantage* (Springer)
[2] Murach M, Wagner H, Kim J and Park D 2022 Trajectories to high income: comparing the growth dynamics in China, South Korea and Japan with cointegrated VAR models *Struct. Change Econ. Dyn.* **62** 492–511
[3] Carsten A H 2008 China's economic growth 1978–2025: what we know today about China's economic growth tomorrow *World Dev.* **36** 1665–91
[4] Ohashi H 2018 The Belt and Road Initiative (BRI) in the context of China's opening-up policy *J. Contemp. East Asia Stud.* **7** 85–103
[5] Brainerd E 1998 Winners and losers in Russia's economic transition *Am. Econ. Rev.* **88** 1094–116 (available at: www.jstor.org/stable/116862)
[6] Mahajan N and Verma S 2015 International financial integration and economic growth in India: an empirical investigation *Eurasian J. Business Econ.* **8** 121–37
[7] Dudzevičiutė G, Šimelytė A and Liučvaitienė A 2018 Government expenditure and economic growth in the European Union countries *Int. J. Soc. Econ.* **45** 372–86
[8] Faber B 2014 Trade integration, market size and industrialization: evidence from China's National Trunk Highway system *Rev. Econ. Stud.* **81** 1046–70
[9] De Haan J and Sturm J E 2000 On the relationship between economic freedom and economic growth *Eur. J. Political Econ.* **16** 215–41
[10] Acemoglu D and Autor D 2012 What does human capital do? A review of Goldin and Katz's The race between education and technology *J. Econ. Literat.* **50** 426–63
[11] Maradana R P, Pradhan R P, Dash S, Gaurav K, Jayakumar M and Chatterjee D 2017 Does innovation promote economic growth? Evidence from European countries *J. Innovation Entrepreneurship* **6** 1–23
[12] Solow R M 1957 Technical change and the aggregate production function *Rev. Econ. Statist.* **39** 312–20
[13] North D C 2005 *Understanding the Process of Economic Change* (Princeton University Press)
[14] Romer P M 1990 Endogenous technological change *J. Political Econ.* **98** S71–S102
[15] Schmookler J 1966 *Invention and Economic Growth* (Harvard University Press)
[16] Bourguignon F 2015 Revisiting the debate on inequality and economic development *Revue d'économie Politique* vol 125 pp 633–63
[17] Hidalgo C A and Hausmann R 2009 The building blocks of economic complexity *Proc. Natl Acad. Sci. USA* **106** 10570–5
[18] Tacchella A, Cristelli M, Caldarelli G, Gabrielli A and Pietronero L 2012 A new metrics for countries' fitness and products' complexity *Sci. Rep.* **2** 723
[19] Tacchella A, Mazzilli D and Pietronero L 2018 A dynamical systems approach to gross domestic product forecasting *Nat. Phys.* **14** 861–5
[20] Fagerberg J and Verspagen B 2021 Technological revolutions, structural change & catching-up *New Perspectives on Structural Change: Causes and Consequences of Structural Change in the Global Economy* ed L Alcorta, N Foster-McGregor, B Verspagen and A Szirmai (Oxford University Press)
[21] Ren Z-M, Zeng A and Zhang Y-C 2020 Bridging nestedness and economic complexity in multilayer world trade networks *Humanit. Social Sci. Commun.* **7** 156
[22] Ye Y, Xu S, Mariani M S and Lü L 2022 Forecasting countries' gross domestic product from patent data *Chaos Solitons Fractals* **160** 112234
[23] Wang C, Wang L, Xue Y and Li R 2022 Revealing spatial spillover effect in high-tech industry agglomeration from a high-skilled labor flow network perspective *J. Syst. Sci. Complex.* **35** 839–59
[24] Bettencourt L M, Lobo J, Strumsky D and West G B 2010 Urban scaling and its deviations: revealing the structure of wealth, innovation and crime across cities *PLoS One* **5** e13541
[25] Lobo J, Bettencourt L M A, Strumsky D and West G B 2013 Urban scaling and the production function for cities *PLoS One* **8** e58407
[26] Fang X-J, Cui J-Y, Hu D-D and Han X-P 2020 Relativistic regional innovation index and novel business cycle *Acta Phys. Sin.* **69** 088905
[27] Hu D-D, Fang X-J and Han X-P 2022 Advantage of populous countries in the trends of innovation efficiency *Chin. Phys.* B **31** 068903
[28] Ding C and Niu Y 2019 Market size, competition and firm productivity for manufacturing in China *Reg. Sci. Urban Econ.* **74** 81–98
[29] Head K and Mayer T 2011 Gravity, market potential and economic development *J. Econ. Geogr.* **11** 281–94
[30] Rahim S, Murshed M, Umarbeyli S, Kirikkaleli D, Ahmad M, Tufail M and Wahab S 2021 Do natural resources abundance and human capital development promote economic growth? A study on the resource curse hypothesis in Next Eleven countries *Resour. Environ. Sustain.* **4** 100018

[31] Gao J, Jun B, Pentland A S, Zhou T and Hidalgo C A 2021 Spillovers across industries and regions in China's regional economic diversification *Reg. Stud.* **55** 1311–26

[32] Dong J, Li W, Cao Y and Fang J 2016 How does technology and population progress relate? An empirical study of the last 10,000 years *Technol. Forecast. Soc. Change* **103** 57–70

[33] Hong I, Frank M R Rahwan I, Jung W-S and Youn H 2020 The universal pathway to innovative urban economies *Sci. Adv.* **6** eaba4934