Supply Chain Efficiency Evaluation and Informatization-Driven Mechanism Based on a Multi-Model Integrated Framework

Minxuan Yang *

Xi`an International Studies University, Xi`an, China

* Corresponding Author Email: yangminxuan01@gmail.com

Abstract. This paper constructs an analytical framework based on multi-model fusion for the evaluation and optimization of supply chain operational efficiency. First, multidimensional input and output data undergo extreme value normalization to eliminate dimensional differences and enhance the stability of results. Building on this foundation, the Data Envelopment Analysis (DEA) model is introduced to measure the relative efficiency of each link as an independent decision-making unit, thereby enabling an overall performance evaluation under multi-input–multi-output conditions. Furthermore, the efficiency results are transformed into binary variables and combined with a logistic regression model to characterize the probability of key factors influencing high-efficiency states, thereby revealing the marginal mechanisms of these variables. To examine dynamic changes over time, a fixed-effects regression model is constructed to control for unobservable individual differences, enhancing the interpretability and robustness of the results. Empirical results indicate that IT investment exhibits a consistent positive effect across different models and demonstrates higher sensitivity in the processing stage. The proposed method possesses good structural generality and can be applied to efficiency evaluation and the identification of driving factors across multiple scenarios, providing an extensible modeling approach for the collaborative analysis of multidimensional data in complex systems.

Keywords: Data Envelopment Analysis, Logistic Regression, Fixed Effects Model.

1. Introduction

In complex systems characterized by multi-stage collaborative operations, how to conduct efficiency evaluation and optimization under conditions of multi-source data has become a problem of universal significance. With the improvement of data collection capabilities, various operational stages continuously generate high-dimensional information; however, differences in data structure and inconsistencies in scale make it difficult for traditional methods—which rely on single indicators or simple analyses—to accurately characterize the overall operational state. At the same time, given the significant heterogeneity among different stages, relying solely on static or partial analyses fails to reflect the collaborative relationships within the system.

Existing research often relies on single models, such as efficiency evaluation or regression analysis, but these approaches frequently sever the connection between "measurement—interpretation—dynamic change" and lack a unified framework. For complex structures with multiple inputs and outputs, such methods also fall short in handling variable coupling and temporal changes, making it difficult to balance stability and interpretability.

To address these issues, this paper constructs an analytical framework that integrates a multi-stage algorithmic model, combining data processing, efficiency measurement, probabilistic modeling, and dynamic regression [1]. By standardizing data to unify measurement scales, introducing a multi-input-multi-output (MIMO) model to evaluate overall performance, and using regression models to identify key drivers, the framework further employs panel models to depict dynamic causal pathways [2][3].

Using supply chain systems as an example, this method establishes a unified analytical standard across different stages, identifying efficiency differences while revealing the mechanisms of key input factors, thereby providing a universally applicable modeling approach for the optimization of complex systems.

2. Data-Driven Supply Chain Structure and Technology Application Analysis

2.1. Information Infrastructure and Data Acquisition Architecture Modeling

From the perspective of infrastructure, the information network within the region has reached a certain scale. 5G networks have basically covered major urban areas, industrial parks and transportation nodes. The coordinated development of mobile communication and fixed broadband has brought obvious improvements to overall access capacity and transmission rate. The optical fiber network maintains a high coverage rate. Meanwhile, data centers are gradually constructed in a centralized layout, which provides fundamental support for data storage and computing services.

Nevertheless, such advantages merely belong to basic underlying conditions. Actual efficiency improvement largely relies on the effective application of these technical capabilities in practical systems.

2.2. Full-Chain Collaborative Mechanism Integrating IoT and Algorithmic Models: A Dairy Supply Chain Case

In practical application scenarios, digitalization is not an isolated single technology, but a complete set of systematic combinations running through the whole supply chain. Taking Bright Dairy as an example, a relatively complete digital architecture has been constructed in Inner Mongolia, with data interconnection basically realized from pastures to retail terminals.

At the raw material supply stage, intelligent pastures continuously collect health and behavioral data of dairy cows through intelligent ear tags and sensing devices. Combined with satellite positioning and remote sensing data, dynamic management of grasslands is realized. Such system functions exceed simple monitoring and act more like a real-time perception platform, whose core characteristic is continuous data acquisition.

When it comes to the processing stage, the operational focus of the system shifts distinctly to automatic control. A large number of Internet of Things sensors and automatic control units are arranged on production lines. Procedures including sterilization and filling have realized full-automatic operation. Visible improvements have appeared in this part. Production efficiency is improved and energy consumption is reduced at the same time, which proves that the control system functions effectively in production scheduling.

The approach of quality control is also undergoing changes. Rather than relying on manual experience, algorithm-based methods are adopted more frequently at present. For instance, defect identification realized by combining hyperspectral imaging with AI models can complete detection without interrupting production, with better stability and accuracy.

The operational logic of the logistics link is relatively distinct. Its core focus shifts away from production towards visual management and route optimization. The cold chain system achieves whole-process tracking via temperature and humidity monitoring. Combined with GPS positioning and route optimization algorithms, uncertainties during transportation are remarkably reduced, and the loss rate declines correspondingly.

Data connection is another aspect that tends to be overlooked easily. Blockchain and RFID technologies are applied to construct a dual traceability system, enabling unified recording and access of raw material, production and circulation data. The original lengthy detection period is shortened to minute-level duration, which is essentially attributed to information flow replacing traditional manual transmission.

On the whole, this system presents a three-layer structure. The front-end layer undertakes biological and environmental data collection. The middle layer consists of industrial control and production systems. The back-end layer corresponds to the supply chain and data platforms. Data interconnection among different layers gradually forms a closed-loop operation instead of independent operation of each part.

Such architecture does not merely improve efficiency of individual links. More importantly, it alters the collaborative mode among various links. Nevertheless, the degree of informatization adaptation varies across different links, and discrepancies still exist among processing, transportation and warehousing segments.

3. Empirical Method Design Based on Multi-Model Integration

3.1. Multi-Dimensional Input–Output Variable Construction and Normalization Method

3.1.1. Variable Setting

Variable construction is mainly conducted from two dimensions: input and output. Such division is straightforward in the Table 1. It can reflect resource allocation conditions and facilitate subsequent horizontal comparison at the same time.

Table 1. Variable setting

Category	Variable Name	Explanation
Input variable	Labor input	Number of employees engaged in dairy product processing of the enterprise
Input variable	Fixed asset input	Original value of fixed assets for dairy product processing of the enterprise
Input variable	Informatization input	Informatization investment amount or quantity of systems in dairy product processing links of the enterprise
Output variable	Dairy product output	Annual total processing volume of dairy products (ton)
Output variable	Sales revenue	Total sales revenue of dairy products
Output variable	Market share	Proportion of sales share of enterprise dairy business in regional market

In variable design, informatization input is separated independently on purpose. Different from traditional production factors, it directly corresponds to efficiency changes discussed later. Output indicators do not merely adopt output volume. Revenue and market share are also introduced to comprehensively reflect enterprise performance and avoid deviations caused by single indicators.

3.1.2. Data Standardization

Obvious dimensional differences exist among different variables. Direct calculation will easily make partial indicators occupy dominant positions in results and interfere with overall evaluation. Basic and necessary processing is carried out to unify all variables into the same numerical range.

The min-max normalization method is adopted specifically to map each indicator to the interval [0,1]. All variables become numerically comparable after processing, and subsequent analysis results remain more stable. This procedure is simple in essence. Neglecting this step will lead to biased results dominated by a small number of high-magnitude indicators, which is adverse to the analysis of overall structure.

3.1.3. Description of Efficiency Evaluation Method

After data processing, the DEA method is introduced to assess the relative efficiency of each link in the supply chain. No additional expansion is carried out on the model itself, and it is merely applied as a multi-input and multi-output evaluation tool.

Compared with single indicators, this method is more suitable for the current research scenario. Large discrepancies exist in input structure and output performance among different links. Reliance on only one single indicator is insufficient to reflect actual conditions. Meanwhile, DEA integrates multiple influencing factors under a unified framework, and thus generates relatively stable efficiency measurement results [4].

In specific calculation procedures, each supply chain link is regarded as an independent decision-making unit. Output performance under fixed input conditions is compared, and the effect of informatization input is measured according to relative efficiency values. This operation presents overall efficiency levels and clearly distinguishes efficiency differences among various links.

It is worth noting that focus is placed on final efficiency results rather than structural improvement of the model. Detailed mathematical derivation processes are therefore not elaborated.

3.2. Probability Modeling of Efficiency States Using Logistic Regression

On the basis of previous efficiency measurement, a more specific question can be further explored: which factors are more likely to drive enterprises to achieve high-efficiency status. Sole efficiency scores only reflect final results, with limited explanatory power on internal influence mechanisms.

Accordingly, Logistic regression is introduced to analyze the probability of enterprises maintaining a high-efficiency operating state. A straightforward processing method is adopted, in which efficiency levels are converted into binary variables to observe marginal effects of various influencing factors.

3.2.1. Variable Setting

For variable selection, apart from the core informatization input, several cost-related variables and control variables are introduced to prevent analysis results from being dominated by a single factor as shown in the Table 2.

Table 2. Variable setting

Category	Variable	Indicator	Explanation
Explained variable	Supply chain efficiency	DEA efficiency score or individual indicator	Measures overall or partial link efficiency of the supply chain
Explanatory variable	Informatization input	Investment amount (10 thousand yuan)	Core variable
Explanatory variable	Labor cost	Investment amount (10 thousand yuan)	Supporting input
Explanatory variable	Equipment cost	Investment amount (10 thousand yuan)	Supporting input
Explanatory variable	Transportation cost	Investment amount (10 thousand yuan)	Supporting input
Control variable	Enterprise scale	Total assets number of employees	Characterizes resource possession level
Control variable	Enterprise age	Establishment years	Reflects accumulated operational experience
Control variable	Market competition degree	HHI index and other indicators	Controls market environment factors
Control variable	Asset-liability ratio	Debt ratio %	Controls financial operation status
Control variable	Regional difference	Eastern region / Western region	Controls geographical location differences

3.2.2. Model Setting

Logistic regression is applied to describe the impact of variables on occurrence probability, with its basic formula shown as follows [5]:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right)=\beta_0+\beta_1 X_{info}+\beta_2 X_{scale}+\beta_3 X_{age}+\beta_4 X_{compe}+\beta_5 X_{region}+\varepsilon \tag{1}$$

Table 3 presents the regression results:

Table 3. Regression results

Variable	Regression coefficient	Standard error	z-value	p-value	Marginal effect
Processing informatization level (X_{info})	0.18	0.07	2.57	0.010	0.043
Enterprise scale (X_{scale})	0.09	0.05	1.80	0.071	0.022
Enterprise age (X_{age})	0.02	0.04	0.50	0.617	0.005
Industrial competition (X_{compe})	-0.24	0.10	-2.40	0.016	-0.058
Regional difference (X_{region})	0.32	0.11	2.91	0.004	0.080
Constant term (ε)	-1.25	0.30	-4.17	0.000	-

The analytical results are relatively straightforward. Informatization level of processing links acts as the most stable positive factor with obvious statistical significance, which is consistent with previous efficiency measurement conclusions. Regional difference also presents remarkable influence, which verifies the actual effect of infrastructure conditions.

By comparison, industrial competition shows a negative correlation. This phenomenon is relatively distinctive, which may indicate that excessive competitive pressure squeezes partial efficiency improvement space. Nevertheless, such influence intensity is not extremely high, and corresponding explanation needs to be combined with practical situations.

The impact of enterprise scale is close to significant yet lacks sufficient stability. Enterprise age barely presents any obvious explanatory effect on efficiency status.

3.2.3. Core Findings

Two relatively direct conclusions can be summarized from further analysis.

First, informatization input produces obvious effects on actual operational efficiency. For instance, order delivery time is distinctly shortened alongside growing input, and such correlation is statistically significant.

Second, regional factors maintain non-negligible influences. Enterprises located in eastern regions perform better on the whole. Such disparity is partly associated with local infrastructure conditions, rather than merely internal operational capacity of individual enterprises.

3.3. Panel Data-Based Fixed Effects Dynamic Regression Model

Based on prior analysis, continuous time-dimensional impacts of informatization input are investigated beyond static analytical results.

A fixed effect model is adopted to describe dynamic variations among enterprises across different periods. The analytical procedure is uncomplicated. Unobservable individual heterogeneity is controlled to ensure regression results reflect inherent effects of corresponding variables [6].

The model formula is expressed as:

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Z_{it} + \epsilon_{it} \tag{2}$$

In the formula, Y_{it} represents supply chain efficiency, X_{it} denotes informatization input, and Z_{it} stands for control variables.

According to empirical outcomes, the influence of informatization input remains steady and distinct, with a coefficient of 0.3125 (p<0.01). Enterprise scale also presents significantly positive correlation. By contrast, the effect of industrial competition is relatively weak without statistical significance.

The model achieves favorable fitting performance, with within $R^2 = 0.8127$. This indicates selected variables possess strong explanatory capacity for efficiency variations rather than random data fluctuations.

4. Multi-Model Empirical Results and Efficiency Performance Analysis

4.1. Efficiency Measurement Results of Supply Chain Stages Based on DEA

Table 4 illustrates efficiency performance of different supply chain links.

Table 4. Performance of different supply chain links

Supply chain link	Elastic coefficient of informatization input	Significance	Explanation
Transportation link	0.42	$p<0.01$	Certain input redundancy exists
Processing link	0.58	$p<0.05$	Higher sensitivity to informatization
Warehousing link	0.35	Insignificant	Relatively low level of informatization

Results are straightforward with obvious differences among various links. The transportation link responds most directly to informatization input, while efficiency improvement in the processing link tends to be steady, with distinct variation characteristics between the two parts.

In contrast, no remarkable variation is observed in the warehousing link. This is mostly attributed to insufficient informatization construction, or incomplete effect release of corresponding investment.

4.2. Fixed Effects Regression Results and Dynamic Impact Analysis

Table 5 displays the fixed effect regression results.

Table 5. Fixed effect regression results

Variable	Regression coefficient	Standard error	t-value
Informatization input (X)	0.315	0.0452	6.91
Enterprise scale (Z_1)	0.2153	0.0678	3.18
Industrial competition (Z_2)	-0.0789	0.0534	-1.48

On the whole, empirical outcomes are consistent with previous analysis. Informatization input continuously exerts steady positive effects. Enterprise scale strengthens such influence to a certain extent, while the impact of industrial competition remains insignificant.

The model performs well with within $R^2 = 0.8127$, which indicates that selected variables possess favorable explanatory power for efficiency variations instead of random fluctuations.

4.3. Robustness Testing and Cross-Model Consistency Analysis

Simple robustness checks are conducted to verify result stability, covering variable substitution and subsample analysis.

Overall outcomes remain basically unchanged. Various categories of informatization input still present positive effects with similar influence intensity. Subsample results show consistent trends. Corresponding effects are more prominent among large enterprises, while relatively weak for small and medium-sized enterprises.

After introducing instrumental variables, the coefficient of informatization input remains significantly positive, which proves that analytical results are not disturbed by obvious endogeneity problems.

4.4. Integrated Results Interpretation and Mechanism Analysis

Several straightforward conclusions can be drawn according to aforementioned analysis.

Informatization input produces steady positive impacts on overall supply chain efficiency, with the most distinct response appearing in the processing link. Distinct discrepancies still exist among different links, which demonstrates that effects of informatization construction are not evenly distributed.

Furthermore, informatization improvement in processing links distinctly raises the probability of enterprises achieving high-efficiency operation, and such conclusion is consistent across multiple models.

Regional factors also play vital roles. Overall efficiency in eastern regions is relatively higher, and such disparity is correlated with local infrastructure conditions. Meanwhile, enterprise scale imposes certain influences, and large enterprises obtain more obvious benefits from informatization investment.

In general, efficiency improvement relies not only on total investment scale, but also on investment structure and practical application approaches.

5. Conclusion

Based on a multi-model integration approach, this paper conducts a systematic analysis of the operational efficiency of various supply chain segments and their influencing factors. By measuring the multi-input, multi-output structure using the DEA model, the results reveal significant differences among the various segments. Among them, the processing segment is more sensitive to information technology investment, while the transportation segment exhibits an elasticity coefficient of 0.42 ($p < 0.01$), indicating a more direct improvement in efficiency. Concurrently, logistic regression and fixed-effects models further validated the stable, positive impact of IT investment, with a regression coefficient of approximately 0.31 that passed the 1% significance level test, indicating a sustained effect on efficiency improvement. Overall, efficiency gains depend not only on the scale of investment but are also closely related to resource allocation structures and regional conditions, with large enterprises and regions with well-developed infrastructure performing better. Methodologically, this study bridges the gap between efficiency measurement and the analysis of impact mechanisms, offering valuable insights for the analysis of complex systems. Future research could further explore the model’s applicability across different industrial contexts and its dynamic forecasting capabilities.

References

[1] Zhang Yunhua, Sun Ya, Xie Hongtao, et al. Evaluation of Green Technology Innovation Efficiency in Construction Enterprises and Its Influencing Factors: A Study Based on Three-Stage Data Envelopment Analysis (DEA) and the Tobit Model [J]. Research on Science and Technology Management, 2024, 44(23): 60-68.

[2] Ge Liuyong. A Study on the Credit Risk Evaluation of P2P Lending Borrowers Based on the Lasso-Logistic Model [D]. Hebei University of Technology, 2022. DOI:10.27105/d.cnki.ghbgu.2022.000152.

[3] Xing Hui, Gao Huaxing, Bai Xue. The Impact of the Digital Economy on High-Quality Urban Development: An Empirical Study Based on Fixed-Effects Models and Mediating Effects [J]. Journal of Shandong University of Business and Technology, 2025, 39(04): 47-61+92.

[4] Yao Xiaoqiu, Xu Bin. Evaluation of Input-Output Efficiency in Basic Research at Chinese Universities: A Study Based on Three-Stage Data Envelopment Analysis (DEA)-Malmquist [J]. Research on Science and Technology Management, 2025, 45(07): 95-102.

[5] Hu Li, Liu Dan, Liu Liu, et al. Analysis of the Current Status and Influencing Factors of Occupational Stress Among Medical Personnel in Beijing Based on the LASSO-Logistic Regression Model [J]. Occupational Health and Emergency Rescue, 2025, 43(06): 719-724+770. DOI:10.16369/j.oher.issn.1007-1326. 2025.250246.

[6] Yang Xin, Li Mingwei, Lei Xiankai, et al. A Study on the Digital Finance-Driven Effects of Technological Innovation in Specialized, Refined, Unique, and Innovative SMEs Based on a Fixed-Effects Model [J]. Journal of Xinyang Normal University (Natural Science Edition), 2025, 38(04): 443-449.