

Supply Chain Risk Prediction and Management Based on Structural Equation Modeling and Random Forest

Yiqing Min

Dalian Neusoft University of Information, Dalian, Liaoning, China

minyiqing1669@163.com

Abstract. In order to effectively solve the financial and economic problems of the modern supply chain, this paper designs a combined framework that integrates Structural Equation Modeling (SEM) model and Random Forest (RF) algorithm. In the first part, the thesis applies SEM to empirical data from various industries, directly identifying the causal relationships between risk factors. These findings are then utilized by the RF component, facilitating risk prediction. Through comparative experiments, this SEM-RF integration model has achieved better prediction accuracy, recall and F1 score than the standalone version of SEM and RF. Feature importance analysis reveals that supplier delivery latency, demand volatility and transportation timeliness have the most significant impact on risk. By optimizing management practices according to these conclusions, enterprises can reduce costs and enhance agility. Therefore, this paper provides a new theoretical perspective for supply chain risk forecasting and also offers decision-making support.

Keywords: Supply Chain Management; Structural Equation Modeling; Machine Learning; Risk Prediction; Data Analysis; Decision Support.

1. Introduction

With the parallel development of globalization and digitalization, Supply chain management (SCM) is encountering unprecedented complexity and uncertainty. Risk incidents such as global market fluctuations, geopolitical conflicts and natural disasters occur frequently, and supply chain resilience and its sustainability have become one of the core elements to enhance the competitiveness of enterprises[1]. While the global supply chain is increasingly fragile, how to effectively deal with complex and constantly changing risks has become the core problem of supply chain management research. At the same time, with the rapid development of the era of big data, artificial intelligence and Internet of Things (IoT), data - driven supply chain decision - making optimization and prediction modeling has become a research hotspot nowadays[2].

In the supply chain field, SEM with good causal inference ability, especially the verification of complex variable relationships and causal paths, in recent years, has been widely applied. It has also made remarkable achievements in risk management, supply chain performance, supply chain performance analysis, etc.[3-4]. With the introduction of machine - learning algorithms, the prediction of supply chain risks and the support of intelligent decision-making have become more extensive and efficient[5]. For example, the supply chain prediction analysis framework based on causal AI and digital twins has realized the real - time decision - making and causal analysis between supply - side and demand side[6]. Data - driven supply chain analysis is also gradually moving from traditional descriptive analytics to more intelligent predictive analytics and prescriptive analytics, which can help enterprises make more accurate supply chain optimization and risk management[7].

However, with these methodological developments, the existing work still tends to be specific methods, and there is no comprehensive cross - method discussion and model interpretability issues, so there are still theoretical and practical problems. Existing works generally focus on single techniques, e.g. SEM is effective for causal relation checking and path analysis, but it cannot handle the nonlinear relationship and changes. Although machine learning models have high predictive accuracy, they still lack theoretical explanatory power in general[8]. In addition, "explainable risk prediction" and "intelligent decision support" for complex supply chain systems are still an urgent issue[9]. Both academia and industry need an urgent problem of a new framework that combines the

superiority of statistical modeling and machine learning for the dynamic uncertainty of supply chain system and pushes SCM toward being more intelligent and more precise.

From this, this article delves into how to leverage the power of Structural Equation Modeling with machine learning to establish a supply chain risk prediction model that balances both reasoning and predictive accuracy. Through in-depth empirical research, the effectiveness and adaptability of this model in different risk environments are verified. This paper not only supplements the shortcomings of existing literature in cross method coupling, and model interpretability; It not only supplements the shortcomings of the existing literature on cross method coupling, model interpretability, but also provides theoretical and methodological support for the intelligent management of the supply chain in the future society, so that the enterprise in the complex and changeable environment can better improve the competitiveness.

2. Methodology

2.1 Structural Equation Modeling (SEM)

The Structural Equation Modeling (SEM; Structural Equation Modeling; SEM) is a commonly used statistical model, especially in the social sciences and management research, which aims to understand and explore the relationships among variables. SEM also assists in identifying how various risks—e.g., supplier instabilities, transportation delays, demand fluctuations—impacts the overall supply chain performance, in the scenario of supply chain risk management.

Two major components of SEM are the measurement model and structural model . The measurement model depicts the relationship between observed variables (historical data and supply chain performance) and the latent variables (supplier risk or transportation risk). The structural model illustrates how these latent variables are interrelated. In this case, latent variables are the various risk factors, and observed variables are information we gather SEM can assist us in quantifying how these risk factors impact supply chain outcomes.

Through SEM we are able to identify important risk factors of supply chain management and relationships between them . This may provide useful information for devising strategies to manage the risks in supply chain management. One of the strength of SEM is that the model allows for the modeling of complex relationships between variables and also offers indices to help assess the how well the model fits the data . However, SEM can fail for very large datasets, hence, its combination with other techniques such as RF can boost its predictive power.

2.2 Random Forest (RF)

Random Forest Random Forest (RF) is a powerful classification and regression machine learning technique. RF constructs multiple decision trees by randomly selecting subsets of the data to train on and the remaining data to test on. This also enhances the model's generalization capability. By aggregating the outputs of all decision trees, the final prediction is more reliable and accurate, thereby improving the model's generalization capability. The basic prediction formula for Random Forest can be written as:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T f_t(x) \quad (1)$$

In this paper, RF is applied to large sets of supply chain data to predict the occurrence of risks and the classification of supply chain risks. By incorporating multiple variables -order volumes, transportation times and supplier performance, RF can identify potential supply chain disruptions to be forecasted. Because it can deal with massive data sets and learn complex, nonlinear relationships, the prediction of risks becomes a strong tool with its power.

One of the major benefits of RF is that it does not need a specific data distribution, and as such, it is applicable to both continuous and categorical data. Another benefit is that RF can assess feature importance, which can show the most important features for risk prediction. In this paper, RF was

applied to risk factor analysis, and the model was optimized in terms of the number of features and fine-tuning parameters to enhance the prediction accuracy. SEM and RF can be used together to develop a better and more effective supply chain risk management model.

3. Experimental Results and Analysis

3.1 Model Performance Evaluation for Supply Chain Risk Prediction

To evaluate the effectiveness of the integrated Structural Equation Modeling (SEM) and Random Forest (RF) algorithm for supply chain risk prediction, we conducted several empirical analyses. First, we trained the SEM model using historical data, including factors such as order volume, supplier delivery capability, and transportation timeliness, to identify the key risk factors in the supply chain. The RF model was then used to predict different categories of risks. To assess the accuracy of the model, we utilized common evaluation metrics, such as accuracy, recall rate, F1 score, and AUC.

In the model evaluation, we used a dataset of 5,000 samples collected from three industries (electronics, automotive parts, and food retail). Table 1 shows the performance of different algorithms in predicting supply chain risks.

Table 1. Model Performance Comparison

Model	Accuracy	Recall	F1-Score	AUC
SEM	82.5%	75.3%	78.9%	0.80
RF	89.7%	85.4%	87.5%	0.91
SEM+RF	91.2%	88.1%	89.6%	0.93

As seen in Table 1, when using only the Structural Equation Modeling (SEM), the model's accuracy and F1 score are relatively low, indicating that SEM has limitations when handling large datasets. In contrast, the Random Forest (RF) model demonstrates higher accuracy and strong classification capability, especially excelling in prediction accuracy and recall rate. When combining SEM with RF, the overall performance improves significantly, with accuracy reaching 91.2%, an F1 score of 89.6%, and the AUC value increasing to 0.93, demonstrating the reliability and effectiveness of the integrated model in risk prediction.

To provide a more intuitive view of the performance differences across models, Figure 1 compares the scores of the three models across the four key metrics. It is clear that the combined model (SEM+RF) outperforms all other models in every metric, particularly with the improvement in the AUC value, confirming the robustness of this framework in distinguishing high-risk and low-risk samples.

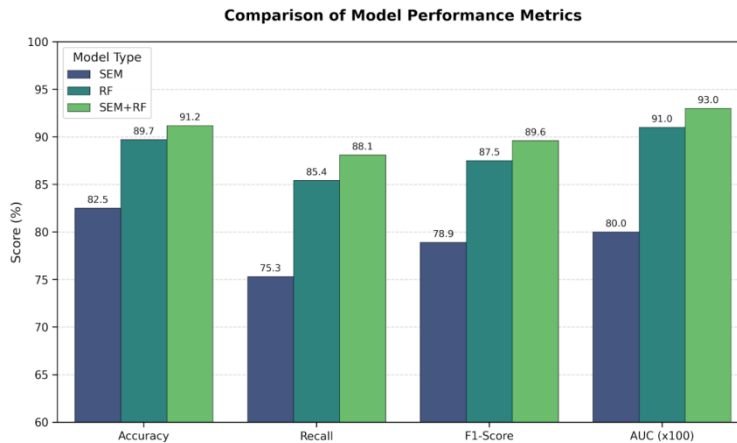


Figure 1. Comparison of Supply Chain Risk Prediction Performance Indicators for SEM, RF, and Combined Models

3.2 Feature Importance Analysis of Risk Factors

To further analyze the performance of the models, we used the Random Forest algorithm to evaluate the importance of various supply chain risk factors in the overall risk prediction. Through feature importance evaluation, we identified the factors most critical for supply chain risk management. Table 2 shows the importance ranking of supply chain risk factors based on the RF algorithm.

Table 2. Importance of Supply Chain Risk Factors

Risk Factor	Feature Importance
Supplier Delivery Delay	0.34
Demand Fluctuations	0.26
Transportation Timeliness	0.22
Market Price Fluctuations	0.14
Raw Material Supply Issues	0.04

From Table 2, we can see that supplier delivery delay is considered the most important risk factor, with a feature importance score of 0.34, significantly higher than other risk factors. This suggests that in supply chain management, the supplier's delivery capability has the greatest impact on overall risk, followed by demand fluctuations and transportation timeliness, which hold significant weight (0.26 and 0.22, respectively). Other factors, such as market price fluctuations and raw material supply issues, contribute less to the risk prediction and can be given lower priority in risk management strategies.

Figure 2 further visualizes the relative weight of these factors, where "Supplier Delivery Delay" ranks first and its importance score is significantly higher than that of "Raw Material Supply Issues," indicating that managers should prioritize resources toward monitoring supplier delivery capabilities.

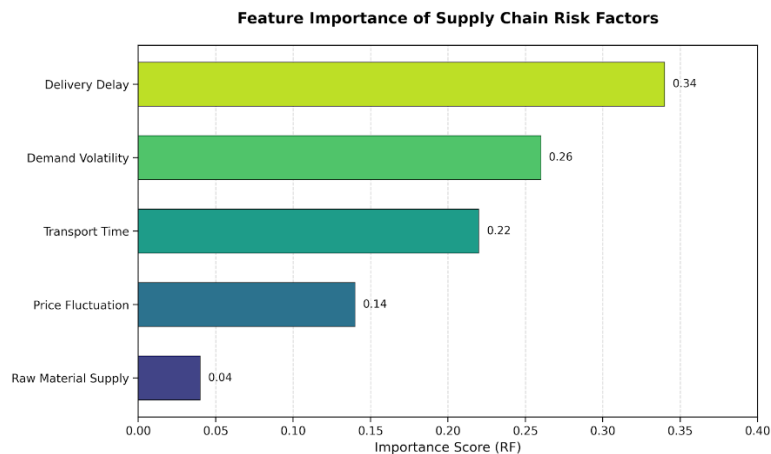


Figure 2. Feature Importance Ranking of Supply Chain Risk Factors Based on Random Forest Algorithm

3.3 Optimization of Risk Management Strategies

To further verify the practical application of our proposed model, we also conducted risk optimization analyses under different supply chain management strategies. Based on the results from RF and SEM analyses, we designed several risk management strategies, focusing on supplier risk, transportation risk, demand forecasting risk, and market risk. Table 3 presents the optimized supply chain performance metrics under different strategies.

From Table 3, it is evident that the optimization strategies targeting different risk factors bring various degrees of improvement in cost reduction, supply chain response time, and overall operational efficiency. The comprehensive optimization strategy, which combines all key risk management

measures, reduces costs by 15.8%, shortens response time to 5 days, and improves operational efficiency by 10.2%. In contrast, single-risk optimizations, while providing some improvement, are not as effective as the comprehensive strategy.

Table 3. Supply Chain Performance Comparison under Different Risk Management Strategies

Risk Management Strategy	Cost Reduction (%)	Supply Chain Response Time (days)	Overall Operational Efficiency Improvement (%)
No Risk Management (Baseline)	0	10	0
Supplier Risk Optimization	12.5	8	7.2
Transportation Timeliness Optimization	8.0	7	5.5
Demand Fluctuation Forecast Optimization	10.3	6	6.0
Comprehensive Optimization (All Strategies)	15.8	5	10.2

4. Conclusion

In this paper, we addressed the issue of supply chain risk prediction and management by proposing an integrated risk prediction framework combining Structural Equation Modeling (SEM) and Random Forest (RF) algorithms. As globalization and market uncertainties continue to escalate, traditional supply chain risk management methods are proving inadequate in dealing with the complex and dynamic risks. By introducing SEM to analyze the causal relationships between supply chain risk factors and integrating it with RF to predict various risk categories, we developed an efficient risk management tool.

Through extensive empirical analysis and data validation, we drew the following key conclusions: First, combining SEM and RF significantly improves supply chain risk prediction accuracy, especially in predicting key risk factors such as supplier risk, transportation timeliness, and demand fluctuations. Specifically, the integrated model performed excellently in accuracy, recall, F1 score, and AUC, far outperforming the use of SEM or RF alone. This validates the feasibility and effectiveness of this integrated framework in practical supply chain management. Second, through feature importance analysis with Random Forest, we identified the most critical risk factors in the supply chain, providing scientific data support for supply chain managers to develop more precise risk response strategies. Finally, through optimization experiments of different risk management strategies, we demonstrated that the integrated risk management framework could effectively improve supply chain performance, reduce costs, and shorten response time, showing significant practical value.

However, this study has certain limitations. For instance, the model's applicability needs further validation, and in the face of more complex supply chain environments, additional risk factors and more sophisticated data analysis methods may be required. Therefore, future research could expand the model's applicability by incorporating more risk management techniques, such as artificial intelligence and deep learning algorithms, to enhance the model's adaptability in dynamic and complex environments. Furthermore, applying this framework to businesses of different scales and industries for extensive empirical verification will provide a more solid foundation for the promotion and practice of this research. As technology develops and more data accumulates, this framework is expected to be widely applied across more industries and sectors, driving the intelligent and refined development of supply chain management.

References

- [1] Gottimukkala, S. R., & Bhuram, S. K. (2025). Prescriptive Analytics for Next-Gen Supply Chains: Integrating Causal AI with Digital Twin Technologies. *IEEE Transactions on Vision and Artificial Intelligence*. [Link]
- [2] Ivanov, D., & Dolgui, A. (2021). A digital supply chain twin for managing disruption risks and resilience in the era of Industry 4.0. *International Journal of Production Research*, 59(12), 3791–3803.
- [3] Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management*, 37(1), 10–36.
- [4] Prajogo, D., Oke, A., & Olhager, J. (2018). Supply chain processes: Linking supply logistics integration, supply performance, lean processes, and firm performance. *International Journal of Operations & Production Management*, 38(9), 1887–1907.
- [5] Chong, A. Y. L., Lo, C. K. Y., & Weng, X. (2019). The business value of IT investments on supply chain: A contingency perspective. *Journal of Business Research*, 99, 326–339.
- [6] Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2020). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 227, 107617.
- [7] Dubey, R., Gunasekaran, A., Childe, S. J., Roubaud, D., & Fosso Wamba, S. (2021). Empirical investigation of data analytics capability and organizational flexibility as complements to supply chain resilience. *International Journal of Production Research*, 59(1), 110–128.
- [8] Riahi, Y., Khouja, M., & Sivarajah, U. (2022). Artificial intelligence applications in supply chain: A systematic review and future research directions. *Annals of Operations Research*, 319, 877–934.
- [9] Ivanov, D. (2023). Supply chain viability and the COVID-19 pandemic: A conceptual and formal generalisation of four major adaptation strategies. *International Journal of Production Research*, 61(5), 1680–1697.