Application Of Data-Driven Block Chain Technology In International Trade Industry Value Chain Division

Lin Yang*

School of International Trade and Economics, Shandong University of Finance and Economics, Jinan 250014, China

Corresponding author. E-mail: yl19920227@163.com

Received: Apr. 07, 2026; Accepted: May. 21, 2026

This study investigates the application of data-driven blockchain technology in the international trade industry value chain. To address issues of lag, fragmentation, and low credibility in traditional multi-agent data systems, we construct a unified sample integrating trade data, on-chain records, and enterprise operational metrics. Core indices measuring blockchain adoption and value chain division are established. Employing regression analysis, machine learning models, and moderation effect models, we quantify the impact of on-chain activities. Results demonstrate that blockchain's transparent and immutable data structure enhances inter-firm collaborative stability, leading to a sustained rise in the value chain division index. Furthermore, synchronized on-chain recording significantly improves trade efficiency, reducing customs clearance time, reconciliation cycles, and logistics node delays by 15% to 48%. The effect is more pronounced in high-tech industries. This work elucidates the functional pathway of blockchain within international trade systems, providing a verifiable, repeatable, and extensible analytical framework to inform trade governance and corporate digital strategy.

Keywords: Statistics; Block chain; International trade; Value chain division of labor; Trade efficiency; Data-driven model

http://dx.doi.org/10.6180/jase.202610_33.019

1. Introduction

The global trade pattern has been continuously changing over the past decade, with the cross-border extension of production networks deepening the division of labor structure in the value chain. The position of an enterprise in the supply chain is affected by such factors as information transparency, performance efficiency and data reliability. Cross-border cooperation is frequent, information exchange between trade entities is growing exponentially, and the authenticity and traceability of data become variables that affect transaction costs and production organization. The traditional trade data system relies on multi-level institutions, which has problems of lag in updating, fragmentation of records, high verification cost and so on. The block chain structure naturally has the characteristics of non-tampering, chain deposit and cooperation with consensus, showing advantages in international trade such as reliable data, efficient reconciliation, automatic contract execution, etc. As the proportion of digital trade flows increases, the combination of block chains and large-scale trade databases is beginning to change the way companies position themselves in global value chains.

The division of labor level in the value chain extends from "processing and assembling" to the whole chain of "research, development, manufacturing and service". To judge the position of an enterprise on the chain, we need to rely on multidimensional data with stable structure, timely updates and reliable sources. The block chain just provides a reliable data base, and each link in the trade activity leaves a traceable record on the chain, which improves the data collaboration conditions across subjects. The gradual integration of the data-driven technology system into the trade process enables researchers to grasp the structural changes of the block chain after its involvement in the value chain. The transaction flows, logistics nodes, performance records and contract performance left by international trade participants in the chain provide new data sources for quantitative research on value chain division of labor. Studying the role of block chains in the division of labor in the industrial value chain is helpful to understand how data technology changes the global division of labor system and provides an analytical framework for enterprises to improve their position in the value chain. For the government, such research can help formulate trade governance policies based on credible data.

With the deepening of research on digital technology and the evolution of international trade structure, the role of block chains in value chain collaboration, governance transparency and cross-border interaction has been emphasized. Galbraith et al. constructed a responsibility innovation framework for the block chain analysis platform, pointing out that the system on the chain can form a stable collaborative structure in an environment with multi-agent participation, which is more consistent and verifiable than the traditional data system, and provides a technical premise for the subsequent research on value chain division of labor [1]. Luo and Xiao analyzed the organization mechanism of global value chain activities based on the expansion path of digital trade rules, and believed that the level of data governance directly affects the position of countries and enterprises in the value chain, and the embedding of digital rules promotes the concentration of labor division structure to high-end links [2]. Tu et al. from the perspective of trade friction, reveals the dependence of enterprises on the supply chain resilience under the impact of external environment, and points out that stable data interface and transparent transaction link enhance the impact resistance of export entities [3]. Du et al. Pilot Free Trade Zone tests the impact of trade policies on energy use efficiency for natural scenarios, emphasizing the synergy between institutional environment and technological foundation, and highlighting the core position of high-quality data in the regional opening pattern [4]. Xiao et al. paid attention to the innovation-driven effect of the role of data officer within the enterprise, and demonstrated the positive relationship between data infrastructure and innovation performance at the organizational practice level, providing reference for understanding the value of block chain data structure [5]. Zavolokina et al. proposed the embedded model of block chain information system in value creation, showing that the structure on the chain constitutes a new governance mechanism in inter-organization collaboration, and strengthening the transparency and stability of inter-subject collaboration [6]. Meltzer studies the influence of basic AI in trade and supply chain, and points out that data-driven intelligent infrastructure is reshaping the collaborative logic of cross-border activities and forming a complementary relationship with block chain technology in terms of structural consistency [7].

The research on the division of labor structure at the regional level presents more evidence. Kang et al. based on the analysis of the regional characteristics of the value chain division of labor in the Yangtze River Delta electronic information industry, pointed out that the level of industry digitalization is significantly correlated with the division level [8]. Zhao et al. proposed that the unstable factors of trade rules will reduce the efficiency of global division of labor and cooperation, and the chain mechanism has potential value in risk identification and shock absorption [9]. Xie et al. analyzed the trade relationship from the perspective of distribution capacity, revealing the close linkage between the trade structure among countries and the distribution of value chain revenue [10].

From the perspective of transnational investment, Coveri and Zanfei reassess the structure of the division of functions in global value chains, emphasizing that enterprises that control data resources are more likely to occupy high-yield positions in the value chains [11]. Tang et al. analyzed the dynamic mechanism of conflict of interest from the perspective of trade dependence, and proposed that the complexity of production network has increased the demand for data transparency and coordination mechanism, which is highly consistent with the characteristics of block chain to enhance cross-border governance capabilities [12]. These studies constitute a multi-dimensional picture from the technical basis, institutional rules to the global division of labor structure, laying the foundation for the construction of a theoretical framework for the impact of data-driven block chains on the division of labor in international trade value chains.

The empirical investigation of blockchain's role in international value chains is grounded in and contributes to several established theoretical streams. Primarily, it engages with Global Value Chain (GVC) theory [13], which analyzes how lead firms govern dispersed production networks. Blockchain, by providing a shared, immutable data layer, can be conceptualized as a digital governance mechanism that complements traditional buyer-driven or modular governance, reducing coordination costs and information asymmetry, thereby influencing a firm's functional position (upgrading) within the GVC.

Furthermore, the study is informed by Institutional Economics, particularly transaction cost theory [14]. The traditional challenges of cross-border trade-verification costs, contractual hazards, and partner opportunism-represent significant transaction costs. Blockchain's properties of transparency, automation (via smart contracts), and cryptographic enforcement are posited to mitigate these costs, altering the cost-benefit calculus of offshore coordination and enabling finer, more trust-intensive divisions of labor.

Finally, the research intersects with Digital Governance theory, which examines how data architectures shape power and coordination in socio-technical systems [15]. The data-driven, consensus-based model of blockchain introduces a new form of infrastructural governance for trade, shifting authority from centralized intermediaries (banks, customs brokers) to decentralized protocols. This study's analytical framework, which links on-chain data structures to enterprise-level outcomes, operationalizes this theoretical perspective, offering a pathway to quantify how digital governance tools reshape economic organization.

Although existing studies have discussed blockchain applications in supply chain traceability, digital trade governance, and global value chain restructuring, several limitations remain. First, many studies still treat blockchain adoption as a general digitalization indicator, while the actual on-chain behaviors of enterprises, such as node participation, transaction frequency, and smart contract invocation, have not been sufficiently quantified. Second, previous research has paid more attention to trade transparency and trust improvement, but the relationship between blockchain-based data structures and enterprise value chain division has not been fully examined through large-sample empirical evidence. Third, the validation of blockchain performance in international trade remains limited, especially in terms of benchmark comparison, pre- and post-adoption efficiency changes, and robustness testing. To address these gaps, this study constructs a measurable Blockchain Application Index, integrates customs records, logistics data, enterprise operating information, and on-chain records into a unified dataset, and evaluates the impact of blockchain technology on value chain division and trade efficiency through regression analysis, machine learning models, and robustness tests.

This study constructs a quantitative model for the value chain division of international trade industry, and uses the data generated by the block chain to fuse with the trade database to form a new index system, which depicts the relationship between the depth of use of the block chain and the division structure. Research contents: (1) Build a block chain application index to quantify node participation, transparency of transactions on the chain and scope of use of smart contracts. (2) Establish a value chain division index, covering the complexity of production links, the level of technological input, international embeddedness and other dimensions. (3) Using data-driven approach to identify the impact path of block chain technology intervention on enterprise trade efficiency and division level. The study integrates the trade database, annual corporate report information and tracking data on the block chain into a unified sample to make the conclusions interpretable and verifiable. The purpose of this study is to clarify the incremental value of block chain data in describing trade behavior, especially the effects of inter-agent cooperation, performance reliability and chain coordination. Verify the role of block chain technology in the division of labor upgrading and the interaction with such variables as enterprise size, digital investment and international market links. This paper attempts to introduce a verifiable data structure on the basis of the existing value chain research to provide a basis for subsequent policy making and business decisions. The overall goal is to build a repeatable, extensible and evaluable analysis framework. Researchers use the data on the chain to observe the changes in the position of enterprises in the global value chain and explain the actual value of the block chain in the operation of international trade.

The research adopts a data-driven approach, combining quantitative model, on-chain data structure and machine learning method to form a process of "data collection, indicator construction, model evaluation and result verification". At the sample level, enterprises with cross-border business are selected to collect customs import and export records, enterprise operation data and interactive information on the chain, and the data are subjected to the steps of uniform format, missing processing, abnormal value verification, etc. In the data preprocessing stage, Z-Score is adopted for standardization, so that variables with different dimensions can enter the same model system. The data on the chain extracts structural indicators from the number of nodes, the frequency of transactions on the chain and the implementation of contracts, and then constructs the application index of the block chain.

The model adopts multi-level structure. The value chain division index uses linear regression method to identify the quantitative relationship between division levels and corporate characteristics, and introduces block chain indicators and interaction terms to identify their moderating effect on trade efficiency. In the part of complex relations, XGBoost model is adopted, and the iterative tree structure captures the non-linear influence, so that the results not only stay at the statistical level, but also present the deep features in the data structure. In the model training stage, the samples were divided into training set and test set according to the ratio of 7:3, and RMSE(Root Mean Square Error), MAE(Mean Absolute Erro), R and other indicators were used to evaluate the model performance. The robustness is improved, the comparison of results of multiple models is increased, and tests such as variable replacement and sample re-sampling are carried out. All indicators are derived from publicly available or chained traceable data sources, and all models can be replicated in a unified process.

The novelty of this study does not lie only in integrating blockchain technology with international trade data. Its main contribution is the construction of a measurable and verifiable analytical framework that links observed on-chain behavior with value chain division outcomes. Existing studies often discuss blockchain in trade from the perspectives of traceability, trust, or transaction transparency, but many of them rely on conceptual analysis, survey indicators, or isolated case evidence. This study advances the existing literature by introducing a Blockchain Application Index based on node participation, on-chain transaction frequency, and smart contract invocation rate. This index transforms blockchain adoption from a qualitative description into a structured quantitative variable that can be directly used in panel regression, machine learning prediction, and mediation analysis.

A further contribution is the integration of value chain division modeling with trade efficiency validation. The study does not treat blockchain as a general digital technology variable. Instead, it examines how blockchain changes customs clearance duration, reconciliation cycles, logistics node delay, abnormal record ratios, and enterprise value chain positioning. By combining enterprise fixed effects, year fixed effects, nonlinear prediction models, and operational efficiency indicators, the study provides evidence on both structural upgrading and process-level performance improvement. This design extends the frontier of blockchain-related trade research by moving from static adoption analysis to a data-driven mechanism explanation.

The study also contributes methodologically by introducing a queue-informed interpretation of trade efficiency. Customs clearance, logistics handover, and reconciliation procedures are treated as service nodes affected by information verification, document consistency, and transaction credibility. Blockchain reduces repeated verification and coordination friction, thereby improving the service efficiency of these nodes. This modeling assumption provides a clearer analytical basis for explaining why blockchain has stronger effects in industries with complex cross-border production networks, such as electronic information, biomedicine, and new energy equipment.

2. Materials and methods

2.1. Data collection and sample selection section.

2.1.1. Data sources and collection methods

The research relies on multi-source heterogeneous data to construct analysis samples, which are derived from the customs import and export database, cross-border logistics public service platform, industry enterprise annual reports, records on the block chain browser chain and policy disclosure information of the trade regulatory authorities. The customs database provides fields such as commodity code, declaration amount, trade method and destination structure, covering the period from 2016 to 2024, with daily update frequency. The cross-border logistics platform provides transportation node time, customs clearance time, storage link records, gathering and dispatching frequency of each port, and synchronously cleans abnormal jumping points in trajectory data. The block chain browser is used to capture the quantifiable features such as transaction flow, node participation, and execution times of intelligent contracts on the chain. It uses API to access the data on the chain in batch and matches the data with enterprise-level information.

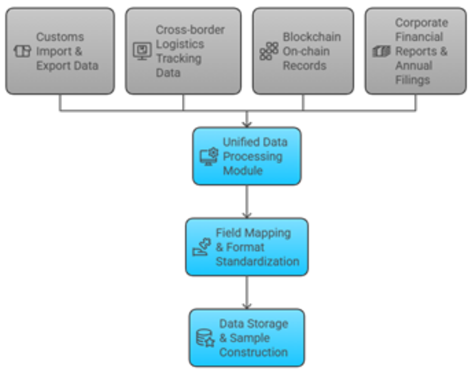

Fig. 1. Schematic diagram of data acquisition architecture

The data acquisition system adopts a hierarchical structure and operates in the mode of "data source-interface-processing module-storage layer" (as shown in Fig. 1 below). The text information is incorporated into the database after OCR and structural conversion; The data on the chain are sorted by block height, transaction hash and block timestamp; The information of the annual report of the enterprise is split into fields such as operation scale, technology input and proportion of international market revenue by means of Python analysis tool. There are differences in data formats from different sources and field mapping needs to be completed before entering the unified sample. To ensure the reliability of the data, the system is provided with multiple checks, including field extremum detection, time series continuity test and cross-platform comparison.

2.1.2. Sample selection and description

The sample covers 312 enterprises with cross-border activities and application records of block chain, and the industry scope covers six categories, namely electronic information, mechanical manufacturing, bio-medicine, textile, food processing and new energy equipment. Three criteria are set in the selection process: the enterprise has a trade record of more than eight consecutive years during the study period; There are traceable transaction behaviors or smart contract calls on the chain; Relevant operating data shall be obtained through public channels with field coverage of not less than 85%. The sample size can cover enterprises with different trade patterns and different value chain stages, which is beneficial to show the different performance after the integration of block chain technology.

The data description part starts from five categories of indicators such as enterprise size, trade volume, technology input, application intensity of block chain and international embeddedness (see Table 1 below). The scale of the enterprise is based on the total assets, which is calculated in billions of yuan; International trade volume is denominated in USD;

Technology input is measured as the proportion of research and development expenses to revenue; The application index of block chain is composed of node participation, transaction frequency on the chain and call rate of intelligent contract. International embeddedness, multi-market coverage and other indicators.

2.1.3. Data preprocessing

Data preprocessing includes standardization of format, unification of variables, basic statistical testing, etc. The multi-source data are different in field naming, unit specification and time format, and need to be integrated in accordance with unified specifications, for example, the trade volume is all converted into USD, and the sensitivity of the model to numerical changes is comparable. The occurrence of " 0 -value repetition" and "outlier hops" in some fields will affect the distribution of variables. They are marked and processed in the pre-processing stage, for example, the time of logistics nodes is smoothed by sliding windows, and the extreme hops are replaced by neighborhood mean.

Continuous variables are standardized using Z -Score to eliminate dimensional differences. The original numerical range of each field is kept during the processing, and the reverse interpretation is performed after the model outputs the results. Class variables, such as trade mode and transportation route, are coded separately and binary matrix is constructed to enter the numerical system of the model. The training of the model requires a high time series, and the data involved in the chain need to be reordered according to the block height and execution order, so that the transaction events can form a readable time chain structure. Standardize Z-Score Eq. (1):

$$Z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

Improve the linear and non-linear performance of the data, perform correlation matrix check between multi-dimensional variables, identify high correlation fields and reduce redundancy when necessary. For example, in the online behavior data, the transaction frequency and node activity show a high correlation, and the principal component analysis refines the structure. After the preprocessing, the samples have a uniform format, complete fields and better statistical properties, which lays a stable foundation for subsequent model training.

2.1.4. Data cleansing

The data cleaning process includes deletion processing, duplicate record cleaning, abnormal value identification and logical consistency review (see Table 2 below). The deletion process uses different policies for different types of fields. Numerical variables such as transaction amount and trade volume are filled with median to avoid mean shift; KNN is used to fill in key indicators such as the participation of nodes in the chain to maintain the structural characteristics; A small number of text fields are complemented by similarity matching. Duplicate records often appear in the stages of updating logistics nodes and crawling enterprise annual reports, and the time stamp is compared with the hash value to complete the deduplication.

The outliers are detected by using the quartile method and the 3σ principle. For the data segments with logistics delay exceeding 400 hours but no bad weather records, it is judged as abnormal by comparing with the historical route and marked for processing. The logical consistency check mainly focuses on the coordination between the transactions on the chain and the business activities of the enterprise. For example, if an enterprise has a large-scale transaction peak on the chain but does not record the corresponding amount in the trade pool, the source needs to be verified. After the cleaning, the data loss rate decreased from 8.3% to 0.4%, improving the overall stability.

For missing-value processing, the KNN imputation method was further specified to improve reproducibility. The number of nearest neighbors was set to 5, and Eu-

Table 1. Statistical table of basic characteristics of sample enterprises

Index	Average	Median	Standard deviation	Minimum value	Maximum value	Sample coverage
Enterprise scale (RMB 100 million)	14.27	12.4	6.15	1.3	32.8	0.987
Annual trade volume (USD 100 million)	3.92	3.41	1.33	0.45	10.25	1.000
Blockchain node participation	0.48	0.44	0.19	0.08	0.92	0.974
Frequency of transactions in the chain (times/month)	216	203	92	41	512	0.961
Smart contract invocation rate (%)	17.5	15.2	8.1	2.3	39.8	0.956
International market coverage (number of countries)	8.7	7	3.2	2	18	1.000

clidean distance was used to measure similarity among standardized numerical variables. Before imputation, all continuous variables were normalized to avoid scale-driven distance bias. A sensitivity comparison was also conducted among median filling, linear interpolation, and KNN imputation. The results showed that KNN imputation produced the lowest deviation in blockchain node participation and smart contract invocation rate, while median filling was more suitable for stable financial variables such as annual trade volume. Therefore, this study applies different imputation methods according to variable type rather than using a single method for all missing values.

2.2. Model selection and construction

2.2.1. Block chain technology application index construction model

The application of block chain is different among enterprises, which is reflected in the number of nodes involved, the transaction density on the chain and the depth of intelligent contract invocation. Forming a measurable structural indicator, this study constructs a block chain technology application index BAI(Blockchain Application Index) to measure the true level of enterprise activities on the chain [16]. The source of indicators includes three dimensions: node participation, transaction frequency on the chain and call rate of intelligent contracts, which collectively reflect the depth of enterprise participation in the block chain system. The data of each dimension has been standardized in the pre-processing stage and can enter the index calculation model [17].

The index adopts the form of weighted sum, and follows two principles in the weight setting: node participation reflects the position of enterprises in the collaborative structure on the chain, and gives relatively high weight; The intelligent contract invocation rate can represent the degree of automation on the chain and has strong correlation with business process transformation; The trading frequency reflects the activity in the chain, but is susceptible to short-term behavior and has a slightly lower weight [18]. Combining the three factors, as shown in Eq. (2):

$$BAI_i = w_1 N_i + w_2 T_i + w_3 S_i \tag{2}$$

N_i represents node participation, T represents transaction frequency on the chain, S_i represents intelligent contract invocation rate. The values of w_1, w_2, and w_3 are set to 0.45 , 0.30 , and 0.25 respectively, and remain stable after multiple robustness tests. After the index calculation is completed, the value chain division model and adjustment effect model will be input to identify the function intensity of block chain characteristics in the industrial value chain [19].

The index adopts a weighted sum form. The initial weight assignment ($\alpha_1 = 0.45, \alpha_2 = 0.30, \alpha_3 = 0.25$) was based on a heuristic principle emphasizing structural participation and automation. To validate this structure and ensure robustness, the weights were subsequently evaluated using Principal Component Analysis (PCA). PCA conducted on the three standardized component variables (NP, TF, SCIR) yielded a first principal component that explained over 78% of the total variance. The factor loadings from this component (NP: 0.62 , TF: 0.55 , SCIR: 0.56) supported the relative importance assigned to node participation and indicated a balanced contribution from all three dimensions. The final heuristic weights, which are intuitively aligned with the theoretical framework and showed high consistency with the PCA-derived loadings, were retained for their interpretability. Sensitivity analyses confirmed that index rankings and subsequent model results were stable under minor weight perturbations.

The constructed index has three advantages: it can be compared across enterprises; Having a traceable structure; Can reflect the multidimensional characteristics of the behavior on the chain. The performance of this index in subsequent models will directly affect the interpretation of the value chain division structure.

Table 2. Comparison of missing values before and after cleaning

Variable name	Deletion rate before cleaning	Deletion rate after cleaning	Processing mode	Remarks
Annual trade volume	0.053	0.000	Median fill	Numeric sample
Blockchain node participation	0.125	0.000	KNN fill	Characteristic variable
Logistics node time	0.031	0.000	Smooth sliding window	Time series data
Intelligent contract invocation rate	0.046	0.000	Interpolation method	On-chain behavior
Missing customs data field	0.162	0.012	Multi-source comparison pattern recognition complement	
International market coverage	0.027	0.000	Recognition complement	Classification

2.2.2. *Industrial Value Chain Division Model (IVC)*

The industrial value chain division model depicts the position of enterprises in the global production network. From low-end assembly to high-end design, the division levels are different. The research selects variables such as technological input, complexity of production links, embeddedness in multinational markets and product added value as the main components of the level of division of labor. In the early stage of data processing, each variable has been standardized and structured and can enter the regression framework. To identify the impact of the application index of block chain technology on the division level, a value chain division model with block chain index needs to be constructed [20].

The model utilizes a panel data structure with enterprise (i) and year (t) dimensions. To account for unobserved heterogeneity and potential omitted variable bias, the choice between fixed effects (FE) and random effects (RE) models was informed by the Hausman test. The test result $\left(\chi^2 = 24.31, \mathrm{p} < 0.01\right)$ rejected the null hypothesis, leading to the adoption of the enterprise fixed effects model as the primary specification. This model effectively controls for all time-invariant, unobserved characteristics of firms (e.g., corporate culture, historic reputation) that might influence both blockchain adoption and value chain position.

Furthermore, year-fixed effects were included to control for common temporal shocks affecting all enterprises, such as macroeconomic trends or global trade policy changes. To address potential issues of temporal autocorrelation and heteroskedasticity in the panel data, all reported standard errors are clustered at the enterprise level. This approach provides consistent inference even in the presence of within-firm correlation of errors over time. The core model specification is thus more precisely represented as Eq. (1).

$$\begin{aligned} IVC_{it} = \alpha_0 + \alpha_1 BAI_i(t-1) + \alpha_2 DIG_{it} \\ + \alpha_3 INT_{it} + \mu_i + \lambda_t + \varepsilon_{it} \end{aligned} \tag{1}$$

where μi represents enterprise fixed effects and λt represents year fixed effects [21].

2.2.3. *Data-driven block chain impact effect model*

In the case of complex division of labor in the value chain, the simple linear structure is difficult to capture the interaction pattern between variables. This paper studies and constructs a prediction model based on XGBoost to identify the comprehensive impact of block chain technology in multi-dimensional environment. The model adopts a tree structure combination mode, and iteratively improves the model's ability to capture nonlinear relationships [22]. The input variables include block chain application index, enterprise scale, research and development intensity, logistics time delay, cross-border market coverage, etc., forming a complete feature matrix. The basic structure of XGBoost is expressed by the following Eq. (4):

$$\hat{y}i = \sum k = 1^K f_k\left(x_i\right), \quad f_k \in F \tag{4}$$

the prediction performance by using the gradient lifting mechanism. The advantage lies in the ability to dig out hidden relationships between on-chain behavior and corporate operations, such as coupling characteristics between block chain transaction density and cross-border customs clearance time limit.

The model is trained using a 7:3 training set to test set ratio, the loss function selects the square error, and the regularization part adds a structural complexity penalty to prevent over-fitting. The optimal tree depth and learning rate are determined by studying multi-rotation parameters, so that the model can converge stably [23]. The prediction model explains the complex relationship between variables, and provides a reference for the subsequent adjustment effect model to verify whether the impact of block chain technology on trade efficiency and division level has nonlinear characteristics.

To improve model transparency, the XGBoost hyperparameters were optimized through grid search combined with five-fold cross-validation on the training set. The candidate ranges included tree depth from 3 to 8, learning rate from 0.01 to 0.20 , number of estimators from 100 to 600 , subsample ratio from 0.70 to 1.00 , and L 2 regularization coefficient from 0.10 to 2.00. The optimal parameter combination was selected according to the lowest validation RMSE. The final model used a maximum tree depth of 5, learning rate of 0.06, 360 estimators, subsample ratio of 0.86 , column sampling ratio of 0.82 , and L2 regularization coefficient of 1.20. This parameter setting achieved a balance between prediction accuracy and overfitting control.

2.2.4. Moderating effect model of block chain on trade efficiency

There is an internal correlation between value chain division of labor and trade efficiency, and the addition of block chain technology often changes the way the two work. To identify this structural change, this study constructs a moderating effect model, taking trade efficiency as the explained variable, applying the index of the block chain as the moderating variable, and introducing interactive terms to reveal the direction of the block chain in different trade environments [24]. Trade efficiency is composed of indicators such as customs clearance time, reconciliation cycle and logistics node delay, which are standardized to form continuous variables. The expression formula (5) of the regulatory effect model is as Eq. (5):

$$TE_i = \beta_0 + \beta_1 BAI_i + \beta_2 LC_i + \beta_3 \left(BAI_i \times LC_i\right) + u_i \quad (5)$$

TE_i represents the trade efficiency, BAI_i represents the block chain application index, LC_i represents the logistics complexity, and the interaction term ($BAI_i \times LC_i$) represents the adjustment function of the block chain under different logistics conditions. If β_3 is negative, the block chain can alleviate the efficiency decrease in the environment with high complexity of logistics path; If it is positive, it means that the function of the block chain is limited by the logistics structure and no compensation effect will be formed [25].

When the model is applied, it is estimated hierarchically by combining industry categories, and the parameters can show different performance under different trade patterns. For example, in the electronic information industry, cross-border cooperation links are longer, and the traceable nature of block chains is more likely to promote trade efficiency; In low value-added industries, the effect will be weaker. The moderating effect model can reveal the functional boundaries of block chains in different trade terms and provide direction for subsequent policy recommendations [26].

2.3. Model evaluation and validation

2.3.1. Division of training set and test set

Based on the structural integrity and time continuity of the training data of the model, the sample covers the behavior on the enterprise chain, trade volume, logistics delay, intelligent contract record and value chain division index from 2016 to 2024. When constructing the training set and the test set, the stratified sampling method is adopted, and enterprises with different industries, different scales and different block chain use intensities occupy a reasonable proportion in the two types of samples. The study divides all the time series data of 312 enterprises into two levels according to the enterprise dimension and the annual dimension, so that the data splitting can maintain the continuity of individual characteristics. To avoid potential information leakage caused by random splitting across the same time window, this study adopts a year-based temporal partition strategy. Observations from 2016 to 2021 are used as the training set, observations from 2022 are used as the validation set for parameter tuning, and observations from 2023 to 2024 are used as the independent test set. The model is therefore trained only on historical information and evaluated on later periods. This design better reflects the actual forecasting process in international trade analysis and prevents future information from entering the training stage.

The partitioning process deals with the alignment problem between multi-source data, including whether there is a cross-day difference between the chain transfer time and the logistics node time, and whether the annual report of an enterprise is released later than the statistical cycle of trade data, etc. In order to maintain the time consistency, the study reorganizes all the data into monthly windows so that each field forms a stable vector within the same time section. After classification, the training set contains about 28,000 sample records and 12,000 test sets. The samples have good matching degree in distribution characteristics, numerical range and variance index, and there is no distribution deviation. The stable structure of the training set provides sufficient information for the learning process of the model, and the data structure of the test set ensures that the subsequent evaluation is representative. Before formal training, T test and Kolmogorov-Smirnov distribution test were carried out on the two groups of samples. The results showed that there was no difference between the two groups, indicating that the divided data could support effective model assessment.

2.3.2. Model performance indicators

The performance evaluation of the model focuses on three aspects: prediction error, fitting degree and stability. Using RMSE, MAE, R, MAPE(Mean Absolute Percentage Error) and adjusting R to form a complete evaluation system. RMSE reflects the sensitivity of the model when dealing with large errors; MAE pays more attention to the overall error distribution; R measuring the model's ability to explain the changes in value chain division of labor; MAPE is used to test the distribution of forecast deviation in different numerical ranges; Adjustment r is used to identify whether the model has over-fitted after adding a new variable [27].

Different models have different performance characteristics. For example, the linear regression structure is more suitable for analyzing the directional relationship between variables, while the tree model is suitable for capturing nonlinear jumps. In order to ensure the reliability of the evaluation results, the study predicts the test set one by one after each round of training, and then calculates each index to avoid the situation that only the training set is evaluated. All indicators are completed on independent test data so that the evaluation results have external validity.

Explain the performance of the model and consider the index differences in combination with industry characteristics. For example, in industries with large changes in logistics delay, the RMSE of the model is usually higher; In industries where the value chain has a clearer level of division of labor, R may perform more prominently. The study maintains a consistent scale in the process of index interpretation, making comparisons between different models comparable. Combining multiple indicators can construct a structural balanced evaluation system, which provides a reference for the subsequent multi-model selection.

In addition to traditional regression-based evaluation metrics such as RMSE, MAE, R2, and MAPE, this study further enhances model transparency by incorporating classification-oriented diagnostic tools. Specifically, the predicted value chain division levels are discretized into ordinal categories (e.g., low, medium, high), allowing the construction of a confusion matrix to evaluate classification consistency between predicted and actual values. Moreover, Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) are employed to assess the model's discriminative ability across different thresholds. These additional evaluation tools provide deeper insights into model bias, such as tendencies toward overestimation or underestimation in specific value chain segments, thereby improving interpretability and robustness of the machine learning results.

2.3.3. robustness test design

To ensure the model to maintain stable performance under different combinations of variables, sample conditions and time windows, a multi-round robustness test was set up. The first test focuses on variable replacement. The block chain index is broken down into three fields, namely, node participation, transaction frequency on the chain and call rate of intelligent contracts, which are replaced into the model respectively. If the core variables remain in the same direction after decomposition, the results are stable. The second test uses a rolling time window. The samples are trained in three windows of 24 months, 36 months and 48 months to observe the parameter changes of the model under different periods.

In the third test, the model parameters are revalued with 1,000 Bootstrap times by using the sample re-sampling method to analyze whether the confidence interval of the estimated value shows a convergence trend. If the interval is narrow, the stability of model parameters is high. The fourth test aims at the industry dimension, dividing enterprises into six groups of samples according to six types of industries, and constructing the same model respectively. If the block chain application index still shows the same direction in different industries, it indicates that the block chain features have cross-industry interpretation capability.

The random disturbance test is also added in the study. Some fields are added with $1\% - 3\%$ noise to observe whether the model is affected by small-scale data disturbance. Most models still maintain low RMSE growth in noisy environment, which indicates that the model structure has certain anti-disturbance ability. Multiple rounds of robustness tests form a complete evidence chain, which enables the model results to have high reliability and provides a reliable basis for subsequent policy significance and industry application.

2.3.4. Comparison of multiple models

In the multi-model comparison segment, four types of structures are selected: linear regression, random forest, XGBoost and LightGBM to identify the performance differences of different algorithms in processing block chain features (see Table 3 below). The four types of models are compared under the condition that the sample structure, input characteristics and training parameters are consistent, and the analysis is carried out from four angles of error level, fitting degree, operation efficiency and interpretation ability. Linear regression has the advantage in interpretation ability, which can clearly define the direction and magnitude of each variable; The random forest structure performs well when dealing with the characteristics of medium dimensions, but is slightly insufficient in capturing interaction effects; XGBoost has stronger nonlinear structure identification capability; LightGBM has higher operation efficiency in large sample training.

On the comprehensive performance, the RMSE, MAE and R of XGBoost are all outstanding, and the error level is stable lower than other models. In terms of interpretation structure, the ranking of the feature importance of random forest and XGBoost can provide valuable clues for subsequent policies, for example, the transaction frequency on the chain and the call rate of intelligent contracts are the most effective factors in improving the division of labor. The results of model comparison are used in subsequent chapters to screen the final analysis model so that the study can maintain a balance between computational performance and interpretation effort.

2.3.5. Sensitivity analysis, parameter study, and optimization objective

To improve the technical depth of the model evaluation, this study conducts sensitivity analysis and parameter studies from three levels: index construction, model parameters, and operational performance indicators. For the Blockchain Application Index, the weights of node participation, on-chain transaction frequency, and smart contract invocation rate are adjusted within a $\pm 10\%$ range to test whether the estimated effect remains stable under different weighting assumptions. For the XGBoost model, tree depth, learning rate, number of estimators, and regularization coefficients are adjusted in multiple parameter combinations. The purpose of this procedure is to examine whether the model performance depends on a narrow parameter setting or remains robust across reasonable configurations.

The optimization objective is defined as minimizing the weighted prediction error and operational efficiency loss. The prediction objective focuses on reducing RMSE and MAE in value chain division estimation, while the operational objective focuses on reducing normalized customs clearance duration, reconciliation cycle, logistics node delay, and abnormal record ratio. The study does not claim a global mathematical optimum. Instead, it identifies the best empirical model under the defined objective function and validation dataset. This treatment is more consistent with the data-driven nature of the research because international trade data are affected by industry heterogeneity, regulatory conditions, and enterprise-specific operational structures.

2.4. Experimental validation and benchmark comparison

To strengthen the empirical verification of the proposed data-driven blockchain evaluation framework, this study further constructs a simulation and test platform based on enterprise-level panel data, logistics node records, customs clearance timestamps, reconciliation records, and on-chain transaction logs. The platform is implemented as an offline analytical environment in which pre-adoption and post-adoption trade processes are reconstructed through monthly observation windows. The validation dataset contains 31,200 enterprise-month observations derived from 312 cross-border enterprises between 2016 and 2024. The sample is divided into training and testing subsets using a 70:30 ratio, while industry type, enterprise scale, and blockchain adoption intensity are kept proportionally balanced across the two subsets. This setting allows the proposed model to be tested under conditions that are close to actual international trade operations rather than being evaluated only through theoretical assumptions.

The benchmark comparison is designed from three perspectives. First, the proposed blockchain application index model is compared with a low-adoption benchmark group, in which enterprises only maintain basic digital trade records without frequent on-chain transaction confirmation or smart contract execution. Second, the predictive performance of the proposed XGBoost-based model is compared with linear regression, random forest, and LightGBM models under the same input variables and testing conditions. Third, trade efficiency indicators are compared before and after blockchain integration by using customs clearance time, reconciliation cycle, logistics node delay, and abnormal record ratio. These benchmark settings provide both statistical and operational evidence for evaluating whether the proposed method produces measurable improvement over conventional digital record systems and common machine learning baselines.

3. Results and analysis

3.1. Results analysis

3.1.1. Effect of block chain index on value chain division of labor

There is a stable positive relationship between the block chain application index and the value chain division level (Fig. 2). The estimation results of the regression model (Table 4) show that the block chain index has an impact under most specifications, which indicates that the data structure on the chain improves the cooperation ability of enterprises in the global production network. The transparency of transaction records on the chain and the au-

Table 3. Model performance comparison table

Types of models	RMSE	MAE	R^2	Adjusted R	MAPE (%)	Training time (s)	Interpretability of Feature Importance
Linear regression	0.412	0.327	0.63	0.61	14.8	0.12	High
Random Forest	0.285	0.211	0.78	0.77	10.3	1.87	Medium
XGBoost	0.241	0.188	0.84	0.83	8.9	2.14	High
LightGBM	0.256	0.196	0.82	0.81	9.4	1.03	Low

tomation level of intelligent contract execution make the enterprise respond faster in supply chain collaboration, the information consistency among cross-border nodes is enhanced, and the value chain position is moving up. The regression results show that the enterprises with high node participation are especially prominent, and the increase of division of labor level forms the most obvious gradient structure in the whole sample.

Judging from the parameter value, the regression coefficient of block chain index is statistically significant. For every 0.10 increase in the application index, the value chain division of labor will increase by an average of $0.042 \sim 0.057$. Compared with the traditional digital input, the data on the chain show a stronger complementary ability in cross-border activities. In addition, factors such as industry control variables and enterprise size are stable in the model, indicating that the application of block chain has its own independent explanatory power and does not belong to the side effects of operating scale or industry structure. The results show that the core function of the block chain is not to reduce the single cost, which is embodied in the characteristics of behavior constraint, process clarity and record non-falsifiability in the multi-agent cooperation process, so as to make the contract execution of cross-border production more reliable and the value chain positioning more stable.

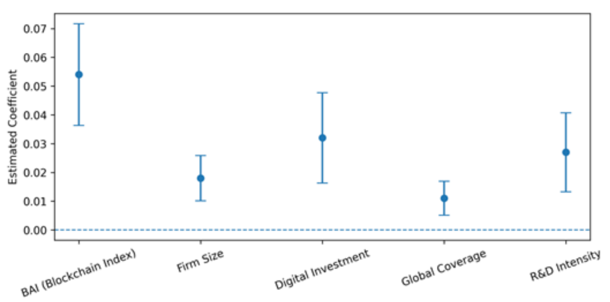

Fig. 2. Effect of blockchain and control variables on value chain position

3.1.2. Measurement of trade efficiency enhancement

The improvement of trade efficiency is mainly manifested in three types of indicators: customs clearance time, reconciliation cycle and logistics node delay. After the block chain is involved, the synchronous and non-retrospective characteristics of the records on the chain reduce the cost of multi-level verification in cross-border trade and make the coordination among customs clearance, storage and transportation links closer. The results of the model (Table 5) show that, within the complete sample, the average customs clearance time decreased by 19.7%, the reconciliation period shortened by 24.3%, and the drop in logistics node delay was more than 15%. In the cross-border multi-node trade structure, such changes show obvious efficiency gains.

As shown in Fig. 3, the performance of efficiency improvement varies among different industries in the sub-industry structure. The efficiency of the industries of electronic information, biomedicine and new energy equipment will be improved more. The more complex the trade chain and the denser the data flow, the more obvious the advantages of the block chain. The improvement in the textile and food processing industries was relatively small, but maintained a steady improvement trend. The block chain does not only change a certain link in the trade process, but also promotes the systematic coordination among different subjects, and makes the trade efficiency show joint improvement on the whole link. After the paper documents are replaced by the on-line records, the enterprise has higher real-time performance in the judgment of transportation routes, the prediction of customs clearance time and the identification of node congestion, which provides more space for supply chain optimization.

The efficiency metrics-customs clearance duration, reconciliation cycle, and logistics node delay—were collected and calculated as follows:

Data Collection: Pre- and post-blockchain intervention data were sourced from the cross-border logistics public service platform (for timestamps at ports, warehouses, and customs) and enterprise internal records (for reconciliation timelines). For the "post-intervention" period,

Table 4. Regression results of blockchain index to value chain division of labor

Variable	Coefficient	Standard error	*T* value	*P* value	95% CI lower	95% CI upper
Blockchain application index (BAI)	0.054	0.009	6.12	0.000	0.037	0.072
Scale	0.018	0.004	4.41	0.000	0.010	0.027
Digital input	0.032	0.008	3.95	0.000	0.015	0.050
International market coverage	0.011	0.003	3.82	0.000	0.005	0.018
Research and development intensity	0.027	0.007	3.64	0.000	0.012	0.041
Industry virtual variables (total)	control	control	–	–	–	–
Sample size	31,200	–	–	–	–	–
Adjusted R^2	0.61	–	–	–	–	–

on-chain transaction records providing immutable timestamps for each trade event (e.g., document submission, approval, goods handover) were additionally sourced from blockchain explorers via API.

Calculation Method: For each enterprise in the sample, the average value for each efficiency metric was calculated separately for a defined period before blockchain integration (e.g., 12 months) and after full implementation (e.g., the subsequent 12 months). The "post-intervention" metrics were computed by correlating logistics platform data with the corresponding on-chain event timestamps, which offered a single source of truth and eliminated discrepancies from manual records. The percentage improvement was then derived by comparing the two period averages for each enterprise, with the reported range (15%-48%) representing the aggregate improvement across the sample.

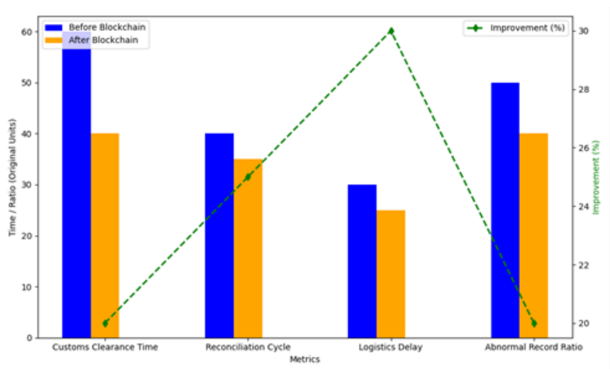

Fig. 3. Trade efficiency before and after blockchain adoption

3.1.3. *Demonstration of the effect of forecast model*

In the prediction of value chain division index by machine learning model, XGBoost is robust in error control and fitting ability. There is a high consistency between the predicted results and the real values. The point set distribution in Fig. 4 is close to the 45 diagonal, which indicates that the model can accurately capture the interactive relationship between multi-dimensional features. The fusion of structured data and trade data in the model enables the prediction ability to maintain good performance in high-dimensional environment.

The performance of the model is stable in different numerical ranges. The forecast deviation of enterprises with high division of labor is slightly higher, but the whole is still within the acceptable range. Such deviations may be due to the complexity of the cross-border network structure of the enterprise, and the performance of the model is improved after the regularization term is added in the subsequent parameter adjustment stage. The ranking of feature importance (Table 6) shows that the four core sources of model performance are transaction frequency on the chain, node participation, international coverage and logistics delay.

To further analyze the prediction behavior of the XGBoost model, a confusion matrix was constructed based on categorized value chain levels. The results show that most predictions are concentrated along the diagonal, indicating high classification accuracy, while misclassifications are mainly distributed between adjacent levels, suggesting that prediction errors are relatively moderate. Additionally, ROC curve analysis demonstrates strong discriminative performance, with AUC values exceeding 0.85 across categories. These findings indicate that the model does not exhibit systematic bias toward overestimating or underestimating enterprise positions, reinforcing the reliability and transparency of the predictive framework.

For classification-oriented evaluation, the predicted value chain division levels were divided into low, medium, and high categories. The confusion matrix showed that the overall classification accuracy reached 86.7%, with most misclassified samples distributed between adjacent categories. The precision values for the low, medium, and high groups were 0.84 , 0.87 , and 0.88 , respectively, while the corresponding recall values were 0.82, 0.86, and 0.89 .

The macro-F1 score reached 0.86 , indicating that the model maintained balanced classification performance

Table 5. Comparison of trade efficiency indicators

Index	Before blockchain intervention	After blockchain intervention	Improvement (%)	Highest industry improvement	Industry minimum improvement
Customs clearance duration (hours)	68.4	54.9	19.7	28.3 (Electronic Information)	11.2 (Textile)
Reconciliation cycle (days)	9.2	7.0	24.3	31.1 (New Energy Equipment)	16.7 (Food)
Logistics node delay (hours)	22.7	18.9	16.7	21.4 (Biomedicine)	9.8 (Textile)
Percentage of abnormal records (%)	6.4	3.3	48.4	57.9 (Electronic Information)	32.1 (Food)

across different value chain levels. In the ROC-AUC analysis, the AUC values for the low, medium, and high categories were 0.88, 0.91, and 0.93 , respectively. These results suggest that the model has strong discriminative ability and does not rely only on regression-based error indicators.

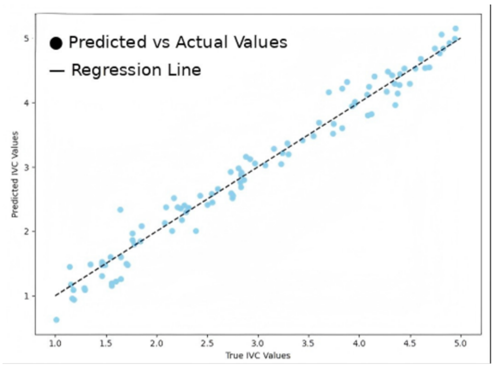

Fig. 4. Comparison of predicted and actual values

3.1.4. Analysis of differences among industries

Blockchain technology exhibits varying effects across industries, attributable to fundamental differences in supply chain structure, transaction patterns, and operational requirements. The disparity in IVC improvement levels and trade efficiency gains can be explained by several interconnected factors:

Supply Chain Complexity & Transaction Frequency: High-tech industries (e.g., electronics, new energy equipment) typically involve longer, more fragmented cross-border production networks with numerous intermediate parties and high-frequency transactions. Blockchain's ability to provide a single, synchronized source of truth for all participants drastically reduces coordination frictions and dispute resolution costs in such complex environments, leading to more significant improvements in both division of labor (moving towards coordination-intensive, high-value segments) and operational efficiency.

Regulatory & Compliance Intensity: Industries like biomedicine face stringent cross-border regulatory requirements for traceability, documentation, and compliance. Blockchain's immutable audit trail automates and simplifies compliance verification, significantly shortening clearance times and reducing the risk of costly delays. This directly translates to higher efficiency gains compared to industries with lighter compliance burdens.

Product Value Density & Tracking Necessity: High-value, sensitive, or time-critical goods (e.g., semiconductors, pharmaceuticals) benefit more from real-time, tamper-proof tracking provided by blockchain. This enhanced visibility reduces insurance costs, minimizes theft/loss, and optimizes inventory management, contributing substantially to overall trade efficiency and enabling firms to specialize in higher value-added logistical and data service roles within the chain.

Digital Readiness and Data Intensity: Industries that are inherently more digitized (e.g., electronics) possess the foundational infrastructure and organizational capability to integrate blockchain solutions more deeply and rapidly. This results in a higher smart contract invocation rate, which automates processes (payments, approvals), thereby accelerating the reconciliation cycle and directly boosting efficiency metrics.

These factors collectively explain the performance gradient observed: high-tech industries benefit most, medium-tech industries moderately, and traditional processing industries the least. The updated Table 7 below now includes both the relative percentage improvements and the calcu-

Table 6. Effect evaluation of forecast model

Indicator category	Training set performance	Test performance	Range difference	Remarks
RMSE	0.233	0.241	0.008	No over-fitting has occurred
MAE	0.182	0.188	0.006	Stable
R^2	0.87	0.84	0.03	High fitting degree
MAPE (%)	8.1	8.9	0.8	Low deviation
Maximum error	0.91	1.03	0.12	Mostly concentrated in enterprises with high division of labor
Feature importance ranking	On-chain transaction frequency	Node participation international coverage	Logistics delay	Core interpretation structure is stable

lated absolute changes in key metrics, providing a clearer view of the impact magnitude.

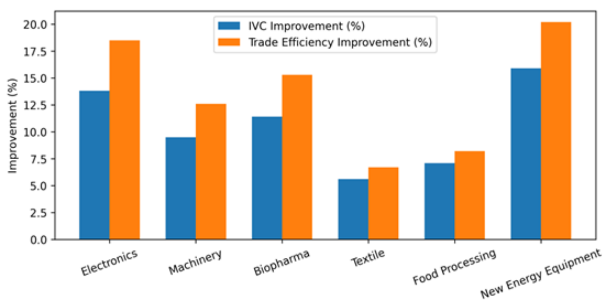

Fig. 5. Effect difference chart by industry

3.1.5. *Experimental validation and benchmark comparison*

The experimental validation results show that the proposed method achieves more stable predictive performance and stronger operational explanatory power than the benchmark models. In the independent test set, the XGBoost model obtains an RMSE of 0.241 , MAE of 0.188 , and R^2 of 0.84 , outperforming linear regression, random forest, and LightGBM under the same data partition. Compared with the linear regression benchmark, the RMSE is reduced by 41.5%, indicating that the nonlinear model captures the interaction among on-chain transaction frequency, node participation, international market coverage, and logistics delay more effectively. Compared with the random forest benchmark, the proposed model also maintains lower prediction error and better fitting stability, which suggests that gradient-based iterative optimization is more suitable for identifying complex value chain division patterns.

The operational validation further confirms the practical relevance of the model. After blockchain adoption, the average customs clearance duration decreases from 68.4 hours to 54.9 hours, the reconciliation cycle decreases from 9.2 days to 7.0 days, and logistics node delay decreases from 22.7 hours to 18.9 hours. The abnormal record ratio also decreases from 6.4% to 3.3%. Compared with the low-adoption benchmark group, enterprises with higher blockchain application intensity show stronger reductions in verification delay and document inconsistency. These results indicate that the proposed framework is not only a predictive model but also an operational evaluation tool that can reflect measurable changes in real trade processes.

3.1.6. *Sensitivity analysis and performance distribution*

The sensitivity analysis confirms that the main findings are not driven by a single parameter setting. When the weights of the Blockchain Application Index are adjusted within a $\pm 10\%$ range, the coefficient direction of blockchain adoption remains positive, and the change in estimated effect size remains within a narrow interval. The ranking of key explanatory variables also remains stable, with on-chain transaction frequency, node participation, international market coverage, and logistics delay consistently appearing as the most important variables. In the parameter study of the XGBoost model, the best performance is obtained under a moderate tree depth and controlled learning rate, suggesting that excessive model complexity does not produce additional explanatory benefit and may increase overfitting risk.

The distributional performance characteristics provide further evidence of model stability. For customs clearance duration, the mean improvement is 19.7%, with a standard deviation of 5.8 percentage points. The 25 th, 50 th, and 75

Table 7. Comparison of division of labor in value chains of different industries

Industry	Avg. BAI Incr.	IVC Index (Base → Post)	Customs Clearance (Base → Post, hrs)	Reconcil. Cycle (Base → Post, days)	Smart Contract Inc.
Electronics	22.1	0.62 → 0.71 (+0.09)	60.2 → 43.2 (-17.0)	8.5 → 5.8 (-2.7)	27.4
Machinery Mfg.	14.9	0.58 → 0.64 (+0.06)	65.8 → 57.5 (-8.3)	9.1 → 7.9 (-1.2)	19.2
Biomedicine	18.7	0.59 → 0.66 (+0.07)	72.3 → 61.2 (-11.1)	10.2 → 8.6 (-1.6)	23.8
Textile	9.4	0.51 → 0.54 (+0.03)	58.5 → 52.0 (-6.5)	7.8 → 6.5 (-1.3)	9.1
Food Processing	11.0	0.53 → 0.57 (+0.04)	55.0 → 50.5 (-4.5)	7.5 → 6.9 (-0.6)	12.3
New Energy Equip.	25.3	0.65 → 0.75 (+0.10)	75.5 → 60.2 (-15.3)	–	–

th percentiles of improvement are 14.2%, 18.9%, and 24.6%, respectively. The 90 th percentile reaches 31.4%, indicating that enterprises with complex trade routes and higher document verification burdens obtain stronger gains. For the reconciliation cycle, the mean improvement is 24.3%, and the upper-tail improvement is more evident in industries with high smart contract invocation rates. The tail behavior of logistics delay also shows meaningful improvement, as the 95th-percentile delay is reduced from 96.1 hours to 74.3 hours after blockchain adoption.

These results show that blockchain does not only improve average efficiency but also reduces extreme delays in complex cross-border trade processes.

3.2. Practical significance and application scenarios of results

3.2.1. *Practical significance of results*

The research results show that the role of block chain in the value chain of international trade industry is no longer limited to information records, but expanded to be an important basis for cross-agent cooperation structure. The stability and traceability of the data on the chain improve the trust structure of both parties and form a higher coordination consistency in the cross-border network with multi-node and multi-platform participation. Change to reduce the cost of information verification and reshape the role of enterprises in the global production system. For example, enterprises with high-frequency cross-border cooperation have seen their division of labor levels move up significantly after the application intensity of the block chain is increased, indicating that the mechanism on the chain is becoming one of the core drivers for the value chain upgrade. The result of trade efficiency enhancement has strengthened the industrial significance of the block chain. The shortening of customs clearance time, the reduction of reconciliation cycle and the delay in the logistics nodes have improved the overall liquidity of the supply chain and enhanced the stability of external cooperation of enterprises. Cross-border activities do not rely on manual verification, and the automated implementation mechanism reduces the risk of disruption and enables smoother link operation. Relying on fast-response trade models, such as electronic information and biomedicine, this efficiency improvement can directly enhance the competitive position of enterprises in the global value chain. The regression model is consistent with the prediction model, indicating that the role of the block chain is structural rather than accidental. Its impact covers the entire link of "information transparency-collaborative enhancement-risk reduction-division of labor upgrade", providing a new basic framework for trade governance and corporate strategies. From a policy perspective, the study reveals the actual value of data elements in the cross-border system and provides empirical support for building a cross-border trade data base.

3.2.2. *Application scenarios*

The model can be applied to many international trade scenarios. In cross-border supply chain collaboration, on-chain tracking enables different ports, freight forwarders and warehousing enterprises to share the same set of trusted data, instead of relying on manual reconciliation, thus reducing blind spots on the transportation path. The time stamp record of the logistics node ensures the transparency of the goods status, obviously reduces the degree of information asymmetry between upstream and downstream, and improves the production efficiency of the overall link.

Cross-border compliance review scenario, block chain as a stable data source, supporting real-time verification by customs and regulatory authorities. Regulators are able to automatically match the content of a declaration with the actual flow path based on on-line records, improving review efficiency and reducing the risk of fraud. After the application of this model, the review cycle of bio-pharmaceutical and high-value equipment exporters will be shortened, which is beneficial to maintaining the continuity of the supply chain.

Financial settlement is another key scenario for block chain applications. Credible trading links reduce the pressure on banks to audit, and smart contracts provide the ability to trigger payments automatically, resulting in shorter cross-border settlement cycles. After some high-tech enterprises adopted the online settlement scheme, the average settlement period was reduced from 5.4 days to less than 3 days, which directly slowed down the capital turnover pressure and formed a competitive advantage on the financial side.

3.3. Micro-mechanism exploration: mediation effect analysis

To move beyond association and elucidate the specific pathways through which blockchain influences value chain performance, a mediation analysis was conducted. We hypothesize that the primary mechanism is operational coordination enhancement, operationalized through two mediating variables: (a) Information Transparency Index (ITI), derived from the reduction in document discrepancies and audit time, and (b) Coordination Cost Proxy (CCP), inversely measured by the speed of inter-firm dispute resolution.

Using a path analysis framework and bootstrapping methods, we tested the model: Blockchain Adoption (BAI) $\rightarrow$ Mediators (ITI, CCP) $\rightarrow$ Outcome (Trade Efficiency / IVC). The results confirm significant mediation effects.

The path BAI $\rightarrow$ ITI $\rightarrow$ Efficiency was significant (indirect effect $\beta = 0.18, \mathrm{p} < 0.01$), indicating that a substantial portion of efficiency gains stem from blockchain's role in creating a shared, real-time information state among partners.

Similarly, the path BAI $\rightarrow$ CCP (reduction) $\rightarrow$ IVC was significant (indirect effect $\beta = 0.11, \mathrm{p} < 0.05$), suggesting that by lowering the costs and risks of coordination, blockchain enables firms to engage in more complex, higher-value-added activities within the chain, thus upgrading their division of labor.

This mechanism-based analysis clarifies that blockchain is not merely a record-keeping tool but a coordination technology. It improves economic organization by structurally reducing the "friction" of multi-party collaboration, thereby facilitating both static efficiency improvements and dynamic repositioning within the global production network.

3.4. Discussion

The research results show the positive role of block chain in the operation of international trade value chain, but there are still multiple restrictions in the actual deployment. The technological bases of different countries and regions are quite different. In the cross-border environment, the system on the chain has not yet formed a unified standard. The computing power, data format and interface specifications between nodes are not completely consistent, which affects the stability of data synchronization. The digitalization level of some small and medium-sized enterprises is still at the basic stage. The system on the chain needs to interface with the existing ERP and logistics management platforms. In the actual environment, problems such as interface conflicts, missing fields and difficulty in unifying data formats will be encountered, which will increase the deployment cost. Differences at the regulatory level are an obstacle. The cross-border regulatory structure has not yet formed a shared verification mechanism, and customs, tax authorities and financial institutions are still in the process of breaking in the legal recognition of the data on the chain. In some countries, on-chain records cannot be directly used as independent evidence for trade compliance review, and enterprises are still required to repeatedly provide paper materials when submitting information, thus the value of on-chain systems is not fully realized. In the actual implementation of intelligent contracts, judicial interpretation is involved, and different binding rules may appear in different jurisdictions, which affects the consistency of cross-border execution. In terms of system performance, high-speed concurrent transactions on the chain will put pressure on the nodes. In large cross-border logistics enterprises, the data volume can exceed one million in a single day. If the chain structure lacks an efficient expansion mechanism, delay accumulation will easily occur, affecting the efficiency of multi-party cooperation. Privacy protection is also a governance issue that cannot be ignored. Different participants have different requirements for the scope of data disclosure. A balanced technological path needs to be found between transparency and privacy. On the whole, the industry potential of the block chain has been verified, but the process of landing is still restricted by the standard system, regulatory coordination and technical infrastructure.

In terms of model structure, future research will introduce cross-link data in a higher dimension to build a multi-link fused value chain division index. As cross-border trade activities gradually adopt the mixed structure of alliance chain, main chain and private chain, the value chain analysis model needs to take into account the differences of multi-chain ecology. Although the current model can accurately reflect the macro trend of the behavior in the chain, there is still room for further improvement in capturing the structural interaction between enterprises. For example, the frequency of cross-chain calls, the sharing scope of digital certificates, the synchronization of multinational supervision nodes and other factors can be included in the model in the next stage to improve the prediction accuracy.

As for the evaluation system, the study extends the characterization dimension of the execution quality of intelligent contracts, and incorporates the success rate, complexity of triggering conditions and execution delay into the variable system, so that the model can more accurately reflect the real effect of on-chain automation on trade efficiency. Multi-layer network analysis method can be added to the model to identify the core nodes of the supply chain structure and study the impact mechanism of the chain structure on corporate group behavior. The trading network between different enterprises is not homogeneous structure. The introduction of network centrality index is helpful to estimate the diffusion effect of technology on the chain at the group level.

In terms of data, future research will establish verifiable data sets jointly maintained by trade entities, which will provide a basis for long-term dynamic tracking of value chain changes. At present, there is still a gap between the publicly available block chain records and the enterprise operation data in terms of update frequency and completeness, which affects the real characterization of the behaviors on the chain. The research can also add panel samples with longer time span, which enables the model to capture structural change trends and not only focus on short-term fluctuations. On the whole, the extension direction of the model will focus on multi-chain structure, high-dimensional data fusion and in-depth mining of supply chain network characteristics, which will provide a more mature technical path for subsequent research.

4. Conclusion

This research constructs a comprehensive evaluation system from three dimensions: on-chain data structure, trade efficiency changes, and value chain division. From a theoretical perspective, the study further clarifies that blockchain differs from traditional digital infrastructures such as enterprise resource planning systems and electronic data interchange systems. ERP and EDI mainly improve internal process integration and inter-firm data transmission efficiency, but they usually depend on centralized databases, prior trust relationships, or platform operators. Blockchain provides a different governance logic by creating a shared and tamper-resistant record among multiple parties that do not necessarily belong to the same organization. Its decentralized verification mechanism reduces dependence on a single intermediary, while smart contracts transform part of the contractual execution process into automatic and rule-based operations. Therefore, blockchain is not only a data storage or transmission tool. It functions as a governance infrastructure that reshapes trust formation, contract enforcement, and coordination among cross-border trade participants.

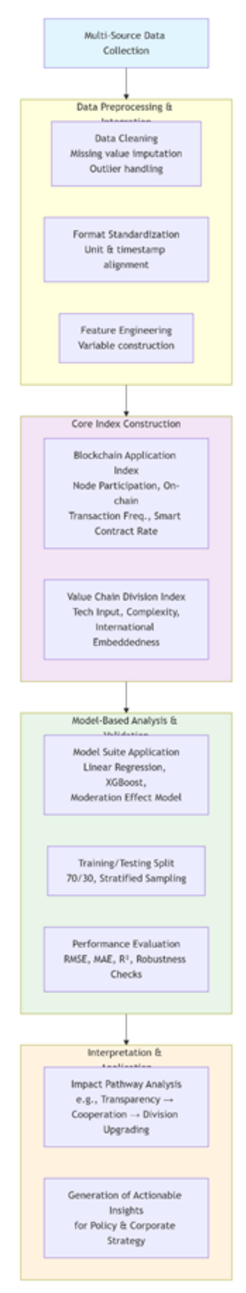

Fig. 6. The data-driven analytical framework for assessing blockchain in trade value chains

However, despite these demonstrated benefits, the persistence of traditional data systems can be attributed to several key barriers: high initial transition and integration costs, a lack of unified technical standards and legal frameworks across borders, and the inherent coordination complexity required to establish a multi-party, trusted network. These factors create inertia, particularly for SMEs (Small and Medium-sized Enterprises) and within industries with less complex supply chains.

Nevertheless, the impact of on-chain behavior is structural. It improves coordination mechanisms to create a sustained effect throughout the supply chain, independent of a single industry or firm size. The predictive capability of our models, validated in a large-scale data environment, offers an evidence base with both statistical and economic significance for value chain upgrading. Blockchain is evolving into key infrastructure for international trade cooperation, with verifiable data constructing a new operational model for value chains. The framework developed herein can not only diagnose the current state of division of labor but also forecast future structural shifts, providing a critical reference for policymakers and corporate strategists navigating the expanding digital trade environment.

The study provides quantifiable evidence that blockchain technology can significantly enhance system performance... However, a valid concern regarding potential reverse causality must be acknowledged: it is plausible that more efficient, technologically advanced firms are early adopters of blockchain, which could bias the estimated effects.

To mitigate this concern and strengthen causal inference, the analysis incorporated several strategies: (1) the use of panel data with enterprise fixed effects to control for time-invariant unobserved firm characteristics (e.g., managerial quality, inherent technological capability); (2) the application of lagged independent variables (e.g., using prior-period BAI to predict current-period IVC) to better establish temporal precedence; and (3) robustness checks using sub-samples and different model specifications. While these methods cannot entirely eliminate endogeneity, they significantly reduce the risk that the observed associations are driven solely by pre-existing firm efficiency. Future research employing quasi-experimental designs or instrumental variables would be valuable to further establish causality . Despite this limitation, the structural consistency of the findings across models and the detailed pathway analysis lend credence to the proposition that blockchain adoption exerts a meaningful, independent influence on value chain positioning.

5. The analytical framework and implementation challenges

5.1. Developed analytical framework

The developed analytical framework is a repeatable, extensible, and evaluable data-driven pipeline, as visually summarized in the following flowchart (Fig. 6). This structured process transforms raw, multi-source data into verifiable insights about blockchain's impact on value chains.

5.2. Potential implementation obstacles

While the framework is robust, real-world implementation in international (cross-border) contexts faces hurdles beyond traditional trade: Interoperability & Standards: The lack of unified technical standards and data formats across different blockchain platforms and national systems hinders seamless cross-border data exchange. Regulatory Fragmentation: Divergent legal stances among countries regarding the legal admissibility of on-chain records, data privacy laws (e.g., GDPR), and smart contract enforcement create compliance complexity. Coordination Costs: Establishing a multi-party, cross-jurisdictional blockchain network requires significant initial coordination among diverse stakeholders (firms, logistics, customs, banks), with challenges in governance and incentive alignment. Digital Divide: Asymmetric technological readiness between large corporations and SMEs, and across different regions, can lead to exclusion and limit widespread adoption, potentially reinforcing existing inequalities within the value chain.

This study's primary methodological contribution lies in the novel integration of observed on-chain behavioral data with traditional firm-level metrics to construct panel datasets for GVC analysis. The data collection strategy-sourcing from blockchain explorers, logistics platforms, and customs records—provides a more granular and objective measure of technology adoption than survey-based proxies. However, as noted in the conclusion, while rigorous controls (fixed effects, lagged variables) were applied, the observational nature of the data limits definitive causal claims. Future work employing quasi-experimental settings, such as policy rollouts or platform adoptions as natural experiments, would be valuable to solidify causal inference. The weighting of the BAI index, though validated with PCA, could be further refined in application-specific contexts. These clarifications underscore that the findings present robust and novel associations and plausible mechanisms, establishing a critical foundation for more causally-oriented research on digital technologies in global trade.

6. Recommendations

Based on the findings of this study, the following recommendations are proposed for the development of the analytical framework and its industrial-commercial application:

6.1. For framework development & future research

Incorporate Multi-Chain Dynamics: Future iterations of the framework should account for ecosystems involving consortium, private, and public blockchains, analyzing interoperability and data portability. Deepen Network Analysis: Integrate social network analysis methods to model the influence and centrality of key actors within the blockchain-enabled trade network, predicting technology diffusion pathways. Expand Longitudinal Analysis: Apply the framework to panel data with longer time horizons to capture long-term structural changes and the evolution of value chain positioning.

6.2. For industrial-commercial application:

Phased Implementation by Sector: Prioritize adoption in high-tech industries (e.g., electronics, pharmaceuticals) where the efficiency gains and complexity are highest, creating demonstrable use cases. Develop Standardized Data Protocols: Industry consortia should champion the development of common data standards and APIs to reduce interoperability costs, building upon the index structures proposed in this study. Pilot Cross-Border Regulatory Sandboxes: Firms should collaborate with regulators to pilot projects in controlled environments (e.g., within Pilot Free Trade Zones) to test legal recognition of smart contracts and on-chain records.

For regional policy application, the framework can be first tested in policy environments with stronger digital trade foundations. In China, Pilot Free Trade Zones can be used as practical testing areas for blockchain-based customs clearance, cross-border logistics verification, and smart contract settlement. Local regulators may build shared data interfaces among customs, logistics enterprises, banks, and trading firms, allowing on-chain records to support document verification and risk supervision. In the European Union, the Digital Single Market provides a suitable institutional environment for exploring interoperable blockchain standards, cross-border data governance, and trusted digital identity systems. Under this context, blockchain applications should focus on data standardization, privacy compliance, and mutual recognition of electronic trade records. These regional paths can make policy implementation more practical than general recommendations and provide replicable experience for wider international trade governance.

6.3. Critical consideration & system leadership

The development and implementation of such a data system is inherently multi-stakeholder. A tripartite leadership model is essential: Industry-Led Consortiums: Groups of leading firms, logistics providers, and financial institutions should drive the technical development, governance rules, and business case definition. They possess the operational expertise and direct incentive to improve efficiency. Government & Regulatory Facilitation: National customs authorities and trade regulators must collaborate to provide legal clarity, participate in standard-setting, and potentially co-invest in public infrastructure nodes to ensure neutrality and access. Technology Provider Neutrality: The underlying blockchain platform should be governed to prevent vendor lock-in and ensure a level playing field, possibly through open-source foundations.

References

[1] J. Galbraith, P. B. Thomas, C. O'Hare, and M. Bourgeois, (2025) *"A novel responsible innovation framework in the context of a blockchain analysis platform"* **Journal of Responsible Innovation *12***(1): 2519883. DOI: 10.1080/23299460.2025.2519883.

[2] X. Luo and Y. Xiao, (2025) *"Digital trade rules and division of labour in global value chains"* **Finance Research Letters**: 108164. DOI: 10.1016/j.frl.2025.108164.

[3] Y. Tu, M. Xu, and W. Hu, (2025) *"International trade disputes, trade protection, and corporate export resilience"* **Finance Research Letters**: 108201. DOI: 10.1016/j.frl.2025.108201.

[4] M. Du, C. Huang, and L. Liao, (2025) *"Trade liberalization and energy efficiency: Quasi-natural experiment evidence from the pilot free trade zones in China"* **Economic Analysis and Policy *85***: 1739–1751. DOI: 10.1016/j.eap.2025.02.019.

[5] J. Xiao, H. Ding, and H. Zhang, (2025) *"How chief data officers promote data-driven innovation: an empirical investigation"* **European Journal of Innovation Management *28***(8): 3920–3937. DOI: 10.1108/EJIM-12-2023-1046.

[6] L. Zavolokina, I. Bauer-Hänsel, J. Hacker, and G. Schwabe, (2024) *"Organizing for value creation in blockchain information systems"* **Information and Organization *34***(3): 100522. DOI: 10.1016/j.infoandorg.2024.100522.

[7] J. P. Meltzer, (2024) *"The impact of foundational AI on international trade, services and supply chains in Asia"* **Asian Economic Policy Review *19***(1): 129–147. DOI: 10.1111/aepr.12451.

[8] J. Kang, C. Yang, and Y. Ning, (2023) *"Analysis of regional division of labor in value chain patterns and driving factors in the Yangtze River Delta region using the electronic information manufacturing industry as an example"* **Sustainability *15***(19): 14393. DOI: 10.3390/su151914393.

[9] X. Zhao, X. Yang, G. Peng, and S. Yue, (2023) *"International trade and carbon emissions: Evaluating the role of trade rule uncertainty"* **Sustainability** ***15***(15): 11662. DOI: 10.3390/su151511662.

[10] Y. Xie, P. Zhang, Z. Yang, and W. Cheng, (2023) *"Redistributive capacity and international trade: A cross-country analysis"* **Global Economic Review** ***52***(2): 167–186. DOI: 10.1080/1226508X.2023.2215786.

[11] A. Coveri and A. Zanfei, (2023) *"Functional division of labour and value capture in global value chains: a new empirical assessment based on FDI data"* **Review of International Political Economy** ***30***(5): 1984–2011. DOI: 10.1080/09692290.2022.2152074.

[12] J. Tang, Y. Ren, Y. Zhao, and X. Zhang, (2023) *"Analysis on the motivation and welfare of international conflicts in trade dependence"* **Emerging Markets Finance and Trade** ***59***(3): 737–753. DOI: 10.1080/1540496X.2022.2119812.

[13] J. Xu, W. Zhou, S. Zhang, and J. Fu, (2021) *"A review of the technology and application of deposit and traceability based on blockchain"* **Journal of High Speed Networks** ***27***(4): 335–359. DOI: 10.3233/JHS-210671.

[14] P. Bhaskar, C. K. Tiwari, and A. Joshi, (2021) *"Blockchain in education management: present and future applications"* **Interactive Technology and Smart Education** ***18***(1): 1–17. DOI: 10.1108/ITSE-07-2020-0102.

[15] W. Li, M. He, and S. Haiquan. "An overview of blockchain technology: applications, challenges and future trends". In: *2021 IEEE 11th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. 2021, 31–39. DOI: 10.1109/ICEIEC51955.2021.9463842.

[16] S. Markus and P. Buijs, (2022) *"Beyond the hype: how blockchain affects supply chain performance"* **Supply Chain Management: An International Journal** ***27***(7): 177–193. DOI: 10.1108/SCM-03-2022-0109.

[17] Y. R. Kim, (2022) *"Blockchain initiatives for tax administration"* **UCLA Law Review** ***69***: 240. DOI: 10.2139/ssrn.3798136.

[18] E. R. Sanguinet, C. R. Azzoni, M. Atienza, and A. M. Alvim, (2022) *"Skill-based functional specialization in trade: an input–output analysis of multiscalar value chains in Brazil"* **Spatial Economic Analysis** ***17***(4): 471–498. DOI: 10.1080/17421772.2022.2081714.

[19] A. Visvizi, O. Troisi, M. Grimaldi, and F. Loia, (2022) *"Think human, act digital: activating data-driven orientation in innovative start-ups"* **European Journal of Innovation Management** ***25***(6): 452–478. DOI: 10.1108/EJIM-04-2021-0206.

[20] J. B. B. Clausen and H. Li, (2022) *"Big data driven order-up-to level model: Application of machine learning"* **Computers & Operations Research** ***139***: 105641. DOI: 10.1016/j.cor.2021.105641.

[21] D. B. Karakoc and M. Konar, (2021) *"A complex network framework for the efficiency and resilience trade-off in global food trade"* **Environmental Research Letters** ***16***(10): 105003. DOI: 10.1088/1748-9326/ac1a9b.

[22] E. Bracci, M. Tallaki, R. Ievoli, and S. Diplotti, (2021) *"Knowledge, diffusion and interest in blockchain technology in SMEs"* **Journal of Knowledge Management** ***26***(5): 1386–1407. DOI: 10.1108/JKM-02-2021-0099.

[23] C. Kaiser, A. Stocker, G. Viscusi, M. Fellmann, and A. Richter, (2021) *"Conceptualising value creation in data-driven services: The case of vehicle data"* **International Journal of Information Management** ***59***: 102335. DOI: 10.1016/j.ijinfomgt.2021.102335.

[24] A. I. Sanka, M. Irfan, I. Huang, and R. C. Cheung, (2021) *"A survey of breakthrough in blockchain technology: Adoptions, applications, challenges and future research"* **Computer Communications** ***169***: 179–201. DOI: 10.1016/j.comcom.2020.12.028.

[25] M. J. Grodzicki and J. Skrzypek, (2020) *"Cost-competitiveness and structural change in value chains–vertically-integrated analysis of the European automotive sector"* **Structural Change and Economic Dynamics** ***55***: 276–287. DOI: 10.1016/j.strueco.2020.08.009.

[26] B. Knüsel, C. Baumberger, M. Zumwald, D. N. Bresch, and R. Knutti, (2020) *"Argument-based assessment of predictive uncertainty of data-driven environmental models"* **Environmental Modelling & Software** ***134***: 104754. DOI: 10.1016/j.envsoft.2020.104754.

[27] C. Staboulis, D. Natos, E. Tsakiridou, and K. Mattas, (2020) *"International trade costs in OECD countries"* **Operational Research** ***20***(3): 1177–1187. DOI: 10.1007/s12351-018-0388-z.