

An Integrated Multivariate Statistical Framework for Uncovering Sectoral Heterogeneity and Classifying Economic Performance in Emerging Economies

Bassey E. Effiong¹ and Maxwell A. Ijomah²

¹Dept. of Statistics, University of Calabar, Nigeria.

²Dept. of Maths/Statistics, University of Port Harcourt, Nigeria

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150900036>

Received: 29 August 2026; Accepted: 03 September 2026; Published: 03 October 2026

ABSTRACT

This study examines sectoral segmentation and classification in emerging economies using a multivariate analysis framework. The study focuses on selected economic indicators, including gross domestic product (GDP), transport and communication, education, agriculture and natural resources, health, investment, government expenditure, employment, and tourism. Factor Analysis (FA), Principal Component Analysis (PCA), K-Means clustering, Linear Discriminant Analysis (LDA), and correlation analysis were employed to identify underlying dimensions, segment observations according to economic characteristics, and evaluate the distinctiveness of the resulting groups. The results indicate that the data are suitable for factor analysis, with a Kaiser-Meyer-Olkin value of 0.620 and a statistically significant Bartlett's Test of Sphericity ($\chi^2 = 130.674$, $p < 0.001$). Factor analysis identifies a dominant economic development dimension strongly associated with GDP, investment, government expenditure, and transport. Similarly, the first three principal components explain 87.75% of the total variance, with the first component accounting for 40.89%. K-Means clustering produces three distinct groups representing relatively high-, moderate-, and low-performing economic profiles. Linear Discriminant Analysis achieves an overall classification accuracy of 86%, indicating meaningful separation among the identified groups. Furthermore, employment and tourism exhibit a moderate positive relationship ($r = 0.54$). The findings demonstrate that multivariate techniques provide complementary insights into the structural heterogeneity of economic sectors in emerging economies. However, given the relatively small number of observations, the findings should be interpreted as exploratory. Future studies should validate the identified segmentation using larger datasets, alternative clustering criteria, and independent or cross-validated classification procedures.

Keywords: Sectoral segmentation; emerging economies; multivariate analysis; factor analysis; principal component analysis; cluster analysis; discriminant analysis; economic development.

INTRODUCTION

Emerging economies play an increasingly important role in global economic growth, contributing substantially to output, employment and international trade. However, many emerging economies continue to experience structural imbalances characterised by dependence on primary commodities, uneven distribution of investment, inadequate infrastructure and differences in productivity across economic sectors (Rodrik, 2016; World Bank, 2023; IMF, 2023). These challenges make it important to understand the structure and performance of different sectors because economic activities do not respond uniformly to changes in the economic environment. While some sectors may remain relatively resilient, others may experience substantial changes in output, employment and investment. Consequently, aggregate economic indicators may conceal important differences in sectoral performance and development potential.

Traditionally, economic activities have been classified into broad categories such as agriculture, industry and services. Although this classification provides a useful overview of economic structure, it may not adequately

capture the considerable heterogeneity that exists within and across sectors. Differences in productivity, investment, employment, technological development and economic contribution can exist among activities that belong to the same broad sector (Todaro & Smith, 2020). A more comprehensive understanding therefore requires analytical methods capable of examining several economic indicators simultaneously. Indicators such as gross domestic product, transport and communication, education, agriculture and natural resources, health, investment, government expenditure, employment and tourism represent different dimensions of economic activity and may collectively reveal patterns that are not apparent when each variable is examined independently.

Multivariate statistical techniques provide useful tools for addressing this complexity. Factor Analysis can reduce a large set of correlated variables into a smaller number of meaningful underlying dimensions, thereby revealing latent patterns in economic performance (Abdi & Williams, 2010; Hair et al., 2019). Principal Component Analysis complements this approach by transforming correlated variables into a smaller number of components that explain a substantial proportion of the total variation in the data (Jolliffe & Cadima, 2016). Cluster Analysis can subsequently classify observations into relatively homogeneous groups based on their similarities, while Discriminant Analysis can assess the extent to which the identified groups are statistically distinguishable and determine the variables contributing most to their separation (Kaufman & Rousseeuw, 2009; Huberty & Olejnik, 2006).

Despite the usefulness of these techniques, they are often applied independently, limiting the ability to obtain an integrated understanding of economic-sector structures. This study therefore develops an integrated multivariate framework combining Factor Analysis, Principal Component Analysis, Cluster Analysis and Discriminant Analysis to identify underlying dimensions, classify observations and validate the resulting groups. Using selected economic indicators, the study seeks to provide a more comprehensive understanding of sectoral heterogeneity and generate evidence that can support economic planning, resource allocation and sector-specific policy formulation.

LITERATURE REVIEW

The literature on economic-sector structure emphasises the importance of understanding differences in productivity, investment, employment, economic complexity, and development potential across sectors. Although the conventional classification of economies into agriculture, industry, and services provides a useful broad description, it may conceal considerable heterogeneity within sectors. Recent evidence suggests that structural transformation in developing economies is not always characterised by a straightforward movement of labour from agriculture to high-productivity manufacturing and services. Instead, labour may move between sectors with limited productivity gains, particularly where informal employment remains dominant (Azenui, 2024; International Labour Organization [ILO], 2024). Structural transformation and productivity growth also vary considerably across countries, suggesting that aggregate economic indicators may not adequately capture sector-specific differences (Erumban & de Vries, 2024).

Recent studies further highlight the relationship between sectoral composition and economic complexity. Ketu and Ningaye (2024) found that sectoral employment composition is associated with economic complexity across African countries, with industrial and advanced-service employment generally contributing more strongly to productive capabilities than agricultural employment. Similarly, Ogbuabor et al. (2025) show that economic complexity and institutional quality are relevant to sectoral performance in Africa. These findings suggest that sectors differ substantially in their capacity to generate productive employment, technological capabilities, and economic upgrading. Evidence from Kazakhstan also indicates that aggregate productivity can be strongly influenced by a relatively small number of firms and sectors, demonstrating the granular nature of emerging-market economies (Konings et al., 2023).

Global integration and technological change provide further evidence of sectoral heterogeneity. Kumar (2023) reports that participation in global value chains does not necessarily produce uniform structural transformation across developing economies. Likewise, Inklaar et al. (2024) demonstrate that differences in sectoral productivity are associated with factors such as tradability and production structures. Climate change and

digitalisation may also affect sectors differently. Nguyen and Pienknagura (2024) find heterogeneous effects of rising temperatures across agricultural, manufacturing, and service activities, while Cook and Rani (2023) highlight the potential of digital platforms to alter employment and structural transformation in developing economies. Gupta (2024) similarly identifies business services as an increasingly important component of structural transformation in selected developing countries.

Given these complexities, multivariate statistical techniques provide valuable tools for examining several economic indicators simultaneously. Factor Analysis (FA) identifies latent dimensions underlying correlated variables and reduces numerous observed indicators into interpretable factors (Abdi & Williams, 2010; Hair et al., 2019). Principal Component Analysis (PCA) provides a complementary dimensional-reduction technique by transforming correlated variables into a smaller number of components that explain a substantial proportion of total variation (Jolliffe & Cadima, 2016). Tsoulfidis and Athanasiadis (2022), for example, demonstrate the usefulness of PCA for identifying economically important industries.

Cluster Analysis groups observations with similar characteristics and can reveal structural patterns that may not be evident from individual indicators (Kaufman & Rousseeuw, 2009). Its application to Nigerian economic data has also demonstrated the possibility of identifying meaningful sectoral groupings (Etuk, 2021). Discriminant Analysis complements clustering by evaluating the extent to which identified groups can be distinguished and by identifying variables contributing to group separation. Recent productivity evidence further confirms persistent differences in sectoral productivity, investment, labour productivity, and resource allocation across economies (OECD, 2024).

Despite these contributions, the literature frequently applies dimensional reduction, clustering, and classification techniques independently. This creates a methodological gap because the identification of latent structures, sectoral segmentation, and validation of group membership are not consistently integrated within a single framework. The present study addresses this gap by combining Factor Analysis, PCA, Cluster Analysis, and Discriminant Analysis to examine sectoral heterogeneity and economic performance. The integrated framework provides complementary evidence on underlying economic dimensions, homogeneous sectoral groups, and the statistical distinctiveness of those groups, thereby offering a more comprehensive approach to understanding sectoral structures in emerging economies.

METHODOLOGY

The study adopts a quantitative research design based on multivariate statistical techniques to examine the underlying structure and classification of selected economic indicators in Nigeria using time series data from CBN, NBS, and World Bank Databases. Secondary data from 1981 to 2023 comprising gross domestic product, transport and communication, education, agriculture and natural resources, health, investment, government expenditure, employment and tourism are analysed. The variables are standardized prior to analysis to eliminate differences arising from their measurement scales and ensure that no variable disproportionately influences the results. Factor Analysis is first employed to identify the underlying dimensions among the correlated economic indicators, with the suitability of the data assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett’s Test of Sphericity. Principal Component Analysis (PCA) is subsequently applied to reduce dimensionality and identify components explaining the largest proportion of total variance. The extracted component scores are then subjected to Cluster Analysis to classify observations into relatively homogeneous groups based on their economic characteristics. Finally, Linear Discriminant Analysis (LDA) is employed to validate the identified clusters and assess the extent to which the groups can be accurately distinguished using the selected indicators. The integrated application of these techniques provides a systematic framework for identifying latent dimensions, reducing data complexity, classifying observations and validating the resulting economic-sector groupings.

RESULTS AND DISCUSSION

Factor Analysis

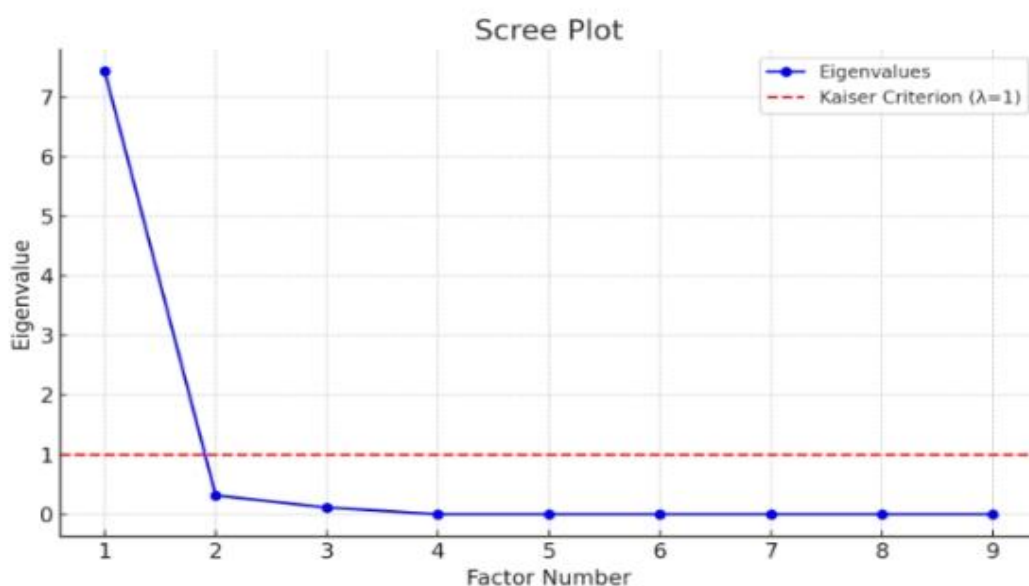
Factor Analysis (FA) was conducted to examine the underlying structure among the selected economic

indicators, namely GDP, transport and communication, education, agriculture and natural resources, health, investment, government expenditure, employment, and tourism. The analysis was intended to reduce the dimensionality of the dataset and identify latent dimensions that explain the relationships among the observed variables.

The suitability of the data for factor analysis was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's Test of Sphericity. The KMO value of 0.620 indicates an acceptable level of sampling adequacy, suggesting that the correlation structure is suitable for factor extraction. Similarly, Bartlett's Test of Sphericity was statistically significant ($\chi^2 = 130.674$, $p < 0.001$), indicating that the correlation matrix differs significantly from an identity matrix. Thus, the variables share sufficient common variance to justify factor analysis.

Table 1: Summary of Factor Loadings

Variable	Factor 1	Factor 2	Factor 3
GDP	0.9934	-0.0681	0.0670
Transport	0.9546	-0.2931	-0.0358
Energy	0.8786	-0.2352	0.2754
Agriculture	0.8900	-0.4049	0.0553
Health	0.8250	-0.4443	0.2707
Investment	0.9992	0.0389	-0.0023
Government Expenditure	0.9802	-0.1133	0.1202
Employment	0.7646	-0.2497	0.3180
Tourism	0.4151	-0.3721	0.0588



The factor loadings reveal that Factor 1 is the dominant dimension. It has particularly strong positive loadings on investment (0.9992), GDP (0.9934), government expenditure (0.9802), transport (0.9546), agriculture (0.8900), energy (0.8786), and health (0.8250). This factor can therefore be interpreted as a broad economic development and investment factor, reflecting the extent to which economic growth is associated with infrastructure, public expenditure, investment, and productive-sector activities. Factor 2 has relatively moderate negative loadings on health (-0.4443), agriculture (-0.4049), tourism (-0.3721), transport (-0.2931), and employment (-0.2497). It may therefore represent a social and sectoral development dimension, distinguishing observations according to differences in social infrastructure and productive activities.

Factor 3 has comparatively weak loadings, although employment (0.3180), health (0.2707), and energy (0.2754) contribute relatively more strongly. This suggests that Factor 3 captures sector-specific variations, particularly those associated with labour and social-sector activities.

The factor loading pattern is further supported by the scree plot, which shows a pronounced decline after the first factor. This indicates that Factor 1 accounts for a substantial proportion of the common variation in the indicators and represents the principal underlying dimension of the dataset.

Principal Component Analysis

Principal Component Analysis (PCA) was subsequently performed to validate the factor structure and determine the extent to which the principal components explain the total variability in the data.

Table 2: Principal Components of Selected Economic Indicators

Variable	PC1	PC2	PC3
G.D.P.	0.8161	0.4228	0.3906
Transport and communication	0.7188	0.4391	0.5194
Education	0.5764	0.5900	0.4817
Agriculture and natural resources	0.6041	0.4949	0.5702
Health	0.4359	0.5665	0.6339
Investment	0.8815	0.3783	0.2816
Government Expenditure (Gex)	0.7466	0.5652	0.3463
Employment	0.4678	0.6668	0.3516
Tourism	0.2388	0.5318	0.1926
Variance Explained	40.89%	27.51%	19.35%

The first three principal components jointly explain 87.75% of the total variance, indicating that a large proportion of the information contained in the original variables can be represented by three components. PC1 alone explains 40.89%, making it the most important component. It has particularly high loadings on investment (0.8815), GDP (0.8161), government expenditure (0.7466), and transport and communication (0.7188). Consequently, PC1 can be interpreted as an economic growth and investment component. PC2 explains 27.51% of the variance and is strongly associated with employment (0.6668), education (0.5900), health (0.5665), government expenditure (0.5652), and tourism (0.5318). It therefore represents a social and institutional development component. PC3 explains an additional 19.35%, with relatively strong contributions from health (0.6339), agriculture (0.5702), and transport (0.5194), indicating a sector-specific development component. The PCA results closely correspond with the factor analysis findings. Both techniques identify economic growth, investment, government expenditure, and infrastructure as major dimensions underlying the dataset. The agreement between FA and PCA provides additional evidence that the extracted structure is meaningful.

Discriminant Analysis and Cluster Classification

Linear Discriminant Analysis (LDA) was employed to determine how effectively the identified groups could be distinguished based on the selected economic indicators. The analysis produced an overall classification accuracy of 86%, indicating that the model has strong discriminatory performance.

Table 3: Classification Performance

Class	Precision	Recall	F1-Score	Support
0	0.71	1.00	0.83	5
1	1.00	0.67	0.80	6
2	1.00	1.00	1.00	3
Overall Accuracy			0.86	14

Class 2 was perfectly classified, with precision, recall, and F1-score all equal to 1.00. Class 0 also recorded perfect recall, although its precision was lower at 0.71, suggesting that some observations classified as Class 0 belonged to other groups. Class 1 achieved perfect precision but a lower recall of 0.67, indicating that some observations belonging to this group were classified elsewhere.

The overall accuracy of 86% suggests that the clusters possess meaningful economic characteristics rather than representing purely statistical groupings. The discriminant visualization further demonstrates a reasonable separation among the groups. Tourism appears to be particularly influential in distinguishing the clusters, while investment and government expenditure also contribute substantially to group separation.

Integrated Cluster Analysis and Factor Scores

Cluster analysis was performed using the factor structure to identify groups of observations with similar economic characteristics. Initial clustering suggested substantial similarity among observations; however, refinement using K-Means produced three distinct clusters.

Table 4: Cluster-wise Sectoral Summary

Cluster	0	1	2
G.D.P.	8,791.76	93,960.64	193.18
Transport & Comm	8,018.81	106,162.05	939.01
Education	12,474.22	58,656.19	379.62
Agric & Natural Resources	16,289.15	110,934.47	724.69
Health	10,818.48	53,722.98	109.22
Invest	456.21	14,050.53	11.13
Gex	771.7	5,948.43	13.36
Employ	41,119,250	62,518,880	6,661,836
Tourism	0.12	0.23	0.05

The cluster means reveal substantial differences in economic performance. Cluster 1 represents the high-performing group, recording the highest values for GDP, transport and communication, agriculture, investment, government expenditure, employment, and tourism. This indicates that strong investment and public expenditure are associated with greater economic activity and employment. Cluster 0 represents a moderate-performing group, with intermediate values across most indicators. Although its GDP and investment levels are considerably below those of Cluster 1, its employment, education, health, and agricultural indicators demonstrate a relatively developed economic structure. Cluster 2 represents the low-performing group, characterized by very low GDP, investment, government expenditure, infrastructure, employment, and tourism. The substantial differences between Cluster 1 and Cluster 2 highlight considerable structural variation in economic performance. Factor scores provide further evidence of the differences among observations.

Table 5: Factor Scores

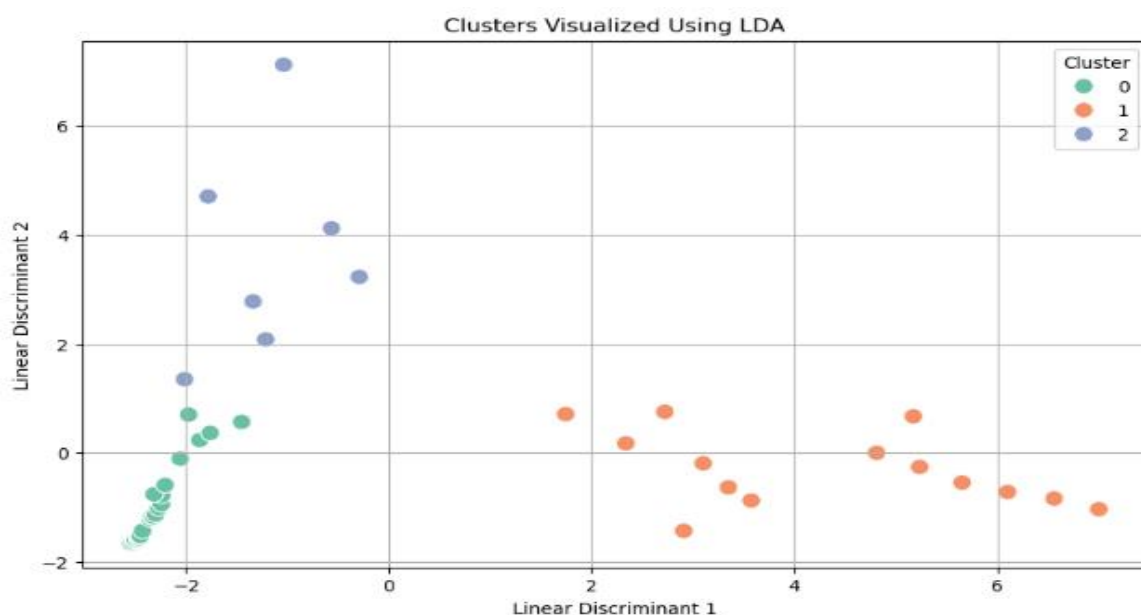
	Factor 1	Factor 2	Factor 3
0	-0.7347	-0.463	-0.8597
1	-0.736	-0.4907	-0.8081
2	-0.7365	-0.4999	-0.7679
3	-0.7367	-0.5054	-0.7501
4	-0.7385	-0.542	-0.6871

The factor scores are negative across all three factors, indicating that these observations generally fall below the standardized mean of the overall sample. Factor 1 exhibits the greatest variation and remains the most important dimension of sectoral performance, consistent with the factor and PCA results.

Correlation and Cluster Validation

As a further validation of the cluster structure, the relationship between employment and tourism was examined. The analysis produced a moderate positive correlation of $r = 0.54$, indicating that higher tourism activity tends to be associated with higher employment. This relationship is economically plausible because tourism generates employment opportunities directly through hospitality, recreation, and tourism services and indirectly through transportation, retail, food services, and other supporting activities. The cluster results reinforce this relationship: clusters with higher tourism levels also tend to exhibit higher employment.

The combined results from factor analysis, PCA, discriminant analysis, K-Means clustering, factor scores, and correlation analysis reveal a consistent underlying structure in the economic indicators. Economic growth, investment, government expenditure, infrastructure, and employment constitute the principal dimensions of economic performance, while tourism and social-sector indicators contribute importantly to the differentiation of economic groups. The results demonstrate substantial heterogeneity among the observed sectors, with clear distinctions between high-, moderate-, and low-performing groups. These findings suggest that policies aimed at improving investment, infrastructure, government expenditure, employment, and tourism development could contribute to strengthening overall economic performance.



CONCLUSION

The study examined the underlying structure and relationships among selected economic indicators using factor analysis, principal component analysis, cluster analysis, discriminant analysis, and correlation analysis. The findings indicate the presence of a dominant economic dimension largely associated with gross domestic product, investment, government expenditure, and transport. This suggests that these indicators constitute important components of the overall economic structure represented in the dataset. The cluster analysis further identified three distinct groups with different average economic profiles, indicating substantial heterogeneity among the observations. The identified groups broadly reflect high-, moderate-, and low-performing economic profiles. The results therefore suggest that economic performance is not uniform across the analysed observations, but varies according to differences in investment, infrastructure, government expenditure, employment, and other sectoral indicators.

The consistency between the factor analysis and principal component analysis provides additional evidence of a common underlying structure in the data. Similarly, discriminant analysis demonstrated that the identified groups could be distinguished with a relatively high level of classification accuracy. The positive association between employment and tourism further indicates that tourism-related activities may contribute to employment generation and broader economic activity. However, the findings should be interpreted with

caution. The classification analysis was based on a relatively small number of observations, which may limit the generalisability and statistical stability of the results. In addition, the reported analytical outputs require further consistency and robustness checks before firm conclusions can be drawn. In particular, future applications should employ clearly documented datasets, formally justified criteria for determining the optimal number of clusters, and independent or cross-validated procedures for assessing classification performance.

REFERENCES

1. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
2. Azenui, N. B. (2024). Why is labor in the SSA LDCs moving from one low productivity sector to another? *European Journal of Development Research*, 36(1). <https://doi.org/10.1057/s41287-023-00599-5>
3. Cook, S., & Rani, U. (2023). Platform work in developing economies: Can digitalisation drive structural transformation? *International Labour Organization*.
4. Erumban, A. A., & de Vries, G. J. (2024). Structural change and poverty reduction in developing economies. *World Development*, 181.
5. Etuk, E. H. (2021). Cluster analysis of Nigerian real GDP. *Ianna Journal of Interdisciplinary Studies*, 2(1), 119–132.
6. Gupta, I. (2024). The rise of business services and structural transformation in India, Kenya and South Africa: Role of female labour supply. *Foreign, Commonwealth & Development Office*.
7. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
8. Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). John Wiley & Sons.
9. Inklaar, R., Marapin, R., & Gräler, K. (2024). Tradability and sectoral productivity differences across countries. *IMF Economic Review*. <https://doi.org/10.1057/s41308-024-00271-w>
10. International Labour Organization. (2024). The centrality of structural transformation. *International Labour Organization*.
11. International Monetary Fund. (2023). Regional economic outlook: Sub-Saharan Africa—The big funding squeeze. *International Monetary Fund*.
12. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Article 20150202. <https://doi.org/10.1098/rsta.2015.0202>
13. Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.
14. Ketu, I., & Ningaye, P. (2024). Sectoral employment shares shape economic complexity: Empirical evidence from African countries. *Global Journal of Emerging Market Economies*, 16(2), 168–187. <https://doi.org/10.1177/09749101231169857>
15. Konings, J., Sagyndykova, G., Subramanian, V., & Volckaert, A. (2023). The granular nature of emerging market economies: The case of Kazakhstan. *Economics of Transition and Institutional Change*, 31(2), 429–464. <https://doi.org/10.1111/ecot.12346>
16. Kumar, R. (2023). Global value chains and structural transformation: Evidence from the developing world. *Structural Change and Economic Dynamics*, 66, 285–299. <https://doi.org/10.1016/j.strueco.2023.05.006>
17. Nguyen, H., & Pienknagura, S. (2024). Rising temperature, heterogeneous effects: Evidence from seasonal and sectoral data (IMF Working Paper No. 2024/202). *International Monetary Fund*.
18. OECD. (2024). *OECD compendium of productivity indicators 2024*. OECD Publishing. <https://doi.org/10.1787/b96cd88a-en>
19. Ogbuabor, J. E., Agu, C., Onah, B. U., & Urama, C. E. (2025). Economic complexity and sectoral performance in Africa: The role of economic governance institutions. *Sustainable Development*, 33, 4557–4576. <https://doi.org/10.1002/sd.3363>
20. Rodrik, D. (2016). Premature deindustrialization. *Journal of Economic Growth*, 21(1), 1–33.

<https://doi.org/10.1007/s10887-015-9122-3>

21. Todaro, M. P., & Smith, S. C. (2020). *Economic development* (13th ed.). Pearson.
22. Tsoulfidis, L., & Athanasiadis, I. (2022). A new method of identifying key industries: A principal component analysis. *Journal of Economic Structures*, 11, Article 2. <https://doi.org/10.1186/s40008-022-00261-z>
23. World Bank. (2023). *Global economic prospects*, June 2023. World Bank. <https://doi.org/10.1596/978-1-4648-1954-8>