An Integrated Evaluation Using Enhanced Panel Factor Model and Machine Learning: Assessing the Level and Structure of Regional Coordinated Development in the Guangdong-Hong Kong-Macao Greater Bay Area

Li Shi[1], Ting Nie[2]*
Guangzhou Huashang College, Guangzhou 51130, China[1]
Guangdong Huashang Institute for Fintech Research, Guangzhou 511300, China[2]

***Abstract*—Regional sustainable and coordinated development has become a central issue in the backdrop of a reshaped global economic landscape. Therefore, it is particularly important to evaluate the level of regional coordinated development effectively. This study aimed to validate and assess the effectiveness of machine learning algorithms and the Enhanced Panel Factor Model for evaluating regional coordinated development. To this end, based on panel data from 11 cities in the Guangdong–Hong Kong–Macao Greater Bay Area for 2005–2023, we constructed a four-dimensional composite indicator system covering economic growth, structural optimization, innovation-driven development, and social development. First, we employ a factor model to achieve dimensionality reduction and extract latent factors. SPSS and the JiekeLi platform are used for visualization, and finally, we combine LASSO regression with linear regression to build predictive models to verify the explanatory power of key factors for regional coordination. The findings indicate that the traditional factor model performs robustly in structural identification, whereas machine learning methods have advantages in variable selection and fitting accuracy. The empirical results show that the overall level of coordination in the Greater Bay Area has steadily improved; however, substantial disparities among cities remain. This study demonstrates a new pathway that integrates econometrics and machine learning for the comprehensive evaluation of regional development levels. It also conducts a comparative analysis of the applicability and effectiveness of these two methods, thereby offering significant theoretical and practical value.**

Keywords*—Comprehensive evaluation; machine learning; regional coordinated development; comparative analysis

I. Introduction

Under the new landscape of the world economy, economic systems exhibit the characteristics of complex systems, evolving from discrete blocks to system linkages and dynamic equilibrium. Building a vibrant and internationally competitive first-class bay area and world-class city cluster, serving as a demonstration zone for deep cooperation between Mainland China and Hong Kong and Macao, and establishing a model for high-quality development is the key mission of the Guangdong–Hong Kong–Macao Greater Bay Area (the Greater Bay Area). The Outline Development Plan for the Guangdong–Hong Kong–Macao Greater Bay Area explicitly calls for building an open, regionally coordinated development community characterized by connectivity and institutional alignment. Establishing a high-quality coordinated development community in the Greater Bay Area carries dual strategic significance against the backdrop of China's transition to high-quality economic development. Regional coordinated development is a dynamic process characterized by the gradual narrowing of internal disparities and achievement of orderly growth. It involves coordinated and adaptive development of economic, social, ecological, and institutional systems that evolve harmoniously. The result of synergistic development is not just a narrowing of the economic output gap, but also a dynamic and balanced development process in which the overall regional economy converges in terms of intrinsic structure and development capacity [1]. A clear assessment of the grid relationships and structures of regional development, along with identifying the grid coordination capacity and stability, is a fundamental basis for exploring pathways to sustainable, coordinated regional development.

Based on the theory of coordinated development, numerous scholars have conducted extensive empirical research on regional coordinated development. These studies mainly focus on measuring the level of regional coordination [1–9] and analyzing the constituent elements of regional coordination [10–22].

Methods for evaluating regional coordinated development levels primarily include the composite system synergy model, the coupling coordination degree model, system dynamics models, and the entropy-weight method.Shi et al. (2025) employed the Dagum Gini coefficient and a composite system synergy model to measure China's nationwide level of new-quality productive forces, and compared levels across the eastern, central, western, and northeastern regions [2]. Li et al. (2023) constructed a system dynamics model to analyze the driving factors of regional growth and the coupling–synergy effects in Yunnan's border areas [3]. Xiong et al. (2022) drawing on the Haken model and provincial panel data, measured the synergy level of the composite system linking fintech and the green transformation of manufacturing in China [4]. Huang et al. (2020) adopted a composite synergy model to examine the orderliness and composite system synergy degree between the regional innovation input system and the innovation output system in Henan Province [5]. Ma (2019) used output synergy indicators and applied a composite system synergy model to measure the level of economic coordinated development in the Beijing–Tianjin–Hebei region [6]. Chen et al. (2022) utilized the entropy-weight method and exploratory spatial data analysis (ESDA) to assess the high-quality development levels of seven urban agglomerations along the Yangtze River Economic Belt from 2009 to 2018, and employed geographically weighted regression to analyze influencing factors [7]. Yi et al. (2022) investigated the sustainable development performance of cities in Liaoning Province following a multi-criteria decision-making process, proposing a novel synergetics-based weighting method that accounts for interrelationships among sustainability criteria [8]. Zhang et al. (2021) used a coordination degree index to measure the degree of coordinated development across China's four major metropolitan areas from 2008 to 2019 [9].

In reviewing methods for assessing regional coordinated development levels, the composite system synergy model is the most widely used approach. Meanwhile, a variety of methods are continually being explored depending on the research objectives. The application of machine learning has yet to be substantively reflected in these literatures.

Analyses of regional coordination elements encompassed industrial synergy, innovation-factor synergy, efficiency synergy, comprehensive evaluation of regional high-quality development levels, and assessments of regional coordinated development levels. Ding and Zhao (2025) measured the level of industrial synergy across 16 metropolitan areas in China for the period 2013–2023 [10]. Zhou et al. (2025) examined the economic effects of different industrial coordination models within regions, including co-constructed industrial parks, headquarters economy, and industrial-chain integration [11]. Sun et al. (2025) measured the level of related industrial diversity among the Beijing–Tianjin, Beijing–Hebei, and Tianjin–Hebei subregions [12]. Qin et al. (2025) analyzed the level of regional coordinated development between the innovation chain and the industrial chain during the integration process of the Beijing–Tianjin–Hebei region, as well as the underlying drivers [13]. Lu et al. (2024) measured the level of new-quality productive forces and regional disparities across 30 provincial-level regions in China [14]. Yan et al. (2022) constructed an indicator system spanning three domains—economy and society, environment, and resource carrying capacity—to analyze the level of high-quality development and regional disparities in the Yellow River Basin [15]. Xue and Cai (2022) employed a multilevel factor analysis approach to evaluate China's overall and regional levels of high-quality development [16]. Li and Ci (2025), using panel data for 30 Chinese provinces, investigated the coupling and coordination relationships among the digital economy, carbon emission efficiency, and high-quality economic development [17]. Chang (2023) applied the entropy-weight method to measure the level of coordinated development in the Beijing–Tianjin–Hebei region across five dimensions: economy, society, ecology, government governance, and coordinated development [18].

From the above literature, analyses of regional coordination elements and structures either concentrate on the synergistic relationships among key elements—such as industry, innovation, and the digital economy—or focus on evaluating the composite level of regional aggregate indicators. Systematic assessments of regional system structures are relatively scarce.

Since the issuance of the Outline Development Plan for the Guangdong–Hong Kong–Macao Greater Bay Area, scholars have explored the coordinated development of the GBA from perspectives such as coordination mechanisms and the effectiveness of industrial synergy. Ye et al. (2022) analyzed the evolution of the GBA's coordinated development stages and the shifts in coordination mechanisms under the contexts of globalization, regional governance systems, and technological innovation [19]. Gao (2022) assessed the degree of industrial synergy in the GBA from the viewpoints of industrial planning and spatial layout [20]. Li (2024) examined inter-urban functional differences and structures in the GBA through the lens of collaborative technological innovation [21]. Wang (2025) constructed a dual-factor linkage network using enterprise investment and liner shipping data to analyze the strength of intercity linkages and synergistic effects within the Bay Area [22]. However, a systematic and quantitative evaluation of coordinated development in the GBA has not yet been clearly demonstrated.

Drawing upon the synthesized literature, current methods for measuring regional coordinated development remain primarily grounded in traditional econometric models, with machine learning algorithms not yet widely applied. Research on synergistic elements within regional systems tends to emphasize validation of single (pairwise) relationships among elements, without considering the systemic interactions among multiple intra-regional factors. Therefore, this study employs an improved spatiotemporal global factor analysis model to quantify the level of multi-element coordinated development within regions, identify the structure of the coordination network, and, concurrently, verify the applicability and effectiveness of machine learning algorithms in research on regional coordinated development.

II. Theoretical Analysis and Research Hypotheses

A. Theoretical Analysis

1) Theory of regional coordinated development: The theory of regional coordinated development holds that various elements such as the economy, population, industry, institutions, and infrastructure should be effectively aligned and interconnected among cities or units within a region, thereby promoting the optimal allocation of resources and improving overall efficiency. Regional coordination is not simply the sum of individual city developments; rather, it emphasizes functional complementarity, structural optimization, and the sharing of development outcomes among cities. At its core, it is a dynamic evolutionary process characterized by multidimensional and integrated optimization. In the case of the Great Bay Area, the region features a unique combination of "one country, two systems," three distinct customs territories, and diverse governance frameworks. These characteristics make coordinated development both challenging and strategically significant. Therefore, it is essential to systematically identify and analyze the evolutionary process and structural level of coordination using scientifically grounded assessment methods.

2) Complex systems theory and the logic of multi-factor matching: From the perspective of systems science, regional development is an open complex system composed of multiple subsystems, such as economic, innovation, social, and ecological systems, which evolve through a pathway of "coupling–coordination–optimization." Systems coordination theory emphasizes that when the interactions among these subsystems become harmonized, the overall order and efficiency of the system are enhanced. In the context of the Great Bay Area, such system coordination is reflected in aspects such as industrial chain collaboration among cities, sharing of innovation resources, and mechanisms for talent mobility. Analyzing a single dimension (such as industry or innovation) in isolation overlooks the complex coupling logic among subsystems. Therefore, adopting a method capable of integrating multiple dimensions and capturing the overall coordination dynamics is of significant importance.

B. Research Hypotheses

To gain an in-depth understanding of the evolutionary trend and structural dynamics of regional coordinated development in the Great Bay Area, this study, grounded in the theories of regional coordinated development, system coordination, and the spatio-temporal global factor analysis model, proposes the following research hypotheses:

- Hypothesis H1: The level of regional coordinated development in the Great Bay Area shows an upward trend over time.
- Hypothesis H2: The level of regional coordinated development in the Great Bay Area can be explained by several latent factors with inherent structural characteristics, forming a structural framework for coordinated development.
- Hypothesis H3: Machine learning methods are applicable to measuring and evaluating the level of regional coordination.

III. Research Design

A. Construction of the Evaluation Index System

The study of synergistic development usually includes the study of synergistic development model (process) and the study of the level (result) of synergistic development. The level of economic synergistic development refers to the result of it. This paper mainly evaluates the level of synergistic development (result) of cities in the Greater Bay Area from four perspectives: the degree of economic growth, the degree of structural optimization, the degree of innovation drive, and the degree of social development to construct the evaluation indicator system of the level of synergistic development of Guangdong, Hong Kong and Macao Greater Bay Area (as shown in Table I).

TABLE I. Evaluation Indicator System of the Level of Synergistic Development of Guangdong, Hong Kong and Macao Greater Bay Area

Primary Indicators	Secondary Indicators	Unit	Symbol	Direction
Degree of Economic Growth	GDP Sum	RMB 100 million	X_1	+
	Fiscal Revenue	RMB 100 million	X_2	+
	Fiscal Expenditure	RMB 100 million	X_3	+
	Total Investment in Fixed Assets	RMB 1 million	X_4	+
	Total Import & Export	100 million USD	X_5	+
Degree of Structural Optimization	Value Added by Secondary Industry as A Percentage of GDP	%	X_6	+
	Value Added by Tertiary Industry as A Percentage of GDP	%	X_7	+
Degree of Innovation Drive	Expenditure on Scientific Research	RMB 10,000	X_8	+
	RMB 10,000	PCS	X_9	+
	Actual Use of Foreign Capital	10,000 USD	X_{10}	+
Degree of Social Development	Per Capita Disposable Income of Urban Residents	RMB	X_{11}	+
	Total Number of Students in Higher Education	Per	X_{12}	+
	Education Expenditure	RMB 10,000	X_{13}	+
	Number of Sickbeds Per 10,000 People	Sheet	X_{14}	+
	Number of Practicing Physicians Per 10,000 People	Per	X_{15}	+

B. Data Sources and Processing

This study employs data from 15 evaluation indicators across 11 cities in the Great Bay Area over the period 2005 – 2023. Data sources included the Guangdong Statistical Yearbook, statistical yearbooks of the respective cities, the website of the Census and Statistics Department of the Hong Kong SAR Government, the website of the Statistics and Census Service of Macao, and the Wind database. Missing data were supplemented using the incremental mean and median imputation methods.

C. Model Construction

1) Enhanced panel factor analysis model: Factor analysis is a multivariate dimensionality-reduction technique that extracts common factors to explain the principal variation patterns of the original variables. The enhanced panel factor analysis model extends the traditional factor analysis approach. Because the traditional factor analysis model cannot reduce the dimensionality and misses more information on the explanatory variables, Lin (2006) [23] proposed an improved factor analysis model based on the traditional factor analysis model, which replaces "the covariance matrix of the error random vector as a diagonal matrix" in the traditional model with the maximum variance contribution of the factor explanatory variables, thus solving the above-mentioned problems. In 2009, Lin Haiming further studied the application steps of factor analysis model L and provided comprehensive evaluation steps for initial factor analysis and rotated factor analysis, respectively [24]. Zhao (2019) and Shi (2020) provided application cases and improved comprehensive evaluation steps when factor analysis model L was applied to cross-sectional data and time series data [25]-[26]. Therefore, this study adopts an Enhanced Pannel Factor Model to evaluate the level of coordinated economic development in the Great Bay Area. The factor analysis model is given by Eq. (1) as follows:

$$\mathrm{X}_{it} = \Lambda \mathrm{F}_t + \epsilon_{it} \quad (i = 1, \dots, N;\ t = 1, \dots, T) \tag{1}$$

where, Xit represents the vector composed of all observed variables at spatial location i and time point t, Λ is the factor loading matrix, F denotes the global common factor vector, and ϵit is the spatiotemporal-specific residual vector.

The specific evaluation steps are as follows:

- Data Preprocessing: Global positive direction transformation and standardization of evaluation indicators. All indicators in this sample are positive indicators. For positive indicators, apply min-max normalization (see Eq.(2)):

$$x_{ijt}^* = \frac{x_{ijt} - \min(x_{\mathrm{ijt}})}{\max(x_{\mathrm{ijt}}) - \min(x_{\mathrm{ijt}})} \tag{2}$$

where, $\mathrm{x}_{\mathrm{ijt}}^*$ represents the data of the i city for the j-th indicator in year t. The normalized data are then further standardized using standard deviation normalization as follows, see Eq. (3):

$$z_{ij} = \frac{x'_{ij} - \mu_j}{\sigma_j} \tag{3}$$

Where μ_j is the mean of the j-th indicator, and σ_j is the standard deviation. Both are calculated over all regions and all time periods.

- Determination of applicability of factor analysis model: The Pearson correlation coefficient matrix R is calculated for all variables to assess whether a factor structure exists，as shown in Eq. (4). Let r_{jk} denote the correlation coefficient between the j-th and k-th indicators, then:

$$r_{jk} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ijt} - \bar{x}_j)(x_{ikt} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ijt} - \bar{x}_j)^2} \cdot \sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ikt} - \bar{x}_k)^2}} \tag{4}$$

If there is a simple linear correlation coefficient between indicators exceeding 0.8, it indicates that there is a common factor between indicators, and a factor analysis model is applied. Furthermore, the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity are employed to evaluate whether the shared variance among variables is statistically significant, thereby confirming the appropriateness of factor analysis.

- The principal factors are extracted using the maximum variance method (which maximizes the explanatory power of the factors). Suppose m factors are to be extracted, such that the cumulative explained variance reaches a predefined threshold (e.g. ⩾ 85%). After extraction, an orthogonal rotation method such as Varimax is applied to enhance the interpretability of the factors, resulting in a rotated factor loading matrix Λ^*.

- Estimation of Factor Scores and Comprehensive Score Calculation: The regression method is used to estimate the score of each sample on each factor. Let F_{ikt} denote the score of city i in year t on factor k. The comprehensive score is then calculated as a weighted sum of the factor scores, see Eq. (5).

$$S_{it} = \sum_{k=1}^{m} w_k F_{ikt} \tag{5}$$

$$\mathrm{w_k} = \frac{\lambda_k}{\sum_{j=1}^{m} \lambda_j} \tag{6}$$

where, the weights being the variance contribution rate of each factor λ_k，see Eq. (6).

- Calculate the scores of the factors, the scores of the combined factors and the ranking: Calculate and rank the score F_i for t×n sample factors and the composite factor score Fsum $==\alpha_1 F_1 + \alpha_2 F_2 + \ldots + \alpha_k F_k$ (α_i is the variance contribution of each factor).

2) Application of machine learning: A Comparative Analysis of Linear Regression and LASSO Regression Models.

To further test the validity and explanatory power of the common factors extracted through factor analysis and to investigate the explanatory ability of factor scores on the level of regional coordinated development, this study introduces two machine learning methods: linear regression and Least

Absolute Shrinkage and Selection Operator Regression (LASSO) regression to validate and compare the modeling results of factor analysis. This approach not only helps to uncover the linear relationship between factor scores and the level of coordinated development but also provides data support for variable selection, thereby enhancing the interpretability and robustness of the model. The specific applications of this method are as follows:

- Linear regression Modeling: The principal factor scores extracted through factor analysis were used as explanatory variables, while the level of regional coordinated development served as the dependent variable. A multiple linear regression model was constructed to verify the statistical significance and explanatory validity of the common factors identified through the factor analysis.
- LASSO Regression: Variable Selection and Model Simplification: Given the potential multicollinearity among certain factor scores, ordinary multiple linear regression may perform unreliably in high-dimensional settings. To enhance explanatory efficiency and achieve feature compression, this study introduces the Least Absolute Shrinkage and Selection Operator (LASSO) regression model for comparison. By incorporating an L1-norm penalty into the loss function, LASSO shrinks some regression coefficients to zero, thus accomplishing variable selection and parameter estimation simultaneously. This preserves model interpretability while effectively reducing complexity and mitigating overfitting.

D. Software Platforms and Tools

This paper uses several advanced software platforms and tools for analysis and modeling, selected for their robustness, user-friendliness, and compatibility with the research objectives.

1) Jiekeli intelligent modeling platform: The Jiekeli Intelligent Modeling Platform integrates various data preprocessing, statistical modeling, and machine learning algorithms. With its high level of automation, interactive visualization, and powerful computing capabilities, the platform provides strong support for complex data modeling and interpretation of the results in this study. Specifically, the main functions applied in this research include automated data preprocessing, which automatically handles missing value imputation, standardization, normalization, and other basic preprocessing tasks, thereby improving data quality and modeling efficiency. Support for linear modeling and variable selection: Equipped with built-in modules for multiple linear regression and LASSO regression, the platform facilitates efficient modeling between factor scores and regional coordinated development levels, helping identify key variables and enhance model interpretability. Result visualization and model evaluation: It offers a variety of visual tools, such as goodness-of-fit plots, residual distribution charts, and coefficient path diagrams, making it easier to evaluate model performance and interpret results, thus increasing the transparency and rigor of the research.

2) SPSS Analytical tool: As a standard tool for professional statistical analysis, SPSS offers powerful capabilities for data reduction and latent structure exploration through its factor-analysis module. With an intuitive graphical interface, SPSS supports exploratory factor analysis (EFA), which was mainly used for the following purposes: variable reduction: core common factors were extracted using Principal Component Analysis (PCA) or Maximum Likelihood Estimation (MLE) to effectively reduce variable dimensionality; structural validity testing: the suitability of the data was assessed through correlation coefficients (r), the Kaiser-Meyer-Olkin (KMO) measure, and Bartlett’s test of sphericity; factor rotation optimization: the software provides various rotation methods, such as Varimax and Direct Oblimin, to enhance the interpretability of the extracted factors.

TABLE II. VARIANCE CONTRIBUTION EXPLAINED BY EACH ROTATED FACTOR

Factor	Variance Contribution λ_i	Variance Contribution (%)	Cumulative Variance Contribution (%)
1	5.047	33.646	33.646
2	3.298	21.988	55.633
3	3.053	20.350	75.984
4	1.494	9.961	85.945
5	1.436	9.576	95.521

IV. EMPIRICAL RESULTS AND ANALYSIS

A. Enhanced Panel Factor Structure Analysis

This study adopts the comprehensive evaluation steps of the enhanced panel factor analysis model to assess the synergistic development level of the Great Bay Area. The dataset comprises p = 15 indicators, denoted as X_1 to X_{15}, covering t = 19 years from 2005 to 2023, with n = 11 cities sampled each year. All indicators X_1 to X_{15} are positive, so no direction adjustment is required. These indicators are globally standardized, with the standardized variables denoted as x_1 to x_{15}. The simple linear correlation coefficients among the indicators were calculated; in particular, the correlations between X_1 and X_2, X_3, and X_4 are 0.874, 0.914, and 0.916, respectively, all exceeding 0.8, indicating the presence of common factors suitable for factor analysis. The rotated factor loading matrices $\boldsymbol{L}_3^{\Gamma},\ldots,\boldsymbol{L}_{15}^{\Gamma}$ were calculated, from which $\boldsymbol{L}_3^{0},\boldsymbol{L}_3^{\Gamma},\ldots,\boldsymbol{L}_{15}^{\Gamma}$ were obtained, and the arithmetic means of the maximum absolute values of each row's elements are denoted as $l_3^{0},l_3^{\Gamma},\ldots,l_{15}^{\Gamma}$.

The explanation of variance for each factor is shown in Table II. The first five factors are selected, the first factor (denoted as F1) explained 33.646%, the second factor (denoted as F2) explained 21.988%, the third factor (denoted as F3) explained 20.350%, the fourth factor (denoted as F4) explained 9.961%, and the fifth factor (denoted as F5) explained 9.576%, with a cumulative extractable variance contribution of 95.521%. F1 is named as "Economic Growth Cofactor" with positive direction; F2 is named as "Innovation Cofactor" with positive direction; F3 is named as "Cofactors of population income and tertiary industry compared with secondary industry", with positive direction; F4 is named as "Cofactor for the number of people in higher education", with a positive direction; F5 is named as "Medical condition cofactor", with a positive direction.

B. Analysis of Enhanced Panel Factor Scores

1) Regional comprehensive scores: Table III is obtained by taking the mean value of the comprehensive scores of each city in the same year as the total evaluation score of the level of synergistic development in the Greater Bay Area over the years, and drawing the corresponding trend line diagram as shown in Fig. 1.

By combining Fig. 1 and Table III, it can be seen that the level of synergistic development (Fsum) of the Greater Bay Area shows an overall upward trend. Among them, the growth is slow around 2008, grows to the average level in 2013, maintains a high level of growth rapidly between 2010 and 2019, and decreases slightly in 2020 due to the epidemic and other factors. It indicates that there is a growing synergy among the Greater Bay Area cities, and the synergistic trend tends to grow steadily.

TABLE III. COMPOSITE SCORE TABLE OF LEVEL OF ECONOMIC SYNERGISTIC DEVELOPMENT OF THE GREATER BAY AREA FROM 2005-2023

Year	F_1	F_2	F_4	F_5	F_{sum}
2005	-0.193	-0.437	0.077	-0.792	-0.864
2006	-0.163	-0.426	0.012	-0.555	-0.699
2007	-0.176	-0.344	-0.001	-0.557	-0.687
2008	-0.144	-0.305	0.055	-0.646	-0.697
2009	-0.210	-0.254	0.095	-0.561	-0.566
2010	-0.112	-0.226	0.089	-0.538	-0.516
2011	-0.032	-0.197	0.069	-0.380	-0.354
2012	0.000	-0.163	0.063	-0.155	-0.122
2013	0.036	-0.113	0.031	0.013	0.032
2014	0.139	-0.135	0.029	0.152	0.182
2015	0.216	-0.034	-0.005	0.318	0.353
2016	0.155	0.189	-0.073	0.523	0.526
2017	0.134	0.341	-0.078	0.648	0.680
2018	0.095	0.583	-0.131	0.778	0.803
2019	0.107	0.699	-0.137	0.896	0.937
2020	0.147	0.821	-0.092	0.855	0.995
2021	0.189	0.943	-0.086	0.924	1.089
2022	0.231	1.065	-0.08	0.993	1.183
2023	0.273	1.187	-0.074	1.062	1.277

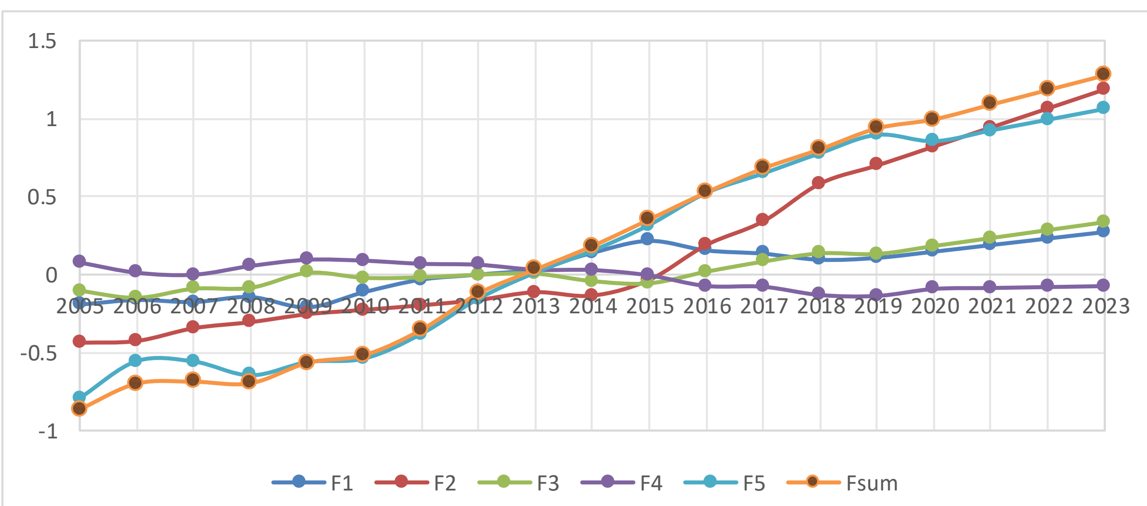

Fig. 1. The trend of level of synergistic development of the Greater Bay Area from 2005-2023.

Economic Growth Cofactor (F1) shows an overall slowly fluctuating upward trend, with the lowest level in 2009, then increasing year by year, reaching the average level in 2012 and the highest level in 2015, and then showing a partial fluctuating downward trend. Innovation Cofactor (F2) shows an overall upward trend, being below average and growing slowly in 2015, but above average and growing faster in 2016 and beyond. Cofactors of population income and tertiary industry compared with secondary industry (F3) show an overall slowly fluctuating upward trend, below average until 2016 (except for 2009 and 2013) and reaching average and slowly increasing trend in 2016 and beyond. Cofactor for the number of people in higher education (F4) shows a small fluctuating overall decreasing trend. Medical condition cofactor (F5) shows an overall upward trend, fluctuating slightly before 2010, increasing faster after 2010, and reaching above average and stable growth in 2013.

The above results, combined with the standard deviation of each factor (results in Table IV), show the changes in each factor. There is an increase and a large change in the F sum and Medical condition cofactor (F5) and Innovation Cofactor (F2); Cofactors of population income and tertiary industry compared with secondary industry (F3) and Economic Growth Cofactor (F1) have increased but not much, indicating that the Greater Bay Area has made relatively small progress in the above two directions; Cofactor for the number of people in higher education (F_4) has decreased but not significantly, indicating a small regression in this direction for the Greater Bay Area.

2) City score: The average comprehensive scores of each city in the Great Bay Area over the years were considered as the overall evaluation scores of their coordinated development levels. The corresponding trend lines are shown in Fig. 2. Three cities have a positive composite score for the level of economic synergistic development, namely Guangzhou, Hong Kong and Zhuhai, with Guangzhou having the highest score for the level of economic synergistic development; the rest of the cities have a negative and below average score. The cities ranked from highest to lowest were Guangzhou, Hong Kong, Zhuhai, Zhongshan, Foshan, Macau, Shenzhen, Dongguan, Jiangmen, Huizhou, and Zhaoqing. In terms of economic growth factor (F1), Hong Kong ranks the highest and Macau the lowest; in terms of Innovation Cofactor (F2), Shenzhen ranked the highest and Hong Kong the lowest; in terms of cofactors of population income and tertiary industry compared with secondary industry (F3), Macau ranks the highest and Foshan the lowest; cofactor for the number of people in higher education (F4); Guangzhou ranks the highest and Shenzhen the lowest; medical condition cofactor (F5), Zhuhai ranks the highest and Zhaoqing the lowest.

Based on the trend graph of scores of Fsum for each city in the Greater Bay Area over the years (see Fig. 3 for details, here the Fsum is used as an example, all other factors can be analyzed similarly), we can see that in general, the level of synergistic development in all cities shows a generally increasing trend year by year, but the levels differ. Guangzhou, Hong Kong, and Zhuhai stand out with higher scores, while other cities show varied growth trends.

TABLE IV. STANDARD DEVIATION OF EACH FACTOR OF ECONOMIC SYNERGY FOR THE GREATER BAY AREA

Factor	F_1	F_2	F_3	F_4	F_5	F_{sum}
Standard Deviation	0.147	0.404	0.093	0.079	0.600	0.650

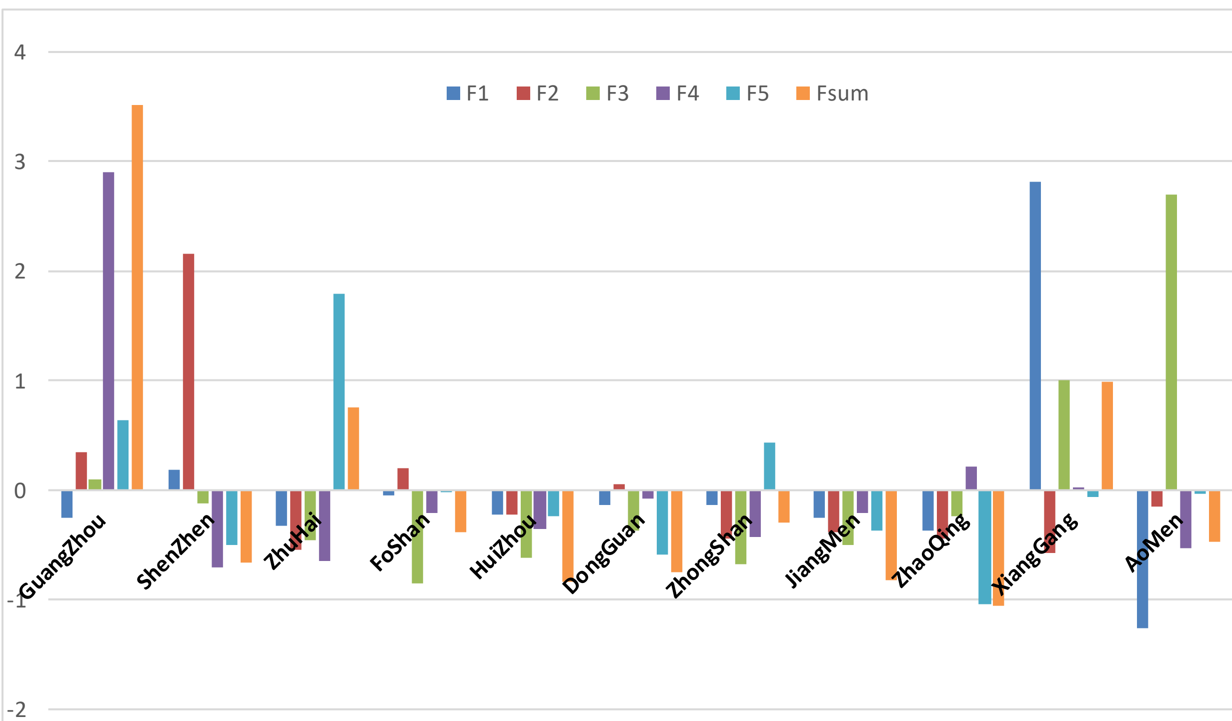

Fig. 2. Comparison of the level of synergistic development by city.

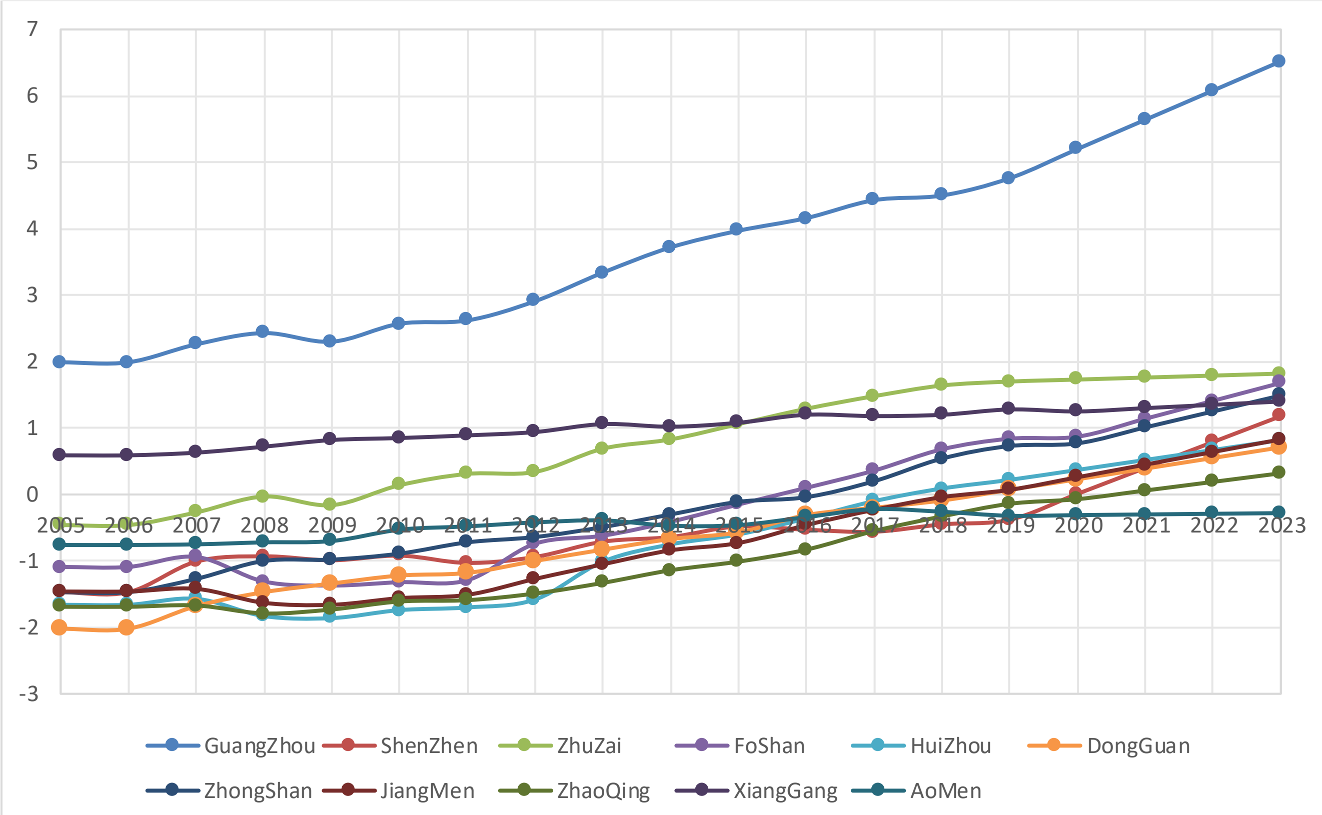

Fig. 3. The composite scores of levels of synergistic development for each city in the greater bay area.

TABLE V. GROWTH RATE AND TOTAL CHANGE OF THE COMPOSITE SCORE OF THE LEVEL OF ECONOMIC SYNERGISTIC DEVELOPMENT OF EACH CITY

City	Growth Rate	No.	Total Change	No.
Guangzhou	0.2256	**1**	3.03	1
Shenzhen	0.0898	9	2.03	6
Zhuhai	0.1656	**3**	2.32	5
Foshan	0.1565	5	1.34	9
Huizhou	0.1796	**2**	2.36	3
Dongguan	0.1491	6	2.36	4
Zhongshan	0.1587	**4**	2.39	2
Jiangmen	0.1442	7	1.77	8
Zhaoqing	0.1355	8	1.85	7
Hong Kong	0.0448	10	0.57	10
Macau	0.0339	11	0.42	11

Note: Growth rate = slope of the regression model with calendar year scores as the dependent variable and time as the independent variable; total change = $F_{sum\,2023}$ - $F_{sum\,2005}$.

By calculating the standard deviation and total change of the level of synergistic development of each city over the years (see Table V), it can be seen that the top three cities with the largest growth rate of the level of synergistic development are Guangzhou, Huizhou and Zhuhai, the smallest are Macau, Hong Kong and Shenzhen, while Zhongshan, Foshan, Dongguan, Jiangmen and Zhaoqing are in the middle level. To conclude, in the process of synergistic development of Guangdong, Hong Kong and Macau, the nine mainland cities have changed more than Hong Kong and Macau, and Hong Kong and Macau have a greater synergistic effect on the nine mainland cities.

C. Evaluation Results of Machine Learning Regression Models

To further verify the scientific validity and accuracy of the comprehensive score Fsum obtained from the factor analysis method and its factor structure, and to enhance the reliability of the research results, this paper introduces two machine learning models - multiple linear regression and LASSO regression - to empirically analyze the relationship between the main feature variables and Fsum. The model evaluation results are as follows:

As shown in Tables VI and VII, both regression models achieve high goodness-of-fit, indicating strong explanatory power of the feature variables for the comprehensive factor score (Fsum). The linear regression model reveals that the number of enrolled university students, hospital beds per 10,000 people, and the proportion of tertiary industry value added are core positive influencing factors of Fsum, with some macroeconomic variables showing statistical significance. Other variables have limited impact. The LASSO regression model automatically selects a small number of highly explanatory variables, further highlighting the key roles of education and healthcare resources and industrial structure on Fsum. The conclusions from both models are highly consistent and mutually validating, enhancing the scientific rigor and robustness of the study results.

TABLE VI. COMPARISON OF OVERALL GOODNESS-OF-FIT FOR REGRESSION MODELS

Model Type	R^2	Adjusted R^2	Mean Squared Error (MSE)
Linear Regression	0.81	0.79	0.41
LASSO Regression	0.85	0.81	0.38

TABLE VII. COMPARISON OF REGRESSION COEFFICIENTS AND THEIR SIGNIFICANCE FOR KEY FEATURES

Variable Name	Linear Regression Coefficient	Linear Regression p-value	LASSO Regression Coefficient	Variable Effect Description
(Intercept)	-5.30	<0.001***	0	Intercept
Number of University Students	0.0000034	<0.001***	0.77	Major positive influencing factor
Hospital Beds per 10,000 People	0.044	<0.001***	0.54	Major positive influencing factor
Share of Tertiary Industry (%)	0.044	<0.001***	0.09	Major positive influencing factor
Total GDP (RMB 100 million)	-0.00013	0.006**	0	Significant in linear regression, shrunk to zero in LASSO
Share of Secondary Industry (%)	0.033	0.002**	0	Significant in linear regression, shrunk to zero in LASSO

Note: *p<0.05, **p<0.01, ***p<0.001. A LASSO coefficient of 0 indicates that the variable was excluded by the model.

V. CONCLUSIONS

This study integrates a panel factor model with machine learning algorithms to assess the level and structural determinants of regional collaborative development in the Guangdong-Hong Kong-Macao Greater Bay Area. The results demonstrate that the factor model effectively synthesizes redundant information, isolates key latent factors, and achieves high explanatory power, with the top three factors accounting for 76% and the top five accounting for 96% of the total variance. Linear regression confirms significant relationships between factors and collaborative development, whereas LASSO regression exhibits superior robustness and sparsity in high-dimensional contexts. Consistency across methods strengthens the scientific validity of the findings, which indicate steady overall progress but persistent regional disparities, suggesting scope for further optimization.

Theoretically, this work demonstrates the value of combining traditional econometric models, characterized by strong theoretical foundations, interpretability, and stability, with machine learning methods, which excel in feature selection, pattern recognition, and predictive accuracy. This integration unites explanatory depth with predictive precision, and offers a more rigorous and robust framework for regional development assessment. Practically, it highlights the importance of addressing both structural drivers and dynamic prediction in strategy design, enabling the identification of industrial, talent, and technological levers, as well as the early detection of disparities and bottlenecks.

The methodological framework presented here is transferable to other economically significant city clusters, such as the Beijing-Tianjin-Hebei region, the Yangtze River Delta, the Tokyo Bay Area, and the London metropolitan area. Beyond regional studies, it can be applied to domains such as real estate market health assessment, industrial chain resilience, energy structure optimization, transportation network efficiency, technological innovation capacity, and cross-regional economic integration.

The machine learning models employed in this study are relatively limited in scope, and the findings may therefore be subject to certain constraints. Future research will broaden methodological diversity and delve more deeply into the integrated application of econometric and machine learning approaches to complex socio-economic development challenges.

ACKNOWLEDGMENT

The authors gratefully acknowledge the intelligent AI assistant Kimi for its valuable assistance in translating several technical terms during the preparation of this manuscript.

The authors would also like to express sincere gratitude to the Guangzhou Huashang College National Scientific Research Project Team (2024HSKT02), the Construction Project of the Core Curriculum Teaching and Research Office for Financial Technology, Guangzhou Huashang College (HS2024ZLGC43), and the Research Project on the Mechanism Innovation and Implementation Path of High-Quality Development of SMEs throughout the Life Cycle Driven by Fintech in the Guangdong-Hong Kong-Macao Greater Bay Area (No. 2022ZDJS125) for their valuable contributions in topic analysis and journal recommendations.

REFERENCES

[1] Zhang M. Y., Quan R. 2020. Evaluation of coordinated development in the Beijing-Tianjin-Hebei region. Statistics and Decision, 36(04), 72–76. (In Chinese).

[2] Shi X. T., Yu Z. Y. 2025. Measurement of the new quality productivity level, regional differences, and synergy of the composite system. Science & Technology Progress and Policy, 42(07), 9–20. (In Chinese).

[3] Li J. M., Yin Q., Zhang Y. Z., Qi J. J. 2023. Synergistic effect analysis of "open corridor-border port-border regional economy" composite system. Commercial Economy Research, (05), 171–175. (In Chinese).

[4] Xiong X. L., Fan J. 2022. Coordinated evolution mechanism and regional coordination differences between fintech and green transformation of manufacturing: Empirical analysis based on the Haken model. Science & Technology Progress and Policy, 39(11), 63–73. (In Chinese).

[5] Huang G. L., Xu R. Y., Wang X. Z., Li M. B. 2020. Evaluation of regional innovation synergy in Henan Province based on the composite system synergy model. Journal of Henan Agricultural University, 54(06), 1059–1066. (In Chinese).

[6] Ma X. 2019. Evaluation of economic synergy in the Beijing-Tianjin-Hebei region based on the composite system synergy model. Industrial Technology Economy, 38(05), 121–126. (In Chinese).

[7] Chen Y., Miao Q. Q., Zhou Q. 2022. Spatiotemporal differentiation and driving force analysis of the high-quality development of urban agglomerations along the Yellow River Basin. International Journal of Environmental Research and Public Health, 19, 2484.

[8] Yi P., Dong Q., Li W., et al. 2022. Assessment of city sustainability with the consideration of synergy among economy–society–environment criteria. Environment, Development and Sustainability. https://doi.org/10.1007/s10668-022-02364-w

[9] Zhang P., Liu Z., Liu X., et al. 2021. Does functional specialization affect collaborative development? Evidence from four Chinese metropolitan areas. Environmental Science and Pollution Research, 28, 41851–41868.

[10] Ding R. Z., Zhao X. Z. 2025. Measurement of industrial synergy in China's metropolitan areas. Journal of Tongji University (Social Science Edition), 36(03), 89–100. (In Chinese).

[11] Zhou J. K., Han L., Zhang X. Y., et al. 2025. Industrial coordinated development in the Beijing-Tianjin-Hebei region: Model, performance, and policy choices. Economic Geography, 45(03), 238–245. (In Chinese).

[12] Sun T. S. 2025. Industrial diversification and regional economic resilience in Beijing-Tianjin-Hebei: Based on the perspective of intra- and inter-regional industrial related diversity. Tianjin Social Sciences, (01), 107–116. (In Chinese).

[13] Qin P., Zhang Y. J., Xiao Q. Y. 2025. Research on the cross-regional coordination mechanism of innovation chain and industrial chain in Beijing-Tianjin-Hebei—Spatio-temporal evolution analysis based on strategic emerging industries. Journal of Beijing Union University (Humanities and Social Sciences Edition), 23(04), 1–14. (In Chinese).

[14] Lu J., Guo Z. A., Wang Y. P. 2024. Development level, regional differences, and promotion path of new quality productivity. Journal of Chongqing University (Social Science Edition), 30(03), 1–17. (In Chinese).

[15] Yan L. J., Zhao Y. J., Qiu S. K., et al. 2022. Construction and evaluation of high-quality development indicator system in the Yellow River Basin—Taking Henan section as an example. Regional Research and Development, 41(06), 37–43. (In Chinese).

[16] Xue W., Cai C. 2022. Comprehensive evaluation of high-quality development in China based on multilevel factor analysis. Statistics and Decision, 38(18), 22–25. (In Chinese).

[17] Li F. S., Ci F. Y. 2025. Spatiotemporal evolution and driving factors of coupling coordination among China's digital economy, carbon emissions efficiency, and high-quality economic development. Sustainability, 17(14), 6410.

[18] Chang S. X., Xue Q., Xu S. Y. 2023. Dynamic evaluation and comparison of the coordinated development level in the Beijing-Tianjin-Hebei region based on the entropy method. Industrial Innovation Research, (15), 22–24. (In Chinese).

[19] Ye Y. Y., Wang X. Y., Xu J. L., et al. 2022. Connotation and mechanism changes of coordinated development in the Guangdong-Hong Kong-Macao Greater Bay Area in the new era. Tropical Geography, 42(02), 161–170. (In Chinese).

[20] Gao J. Y. 2022. Evaluation of industrial policy effects and industrial synergy degree in the Guangdong-Hong Kong-Macao Greater Bay Area. Science and Technology Management Research, 42(12), 53–58. (In Chinese).

[21] Li Y. C., Peng W. K., Yang Y. H. 2024. Research on urban functional differences and guiding strategies in the Guangdong-Hong Kong-Macao Greater Bay Area: An analysis based on the technological innovation network perspective. Urban Problems, 2024(11), 54–62. DOI:10.13239/j.bjsshkxy.cswt.241107. (In Chinese).

[22] Wang S. B., Qin Y., Lin X., et al. 2025. Network structure and synergy characteristics in the Guangdong-Hong Kong-Macao Greater Bay Area. Applied Sciences, 15(14), 7705. https://doi.org/10.3390/app15147705

[23] Lin H. M. 2006. The exact model of factor analysis and its solution. Statistics & Decision, 2006(14), 4–5. (In Chinese).

[24] Lin H. M. 2009. Improvement and application of factor analysis model. Journal of Applied Statistics and Management, 28(06), 998–1012. (In Chinese).

[25] Zhao H. Q., Zhu J. P. 2019. How to calculate factor analysis application results with SPSS software. Statistics & Decision, 35(20), 72–77. (In Chinese).

[26] Shi L., Lin H. M. 2020. Statistical evaluation and empirical evidence of the degree of economic openness—Guangdong Province as an example. Statistics & Decision, 36(24), 89–93. (In Chinese).