Global Economic Data Research Using Python, Machine Learning, and Big-Data Visualization Techniques

Yuzheng Li
Beijing Academy International Division

Contents

1 Research Background

Since the outbreak of the COVID-19 pandemic in 2020, the global economy has suffered unprecedented shocks. Countries around the world have successively adopted large-scale monetary easing and fiscal stimulus measures to respond. With the end of the pandemic and the recovery of social activities, the global economy is gradually moving toward recovery, but challenges remain severe.

Since the end of the pandemic, the global economy has shown a high degree of divergence. The United States has maintained strong economic momentum, with high technology represented by artificial intelligence leading a new technological revolution. The Eurasian continent has been caught in turmoil, with the Russia–Ukraine crisis and the chaos in the Middle East casting shadows, causing many countries to face serious difficulties in economic growth, and immigration and refugee issues are emerging one after another. As the world's second-largest economy, China took effective measures quickly after the pandemic, implementing a series of policies to restore the economy. In 2023, China's GDP growth rate reached 8.1

The macro policies and business environments of major economies are undergoing profound transformation. These changes will not only affect global financial health but also reshape the future structure of the world economy. Against the background of global economic recovery, the movements of major economies in terms of policy and market environments have attracted the attention of numerous investors and policymakers.

On January 4, 2024, the United Nations released the report World Economic Situation and Prospects 2024, pointing out that persistently high interest rates, further escalation of conflicts, weak international trade, and the increasing number of climate disasters have brought great challenges to global growth. The report predicts that global economic growth will slow from 2.7% in 2023 to 2.4% in 2024, lower than the pre-pandemic growth rate of 3%. The report believes that greater policy coordination and international cooperation are needed. Under the current condition of tight global monetary policy, governments need to expand fiscal support to stimulate economic growth. When balancing inflation, economic growth, and financial stability, countries around the world continue to face difficult trade-offs, and it is especially necessary to deploy a wide range of macroeconomic and macro-prudential policy tools to minimize the adverse spillover effects of monetary tightening in developed economies.

Under the above background, achieving resilient economic expansion is essential to improving the quality of life and promoting overall prosperity. This goal largely depends on the implementation of effective macroeconomic strategies. However, economic development involves navigating the complex interactions of many macroeconomic factors, including inflation, interest rates, exchange rates, employment, and balance of payments (Anaripour, 2011; Ingham, 2013; Jakob, 2015; Kibria et al., 2014; Mamo, 2012; Svyrydenko et al., 2023; Tapsin Hepsag, 2014). In particular, inflation has a significant impact on economic stability and growth, prompting policymakers to adopt strategies aimed at maintaining low inflation levels while promoting strong economic performance (Zhao Brychko, 2023; United Nations Conference on Trade and Development [UNCTAD], 2019).

Because of the interconnection and mutual restriction among macroeconomic indicators, the purpose of macroeconomic regulation is to properly handle the relationships among these aspects. This study aims to conduct quantitative analysis of the overall development level of various countries and the interactions among macroeconomic indicators through a multi-level indicator system, providing useful policy reference for macroeconomic regulation.

Based on the above background, this research identifies the key drivers of economic development and their impact on economic performance, establishes an overall evaluation model of economic development, and studies the complex interactions among key macroeconomic indicators. For important key indicators affecting economic development, further step-by-step exploration is carried out through linear regression and time series analysis to analyze the differentiated driving factors of GDP growth and their future trends across economies. This study uses machine learning algorithms and panel causality tests to measure the relationships between various macroeconomic variables and economic development, attempting to clarify the complex dynamics of economic growth across countries, and to provide decision-making references and predictive models for understanding and utilizing macroeconomic indicators to promote sustainable economic development. By studying the trajectories of economic development across economies and analyzing the changing impacts of various factors, the study provides insights for national economic strategies, enabling countries to improve relevant indicators in a targeted way to narrow these gaps.

2 Construction of a Comprehensive Evaluation Model for Economic Development

2.1 Establishment of the Comprehensive Evaluation Index System

The quality of economic development is a comprehensive evaluation indicator of a country's or region's economic development status. It not only focuses on the effects of economic growth but also emphasizes the social effects closely related to economic growth. It is a value judgment of the quality level based on quantitative economic growth. Therefore, the construction of the economic development quality index should be a comprehensive reflection of growth speed, stability, and improvement of social welfare. It can provide a reference for the formulation of medium- and long-term macroeconomic policies to achieve sustainable economic and social development.

According to the connotation of economic development quality, and with reference to the commonly used indicators of the National Bureau of Statistics of China, the Worldwide Governance Indicators (WGI) of the World Bank, and the Sustainable Development Goals (SDGs) of the United Nations, this study constructs an index system of economic development quality from five dimensions: economic growth, full employment, price stability, balance of payments, and social welfare. Multiple representative secondary indicators were selected and classified under corresponding primary indicators. In order to ensure that each primary indicator has a scientific and reasonable weight, the Entropy Weight Method (EWM) was adopted to weight and calculate the composite scores of both secondary and primary indicators.

2.1.1 Construction of the Index System

When constructing the comprehensive evaluation system, the following indicators were selected according to research needs and data availability:

(1) Economic Growth: includes GDP, GDP growth rate, per capita GDP, per capita GDP growth rate, and Gross National Income (GNI) as secondary indicators.

(2) Full Employment: uses the unemployment rate as a secondary indicator to reflect the employment situation of a country.

(3) Price Stability: uses the inflation rate as a secondary indicator to measure price stability.

(4) Balance of Payments: uses the proportion of current account balance to GDP as a secondary indicator to measure a country's international economic stability.

(5) Social Welfare: uses per capita household consumption expenditure as a secondary indicator to reflect the level of national welfare.

Among them, the unemployment rate and inflation rate are inverse indicators, meaning that lower values represent better development. Therefore, before applying the Entropy Weight Method, these two inverse indicators were inversely processed so that the direction of all indicators remains consistent.

Table 1: Comprehensive Evaluation Index System of Economic Development Quality

No.	Primary Indicator	Secondary Indicator	Indicator Attribute
1	Economic Growth	GDP	Positive
2	Economic Growth	GDP Growth Rate	Positive
3	Economic Growth	Per Capita GDP	Positive
4	Economic Growth	Per Capita GDP Growth Rate	Positive
5	Economic Growth	Gross National Income (GNI)	Positive
6	Full Employment	Unemployment Rate	Negative
7	Price Stability	Inflation Rate	Negative
8	Balance of Payments	Current Account Balance as % of GDP	Positive
9	Social Welfare	Per Capita Household Consumption Expenditure	Positive

2.1.2 Calculation of Secondary Indicator Weights and Generation of Primary Composite Data

After establishing the above index system, the Entropy Weight Method (EWM) was first used to calculate the weights of the secondary indicators under each primary indicator, in order to obtain the composite data of each primary indicator. The specific steps are as follows:

(1) Data Standardization. To eliminate the effect of different measurement units, each secondary indicator was standardized. For positive indicators, the standardization formula is:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{1}$$

For negative indicators, the processing formula is:

$$x'_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)} \tag{2}$$

where x_{ij} represents the value of the j-th secondary indicator for the i-th country, and $\max(x_j)$, $\min(x_j)$ denote the maximum and minimum values of indicator j, respectively.

(2) Calculation of Proportion and Entropy Value. For the standardized secondary indicators, the proportion P_{ij} of each country on that indicator is calculated, and the entropy value e_j of the indicator is then obtained:

$$P_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}}, \quad e_j = -k \sum_{i=1}^{n} P_{ij} \ln(P_{ij} + \varepsilon) \tag{3}$$

where $k = 1/\ln(n)$, and ε is a very small constant (e.g., 10^{-6}) to prevent the zero-value issue in logarithmic computation.

(3) Calculation of Weights and Generation of Primary Indicator Composite Data. According to the entropy values, the difference coefficient is calculated as:

$$d_j = 1 - e_j \tag{4}$$

The weight of each secondary indicator is then obtained as:

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{5}$$

Using these weights, each primary indicator's composite value is calculated through weighted summation of its secondary indicators. For example, the composite data of "Economic Growth" is expressed as:

$$S_{\text{growth}} = \sum_{j=1}^{m} w_j x'_{ij} \tag{6}$$

Through the above steps, each primary indicator (such as economic growth, full employment, etc.) obtains comprehensive data that represents its overall situation.

—

2.1.3 Calculation of Primary Indicator Weights and Final Comprehensive Score

After obtaining the composite data of each primary indicator, the Entropy Weight Method is applied again to assign weights to the primary indicators and calculate the final comprehensive score for each economy. The specific steps are as follows:

(1) Standardization of Primary Indicators. Each primary indicator's composite data is standardized to eliminate the effect of scale differences.

(2) Entropy Weight Calculation for Primary Indicators. The entropy and weights are calculated following the same method as for the secondary indicators, yielding a set of weights that reflect the importance of each dimension in the overall evaluation.

(3) Comprehensive Score Calculation. The final comprehensive score of each economy is computed as:

$$E_i = \sum_{k=1}^{p} W_k S_k \tag{7}$$

where S_k is the composite data of the k-th primary indicator, and W_k is its corresponding entropy weight.

Through these steps, this study obtains each economy's comprehensive score across different years, reflecting relative performance in economic growth, employment stability, price stability, balance of payments, and social welfare. The score provides a quantitative foundation for further analysis of the comprehensive development level among countries.

2.2 Comprehensive Evaluation Results of Economic Development Quality

Since this paper focuses on medium- to long-term *structural* economic development rather than cyclical fluctuations, the study period is set to the most recent Kondratiev wave, i.e., 2008–2023 (source: Dutch economist Jacob van Duijn; China Securities). Because most countries and regions have data available in the World Bank, this paper selects 261 economies (including classified dimensions) with data recorded by the World Bank for analysis. After substituting the annual data of each economy into the indicator-system model, the visualization results are as follows.

2.2.1 Evaluation by Geographic Region in 2023

Table 1: Score Summary (2021-2023)

English Region	Score_2021	Score_2022	Score_2023	Average
North America	48.28	48.59	50.06	42.58
Europe & Central Asia	20.31	20.41	21.36	19.48
Small states	17.01	15.97	16.48	19.52
East Asia & Pacific	11.54	11.97	12.58	10.89
Central Europe and the Baltics	9.38	9.21	10.15	8.16
Latin America & Caribbean	7.86	7.87	8.66	7.54
Arab World	6.78	6.80	7.54	6.11
Middle East & North Africa	5.83	5.79	6.75	5.61
Sub-Saharan Africa	4.53	4.10	4.55	5.61
Africa Eastern and Southern	4.11	3.68	4.16	5.14
South Asia	2.47	2.36	2.94	2.29
Africa Western and Central	0.88	0.68	1.24	1.10

Figure 1: Composite scores by geographic region, 2023

From a regional perspective, in 2023 North America achieved the highest overall score, followed by Europe and Central Asia, while Central and Western Africa registered the lowest scores. This pattern is broadly consistent with the World Bank's regional classification by income level (see Figure 2), and in general aligns with the actual development status of world economic regions.

2.2.2 Evaluation by Country Using the Average Index Across Years

From a country perspective, to mitigate the influence of outlier years, the multi-year average (2008–2023) is used for cross-country comparison. The top 20 include the United States, Switzerland, Luxembourg, Australia, the United Kingdom, and other advanced economies; China ranks 49th, with a certain gap from developed countries in terms of the evaluation score. The bottom 20 include South Sudan, Afghanistan, Liberia, and others, reflecting the significant impact of war, conflict, and historical legacy on national economic development.

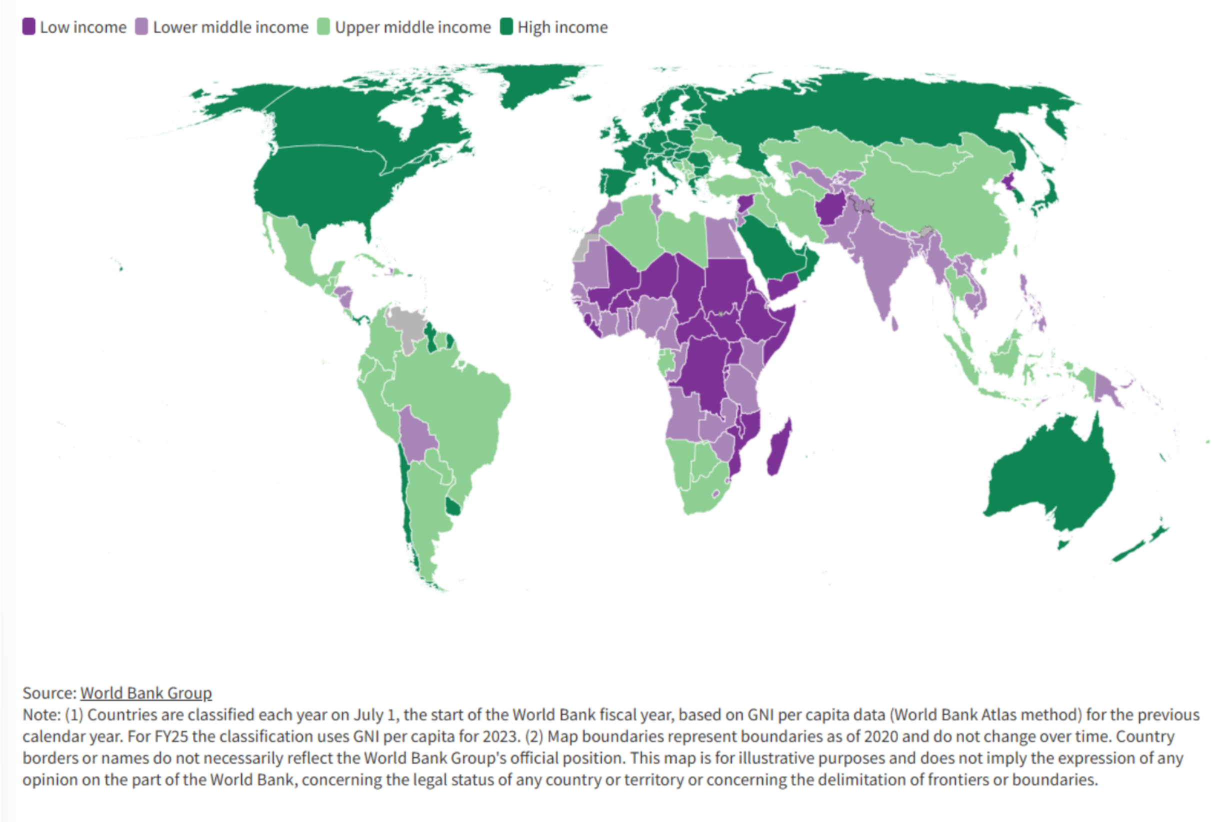

Figure 2: World Bank (2023) income-based regional classification

Table 2: Top 20 countries' Average Score

Country Name	Average
Virgin Islands (U.S.)	67.76
United Arab Emirates	30.54
Greenland	26.15
European Union	20.47
Puerto Rico	18.06
United States	17.58
Northern Mariana Islands	16.15
Faroe Islands	15.16
Switzerland	15.05
Bermuda	14.89
Luxembourg	12.87
Norway	11.60
Australia	11.11
United Kingdom	10.30
Iceland	9.61
Denmark	9.32
Hong Kong SAR, China	9.22
New Caledonia	9.18
Canada	9.12
Germany	8.68

Figure 3: Country rankings by multi-year average composite score (2008–2023)

Table 3: Top 49-60 countries' average

Country Name	Average
China	4.37
Monaco	4.14
Palau	4.11
Liechtenstein	4.07
Bahrain	4.01
Malta	3.95
Curacao	3.81
Seychelles	3.73
Slovak Republic	3.69
Estonia	3.63
Czechia	3.58
Lithuania	3.55

Figure 4: Composite scores by economic-organization category, 2023 (OECD, Euro Area, Fragile/Conflict-Affected, LDCs, etc.)

Table 4: Last 20 countries average score (2023)

Country Name	Average
Nigeria	0.80
Micronesia, Fed. Sts.	0.80
Lesotho	0.79
St. Lucia	0.79
Niger	0.79
Zambia	0.77
Maldives	0.77
Congo, Dem. Rep.	0.77
Madagascar	0.74
Papua New Guinea	0.72
Suriname	0.72
St. Vincent and the Grenadines	0.69
Kiribati	0.66
Burundi	0.64
Eritrea	0.61
Malawi	0.61
Liberia	0.60
Sao Tome and Principe	0.59
Afghanistan	0.59
Venezuela, RB	0.47
South Sudan	0.40

Figure 5: Composite scores by economic-organization category, 2023 (OECD, Euro Area, Fragile/Conflict-Affected, LDCs, etc.)

2.2.3 Evaluation by Economic-Organization Categories in 2023

Among the economic-organization categories with available data, OECD member countries show relatively high overall scores; the euro area is generally close to the level of high-income countries; countries that are fragile and affected by conflict exhibit economic distress; and there is a large gap between the least developed countries (United Nations classification) and high-income countries.

Table 5: Score Summary (2023)

Region	Score_2023
High income	38.93
OECD members	38.00
Euro area	25.31
European Union	23.35
Middle income	10.72
Fragile and conflict affected situations	3.64
Least developed countries: UN classification	2.98
Heavily indebted poor countries (HIPC)	2.80
Low income	2.75

Figure 6: Composite scores by economic-organization category, 2023 (OECD, Euro Area, Fragile/Conflict-Affected, LDCs, etc.)

The complete table of comprehensive evaluation indices for all economies from 2008 to 2023 is provided in the Appendix.

3 3. Linear Regression Analysis of the GDP Indicator

In the comprehensive index of economic development quality, Gross Domestic Product (GDP) plays a vital role as the primary indicator of a country's economic progress JimSuleiman2023,Mollah2022. This section further develops regression analysis based on this indicator. Using the expenditure approach to GDP calculation and taking the United States as an example, empirical tests are performed on World Bank sample data from 1970 to 2023. Four key variables are selected to measure GDP growth: fixed asset investment, total retail sales of consumer goods, government fiscal expenditure, and net exports. By constructing a multiple linear regression model, eliminating multicollinearity, and employing an ARIMA model to capture the time-series characteristics of GDP growth, this study establishes the GDP growth model. Through model-based prediction, governments can identify the optimal policy mix driving domestic economic growth.

According to the theoretical framework of the expenditure approach, GDP is given by:

$$GDP = C + I + G + NX$$

where C represents consumption, I represents investment, G represents government spending, and NX represents net exports. In this study, the dependent variable GDP (Y) is modeled as a function of the four explanatory variables: I (fixed asset investment), R (total retail sales of consumer goods), G (government fiscal expenditure), and NX (net exports).

—

3.1 Research Method and Procedures

Through a stepwise exploration process combining linear regression and time-series analysis, the driving factors of a country's GDP growth and its future trends are analyzed. First, a multiple linear regression model is constructed to identify major influencing factors. Then, insignificant variables are excluded to mitigate multicollinearity, and finally, an ARIMA model is employed to capture the temporal dynamics of GDP growth. The research steps and results are as follows.

3.1.1 Initial Linear Regression Analysis

A multiple linear regression model is constructed to analyze the drivers of GDP growth. Taking the United States as an example—the world's largest economy—empirical tests are performed using GDP data from 1970 to 2023. The dependent variable is the logarithm of GDP ($\ln Y$), and the independent variables are the logarithms of fixed asset investment ($\ln I$), total retail sales ($\ln R$), government expenditure ($\ln G$), and net exports ($\ln NX$). The model is specified as follows:

$$\ln Y_t = \beta_0 + \beta_1 \ln I_t + \beta_2 \ln R_t + \beta_3 \ln G_t + \beta_4 \ln NX_t + \varepsilon_t \tag{8}$$

Using Ordinary Least Squares (OLS), the regression results are obtained as shown in Table 2.

Table 2: OLS Regression Results for GDP Growth Model (1970–2023)

Variable	Coefficient	Std. Error	t-Statistic	P-Value
Constant (β_0)	-3.8257	2.722	-1.405	0.166
$\ln I$	0.1489	0.047	3.183	0.003
$\ln R$	1.0274	0.006	184.298	0.000
$\ln NX$	0.0274	0.008	3.565	0.001
$\ln G$	0.0888	0.099	0.901	0.372
R^2		0.999		

The model exhibits a high explanatory power ($R^2 = 0.999$). The coefficients of $\ln R$ and $\ln NX$ are statistically significant at the 5% level, while $\ln I$ and $\ln G$ are not significant.

—

3.1.2 Treatment of Multicollinearity

Although the initial regression model shows high goodness of fit, the Variance Inflation Factor (VIF) is computed to diagnose multicollinearity. Results indicate high VIF values for $\ln I$ (94.53) and $\ln R$ (81.64), implying strong multicollinearity. To alleviate this, the insignificant variable $\ln I$ is removed. The regression results after exclusion are shown in Table 3.

Table 3: OLS Regression Results After Removing $\ln I$

Variable	Coefficient	Std. Error	t-Statistic	P-Value
Constant (β_0)	-3.8257	2.772	-1.380	0.174
$\ln R$	1.0274	0.008	130.624	0.000
$\ln NX$	0.0274	0.009	3.215	0.002
$\ln G$	0.0888	0.098	0.904	0.370

After the exclusion of $\ln I$, the VIF for $\ln R$ decreases to 2.65, with all variables within acceptable ranges. $\ln R$ and $\ln NX$ remain significant, whereas $\ln G$ still has no significant impact on GDP growth.

—

3.1.3 Residual Autocorrelation Check

After model fitting, the residuals are tested for autocorrelation using the Durbin–Watson statistic. The value obtained is 0.391, indicating significant positive autocorrelation in the residuals. Therefore, an ARIMA model is introduced to better capture the time-series characteristics of GDP growth.

—

3.1.4 Time-Series Analysis and ARIMA Model Construction

To address residual autocorrelation, the ARIMA model is applied for time-series analysis. An Augmented Dickey–Fuller (ADF) test is conducted on the GDP series, and stationarity is achieved after first differencing. An automatic parameter-selection algorithm identifies ARIMA(0,2,1) as the optimal model based on the minimum AIC criterion. The model is expressed as:

$$\Delta^2 \ln Y_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t \tag{9}$$

where Δ^2 denotes the second-order difference operator and θ_1 is the first-order moving average coefficient. Estimation results show $\theta_1 = -0.8455$ with a P-value < 0.05, indicating significant influence. The estimated variance of residuals is $\sigma^2 = 0.0004$, suggesting low model residual variance.

—

3.1.5 Model Fitting and Forecasting

Based on the ARIMA(0,2,1) model, GDP growth data are fitted and five-year forecasts are generated. The predicted logarithmic GDP values are as follows:

Table 4: Forecasted Log GDP Values (Next 5 Periods)

Time Point	Predicted $\ln Y$
52	30.87
53	30.91
54	30.96
55	31.00
56	31.04

These predictions can be used to estimate future GDP growth trends.

—

3.1.6 Residual Diagnostics and Model Evaluation

To ensure the reliability of the ARIMA model, the Ljung–Box test is applied to the residuals, yielding a p-value of 0.562, indicating no significant autocorrelation. The Jarque–Bera test yields a p-value of 0.00, showing that residuals deviate from normality—possibly due to data characteristics—but this does not severely affect forecast robustness. Overall, the ARIMA(0,2,1) model effectively captures the time-series characteristics of $\ln Y$ and produces valid predictions. Residual analysis confirms satisfactory model fitting.

4 Conclusion

Macroeconomic indicators provide the basis for macroeconomic regulation and are fundamental to balancing the relationships among maintaining steady and relatively rapid growth, adjusting the economic structure, and managing inflation expectations. This study constructs a comprehensive evaluation index model of economic development quality based on the Entropy Weight Method (EWM) to examine the complex interactions among macroeconomic indicators. Using panel data from the World Bank for 2008–2023, it evaluates how different macroeconomic variables affect economic development across economies, aiming to inform policymaking on key issues that shape development quality.

To further explore heterogeneous development dynamics across economies, the study builds a multiple linear regression model and employs an ARIMA specification to capture the time-series properties of GDP growth. Using the United States (1970–2023) as an example, the results reveal direct and statistically significant associations between GDP growth and key variables, thereby offering references and insights for government policymakers and stakeholders. By quantifying how key drivers affect national economic performance, policymakers can utilize model-based predictions to identify the optimal combination of instruments that drive domestic growth.

The framework applies to different types of economies and reflects national economic-development conditions from a dynamic perspective, highlighting priority areas for development and providing a basis for formulating policies around the world. By decomposing the development index, the approach helps identify weak links and issue early warnings. Countries can refer to their indicator scores and their GDP linear-regression model to exercise holistic control over economic policies and guide interventions in areas falling short of average performance, adjusting key indicators toward desired ranges. In sum, this study clarifies the complex dynamics of national economic development and provides a policy tool for employing macroeconomic indicators to promote sustainable growth.

5 Limitations and Future Directions

This study also identifies limitations in the comprehensive evaluation indicator system. First, whether comparing across countries or across time, all scores—and a country's scores within a period—are subject to standard errors. These standard errors reflect the number of available data sources for a country and the degree of agreement among them (the more sources and the greater the agreement, the smaller the standard error). Because measurement approaches inherently involve imprecision, caution is required when performing cross-country and long-horizon comparisons.

Second, the current indicator selection is based on data that are more readily obtainable and on simplified operations, which cannot fully and accurately reflect reality. Future research should incorporate more variables and—based on core methodology—develop indicator systems with different levels of precision for different economies, in order to better measure and enable comparability across classification dimensions.

Third, the evaluation results are susceptible to bias caused by extreme maximum and minimum values. Economies differ in demographic structure, economic scale, and industrial composition, and such heterogeneity is difficult to capture fully in the weights of a composite index. Because the approach relates actual values to ideal and minimum values to evaluate relative development levels, changes in ideal or minimum values may alter a country's evaluation score even if its indicators remain unchanged.

As directions for improvement, one may adopt clustering methods to design indicator systems with reference values tailored to similar groups of economies, thereby adjusting composite-index weights more reasonably. For national policy design, countries can also estimate economy-specific coefficients more accurately via linear regression, to determine macro-regulatory targets and identify optimal policy trade-offs.

References

[1] United Nations (2024). *World Economic Situation and Prospects 2024*. 4 January 2024.

[2] International Monetary Fund (2024). *World Economic Outlook*. July 2024.

[3] World Bank (2024). *Global Economic Prospects*. June 2024.

[4] Luo, S., Singham, R., & Cernadas, G. (2024). Big-data across 210 economies: decoding patterns in world and Chinese economic growth. Chongyang Institute for Financial Studies, Renmin University of China (in Chinese). 5 October 2024.

[5] Chen, J., & Zhang, Y. (2013). *Spillover Effects of Macroeconomic Policies in G20 Economies*. World Economy Studies, 2013-12-18 (in Chinese).

[6] Almeida, L. M. L., & Balanco, P. A. F. (2020). Application of multivariate analysis as a complementary instrument in studies about structural changes: An example of the multipliers in the US economy. *Structural Change and Economic Dynamics*, 53, June.

[7] Pourghaz, A., Samani, E. B., & Shokri, B. (2023). Analysis of the impact of research output on economic growth using a multivariate random-effects model. *Scientometrics*, 128(4).

[8] Liu, X. *Theory and Methods of Multi-factor Impact Analysis*. School of Economics and Management, Yanshan University (in Chinese).

[9] Li, W., & Li, J. (2022). An empirical analysis of China's GDP growth based on a linear regression model. *Statistics and Applications*, 11(6) (in Chinese).

[10] Guo, W., & Huang, F. (2019). Can increasing economic freedom improve the quality of economic growth? Evidence from G20 panel data. *Journal of International Trade*, 2019(12) (in Chinese).

[11] Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2023). A Survey of Large Language Models. arXiv:2303.18223v13 [cs.CL], 24 Nov 2023.

[12] Zhou, Z. (2016). *Machine Learning*. Tsinghua University Press (in Chinese).

[13] Harrington, P. (2012). *Machine Learning in Action*. Manning Publications.

[14] Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning* (2nd ed.). Packt Publishing.

[15] Mueller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media.

[16] VanderPlas, J. *Python Data Science Handbook*. `https://jakevdp.github.io/PythonDataScienceHandbook/`