Journal of Engineering, Project, and Production Management
2026, 16(1), 2025-198

International Economic Analysis Based on FP-Growth and Pearson Model

Gaihong Lu
Associate Professor, School of Economics and Management, Guangdong Open University, Guangzhou 510091, China,
E-mail: jkpasl66@outlook.com

Project Management
Received September 4, 2025; revised October 24, 2025; accepted October 30, 2025
Available online December 8, 2025

Abstract: In the context of global economic integration and a complex, volatile international situation, accurately identifying and quantifying the core factors and constraints influencing the development of the international economy holds significant practical and research value for economists, policymakers, data scientists. However, existing analytical methods have limitations such as excessive deviation, insufficient accuracy, and low efficiency. To address these issues, this study utilizes the hypergraph frequent pattern mining algorithm to optimize the frequent pattern growth algorithm, enhancing its efficiency in mining association rules for high-dimensional sparse economic data. Additionally, the Spearman’s rank correlation coefficient was introduced to improve the Pearson model, enabling a comprehensive quantification of both linear and nonlinear economic variable relationships. Meanwhile, by integrating the improved frequent pattern growth algorithm with the improved Pearson model, an analysis framework and model for international economic development factors were constructed. The experimental results show that this model achieves an accuracy of 98.4% and a precision of 89.7% in high-noise environments, with a theoretical consistency of 0.889, and the computation time and memory usage were only 147.3 ms and 138.4 MB, respectively. The average sensitivity across different datasets was 94.33%, the support degree of relationship recognition was 95.2%, the data coverage rate was 93.9%, a variable correlation consistency rate of 96.5%, and an error rate of only 0.41%. The robustness and generalization of the model were verified. All these results are superior to those of the comparison model, fully demonstrating the feasibility and superiority of the proposed model. It provides economists with a more precise tool for mining the correlation of international economic variables and offers precise data support for policymakers to formulate economic regulatory policies, which is conducive to promoting the stable development of the international economy.

Keywords: FP-Growth, H-Mine, Pearson correlation coefficient, Spearman's rank correlation coefficient, international economy, factor analysis of development.

DOI 10.32738/JEPPM-2025-198

1. Introduction

With the deepening integration of the global economy, accurately identifying the key factors affecting international economic growth has become a top priority for many countries in formulating their economic policies (Zhang et al., 2022). In recent years, the international landscape has changed dramatically, with growing political conflicts and the continuous rise of emerging economies, making the analysis of international economic development factors increasingly complex and dynamic (Liu et al., 2023). Against this background, scholars worldwide have conducted numerous studies on economic development. For example, Chen et al. (2023) developed an evaluation index system for digital economic development to analyze the economic downturn caused by the pandemic. Using panel data from 31 provinces in China from 2010 to 2019, they found that the level of digital economic development in China showed a steady upward trend during this period. To explore the impact of industrial structure on digital economy and low-carbon development, Tan et al. (2024) proposed a regression model using urban data. Their study revealed that industrial structure drives the digital economy to promote low-carbon, sustainable development, and that the digital economy, in turn further enhances low-carbon development. Nguyen et al. (2025) proposed a club convergence among Organisation for Economic Co-operation and Development (OECD) countries.By applying the Granger causality test to panel data from 1997 to 2021, they found no evidence of overall convergence, but identified several convergence clubs across individual indicators. To investigate the relationship between social entrepreneurship and socioeconomic development, Ahmad and Bajwa (2023) conducted a meta-analysis using bibliometric and content analysis methods. They analyzed relevant literature from 45 countries between 2005 and

2020 and found that the United States and the United Kingdom are the most productive countries in the field of socioeconomic development.

Although progress has been made in economic development research, the existing studies still have two main limitations. First, current analytical methods struggle to balance efficiency and accuracy. For example, the multiple linear regression in econometrics cannot capture complex nonlinear relationships [7]. Second, there is a lack of integrated research that both mines correlation patterns and quantifies the strength of variable relationships, making it difficult to determine which international economic factors are created and to what degree (Hebbi and Mamatha, 2023). The Frequent Pattern Growth (FP-Growth) algorithm compresses large-scale data by constructing a frequent pattern tree, which avoids generating numerous candidate itemsets. It has been widely applied in association rule mining, financial risk control, and other fields (Wei et al., 2023). The Pearson Correlation Coefficient (PCC) model calculates the strength and direction of linear co-movement between variables and is widely used in statistics and economics (Wang et al., 2022). Therefore, this study introduces the Hypergraph-based Frequent Pattern Mining (H-Mine) algorithm to optimize FP-Growth. This aims to solve the traditional FP-Growth algorithm's problems of insufficient processing capacity and poor dynamic adaptability with high-dimensional sparse economic data. The Spearman's Rank Correlation Coefficient (SRCC) was introduced to improve the PCC, aiming to comprehensively quantify correlations between both linear and nonlinear variables in the international economy and to enhance the PCC stability in high-noise environments. These two improved methods are integrated into a model for analyzing the factors behind international economic development. The proposed approach aims to solve problems such as long analysis time and insufficient accuracy in current methods while intuitively identifying the key factors affecting global economic growth.

This paper is organized into five sections: Section 1 introduces the background, reviews related literature, and analyzes the current state of international economic development. Section 2 presents the advantages of the H-Mine-FP-Growth hybrid algorithm and the SRCC-PCC model for analyzing development factors and constructs an integrated analytical model. Section 3 verifies the model's performance through comparative experiments on large-scale economic data. Section 4 discusses the experimental results, compares them with previous studies, and addresses the reliability of those findings. Section 5 concludes the study, identifies its limitations, and provides suggestions for future research.

2. Methods and Materials

2.1. Design of the H-Mine-FP-Growth Hybrid Algorithm

The analysis of international economic development factors aims to systematically explore the driving and limiting elements behind global economic growth from a global perspective. This process encompasses multi-dimensional variables, including institutional, technological, and other diverse factors (Praveen et al., 2023). As a classic data mining algorithm, FP-Growth compresses data structures using a divide-and-conquer strategy to address the combinatorial explosion of candidate itemsets in traditional association rule mining. It is particularly suitable for processing and analyzing large-scale data (Yue, 2024). The structure of FP-Growth used for analyzing international economic development factors is shown in Fig. 1.

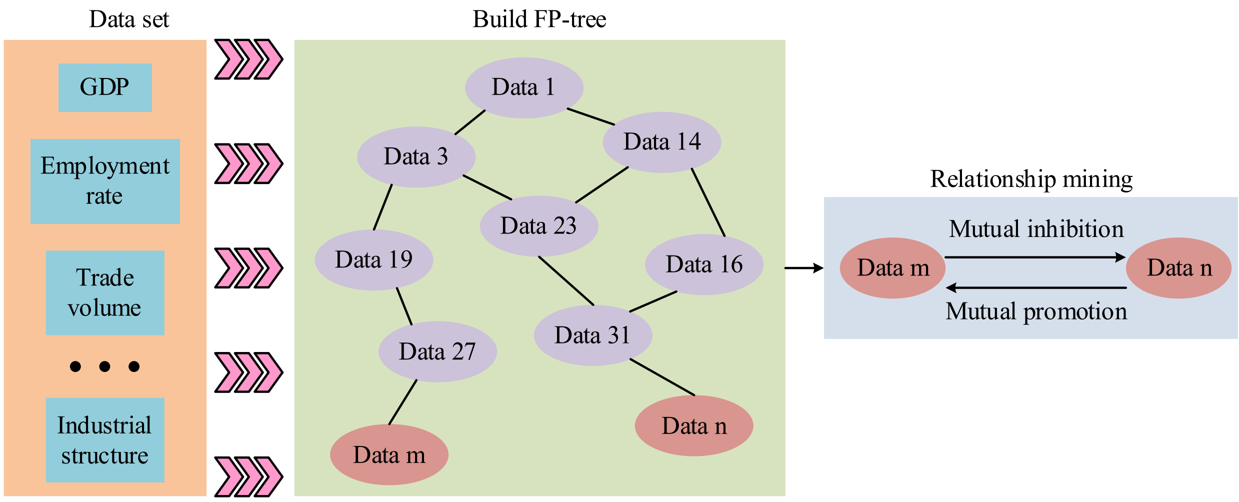


Fig. 1. FP-Growth structure for analyzing the influencing factors of international economic development

In Fig. 1, FP-Growth first scans economic indicators such as Gross Domestic Product (GDP), employment rate, and trade volume, then ranks them by frequency to automatically filter out infrequent data. Based on this ranking, it constructs the Fp-tree and continues to create new nodes until it identifies the most relevant factors for international economic development, finally extracting the relationships among them. The frequency of relevant data appearing in a dataset is represented by its support, defined by Eq. (1) (Xu, 2024).

$$\mathrm{Support}(X)=\frac{Count(X)}{|D|} \tag{1}$$

In Eq. (1), $\mathrm{Support}(X)$ represents the support of data X related to international economic development, and D is the total dataset or database. FP-Growth determines whether the data is frequent—i.e., a frequent itemset—based on the condition shown in Eq. (2) (Saputra et al., 2024).

$$X\in F\Leftrightarrow\frac{\left|\{t\in D|X\subseteq t\}\right|}{|D|}\geq\sigma \tag{2}$$

In Eq. (2), F is the set of all frequent itemsets, t is an itemset in the database, and σ is the minimum support threshold used to identify frequent itemsets. After filtering frequent itemsets, FP-Growth generates a header table for building the Fp-tree. The construction process is described in Eq. (3) (Hasan et al., 2023).

$$H(D,\sigma)=\left\{(p,\left|\{t\in D|p\in t\}\right|)\middle|\frac{\left|\{t\in D|p\in t\}\right|}{|D|}\geq\sigma\right\} \tag{3}$$

In Eq. (3), $H(D,\sigma)$ is the header table function that takes the full database D and the support threshold σ as inputs, and $\left|\{t\in D|p\in t\}\right|$ is the support count of a single itemset. However, the standalone FP-Growth algorithm has limitations in handling high-dimensional or sparse data and performs poorly when adapting to dynamic datasets. Its robustness is also insufficient. To address these issues, this study introduces the hypergraph-based H-Mine algorithm to optimize FP-Growth. H-Mine uses a hypergraph structure to retain rule-mining capability while lowering the minimum support threshold and reducing redundant computations, thus efficiently discovering multi-item association relationships (Qu et al., 2023). The structure of the H-Mine-FP-Growth hybrid algorithm is shown in Fig. 2.

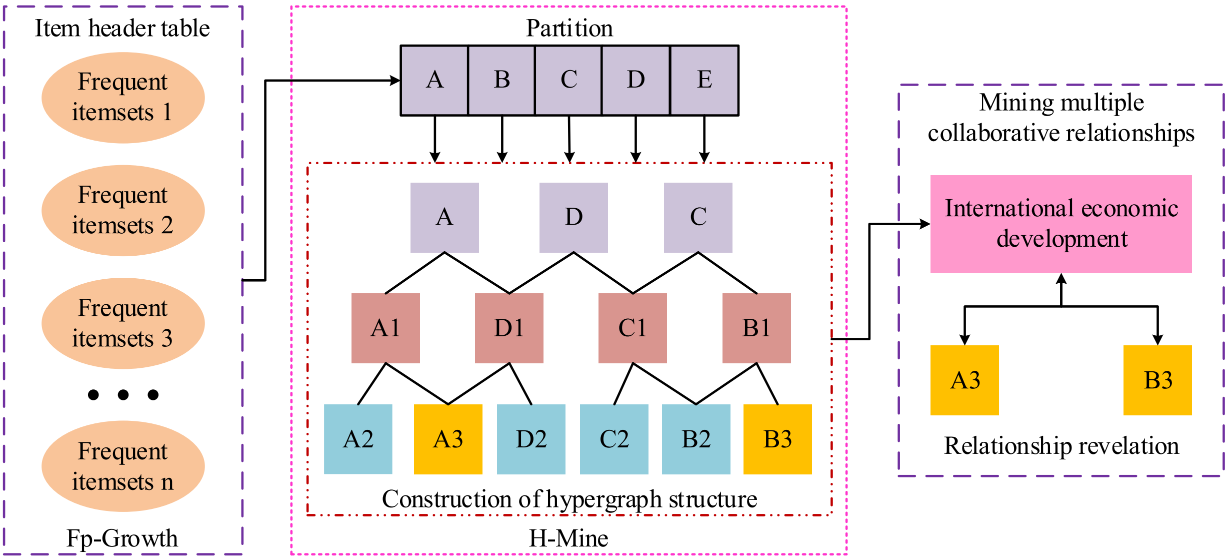


Fig. 2. Structure diagram of the H-Mine-FP-Growth hybrid algorithm

As shown in Fig. 2, the optimization introduced by H-Mine primarily improves the Fp-tree. After H-Mine reduces the support threshold, FP-Growth scans a large volume of economic indicators to generate a header table. Then, H-Mine partitions the data by grouping similar types together, which helps reduce computational complexity. Using these partitions, it constructs a hypergraph structure to replace the original Fp-tree. The hyperedges reduce noise and filter unrelated data while enabling deeper analysis of multi-item associations. This process continues until the most relevant economic development factors are identified and clearly presented. The expression for support after incorporating the reduced threshold from H-Mine is shown in Eq. (4).

$$\mathrm{Support}_{HFP}(X)=\frac{\sum_{i=1}^{m}Count_i(X)}{\sum_{i=1}^{m}|D_i|} \tag{4}$$

In Eq. (4), $\text{Support}_{HFP}(X)$ represents the hybrid support from H-Mine and FP-Growth. D_i is the database in the i-th partition, m is the total number of partitions, and $Count_i(X)$ is the frequency of itemset X in partition i. H-Mine transforms the data into a vertical format to construct the hypergraph, and this transformation process is described in Eq. (5) (Lin et al., 2023).

$$\text{Vertical DB}(D) = \{(p, T_p) | p \in I\} \tag{5}$$

In Eq. (5), $\text{Vertical DB}(D)$ defines the vertical representation of the database D, I is the set of all itemsets, and T_p is the vertical symbol for itemset p. H-Mine uses a recursive mining function to dig into complex relationships, as shown in Eq. (6) (Jamsheela and Raju, 2023).

$$\text{HM}(P,\alpha) = \begin{cases} \{\alpha\} & \text{if } |P| = 1 \\ \bigcup_{p\in P} \text{HM}(\text{Project}(P,p), \alpha \cup \{p\}) & else \end{cases} \tag{6}$$

In Eq. (6), $\text{HM}(P,\alpha)$ represents the recursive mining function of H-Mine. P is the current projected database, α is the prefix pattern, and $\text{Project}(P,p)$ is the conditional projection operation based on itemset p.

2.2. Construction of the SRCC-PCC Model for Relation Identification and Quantification

Although the H-Mine-FP-Growth algorithm is powerful for discovering associations, it cannot quantify the strength of relationships among variables. To address this limitation, a complementary model is needed for a more accurate analysis of international economic development factors. The PCC model is a classical method used to quantify the linear relationship between two variables. It measures correlation through the covariance of standardized data (Zhang et al., 2024). The workflow of the PCC model is illustrated in Fig. 3.

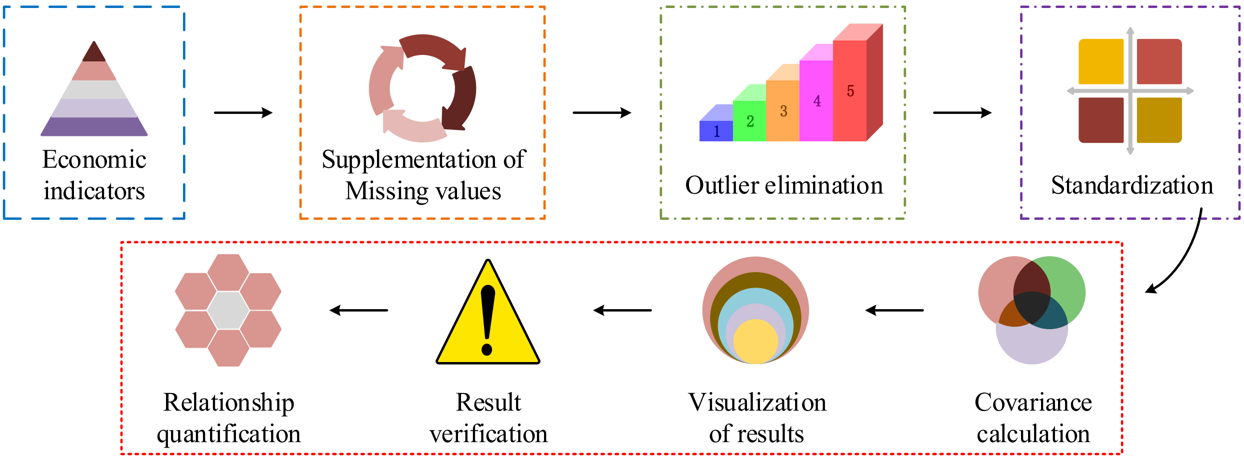


Fig. 3. Flowchart of the PCC model for quantifying the relationships among economic variables

As shown in Fig. 3, the PCC model begins by collecting relevant economic data and preprocessing it, which includes filling in missing values, removing outliers, and standardizing data to eliminate the effects of differing units. Next, it calculates the PCC and uses visualization tools to reveal the relationships between variables. It then applies t-tests and robustness checks to verify the reliability of the results, finally quantifying the relationships. The function for calculating the PCC is given in Eq. (7) (Shi et al., 2023).

$$\rho(x,y) = \frac{\text{Cov}_n(x,y)}{\sigma' x \cdot \sigma' y} \tag{7}$$

In Eq. (7), $\rho(x,y)$ is the PCC between variables x and y. n is the total sample size, and σ' refers to the linear correlation strength and direction between the variables. The t-test used by the PCC model to check significance is described in Eq. (8) (Bu and Sun, 2024).

$$t^* = r\sqrt{\frac{n-2}{1-r^2}} \; (n-2) \tag{8}$$

In Eq. (8), t^* represents the t-test for determining whether the correlation coefficient is significantly different from zero. n is the total sample size, and r is the observed correlation coefficient. The PCC model also uses confidence intervals to measure the error of the results, as defined in Eq. (9) (Ballukja et al., 2025).

$$\mathrm{CI}(\rho)=\left[\tanh(a\tanh(r)-\frac{z/2}{\sqrt{n-3}}),\tanh(a\tanh(r)+\frac{z/2}{\sqrt{n-3}})\right] \tag{9}$$

In Eq. (9), $\mathrm{CI}(\rho)$ is the confidence interval for ρ. a is a constant used for constraint, and z is a specific sample. However, the PCC model has limitations when used alone. It cannot capture nonlinear relationships, and is highly sensitive to outliers and units of measurement. To overcome these problems, the SRCC algorithm is used. SRCC evaluates monotonic relationships between variables by ranking the data and calculating correlations based on these ranks (Guan et al., 2024). This study incorporates SRCC into the PCC model to form the SRCC-PCC model. The process of capturing complex relationships among international economic development factors is shown in Fig. 4.

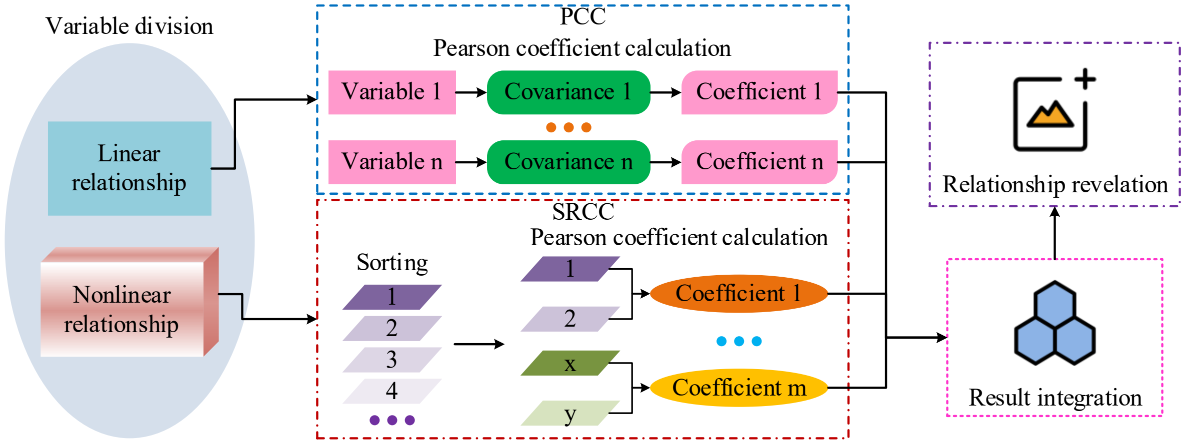


Fig. 4. Process of the SRCC-PCC model for quantifying variable relationships (Icon source from: https://iconpark.oceanengine.com/home)

In Fig. 4, SRCC-PCC processes both linear and nonlinear relationships based on variable characteristics. PCC computes the coefficients using raw data, while SRCC first ranks the data and then applies the PCC to the ranks, generating multiple sets of values. Comparing the coefficients from PCC and SRCC indicates whether the relationship is linear and whether the data are free of outliers. This process ultimately enables the quantification of variable relationships. The rank-based coefficient from SRCC is expressed in Eq. (10) (Amman et al., 2023).

$$\rho'=1-\frac{6\cdot\sum_{n=1}^{n}\Delta d}{n(n^2-1)} \tag{10}$$

In Eq. (10), ρ' is the rank-based PCC from SRCC. n is the sample size, such as the number of countries or total GDP over recent years. Δd is the rank difference between the two variables. SRCC can also calculate the PCC using the covariance of ranks. This process is described in Eq. (11) (Yu and Hutson, 2022).

$$\rho^{*}=\frac{\mathrm{Cov}(\mathrm{Rank}(X'),\mathrm{Rank}(Y'))}{\sqrt{\mathrm{Var}(\mathrm{Rank}(X'))\cdot\mathrm{Var}(\mathrm{Rank}(Y'))}} \tag{11}$$

In Eq. (11), ρ^{*} is the PCC based on the covariance of variables in SRCC. $\mathrm{Rank}(\cdot)$ represents the conversion of variables X' and Y' into rank sequences for sorting.

2.3. Construction of the International Economic Development Factor Analysis Model

The H-Mine-FP-Growth hybrid algorithm not only inherits H-Mine's efficient processing capability for high-dimensional and sparse data but also retains FP-Growth's high efficiency in mining internal relationships among variables. It balances mining depth with computational efficiency. Meanwhile, the SRCC-PCC model accurately captures linear correlations through PCC and effectively analyzes monotonic relationships in nonlinear and non-normal distributions through SRCC. By comparing the results of the two algorithms, it quantitatively evaluates relationships between variables in a comprehensive and precise manner. Based on these advantages, this study integrates the H-Mine-FP-Growth hybrid algorithm with the SRCC-PCC model to comprehensively analyze and visualize the relationships between international economic development factors. The structure of this integrated framework is shown in Fig. 5.

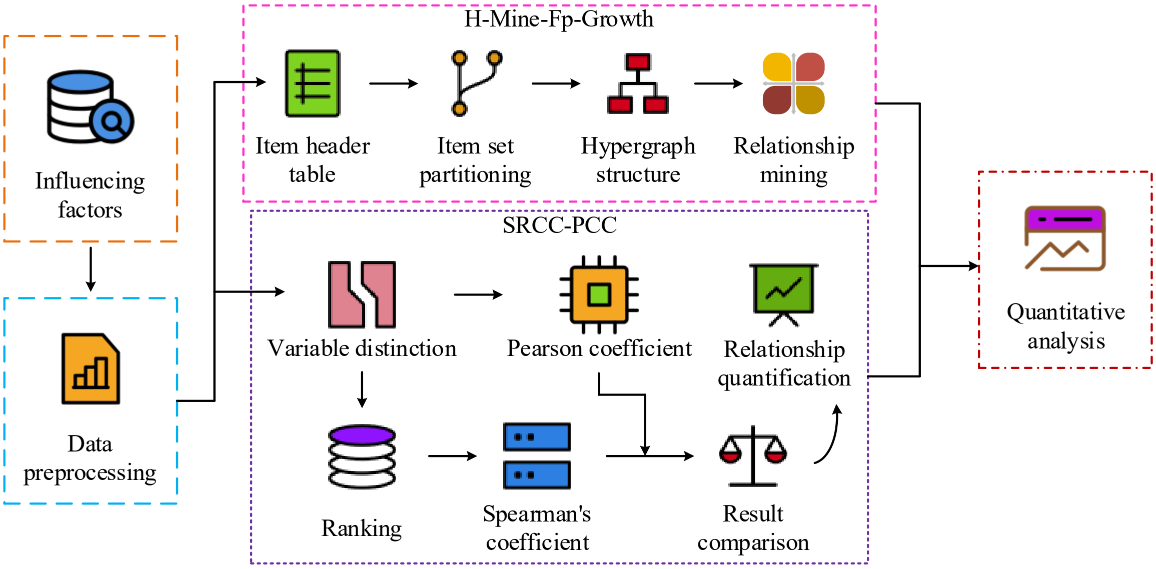


Fig. 5. Process of analyzing the relationships among the influencing factors of international economic development (Icon source from: https://iconpark.oceanengine.com/home)

As shown in Fig. 5, the integrated framework first preprocesses various economic indicators, separately performs relationship mining, and correlation quantification. The H-Mine-FP-Growth algorithm conducts in-depth association mining by generating a header table, partitioning variables, and constructing the hypergraph structure. It then performs a comparative analysis of the results to quantify the correlations among economic indicators. Finally, the results are integrated to provide a comprehensive quantitative analysis of international economic development factors. H-Mine reduces computational complexity through data partitioning, and its partition memory allocation function is shown in Eq. (12).

$$\text{M- Allocation}(D_i) = \frac{|D_i| \cdot \max_{\text{M}}}{\sum_{j=1}^{m} |D_j|} \tag{12}$$

In Eq. (12), $\text{M- Allocation}(D_i)$ is the partition memory allocation function of H-Mine, D is the total database, i and j represent specific partitions, m is the total number of partitions, and $\max_{\text{M}}$ defines the maximum memory allocated per partition. The SRCC-PCC model automatically adjusts the correlations among variables using a nonlinear correction equation to avoid deviations caused by linear assumptions. The equation is shown in Eq. (13).

$$\rho_{PS} = \rho_P \cdot (1 + \beta \cdot |\rho_P - \rho_S|) \tag{13}$$

In Eq. (13), ρ_{PS} is the corrected variable coefficient, and β is a constant with a value range of [0,1]. Compared with standalone H-Mine-FP-Growth and SRCC-PCC models, the integrated framework combines strong association mining capabilities with visual correlation quantification. It also shows greater robustness against noise and interference. Based on this framework, the study constructs a model for analyzing international economic development factors. The analytical process of the model is illustrated in Fig. 6.

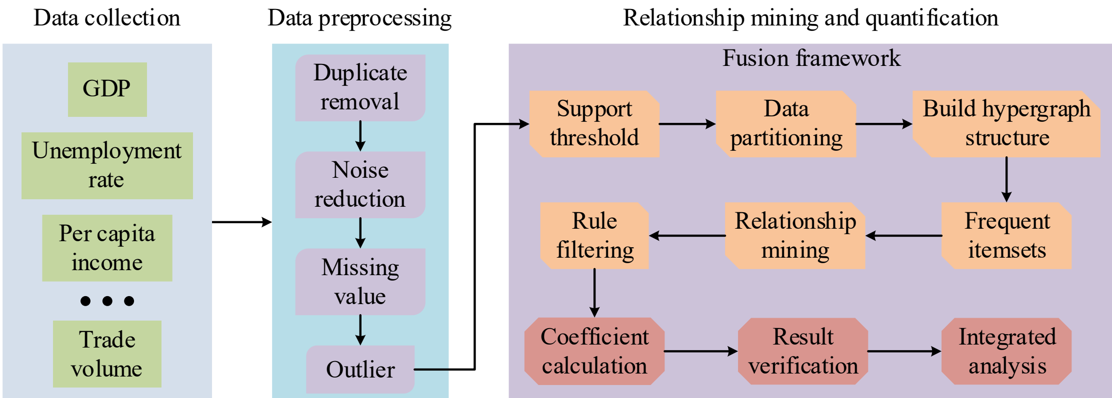


Fig. 6. Flowchart of analyzing factors of international economic development

As seen in Fig. 6, the model analyzes international economic development factors in three steps. First, it collects economic indicators from major countries in recent years through various sources such as the World Bank and national statistics bureaus. Second, it preprocesses the data by removing duplicates and noise, filling missing values, eliminating outliers, and standardizing the data to eliminate the effects of units and dimensions. Third, it mines relationships among economic indicator variables and quantifies their correlation strengths. H-Mine-FP-Growth sets a minimum support threshold and constructs a hypergraph structure to extract association rules with high confidence. By comparing these results, it quantitatively analyzes international economic development factors. During result validation, PCC generates a correlation coefficient matrix, which is shown in Eq. (14) (Li et al., 2024).

$$R = \begin{bmatrix} 1 & \rho_{17} & \cdots & \rho_{1k} \\ \rho_{23} & 1 & \cdots & \rho_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1} & \rho_{k2} & \cdots & 1 \end{bmatrix} \tag{14}$$

In Eq. (14), R is the correlation coefficient matrix of the variables, k is the total number of variables. When analyzing the relationships between variables, PCC usually introduces control variables to obtain partial correlation coefficients. The corresponding function is defined in Eq. (15).

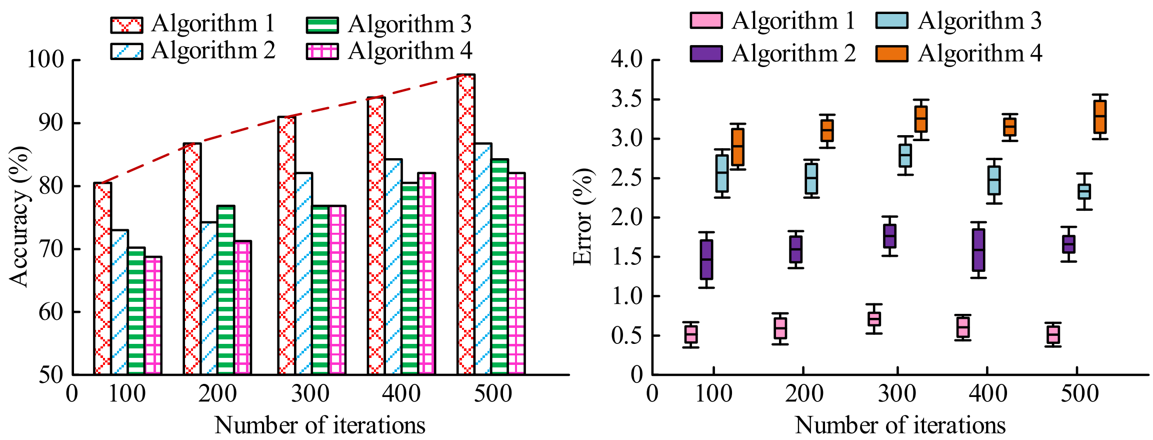
$$\tilde{\rho}_{MN,Z} = \frac{\rho_{MN} - \rho_{MZ} \cdot \rho_{NZ}}{\sqrt{(1-\rho_{MZ}^2)\cdot(1-\rho_{NZ}^2)}} \tag{15}$$

In Eq. (15), $\tilde{\rho}_{MN,Z}$ is the partial correlation coefficient between variables M and N after introducing the control variable Z. In addition, to address issues related to diverse data sources and dynamic real-time expansion, this research incorporates a modular data source access layer and a lightweight real-time computing engine to the proposed model. The former uses standardized interfaces and multi-rule cleaning libraries to support data sources with different structures, such as World Bank LDC data and IEA energy data. The latter adopts an incremental hypergraph update mechanism and simplified quantization logic, reducing real-time data update time to one-fifth of full computation time and significantly decreasing single-variable correlation quantization time, laying the foundation for dynamic scene analysis.

3. Results

3.1. Performance Validation of the H-Mine-FP-Growth Hybrid Algorithm

To evaluate the performance of the H-Mine-FP-Growth hybrid algorithm, the study compared it with the Improved Apriori Algorithm (IAA), Equivalence Class Transformation (ECT), and Improved Genetic Algorithm (IGA). The experiments were conducted on an advanced computer equipped with an Intel Core i5-14600KF CPU, NVIDIA RTX 5060ti 8GB GPU, Windows 11 Professional operating system, 36GB DDR5 4800MHz RAM, and a 2TB solid-state drive combined with a 2TB hard disk. All programs were developed using Python 3.8. The OECD dataset, which contains socioeconomic information from all member countries of the Organization for Economic Cooperation and Development, was selected as the data source to analyze global economic trends and influencing factors. The four algorithms were named Algorithm 1 through Algorithm 4. First, a comparative experiment was conducted to evaluate their accuracy and error rates in mining the relationships among variables. The results are shown in Fig. 7.



(a) Comparison of the accuracy rates of relationship mining among four algorithms

(b) Comparison of accuracy errors in relationship mining among four algorithms

Fig. 7. Accuracy and error of variable relationship mining using four algorithms

As shown in Fig. 7(a), Algorithm 1 achieved a maximum accuracy of 98.4% in identifying associations among economic variables, which was significantly higher than the 87.1% of Algorithm 2, 84.3% of Algorithm 3, and 81.8% of Algorithm 4. This indicated that Algorithm 1 was more compatible with economic data and more effective at capturing its features. For instance, when exploring the correlation among GDP, trade volume and employment rate, Algorithm 1 filtered non-core interfering data such as short-term exchange rate fluctuations, accurately captured the long-term synergy patterns among the three, and provided reliable data support for policymakers. Fig. 7(b) showed that Algorithm 1 had an average error of only 0.59%, which was much lower than Algorithm 2's 1.76%, Algorithm 3's 2.58%, and Algorithm 4's 3.37%. This suggested that the mining results of Algorithm 1 were more reliable and its identification of economic variable relationships was closer to reality. The error of Algorithm 1 peaked at 0.65% during the 300th iteration and gradually stabilized thereafter. Next, the study compared the time consumption and computational complexity of the algorithms during rule mining. The results are presented in Table 1.

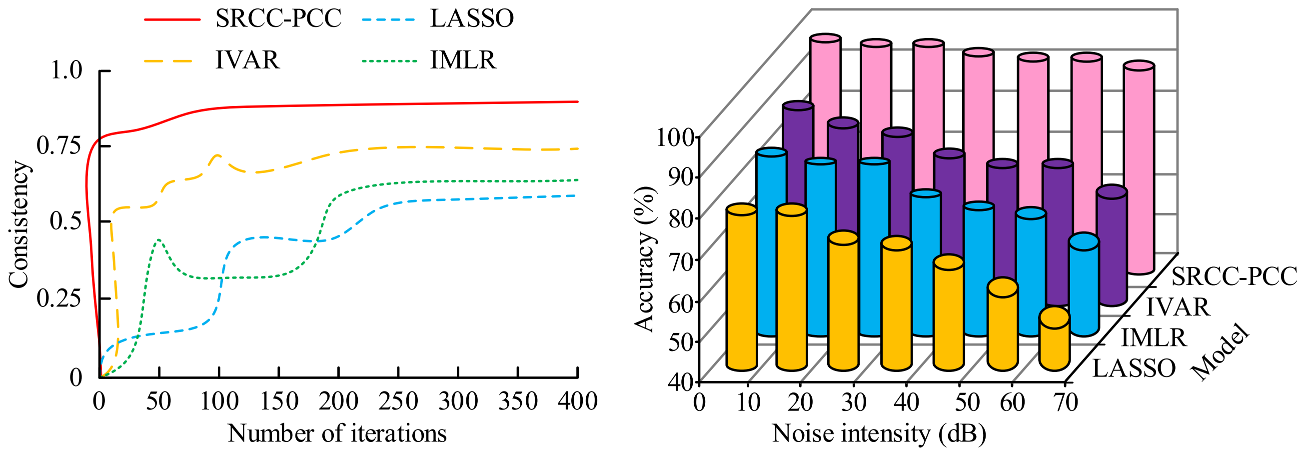
Table 1. Time and complexity comparison of different algorithms in rule mining

Experimental Indicators	Number of Iterations	Algorithm Category			
		Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4
Consume Time (ms)	Iterate 100 times	83.6	137.1	150.7	184.2
	Iterate 200 times	97.2	168.4	173.8	216.7
	Iterate 300 times	118.5	193.5	197.4	243.9
	Iterate 400 times	129.0	224.6	238.5	277.0
	Iterate 500 times	147.3	251.2	272.6	302.4
Memory Usage (MB)	Iterate 100 times	78.4	116.7	164.2	146.9
	Iterate 200 times	93.6	153.2	201.5	173.1
	Iterate 300 times	107.2	187.9	236.8	227.0
	Iterate 400 times	126.9	213.0	255.1	269.6
	Iterate 500 times	138.4	257.6	283.4	302.7

In Table 1, Algorithm 1 reached its maximum time consumption of 147.3 ms at the 500th iteration, which was significantly lower than the 251.2 ms of Algorithm 2, 272.6 ms of Algorithm 3, and 302.4 ms of Algorithm 4. This demonstrated Algorithm 1's superior adaptability to high-dimensional and large-scale data and its higher efficiency in resource utilization. Furthermore, the increase in time consumption with more iterations was the smoothest in Algorithm 1, with only a 63.7 ms increase after 500 iterations, again outperforming the other three algorithms. The maximum memory usage of Algorithm 1 was only 138.4 MB, which was lower than the 257.6 MB of Algorithm 2, 283.4 MB of Algorithm 3, and 302.7 MB of Algorithm 4. This showed that Algorithm 1 had a more reasonable structural design, resulting in higher potential for robustness and resource efficiency. In the real-time scenario of sudden changes in the international economic situation, Algorithm 1 could quickly process new high-dimensional sparse economic data (such as export data of new trading partners and industry output data under price fluctuations), complete the update of core association rules within minutes, and thus avoid economic losses caused by lagging policy responses. Overall, Algorithm 1, the H-Mine-FP-Growth hybrid algorithm, efficiently and accurately mined the hidden relationships among economic indicators with minimal error, ensuring high result reliability.

3.2. Performance Testing of the SRCC-PCC Model

To test the performance of the SRCC-PCC model in quantifying variable correlations, this study compared it with the Improved Vector Autoregression (IVAR), Least Absolute Shrinkage and Selection Operator (LASSO), and Improved Multiple Linear Regression (IMLR). The experimental settings, parameters, and dataset remained unchanged. First, the study compared the theoretical consistency and anti-interference capability of the models. The results are shown in Fig. 8.



(a) Theoretical consistency comparison of the quantification of the relationships among the four models

(b) Comparison of the accuracy of relational quantization under noise interference of four models

Fig. 8. Comparison of theoretical consistency and anti-interference performance

As shown in Fig. 8 (a), the theoretical consistency of SRCC-PCC in quantifying variable relationships rose to 0.773 within the first 50 iterations and reached its peak of 0.889 at the 113th iteration, remaining stable thereafter. This peak was significantly higher than the 0.752 of the LASSO model, 0.637 of the IVAR model, and 0.586 of the IMLR model. These results indicated that SRCC-PCC provided outputs more closely aligned with established economic theory and scholarly interpretations, thus offering higher reliability. Fig. 8 (b) shows that SRCC-PCC maintained an accuracy of 89.7% under 70 dB noise interference, while the accuracy of the LASSO model dropped to 53.8% under the same conditions, demonstrating much lower stability. As the noise level increased from 0 dB to 70 dB, the accuracy of SRCC-PCC only decreased by 4.6%, outperforming all three baseline models. This demonstrated SRCC-PCC's superior capability to extract meaningful information and its robustness in noisy environments. This study then introduced the UNCTAD and IMF datasets to further verify the generalization ability of the model. The UNCTAD dataset focused on trade and investment, while the IMF dataset included exchange rate and macroeconomic data. The models' sensitivity was tested on OECD, UNCTAD, and IMF datasets, and the results were validated with t-tests. The results are shown in Table 2.

Table 2. Sensitivity comparison and t-test results of four models on different datasets

Dataset	Number of Iterations	Sensitivity (%)			
		SRCC-PCC	IVAR	LASSO	IMLR
OCED	Iterate 100 times	92.8	80.6	70.8	86.2
	Iterate 200 times	93.4	82.9	73.1	89.1
	Iterate 300 times	94.6	84.3	80.6	87.5
	Iterate 400 times	96.3	83.5	81.2	90.4
	Iterate 500 times	97.2	85.8	81.7	88.6
UNCTAD	Iterate 100 times	91.7	73.9	82.5	64.8
	Iterate 200 times	92.5	77.4	83.3	67.6
	Iterate 300 times	94.1	76.3	85.2	69.8
	Iterate 400 times	95.6	75.8	84.6	72.5
	Iterate 500 times	96.9	78.2	88.0	74.7
IMF	Iterate 100 times	90.4	85.0	77.2	79.7
	Iterate 200 times	92.3	87.6	79.8	83.5
	Iterate 300 times	94.0	89.3	76.5	86.8
	Iterate 400 times	95.7	90.5	80.6	85.2
	Iterate 500 times	97.5	88.4	82.3	87.0

Note: *, **, *** indicate significance levels of 0.1, 0.05, and 0.01, compared to SRCC-PCC.

As shown in Table 2, SRCC-PCC achieved average sensitivity values of 94.86%, 94.16%, and 93.98% on the three datasets, significantly higher than those of the other models. The overall variance in SRCC-PCC's sensitivity values across the three datasets was minimal (the largest difference was only 0.88%), indicating better generalization and adaptability. Additionally, only the LASSO model showed a significant difference from SRCC-PCC on the OECD dataset. On the UNCTAD dataset, both IVAR and IMLR exhibited significant differences across all iteration intervals. On the IMF dataset, only the LASSO and IMLR models showed significant differences within the first 200 iterations. These findings confirmed that SRCC-PCC delivered more reasonable and reliable results in quantifying economic variable relationships.

3.3. Application Evaluation of the International Economic Development Factor Analysis Model

After verifying the performance of both the H-Mine-FP-Growth and SRCC-PCC models, the study further tested the feasibility of the proposed model by comparing it with three alternative models: one based on attention mechanisms and long short-term memory networks, the second based on Bayesian networks, and a third based on graph neural networks. These models were labeled as Model 1 through Model 4. Economic indicators from China, the United States, the United Kingdom, and other countries during 2015–2024 were selected as the experimental dataset. First, the study compared the quantification results of variable correlations, as shown in Fig. 9.

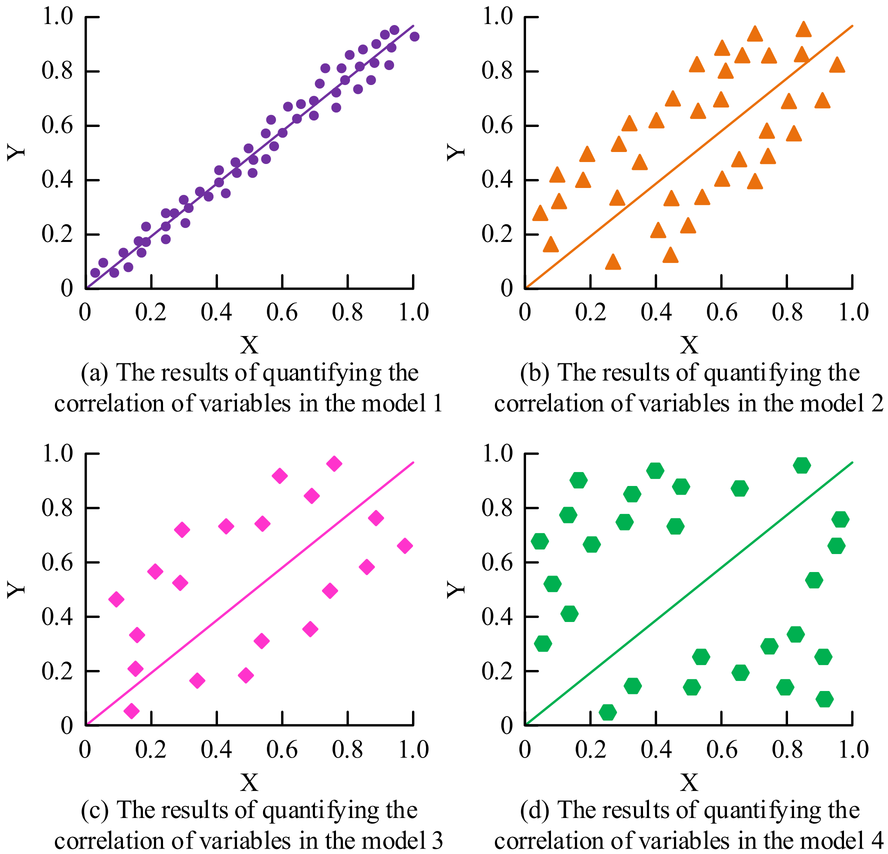
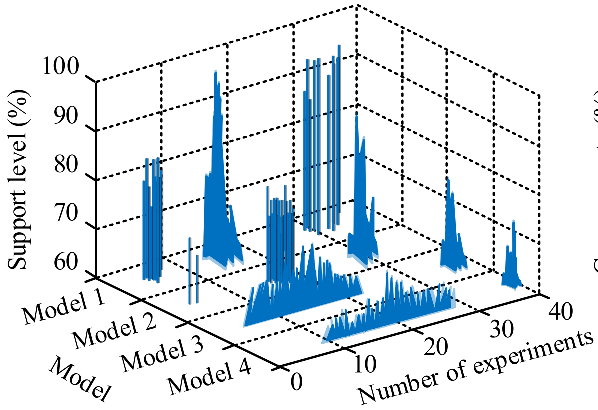
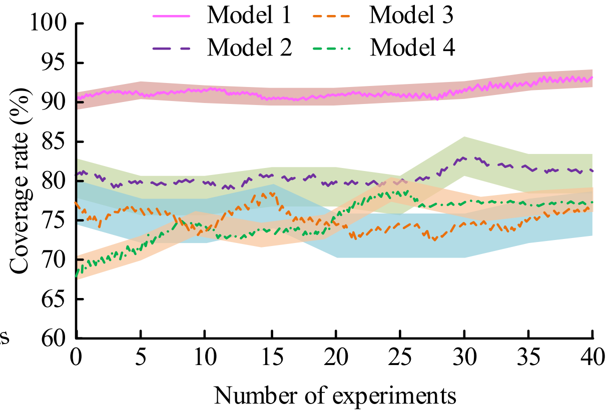


Fig. 9. Correlation quantification of economic indicators using four models

As shown in Fig. 9 (a), Model 1 successfully quantified multiple sets of variable correlations, and its results were closest to 1. This indicated that Model 1 had stronger recognition capability for large and complex datasets and produced more reliable outcomes. It also demonstrated the ability to capture dynamic relationships among variables. Figs. 9 (b) through 9 (d) showed that Model 3 had the weakest performance in describing variable relationships, and Model 4 had the lowest degree of correlation quantification. Although Model 2 outperformed Models 3 and 4 in several aspects, it was still clearly inferior to Model 1 in terms of quantifying variable relationships. Next, the study compared the models' support for identifying development factors and their data coverage, as shown in Fig. 10.



(a) Comparison of the support levels for identifying development factors among various models



(b) Comparison of the coverage rates of frequent item sets among various models

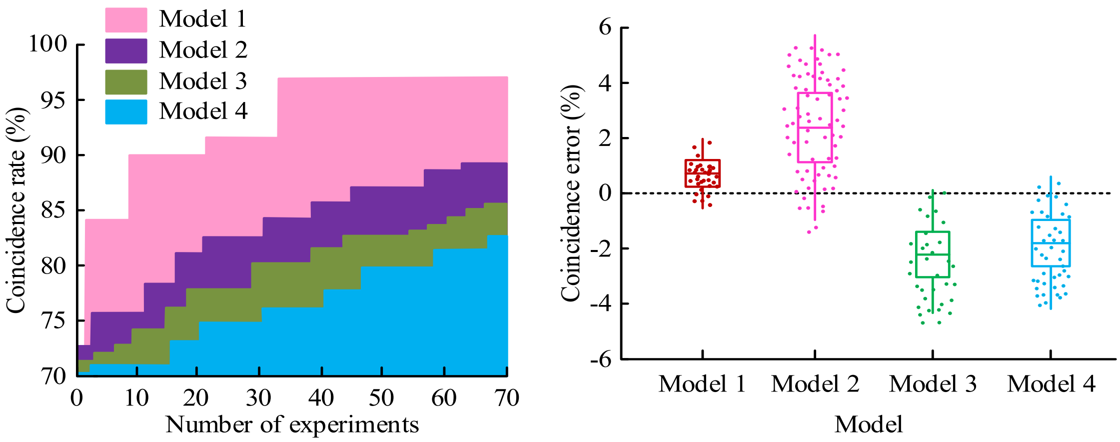
Fig. 10. Support and data coverage of four models in identifying development factors

As shown in Fig. 10 (a), Model 1 reached the highest support value of 95.2% in the 18th experiment, significantly surpassing the 84.6% of Model 2, 73.8% of Model 3, and 69.7% of Model 4. This indicated that the combinations of variables it has mined, such as cross-border capital flows, exchange rates, and foreign exchange reserves, frequently appear in the data of major global economies, possess universal applicability, and are conducive to alleviating the fluctuations and inflationary pressures in the global industrial chain. Fig. 10 (b) showed that Model 1 achieved a data coverage rate of 93.9%, much higher than Model 2's 82.8%, Model 3's 78.4%, and Model 4's 81.2%. This suggested that Model 1 had a stronger ability to cover relevant data relationships without missing key variable combinations. For instance, when analyzing the relationship between agricultural subsidies and the reduction of the rural poor population, Model 1 could not only capture the correlations of high subsidies, high mechanization, and low poverty rates in developed countries, but also the distinctive correlations of precise subsidies, improved agricultural production efficiency, and decreased poverty rates in developing countries, thereby providing differentiated references. The study then extended the time range to 2005–2024 and divided it into two periods: 2005–2014 and 2015–2024. Countries were also grouped into developed and developing nations. The quantification accuracy of the models was compared across these different samples, and the results are shown in Table 3.

Table 3. Quantification accuracy of variable relationships across different samples

Sample Category	Number of Experiments	Precision (%)			
		Model 1	Model 2	Model 3	Model 4
Country's	Experiment 10	90.4	82.2	76.6	79.3
	Experiment 20	92.3	82.8	78.2	82.6
	Experiment 30	93.5	84.1	80.3	83.7
	Experiment 40	95.7	83.0	83.4	85.2
	Experiment 50	95.9	86.7	84.8	89.1
Times	Experiment 10	91.6	76.3	82.5	62.8
	Experiment 20	92.1	77.9	80.7	64.3
	Experiment 30	93.3	79.2	83.4	65.7
	Experiment 40	94.2	81.5	84.9	70.6
	Experiment 50	96.4	82.8	86.3	73.9

As shown in Table 3, Model 1 achieved the highest accuracy of 95.9% and 96.4% across country and time samples, respectively, and its average accuracy was highest at 93.56% and 93.52%. The above experimental data indicated that Model 1 had a strong generalization ability and could support policy planning across periods and economies. For instance, during the traditional industrialization period, Model 1 could precisely quantify the relationship between industrial investment and economic growth, providing a basis for countries to formulate industrialization policies at that time. In the digital economy era, Model 1 could further quantify the correlation between digital infrastructure and economic growth, thereby guiding countries to shift towards digital economy investment. Finally, the study compared the four models in terms of overlap rate and error when quantifying correlations among large-scale indicators. These results are shown in Fig. 11.

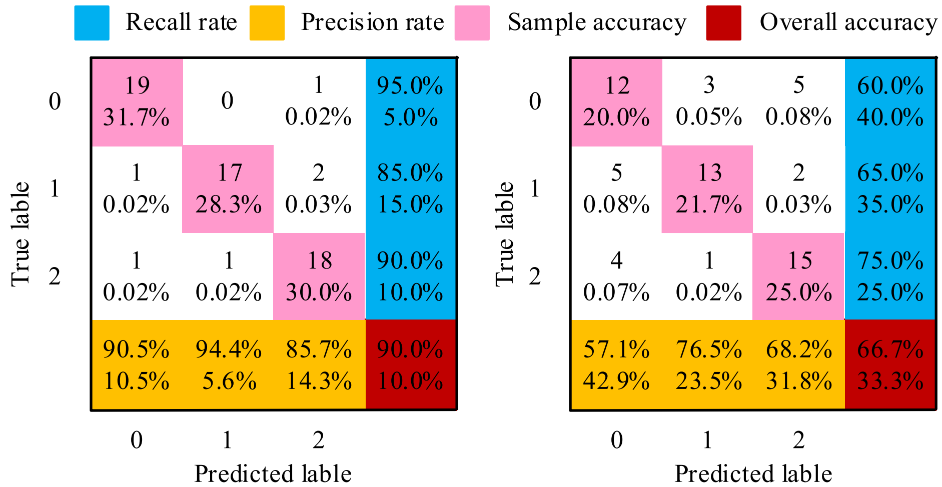


(a) Comparison of the correlation and coincidence rates of quantitative indicators

(b) Comparison of correlation degree, coincidence rate and error of quantitative indicators

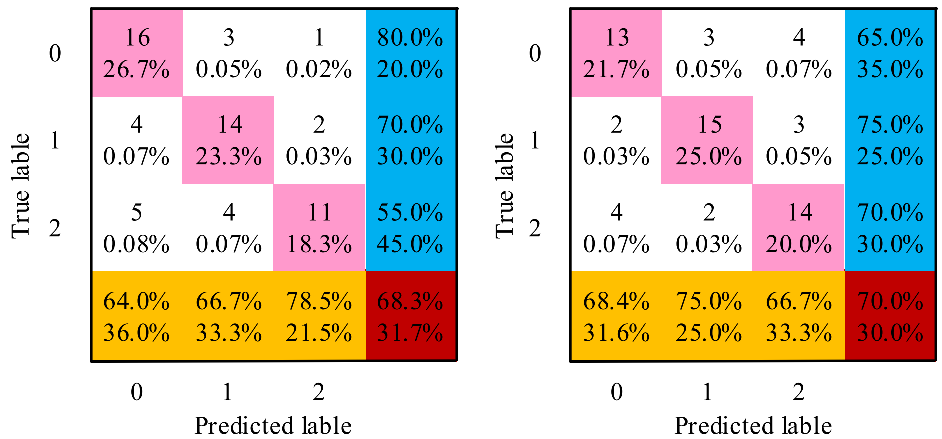
Fig. 11. Overlap rate and error of correlation quantification using four models

In Fig. 11 (a), Model 1 achieved an overlap rate of 96.5% in quantifying indicator correlations, significantly higher than the 89.4% of Model 2, 85.8% of Model 3, and 83.1% of Model 4. This indicated that Model 1 more precisely identified relationships among core variables and demonstrated better adaptation to large-scale data distributions. Fig. 11 (b) showed that the average error in overlap rate for Model 1 was only 0.41%, lower than the 2.17% of Model 2, -2.05% of Model 3, and -1.94% of Model 4. Model 1 also had the fewest outliers and the highest number of valid quantification results, indicating its more reasonable structure and greater reliability in explaining economic indicators. The study further divided 60 economic indicators into high, medium, and low correlation groups, with 20 samples each labeled as 0, 1, and 2. The models' performance in identifying these groups was compared using confusion matrices, as shown in Fig. 12.



(a) Analysis results of Model 1

(b) Analysis results of Model 2



(c) Analysis results of Model 3

(d) Analysis results of Model 4

Fig. 12. Confusion matrix of correlation level classification

In Fig. 12 (a), Model 1 achieved recall rates of 95.0%, 85.0%, and 90.0% for the high, medium, and low correlation variables. Its overall accuracy was 90.0%, and it correctly identified 19 of the 20 highly correlated variables, misclassifying only one as low. Fig. 12 (b) showed that Model 2 had the lowest precision (57.1%) for high-correlation variables and the lowest overall accuracy at 66.7%. Model 3 achieved only 55.0% recall for low-correlation variables, with an accuracy just 1.6% higher than Model 2, both underperforming compared to Model 1. The precise identification ability of Model 1 could help policymakers focus on the core of economic development, prioritize the investment of limited resources to enhance policy effectiveness, thereby reducing unnecessary regulation and avoiding the waste of policy resources.

4. Discussion

To address the limitations of existing international economic factor analysis methods, such as low efficiency, insufficient accuracy, and large result errors, the study introduced H-Mine and SRCC to optimize the FP-Growth and PCC model. This resulted in the H-Mine-FP-Growth hybrid algorithm and the SRCC-PCC model. Their performance was tested through experimental validation. The results showed that the H-Mine-FP-Growth hybrid algorithm achieved an accuracy of 98.4% in mining associations among economic variables, with an average error as low as 0.59%. This is similar to the result obtained when Trabelsi MA explored the impact of artificial intelligence on economic development. Trabelsi M. A analyzed the impact of artificial intelligence on economic development by introducing two factors, labor and capital, and argued that it could enhance the efficiency of economic development by analyzing a large amount of data and significantly improve the implementation process of economic decisions. (Trabelsi M A, 2024). Additionally, the hybrid algorithm consumed only 147.3 ms of processing time and 138.4 MB of memory during the mining process, outperforming other algorithms. These outcomes were consistent with the experimental findings of E. A. Widjaja on the effectiveness of Apriori and FP-Growth (Widjaja, 2024). The reason behind this performance is that the H-Mine-FP-Growth hybrid algorithm leverages the data compression structure of FP-Growth to reduce the data scale, while it sets reasonable thresholds to construct the FP-tree and improve result accuracy. Meanwhile, H-Mine efficiently processes low-frequency items, reduces memory usage, and lowers computational complexity, thus significantly enhancing computational efficiency.

The performance validation of the SRCC-PCC model showed that its theoretical consistency in quantifying variable correlation reached 0.889, and its accuracy remained at 89.7% even under strong noise interference of 70 dB. When tested across three different datasets, its average sensitivity reached 94.86%, 94.16%, and 93.98%, with sensitivity differences of only 0.7%, 0.88%, and 0.18%. This idea is similar to that of scholar Qin Y,, who has used intelligent algorithms to analyze the trend of economic development. Qin Y conducted a qualitative clustering analysis of 2,211 economic-related literatures through different measurement methods and intelligent algorithms, and found that the mining of economic data correlation relationships is a key direction for future economic development research. (Qin Y, et al., 2024). In practical applications, the proposed model based on H-Mine-FP-Growth and SRCC-PCC also showed excellent performance. In the experiment on quantifying the correlation of economic variables, the model demonstrated stronger recognition ability for large-scale complex data and better capability in capturing dynamic relationships among variables. The model achieved a support value of 95.2% and a data coverage rate of 93.9% in identifying international economic development factors. It further reached an accuracy of 95.9% for country samples and 96.4% for time samples. Moreover, the model achieved an overlap rate of 96.5% when quantifying indicator correlations, with an error rate as low as 0.41%. These findings were similar to the experimental results of the team led by Yu and Hutson (2024), who used a weighted PCC model to evaluate t-test errors in statistical software packages. This is because the PCC model calculates the linear correlation between variables while removing irrelevant factors, which helps reduce the impact of unrelated variables on the results. It also filters key variables to enhance data consistency and reliability, thereby reducing experimental errors. In the comparison of international economic development factor analysis results, the recall rates for high, medium, and low correlation variables using the proposed model were 95.0%, 85.0%, and 90.0%, respectively. All of these performance indicators outperformed the three baseline analysis methods. The model provides a more comprehensive basis for decision-making in the formulation of economic policies (Purnomo S, et al., 2024).

Although the performance verification of the H-Mine-FP-Growth hybrid algorithm and the SRCC-PCC model used the OECD dataset as the core benchmark, the UNCTAD dataset (focusing on international trade and investment) and the IMF dataset (covering exchange rates and global economic statistics) were subsequently introduced for extended tests. The adaptability of the proposed model to the relevant data of non-OECD economies was briefly verified. For example, the average sensitivity of SRCC-PCC under the UNCTAD dataset reached 94.16%, which was only 0.70% lower than the average sensitivity of 94.86% under the OCED dataset. Moreover, when mining correlations such as primary product exports, foreign exchange reserves, and regional trade agreements of developing countries, the data coverage rate of SRCC-PCC still remained at 92.3%, close to 93.9% of the OECD dataset. This indicates that the difference in the proposed model's ability to capture the core economic relations of developing economies and its data processing performance for OECD economies is relatively small. However, the study did not conduct experiments on the characteristic economic indicators of some low-income non-OECD countries, such as those in Africa and South Asia. This may lead to insufficient generalization of the model in economies with low economic complexity [32]. Future research should supplement special datasets containing data from low-income countries to further test the effectiveness of the model when applied to typical economic chains in developing countries. At the same time, future research should clarify the applicable boundaries of the proposed model at different economic development stages by comparing the quantitative results of variable correlations between emerging economies and traditional industrialized countries, in order to meet the needs of cross-economic policy analysis.

The contributions of this study are mainly reflected in the following three aspects: first, The study optimized the traditional FP-Growth algorithm and PCC model by introducing H-Mine and SRCC, resulting in the improved H-Mine-FP-Growth hybrid algorithm and SRCC-PCC model, second, The study combined H-Mine-FP-Growth and SRCC-PCC into a unified framework and built a new international economic development factor analysis model based on this framework., third, The study successfully applied the proposed model to real-world international economic development factor analysis and verified its feasibility and advantages through direct comparisons with three existing analysis models. These three contributions provide new perspectives and approaches for related fields and support further exploration of international economic development trends.

5. Conclusion

In response to the limitations of current international economic factor analysis methods, this study proposed an integrated framework based on the H-Mine-FP-Growth hybrid algorithm and the SRCC-PCC model. On this basis, a comprehensive analysis model for international economic development factors was constructed. The results showed that the proposed model not only accurately identified association rules among economic indicator variables but also intuitively quantified the degree of correlation between variables. Although the comparative experiments in the research were mainly based on the OECD dataset, the cross-economic application potential of the model was verified through the expansion of multiple datasets. Especially when dealing with economic data related to international trade and exchange rates of emerging economies and developing countries, the model demonstrated strong robustness and generalization. Although the proposed model performed very well in practical applications, the experiment did not break down all economic indicators, and the validity of the analysis of characteristic economic indicators of low-income non-OECD countries needs to be further verified. Future research should integrate more data from non-OECD economies to enhance the model's adaptability to economies with different levels of economic development and industrial structures, thereby providing more comprehensive data support for the formulation of global economic policies.

Funding

This research was supported by the Guangdong Province Lifelong Learning Society Construction (Continuing Education) Quality Improvement Project, grant number [JXJYGC2024D227], and the Guangdong Education Science Planning Project, grant number [2024GXJK142].

Institutional Review Board Statement

Not applicable.

Reference

Ahmad, S. and Bajwa, A. I. (2023). The role of social entrepreneurship in socio-economic development: a meta-analysis of the nascent field. *Journal of Entrepreneurship in Emerging Economies*, 15(1), 133–157. https://doi.org/10.1108/JEEE-04-2021-0165

Amman, M., Rashid, T., and Ali, A. (2023). Fermatean fuzzy multi-criteria decision-making based on Spearman rank correlation coefficient. *Granular Computing*, 8(6), 2005–2019. https://doi.org/10.1007/s41066-023-00421-x

Ballukja, E., Niewiadomski, K., Koch, P., Bojarski, J., Evans, P., and Moonen, N. (2025). A Pearson's random walk method of estimating the electromagnetic emissions of N parallel connected power electronic converters. *IEEE Transactions on Electromagnetic Compatibility*, 67(3), 1004–1015. https://doi.org/10.1109/TEMC.2025.3556297

Bu, H. and Sun, M. (2024). Clopper-Pearson algorithms for efficient statistical model checking estimation. *IEEE Transactions on Software Engineering*, 50(7), 1726–1746. https://doi.org/10.1109/TSE.2024.3392720

Chen, Y., Xu, S., Lyulyov, O., and Pimonenko, T. (2023). China's digital economy development: Incentives and challenges. *Technological and Economic Development of Economy*, 29(2), 518–538. https://doi.org/10.3846/tede.2022.18018

Gholami, A. and Srivastava, A. K. (2023). ORCA: Outage Root Cause Analysis in DER-Rich power distribution system using data fusion, hierarchical clustering and FP-Growth rule mining. *IEEE Transactions on Smart Grid*, 15(1), 667–676. https://doi.org/10.1109/TSG.2023.3281489

Guan, J., Jiang, X., Liu, N., Ding, H., Dong, Y., and Liu, T. (2024). Target detection in sea clutter based on feature re-expression using Spearman's correlation. *IEEE Sensors Journal*, 24(19), 30435–30450. https://doi.org/10.1109/JSEN.2024.3440065

Guo, C., Song, Q., Yu, M. M., and Zhang, J. (2024). A digital economy development index based on an improved hierarchical data envelopment analysis approach. *European Journal of Operational Research*, 316(3), 1146-1157. https://doi.org/10.1016/j.ejor.2024.02.023

Hasan, N. F., Aziz, S. A., and Nofendri, Y. (2023). Utilization of data mining on msmes using FP-Growth Algorithm for menu recommendations. *Matrik: Jurnal Management, Teknik Informatika dan Rekayasa Komputer*, 22(2), 261–270. https://doi.org/10.30812/matrik.v22i2.2166

Jamsheela, O. and Raju, G. (2023). An improved frequent pattern tree: the child structured frequent pattern tree CSFP-tree. *Pattern Analysis and Applications*, 26(2), 437–454. https://doi.org/10.1007/s10044-022-01111-1

Li, J. J., Zhou, J. H., Bickel, J. P., and Tong, X. (2024). Dissecting gene expression heterogeneity: generalized pearson correlation squares and the k-lines clustering algorithm. *Journal of the American Statistical Association*, 119(548), 2450–2463. https://doi.org/10.1080/01621459.2024.2342639

Lin, J., Lu, M., Jiang, Y., Fu, R., Peng, X., and Wu, E. Q. (2023). Operating optimization of steam turbine unit based on big data parallel association rule mining. *IEEE Transactions on Automation Science and Engineering*, 21(3), 4226–4236. https://doi.org/10.1109/TASE.2023.3293843

Liu, R., Sun, W., and Hu, W. (2023). Placement of high availability geo-distributed data centers in emerging economies. *IEEE Transactions on Cloud Computing*, 11(3), 3274–3288. https://doi.org/10.1109/TCC.2023.3280983

Nguyen, T., Nghiem, S., and Doan, T. A. (2025). Energy diversification, financial development and economic development: an examination of convergence in OECD countries. *China Finance Review International*, 15(2), 337–362. https://doi.org/10.1108/CFRI-07-2024-0427

Praveen, B. K., Padmavathy, T., Muthunagai, U. S., and Paulraj, D. (2023). An optimized fuzzy based FP-Growth algorithm for mining temporal data. *Journal of Intelligent & Fuzzy Systems*, 46(1), 41–51. https://doi.org/10.3233/JIFS-2230

Purnomo, S. and Darodjat, T. A. (2024). An analysis of potential factors influencing the sustainability in global economic sectors. *Asia Pacific Journal of Business Economics and Technology*, 4(5), 74-87. https://doi.org/10.98765/apjbet.v4i05.271

Qin Y., Xu Z., Wang X., and Skare, M. (2024). Artificial intelligence and economic development: An evolutionary investigation and systematic review. *Journal of the Knowledge Economy*, 15(1), 1736-1770. https://doi.org/10.1007/s13132-023-01183-2

Qu, J.-F., Fournier-Viger, P., Liu, M., Hang, B., and Hu, C. (2023). Mining high utility itemsets using prefix trees and utility vectors. *IEEE Transactions on Knowledge and Data Engineering*, 35(10), 10224–10236. https://doi.org/10.1109/TKDE.2023.3256126

Rosadi, M. and Hasibuan, S. M. (2024). Comparison of apriori and FP-Growth algorithms in analyzing association rules. *PIKSEL*, 12(2), 399–408. https://doi.org/10.33558/piksel.v12i2.9965

Saputra, B. P. J., Rahayu, A. S., and Hariguna, T. (2024). Market basket analysis using FP-Growth algorithm to design marketing strategy by determining consumer purchasing patterns. *Journal of Applied Data Science*, 4(1), 38–49. https://doi.org/10.47738/jads

Shi, Z., Xiao, J., Jiang, J., Zhang, Y., and Zhou, Y. (2023). Identifying reliability high-correlated gates of logic circuits with Pearson correlation coefficient. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 71(4), 2319–2323. https://doi.org/10.1109/TCSII.2023.3334390

Tan, L., Yang, Z., Irfan, M., Ding, J. C., Hu, M., and Hu, J. (2024). Toward low-carbon sustainable development: Exploring the impact of digital economy development and industrial restructuring. *Business Strategy and the Environment*, 33(3), 2159–2172. https://doi.org/10.1002/bse.3584

Trabelsi, M. A. (2024). The impact of artificial intelligence on economic development. *Journal of Electronic Business & Digital Economics*, 3(2), 142-155. https://doi.org/10.1108/JEBDE-10-2023-0022

Wang, S., Hou, D., Guo, Y., and Dai, L. (2022). The mediation effect of innovation in the domestic and international economic development circulation. *Technology Analysis & Strategic Management*, 36(5), 989–1001. https://doi.org/10.1080/09537325.2022.2069003

Wei, X., Hua, B., Wu, Y., and Chen, Z. (2023). A Pearson-type VII distribution with adaptive parameters selection-based interacting multiple model Kalman filter. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 70(8), 3204–3208. https://doi.org/10.1109/TCSII.2023.3252597

Xu, R. (2024). The evaluation of ethnic costume courses based on FP-Growth algorithm. *Scalable Computing: Practice and Experience*, 25(1), 313–326. https://doi.org/10.12694/scpe.v25i1.2297

Yu, H. and Hutson, D. A. (2022). A robust Spearman correlation coefficient permutation test. *Communications in Statistics - Theory and Methods*, 53(6), 2141–2153. https://doi.org/10.1080/03610926.2022.2121144

Yu, H. and Hutson, D. A. (2024). Inferential procedures based on the weighted Pearson correlation coefficient test statistic. *Journal of Applied Statistics*, 51(3), 481–496. https://doi.org/10.1080/02664763.2022.2137477

Yue, L. (2024). E-commerce return data based on frequent itemset mining and time series symbolization clustering. *Journal of Computational Methods in Sciences*, 25(3), 2024–2039. https://doi.org/10.1177/14727978241309189

Zhang, X. J., Cheng, W. J., Philbin, P. S., Ballesteros-Perez, P., Skitmore, M., and Wang, G. (2022). Influencing factors of urban innovation and development: a grounded theory analysis. *Environmental Development and Sustainability*, 25(3), 2079–2104. https://doi.org/10.1007/s10668-022-02151-7

Zhang, Y., Zheng, Z., Yan, Z. L., Chang, Z. M., Jin, C. T., and Yuan, L. Y. (2024). A high-accurate H2S recognition method based on Pearson correlation coefficient feature extraction of dynamic response. *IEEE Sensors Journal*, 24(20), 31712–31719. https://doi.org/10.1109/JSEN.2024.3407741

Zhao, S., Peng, D., Wen, H., and Wu, Y. (2022). Nonlinear and spatial spillover effects of the digital economy on green total factor energy efficiency: Evidence from 281 cities in China. *Environmental Science and Pollution Research*, 30(34), 81896–81916. https://doi.org/10.1007/s11356-022-22694-6

Gaihong Lu obtained her master's degree in Economics (2007) from Zhongnan University of Economics and Law. She is currently an Associate Professor in the Department of World Economics at Guangdong Open University. She has published more than 20 papers in renowned journals both domestically and internationally. Her research interests include international trade theory and policy, global value chains, and cross-border e-commerce.