

# Research on cost control and resource optimisation in supply chain link based on combination prediction model

Zhicheng Ma<sup>1,✉</sup>

<sup>1</sup> Songyuan Yongsheng Construction Company, Songyuan, Jilin, 138000, China

## ABSTRACT

The supply chain applies large number according to the technology, can reduce the cost of each link, optimize the resource allocation, increase the enterprise benefit. In this paper, the supply chain cost control program based on large number according to the previous forecast, the control of the event and the analysis of the three levels of the analysis of the supply chain. The combination time series model and the multivariate regression model, the joint CPFR concept, the establishment of the CPFR sales combination demand prediction model, the design form according to the sales prediction system, the resource optimization plan of the supply chain inventory in real time. Analyze the prediction effect of the combination prediction model, predict the product sales in the week, calculate the product safety inventory and the remaining inventory. The analysis is based on the cost control effect of the enterprise supply chain according to the sales forecast. The cost of purchasing the supply chain costs less than the operating income, which fell to 0.5107 in 2023. The gross margin of gross profit was 0.53666 in 2023, which was controlled by the gross margin, and the gross profit was improved. It is said that the enterprise is using large number according to the technology to the supply chain resource optimization in the supply chain cost control, the cost control effect is better.

*Keywords:* time series model, multiple regression model, supply chain cost, big data analysis

## 1. Introduction

Enterprises face enormous competitive pressure in the digital era, and in order to adapt to this change, various industries have carried out digital transformation. The supply chain, as the main component of enterprise operation, utilizes advanced technologies such as the Internet of Things, big data analysis, artificial intelligence, etc., in the digital transformation process to monitor, analyze and

✉ Corresponding author.

*E-mail address:* [mzc196304@163.com](mailto:mzc196304@163.com) (Z. Ma).

Received 18 June 2024; Revised 10 August 2024; Accepted 16 December 2024; Published Online 30 March 2025.

DOI: [10.61091/jcmcc125-27](https://doi.org/10.61091/jcmcc125-27)

© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

apply the data of each link in real time, which realizes the intelligent management and optimization of the supply chain [13, 17].

For the majority of manufacturing industries, it is essential for them to increase the importance of supply chain cost control [12]. Supply chain cost management includes the costs that need to be invested in material procurement, the costs that need to be invested in the process of product marketing and promotion, and the costs that need to be invested in the process of production and manufacturing [3, 11]. The existence of supply chain cost management, to a large extent, can promote the extension of internal cost control work, and ultimately develop into a management behavior across multiple departments in the enterprise, which can have a direct impact on the enterprise's transaction and cooperation object selection [20, 23, 8, 15]. In addition, digital transformation of an enterprise requires it to adjust the way it allocates resources according to the digital technology paradigm [14]. Digitally transformed enterprises are characterized by flexibility and ease of change, and can adapt quickly to reconfigure resources and develop new capabilities [5, 4]. In particular, they establish strong new connections with other stakeholders through digital platforms or networks, expand into new areas, identify new market opportunities and create value through strategic approaches [9, 10, 21].

Tiwari et al. [17] showed that big data analytics is one of the best techniques for supply chain professionals to deal with the massive amount of data generated in end-to-end supply chain management practices by investigating and analyzing the data to provide valuable insights to the industry. Gunasekaran et al. [7] investigated the impact of big data predictive analytics (BDPA) on supply chain (SCP) and organizational performance (OP) and clarified the mediating effect of connectivity and information contributing resources in executive commitment based on a resource-based perspective. Gopal et al. [6] analyzed the impact of big data analytics capabilities on the development of retail supply chains, and found that big data practices can provide retailers who are caught in a dilemma between customer loyalty and cost with importance and dominance assessment based on the optimal decision-making. Zhang et al. [22] proposed a dynamic scheduling method for logistics supply chain based on adaptive ant colony algorithm, constructed a dynamic scheduling model of the supply chain based on the constraints, and introduced the adaptive ant colony algorithm to solve the optimal dynamic scheduling scheme for logistics supply chain. The cost control method of mobile e-commerce under the information of big data, and utilized the intelligent and digital means to analyze the internal and external problems of the supply chain and establish an enterprise supply chain cost control model based on this. Xue et al. [19] used cloud genetic algorithm as the basis and means of quality management and cost control in the long-term operation of logistics service supply chain, and proposed suitable supply chain construction method and cost optimization method, aiming to construct and operate logistics service supply chain in the form of high quality with lower cost.

This paper analyzes the necessity of using big data to optimize supply chain cost control, takes big data technology as the technical support for enterprise cost management, and designs the supply chain cost control procedure based on big data. The intelligent sales forecast is constructed, the CPFR theory is introduced, and the CPFR sales combination forecast model is formed by combining the time series forecast model and the multiple regression model. For the inventory optimization based on supply chain coordination, safety stock design is added to form the supply chain inventory optimization program for product demand forecast. The time series forecasting model and multiple regression forecasting model are validated separately, and the forecasting results of the two are synthesized to obtain the combined forecasting effect. Based on the sales forecast, inventory surplus calculation is carried out and flexible ordering is implemented. The optimization effect of the inven-

tory optimization scheme based on big data sales forecast on supply chain cost control is analyzed in terms of procurement cost and operation cost.

## 2. Measures to control supply chain costs using big data technologies

### 2.1. *The need for optimal control of big data*

Supply chain in enterprise development occupies an increasingly important position, is the enterprise “the third source of profit”, supply chain is not only the point of enterprise cost generation but also the focus of cost reduction [2, 1].

Supply chain cost control not only includes the internal cost of the enterprise, but also includes the cost between the supply chain and other enterprises. The ultimate goal of supply chain cost control is the optimization and reduction of the total cost of the whole supply chain. The application of big data technology in each link of the supply chain can reduce the cost of each link and increase the efficiency of the enterprise, thus improving the comprehensive competitiveness of the company.

For the supply chain to play its due role, it is not only necessary to see whether the whole supply chain can reduce costs for the company, but also to see whether the company’s supply chain is reasonable, i.e., whether the operation efficiency is efficient.

The company’s supply chain management covers all aspects of R&D, procurement, production, sales and logistics. The complex and changing internal and external operating environment requires the company to make operational decisions quickly. The use of big data technology can help the company accurately locate the balance point of various operational decisions from the complicated influencing factors, so as to reach the long-term, dynamic global optimal wisdom decision. The company can utilize big data technology to improve the existing supplier database, so as to facilitate the rapid screening of suitable suppliers to shorten the procurement time, thus improving the procurement efficiency. At the same time in the production can also use big data technology to predict the amount of inventory, etc. to improve production efficiency. It can be seen that the application of big data technology to all aspects of the company’s supply chain, in order to achieve intelligent prediction, intelligent demand, intelligent procurement, etc., to improve the efficiency of the supply chain operation.

### 2.2. *Supply chain cost control procedures*

With the development of information technology, big data technology provides strong support for enterprise cost management. Based on the reasonable collection of data information generated by various aspects of the business of the enterprise, the cost data information is quantified through professional information processing technology. The company needs to improve the process of big data supply chain cost control, and implement relevant improvement measures according to the standards, so that the supply chain cost control program can be effective. The supply chain cost control program based on big data is shown in Figure 1.

Prior prediction refers to the enterprise before incurring costs, based on the historical cost data of each link of the enterprise supply chain, using big data prediction technology to comprehensively predict and analyze the costs that may be incurred by each link of the enterprise supply chain. From there, it determines the standard cost of each link of the supply chain of the enterprise under normal operation and establishes a standard cost database. As a measure of enterprise supply chain cost overspending or saving standards, so as to realize cost control beforehand.

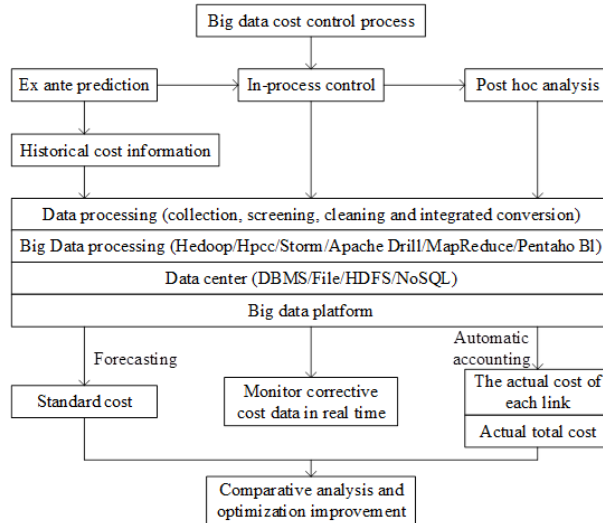


Fig. 1. Based on large number of supply chain cost control procedures

Ex-ante control refers to the enterprise with the help of big data platform, real-time supervision of the cost of each link of the enterprise supply chain and the degree of impact on the total cost of the supply chain, timely detection of problems, real-time supervision and correction of errors. Timely adjustment of solutions according to changes in specific problems. The big data analysis platform shortens the cycle from data extraction to offline analysis and then to report production, eliminating the need to repeat the number of withdrawals, and the marginal cost tends to be close to zero, which significantly reduces the time and labor costs.

Post-facto analysis refers to the fact that after generating costs, the enterprise automatically accounts for the actual costs of each link in the supply chain for a certain period of time with the help of the big data platform, compares the standard costs in the database established in the pre-prediction, analyzes the causes of the discrepancies and pursues the responsibility, reports to the corresponding management and decision-making levels in a timely manner, formulates the solutions, and sums up the experience to optimize and improve the existing measures.

2.3. Sales inventory optimization based on supply chain coordination

2.3.1. Construction of a sales forecasting system. Sales forecasting, as the first part of supply chain management, should be a key concern for enterprises. Reasonable and accurate sales forecasts can help enterprises adjust their short-term development plans and marketing strategies, and rationally deploy their daily management.

Generally speaking, sales forecasting should not only take into account the external market demand, economic environment and other factors, but also consider the enterprise’s own market positioning, product price prediction and promotion strategy and other factors. Therefore, building a sales forecasting system helps to do a good job of strategic planning, adjusting its sales quota, rationally managing cash flow, and then improving the efficiency of resource allocation.

2.3.2. Sales forecasting using big data. The flow of the intelligent sales forecasting system is shown in Figure 2. Using visual analysis tools, the order analysis situation, inventory data and internal and external factors of offline stores are summarized and analyzed. It also extracts the characteristics of each commodity type and constructs the corresponding sales prediction model. The system finally

derives various scenarios of total sales forecast of commodities, single-point sales forecast of different regions, and single product forecast of different types of the same commodity. The above forecasts are then transmitted to the purchasing system for coordination and communication between the purchasing department and suppliers.

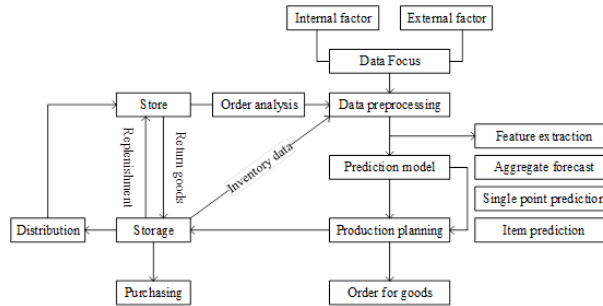


Fig. 2. Intelligent sales forecasting system process

By utilizing big data technology, the Intelligent Sales Forecasting System can build an endless system of indicators based on different merchandise types and scenarios, thus helping management reduce the time it takes to make important business decisions. In addition, from marketers to store managers to purchasing specialists, the forecasting system can provide insightful analytics to members in all segments of the supply chain as a way to help people in all departments make proactive, quick decisions.

2.3.3. CPFR sales mix forecasting model construction. CPFR is collaborative supply chain inventory management, i.e., collaborative planning, forecasting and replenishment, which is a collaborative supply chain inventory management technique and an important strategy of collaborative supply chain management.

CPFR not only requires relevant units upstream and downstream of the supply chain to implement joint forecasting and replenishment, but also the planning of each enterprise’s internal affairs is also jointly developed by each unit of the supply chain.

CPFR implementation path, the supply chain upstream and downstream enterprises to share data in real time, coordinating the supply chain related units of common planning, forecasting and replenishment of inventory, the ultimate goal is to reduce the information does not work to produce the bullwhip effect, reduce inventory, cost savings, reduce the seller’s inventory at the same time, but also to increase the supplier’s sales. Through the implementation of CPFR, it can realize the dynamic balance of the supply chain inventory, reach the maximization of the overall value benefit of the supply chain, and realize the overall cost reduction and efficiency of the relevant units in the supply chain, as well as to reduce costs and reduce expenses.

1) Time Series Analysis Method. Given a time series  $\{X_i, t \in T\}$ , when

$$\int_{-\infty}^{\infty} x dF_i(x) < \infty,$$

where  $F_i(x)$  is the probability distribution function of the random variable  $X_i$ , the mean of  $X_t$  exists and is denoted by:

$$\mu_t = \mathbb{E}[X_t] = \int_{-\infty}^{\infty} x dF_t(x). \tag{1}$$

When  $t$  spans all observed time points, a sequence of mean functions is obtained:

$$\{\mu_t, t \in T\}. \tag{2}$$

Assuming  $\int_{-\infty}^{\infty} x dF_i(x) < \infty$ , the variance function describes the degree of random fluctuation around the mean:

$$D[X_i] = \mathbb{E}[(X_i - \mu_i)^2] = \int_{-\infty}^{\infty} (x - \mu_i)^2 dF_i(x). \tag{3}$$

Over all observed time points  $t$ , this yields a sequence of variance (or covariance) functions  $\{D[X_i], t \in T\}$ .

For any  $t, s \in T$ , the self-covariance function of the sequence  $\{X_i\}$  is defined as:

$$\gamma(t, s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)]. \tag{4}$$

The autocorrelation coefficient of the time series  $\{X_t, t \in T\}$ , denoted by  $\rho(t, s)$ , or ACF, is given by:

$$\rho(t, s) = \frac{\gamma(t, s)}{\sqrt{D[X_t] \cdot D[X_s]}}. \tag{5}$$

This coefficient measures the degree of dependence between different time periods, quantifying how past values influence the present.

A model with the following structure is called an autoregressive moving average model, abbreviated as ARMA( $p, q$ ):

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \\ \phi_p \neq 0, \theta_q \neq 0, \\ \mathbb{E}[\varepsilon_t] = 0, \text{ Var}[\varepsilon_t] = \sigma_t^2, \mathbb{E}[\varepsilon_t \varepsilon_s] = 0, s \neq t, \\ \mathbb{E}[x_s \varepsilon_t] = 0, \forall s < t. \end{cases} \tag{6}$$

If  $\phi_0 = 0$ , the model becomes the centralized ARMA model:

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}. \tag{7}$$

Although less interpretable, differencing is a simple and effective technique for smoothing a time series. It extracts deterministic trends, and most non-stationary sequences become stationary after applying an appropriate order of differencing. Such series can be modeled using the ARIMA model [18, 16].

An ARIMA model with differencing order  $d$  is defined as:

$$\begin{cases} \phi(B) \nabla^d x_t = \theta(B) \varepsilon_t, \\ \mathbb{E}[\varepsilon_t] = 0, \text{ Var}[\varepsilon_t] = \sigma_t^2, \mathbb{E}[\varepsilon_t \varepsilon_s] = 0, s \neq t, \\ \mathbb{E}[x_t \varepsilon_t] = 0, \forall s < t, \end{cases} \tag{8}$$

where  $\nabla^d = (1 - B)^d$  is the differencing operator,  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  is the autoregressive polynomial, and  $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  is the moving average polynomial. This can be simplified as:

$$\nabla^d x_t = \frac{\theta(B)}{\phi(B)} \varepsilon_t, \tag{9}$$

where  $\{x_t\}$  is a zero-mean white noise process.

From the definition of the ARIMA model, it is evident that its essence lies in applying differencing to achieve stationarity and then fitting an ARMA model to the differenced series.

The  $d$ -th order differenced series can be written as:

$$\nabla^d x_t = \sum_{i=1}^d (-1)^d \binom{d}{i} x_{t-i}. \tag{10}$$

2) Multiple Linear Regression. Linear regression is a statistical method used to model the relationship between a dependent variable  $y$  and one or more independent variables  $x$ . When multiple independent variables are involved, the approach is referred to as multiple linear regression. This method estimates how changes in the predictors affect the outcome variable, enabling the prediction of  $y$  values for new observations of  $x$ .

The general form of a multiple linear regression model is:

$$h(x) = h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n, \tag{11}$$

where  $\theta_0$  is the intercept,  $\theta_1, \theta_2, \dots, \theta_n$  are the regression coefficients, and  $x_1, x_2, \dots, x_n$  are the independent variables.

The coefficient  $\theta_n$  represents the expected change in the response variable  $y$  associated with a one-unit increase in predictor variable  $x_n$ , assuming all other variables are held constant. This interpretation is referred to as the marginal effect of  $x_n$  on  $y$ . However, it is important to recognize that such interpretations may be limited for certain types of variables—such as dummy variables or intercept terms—that either cannot be incremented or cannot be held fixed meaningfully.

In practice, marginal and unique effects may diverge. A variable may have a large marginal effect but a negligible unique effect if its contribution is already captured by other covariates in the model. Conversely, a variable may exhibit a significant unique effect while having a small marginal effect if other predictors explain a substantial portion of the variance in  $y$  but in a manner complementary to the information provided by  $x_i$ . Including additional predictors can thus refine the relationship between  $x_i$  and  $y$  by isolating the portion of variance that is specifically attributable to  $x_i$ .

This understanding is essential when interpreting regression results, especially in models with multiple correlated predictors, where multicollinearity may obscure or exaggerate the apparent impact of individual variables.

3) CPFRR Joint Forecasting Model Construction. In the field of economic forecasting—spanning econometrics, time series analysis, and related disciplines—an important objective is the development of realistic models that accurately describe real-world systems. One common approach is combined forecasting, generally expressed as:

$$f_{ct} = w_0 + \sum_{i=1}^n w_i f_{it} + w_{n+1} y_{t-1}, \tag{12}$$

where  $f_{it}$  are individual forecasts and  $w_i$  are the weights assigned to each.

Combined forecasting methods are theoretically grounded in both econometric modeling and combinatorial mathematics. Econometric modeling typically expresses economic relationships through systems of interrelated equations, while combinatorial methods focus on aggregating predictions under conditions such as smoothness or cointegration.

If the forecast object is a smooth series, the predicted values must also remain smooth to prevent error amplification over time. Assume two forecasting methods:

$$\begin{cases} f_{1t} = a_1 + e_{1t}, \\ f_{2t} = a_2 + e_{2t}, \end{cases} \tag{13}$$

where  $e_{1t}$  and  $e_{2t}$  are white noise processes with zero mean and variances  $\delta_1$  and  $\delta_2$ , respectively.

For non-smooth sequences, forecast combinations should include at least one component that is cointegrated with the target variable. Specifically, a cointegration relationship of rank  $r$  exists for vector sequence  $y_t$  if all components are  $I(1)$  and there exists a matrix  $C$  such that  $z_t = Cy_t$  is  $I(0)$ , and  $\text{rank}(C) = r$ .

To test cointegration, one typically uses the residuals from an OLS regression:

$$Y_t = \text{const} + \beta X_t + \varepsilon_t, \tag{14}$$

and verifies if  $\varepsilon_t$  is stationary via the ADF test:

$$\Delta u_t = \gamma_0 u_{t-1} + \sum_{i=1}^k \gamma_i u_{t-i} + \nu_t, \tag{15}$$

where  $\nu_t$  is white noise and  $k$  is the lag order. If the ADF test statistic exceeds the critical value at the 1%, 5%, or 10% significance level, cointegration is confirmed.

This paper proposes a CPFR-based joint forecasting process that integrates time series and regression models into a hybrid framework to improve forecast accuracy and robustness. The proposed model includes criteria for selecting individual forecasting methods and constructing combined models based on practical forecasting requirements of enterprises.

As an illustrative scenario, weekly sales are used as the unit of demand forecasting. A simple and widely used method—moving average—is chosen due to its practicality and cost-effectiveness. For a growing, autocorrelated sales trend, it is represented as:

$$f_u = \frac{1}{n}(y_{t-1} + y_{t-2} + \dots + y_{t-n+1}). \tag{16}$$

Several forms of combined forecasting models are considered:

–Simple average combination:

$$f_{ct}^{(1)} = \frac{1}{n} \sum_{i=1}^n f_{it}, \tag{17}$$

–First-order lagged average combination:

$$f_{ct}^{(2)} = \frac{1}{n+1} \sum_{i=1}^n (f_{it} + y_{t-1}), \tag{18}$$

–Error-weighted combination (without lag):

$$f_{ct}^{(4)} = \left( \sum_{i=1}^n \frac{f_{it}}{e_i^2} \right) / \left( \sum_{i=1}^n \frac{1}{e_i^2} \right), \tag{19}$$

–Error-weighted combination (with lag):

$$f_{ct}^{(5)} = \left( \sum_{i=1}^n \frac{f_{it}}{e_{it}^2} + \frac{y_{t-1}}{e_{n+1}^2} \right) / \left( \sum_{i=1}^{n+1} \frac{1}{e_{it}^2} \right), \tag{20}$$



where  $e_{it}$  represents the mean square prediction error of method  $f_{it}$ .

–Regression-based combination:

$$f_{ct}^{(6)} = \sum_{i=1}^n w_i f_{it}. \quad (21)$$

This final model assumes individual forecasts are unbiased and uncorrelated with past data (i.e., already incorporate  $y_{t-1}$ ). Only when  $w_0 = 0$  does  $f_{ct}$  remain unbiased. The integration of time series and regression-based forecasting provides a comprehensive approach to improving accuracy and reliability in CPFR joint forecasting models.

#### 2.3.4. Inventory optimization scheme based on demand forecasting.

1) Feasibility analysis of CPFR-based inventory management program. In recent years, CPFR theory has been developing rapidly at home and abroad, and has been gradually matured and implemented in many fields of enterprises. CPFR theory is based on the continuous improvement of supply chain collaborative management, and through the practice of forming a scientific theory of collaborative supply chain inventory management, there is a great improvement in the flexibility of the practice of using it, and it is more conducive to the combination of the flexible implementation of different business organizations, and adapt to the flexible application of different business management needs. With the arrival of artificial intelligence, industrial interconnection, 5G and big data era, the rapid development of information technology and system applications, big data and industrial chain Internet data to achieve interoperability and sharing, supply chain coordination to provide technical guarantee support.

2) Inventory optimization based on supply chain coordination. Product sales are affected by a variety of factors, variety differences, market, economic and policy factors may lead to fluctuations in the sales of parts, so inventory management must be based on the changes in sales to set indicators. Safety stock is a reserve stock set up to cope with sales peaks or other contingencies, and is usually used to compensate for demand generated by actual demand exceeding expectations during the order lead time. Timing of purchases means that the total amount of inventory is always being reduced, regardless of actual sales. When the reduction is within a reasonable range, periodic or rationed purchases are required based on demand. However, since purchasing takes time and usually requires logistics, it is too late to make purchases when the inventory level is close to or below the lower limit. Therefore, it is necessary to determine the upper and lower inventory limits as well as the inventory threshold. Inventory thresholds can be set based on past sales and future sales expectations, including upper and lower inventory limits. When the inventory quantity reaches the alert line, it needs to be adjusted according to the inventory situation. The upper inventory limit should be set to avoid being too large, otherwise it will increase the capital utilization and inventory holding cost. At the same time, it is important to pay attention to the fluctuation of market demand and make timely purchases in order to reduce the capital occupation and burden. Enterprises can make use of upstream and downstream information sharing in the supply chain to make purchases in a planned manner to minimize inventory costs.

### 3. Enterprise supply chain cost control and resource optimization analysis

#### 3.1. Empirical evidence of CPFR-based sales mix forecasting

In order to make the research process more feasible and the findings more representative, the secondary supply chain dominated by manufacturer enterprise A and retailer enterprise B is chosen as the research object. (Retailer B piloted the CPFR sales mix forecasting model based on the composition of big data technology.)

Monthly data on the retailer’s sales volume from January 2016 to December 2023 was compiled and summarized by reading a large number of relevant reports on the above companies and their annual and monthly reports. And manufacturers’ product price data from January 2016 through December 2023.

3.1.1. Time series model testing and forecasting. In Eviews by performing a first order difference on the raw data and performing a unit root test. The ADF test is the most commonly used of the six methods of unit root test. If the result of ADF test is not significant, i.e., the original hypothesis is accepted that the series does not have a unit root, it indicates that the time series is smooth.

The results of the order difference test for the monthly data of sales volume are shown in Table 1, from the test results shown in the table, it can be seen that Prob is 0 and the test is not significant. Therefore, the first order difference series of the sales time series is smooth.

**Table 1.** Monthly data order difference test results

		t-Statistic	Prob.*
ADF-Statistic	25.68546	-30.15688	0.0000
	1% level	-8.680711	0.0000
Test-critical values	5% level	-6.556294	0.0000
	10% level	-6.312247	0.0000

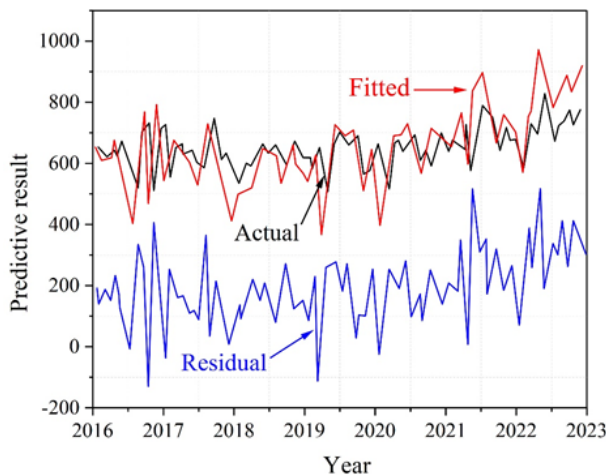
The estimation of the parameters and their significance tests by the least squares method in Eviews system gives specific results. The results of parameter estimation for the ARMA model are shown in Table 2. For the model MA(2) Model  $X_t = \alpha_i + \theta_1\alpha_{t-1} + \theta_2\alpha_{t-2}$ , the constant term is denoted by C. The constant C in the model is estimated to be 1052.3965. The parameter estimates of the two parameters  $\theta_1$  and  $\theta_2$  in the model are denoted by “MA1,1” and “MA1,2”, and the specific estimates are 0.784512 and 0.200453, respectively. The time series model can be expressed as the result retained to three decimal places,  $Y_t = 1052.397 + 0.785\alpha_{t-1} + 0.200\alpha_{t-2}$ .

**Table 2.** Parametric estimation of ARMA model

Variable	Coefficient	Std.Error	t-Statistic	Prob.	Lag
C	1052.3965	25.66349	15.62243	0.0000	0
MA1,1	0.784512	0.245113	7.86594	0.0000	1
MA1,2	0.200453	0.257994	3.66001	0.0044	2

The effect of fitting the time series model to the sample observations is shown in Figure 3. It can be seen that there is a better match between the predicted value and the real value, and the

fluctuation side and amplitude remain basically the same. The time series model predicted curve in 2018, 2019, 2020 appeared larger fluctuations, but keep the same fluctuation direction with the sample observations, the model fitting effect is better.



**Fig. 3.** The time series model is the fitting effect of the sample observation value

3.1.2. Testing and prediction of multiple regression models. Eviews statistical analysis software was applied to analyze the raw data by multiple regression. The regression ANOVA is shown in Table 3. The F of the regression model is 1654.079, and its probability of significance is much less than 0.01. Therefore, the multiple regression model has a good fit to the data, and the regression model has a good overall effect.

**Table 3.** Regression variance analysis

Model	Sum of Squares	df	Mean Square	F	p
Regression	785108.611	25	31404.34444	1654.079	0.0000
Residual	65392.159	205	18.986	/	/
Total	850500.77	230	/	/	/

**Table 4.** Regression coefficient and its inspection

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	893.5221	97.624001	9.152688794	0.0000
Price	-0.869550	0.168593	-5.157687448	0.0000
Advertising	8.960751	0.218996	40.91741858	0.0000
Substitution	-0.718993	0.187763	-3.829258161	0.0000
Complementarity	7.986614	0.973142	8.207038644	0.0000

The regression coefficients and their tests are shown in Table 4, the regression coefficients of product price, advertising input, price of substitutes and price of complementary products on the sales volume are -0.869550, 8.960751 and -0.718993 respectively, and the significance probability level of each value is much less than 0.01. This indicates that all the influencing factors in the data have a significant effect on the sales volume of this product, and that the product’s own price and

the price of substitutes are negatively correlated with sales, while advertising inputs and the price of complementary products are positively correlated with sales. This is basically in line with the economic and social reality, which shows that the model not only has a good fit and prediction accuracy, but also has a good practical application and interpretation ability.

3.1.3. Combined forecasting and evaluation of forecasting effects. After testing and analyzing the time series model and multiple regression model, the combination is synthesized and the results are calculated according to the principle of combination forecasting.

The fit of the combined model to the data is shown in Figure 4, which shows that the overall fit of the model to the data is better than that of the time series sub-model and the multiple regression sub-model, and it can be used for forecasting. The Residual value reaches a minimum in 2020 and is at [-100,0].

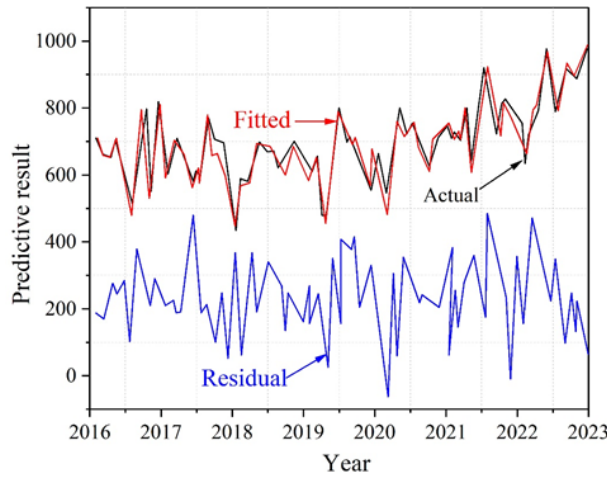


Fig. 4. The combination model log is based on the simulation

Comparing the prediction results of the combined model with the prediction results of the time series sub-model and the multiple regression sub-model, the prediction error of each prediction model can be obtained. Various prediction errors of the combined prediction model  $\varepsilon < \max(\varepsilon_1, \varepsilon_2)$ , where  $\varepsilon$  indicates the prediction error of the combined model,  $\varepsilon_1$  indicates the prediction error of the time series sub-model, and  $\varepsilon_2$  indicates the prediction error of the multiple regression sub-model.  $\max(\varepsilon_1, \varepsilon_2)$  denotes the prediction error of the whole system of supply chain. Since the time series sub-model and the multiple regression sub-model are relative to a single enterprise, the prediction error of the whole supply chain system should take the larger of the sub-model prediction errors. As the combined forecasting model is relative to the whole supply chain system, the forecasting error of the combined forecasting model is much smaller than the forecasting error of the whole supply chain system.

The prediction effect evaluation indexes of each prediction model are shown in Table 5, and the variance of the prediction residuals of the time series sub-model and the variance of the prediction residuals of the multivariate regression sub-model can be obtained by calculation. The variance of the combined model prediction residuals is 2123.175.

Obviously, for the whole supply chain system, the variance of its prediction residuals,  $\delta < \frac{1}{2}\max(\delta_1, \delta_2)$ . Based on the above analysis, it can be seen that the values of various error indicators of the sales combination prediction model based on are significantly lower than the maximum value of the error value of the single prediction model, thus indicating the superiority of the sales combination predic-

tion method based on the accuracy is higher, and the sales prediction accuracy can be effectively improved.

**Table 5.** Prediction model of prediction results

Prediction model	Mean absolute error (MAE)	Mean error (MSE)	Square sum error (SSE)	Mean absolute percentage error (MAPE)
Time sequence submodel	87.26112	74.51361	507263.18	12.54348
Multivariate regression submodel	35.26997	36.24456	56.28034.31	3.52644
Combination prediction model	67.21569	48.01279	205739.48	2.67986

*3.2. Application of sales forecasts to safety stock calculations*

Setting an inventory cordon considers whether the inventory needs to be adjusted or not, and the classic formula for safety stock is used below to make an accurate calculation of the inventory of pharmaceuticals.

The influencing factors affecting safety stock mainly include uncertainty of consumer demand, instability of production process, changes in distribution cycle, and service level. Its calculation requires forecast data and historical demand data, according to the classic safety stock formula there:

$$SS = z * \sqrt{e_d^2 + e_i^2 (d)^2}, \tag{22}$$

where *SS* is the safety stock.

The optimized safety stock equation:

$$SS = ze_T \bar{L}/T, \tag{23}$$

where  $\bar{L}$  is the average value of the lead time. Based on the predictions in this paper, it is useful to set a week as  $T = 7$  days, and the formula can be converted to the following equation for the case where the lead time *L* is unchanged:

$$SS = ze_K \bar{L}/7, \tag{24}$$

where,  $e_K$  represents the standard deviation of weekly demand. And the weekly demand approximates the sales forecast as shown in Table 6, displaying as the occasion sales of the product in each week with the forecast value. As an example, the product inventory balance obtained for a product is 104 in December.

**Table 6.** Sales and forecasts for each week of the product

Month	Period	First week	Second week	Third week	Fourth week	Total
January	Actual sales	89	105	114	134	442
	Predictive value	95	112	125	145	477
February	Actual sales	124	130	132	155	541
	Predictive value	150	166	165	201	682
March	Actual sales	475	550	580	624	2229
	Predictive value	412	470	512	597	1991
.....	.....	.....	.....	.....	.....	.....
December	Actual sales	852	856	891	1025	3624
	Predictive value	878	900	950	1000	3728

Calculate with the forecasted results. Next, this paper to act out the sales data of the first four weeks of each month to get the inventory balance of a product as an example. Assuming that the service level  $CSL = 99\%$  and lead time  $L = 1$ , the sales data of January in the table as an example, using the formula in this paper, to get the safety stock value of  $SS_{January}$  each week. According to the statistical time of the remaining inventory and the actual sales of get the inventory balance of the cycle as shown in Table 7. The inventory balances for the 1st, 2nd, 3rd, and 4th weeks of January are 155, 167, 180, and 201, respectively.

**Table 7.** January safety inventory calculation results

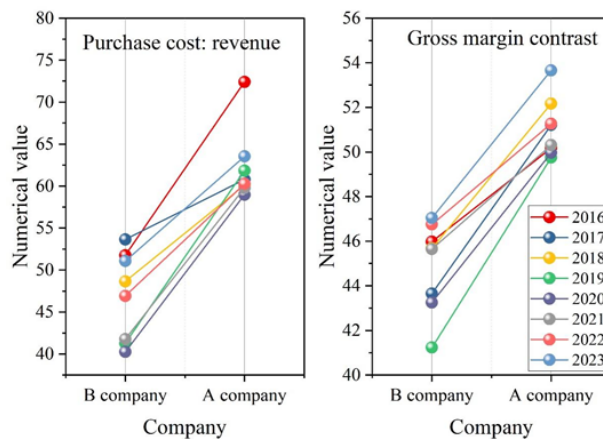
Period	First week	Second week	Third week	Fourth week
$SS_{January}$	112	120	125	142
$SS_{January}/Actual\ sales$	0.7946	0.875	0.912	0.9437
Inventory balance	155	167	180	201

Enterprises can be guided by safety stock, when holding a high level of safety stock, then the balance of inventory can be used to meet the needs of later sales without the need to restock. In practice, managers have to make the right decision between high inventory and out-of-stock in order to effectively control inventory.

3.3. Analysis of supply chain cost control effect

3.3.1. Procurement cost analysis. Retailer Company B, unlike the traditional manufacturing industry, does not have its own production and processing plant and can only purchase goods from manufacturers. Therefore, the procurement cost of purchasing products becomes a major component of Company B’s supply chain costs.

A comparison of the purchasing costs and gross margins of retailer Company B and manufacturer Company A is shown in Figure 5. Retailer Company B sells many categories of products and it is difficult to identify individual products for bookkeeping and accounting, so it is not possible to find the cost of the products purchased based on the categorization. Therefore, companies like Retailer B, Inc. generally use a combination of the cost of goods purchased plus a certain amount of gross profit to determine the selling price of the product, and do not add any other additional prices.



**Fig. 5.** Purchase cost and gross margin contrast

It can be noticed that when comparing Retailer B, Inc. and Manufacturer A from 2016-2023, Retailer B, Inc.’s share of purchasing costs has declined more significantly. It was 0.5175 in 2016 and

decreased to 0.5107 in 2023. Indicating that the utilization of big data technology has effectively controlled the supply chain cost purchasing cost of Retailer B Company. The gross profit ratio of Retailer B also shows an increasing trend, which is 0.4598 in 2016 and grows to 0.5366 in 2023, indicating that its cost can be controlled and gross profit can be increased. Comparing with Manufacturer A, Retailer B Company has lower purchasing cost than Manufacturer A in the process of supply chain cost control using big data technology and it is effective.

3.3.2. Analysis of operating costs. The composition of operating costs of retailer Company B from 2016-2023 is shown in Figure 6, which shows that order processing expenses have always accounted for a larger share of operating costs, followed by technology research and development expenses, then sales and promotion expenses, and finally administrative expenses.

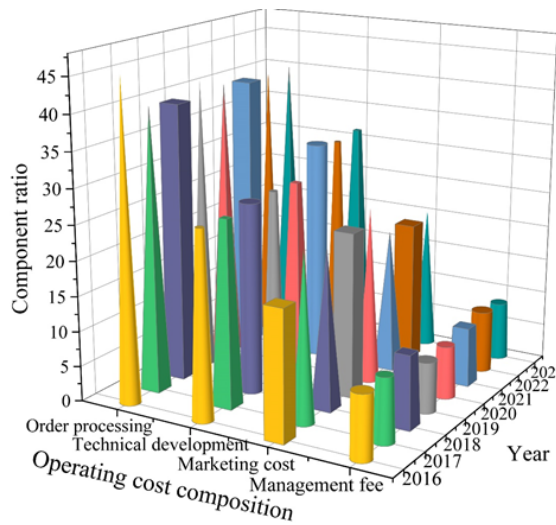


Fig. 6. Operating cost composition in 2016-2023

The amount of order processing costs in 2017-2023 shows a year-on-year decline, but its share of operating costs basically fluctuates up and down around 40%. Order processing cost is the main component of the operating cost of Retailer B, so it is necessary to control it by utilizing big data technology.

The proportion of marketing costs is basically showing an upward trend, and the amount is constantly rising, which has a lot to do with the expansion of Retailer B’s market share. The proportion of technology research and development expenses is maintained at 31% in both 2021 and 2023. After 2019, this proportion is basically rising and the amount is rising, which also proves that Retailer B is also increasing its investment in technology research and development.

The amount of administrative expenses shows a downward trend in 2019, although the amount has risen, the share of administrative expenses in operating costs is decreasing, which is inseparable from the relationship between Retailer B’s utilization of big data technology for supply chain cost control.

## 4. Conclusion

This paper designs a supply chain cost control program based on big data technology, unites CPFR theory, and constitutes an inventory optimization program for CPFR sales mix forecasting. It analyzes product demand, adjusts enterprise ordering, optimizes inventory resources, and realizes enter-

prise supply chain cost control.

The combined forecasting results are synthesized by combining the forecasting results of the time series model and the multiple regression model. For the whole supply chain system, the residual variance of the combined forecasting model is  $\delta < \frac{1}{2}max(\delta_1, \delta_2)$ . The estimation results of the CPFR-based sales combination forecasting model are more accurate and can effectively improve the sales forecasting accuracy.

The concept of safety stock is proposed, and the safety stock of the product cycle based on sales forecast is measured. In response to the inventory balance of the cycle, the enterprise carries out flexible ordering, adjusts the order quantity, and optimizes the supply chain link.

Comparing the effect of supply chain cost control between retailers and manufacturers, the gross profit margin percentage of the enterprise based on the CPFR sales mix forecasting model (i.e., retailer Company B) shows an increasing trend and grows to 0.5366 in 2023, which indicates that its cost can be controlled and its gross profit can be improved. Retailer Company B has reduced and effective purchasing costs in the process of supply chain cost control using big data technology.

## References

- [1] A. M. Abed. Accelerate demand forecasting by hybridizing catboost with the dingo optimization algorithm to support supply chain conceptual framework precisely. *Frontiers in Sustainability*, 5:1388771, 2024. <https://doi.org/10.3389/frsus.2024.1388771>.
- [2] N. Agarwal, S. Modgil, and S. Gupta. Esg and supply chain finance to manage risk among value chains. *Journal of Cleaner Production*:143373, 2024. <https://doi.org/10.1016/j.jclepro.2024.143373>.
- [3] X. Brusset, A. Jebali, D. La Torre, and D. Liuzzi. Production optimization in the time of pandemic: an sis-based optimal control model with protection effort and cost minimization. *Annals of Operations Research*:1–24, 2023. <https://doi.org/10.1007/s10479-023-05206-8>.
- [4] R. Cui, M. Zhang, S. Zhang, and S. Zhang. Cost optimization strategy and robust control strategy for dynamic supply chain system. *Annals of Operations Research*:1–15, 2023. <https://doi.org/10.1007/s10479-021-04385-6>.
- [5] H. Gholizadeh, H. Fazlollahabbar, and M. Khalilzadeh. A robust fuzzy stochastic programming for sustainable procurement and logistics under hybrid uncertainty using big data. *Journal of Cleaner Production*, 258:120640, 2020. <https://doi.org/10.1016/j.jclepro.2020.120640>.
- [6] P. Gopal, N. P. Rana, T. V. Krishna, and M. Ramkumar. Impact of big data analytics on supply chain performance: an analysis of influencing factors. *Annals of Operations Research*, May 2022. [10.1007/s10479-022-04749-6](https://doi.org/10.1007/s10479-022-04749-6).
- [7] A. Gunasekaran, T. Papadopoulos, R. Dubey, S. F. Wamba, S. J. Childe, B. Hazen, and S. Akter. Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, 70:308–317, 2017. <https://doi.org/10.1016/j.jbusres.2016.08.004>.
- [8] M. S. Habib, M. Omair, M. B. Ramzan, T. N. Chaudhary, M. Farooq, and B. Sarkar. A robust possibilistic flexible programming approach toward a resilient and cost-efficient biodiesel supply chain network. *Journal of Cleaner Production*, 366:132752, 2022. <https://doi.org/10.1016/j.jclepro.2022.132752>.
- [9] S. Jabbar, H. Lloyd, M. Hammoudeh, B. Adebisi, and U. Raza. Blockchain-enabled supply chain: analysis, challenges, and future directions. *Multimedia Systems*, 27:787–806, 2021. <https://doi.org/10.1007/s00530-020-00687-0>.



- 
- [10] Y. Kittichotsatsawat, V. Jangkrajarn, and K. Y. Tippayawong. Enhancing coffee supply chain towards sustainable growth with big data and modern agricultural technologies. *Sustainability*, 13(8):4593, 2021. <https://doi.org/10.3390/su13084593>.
- [11] Z. Liu, C. Zhou, H. Chen, and R. Zhao. Impact of cost uncertainty on supply chain competition under different confidence levels. *International Transactions in Operational Research*, 28(3):1465–1504, 2021. <https://doi.org/10.1111/itor.12596>.
- [12] J. Mageto. Big data analytics in sustainable supply chain management: a focus on manufacturing supply chains. *Sustainability*, 13(13):7101, 2021. <https://doi.org/10.3390/su13137101>.
- [13] T. Nguyen, Z. Li, V. Spiegler, P. Ieromonachou, and Y. Lin. Big data analytics in supply chain management: a state-of-the-art literature review. *Computers & Operations Research*, 98:254–264, 2018. <https://doi.org/10.1016/j.cor.2017.07.004>.
- [14] I. A. Omar, M. Debe, R. Jayaraman, K. Salah, M. Omar, and J. Arshad. Blockchain-based supply chain traceability for covid-19 personal protective equipment. *Computers & Industrial Engineering*, 167:107995, 2022. <https://doi.org/10.1016/j.cie.2022.107995>.
- [15] Y. Peng, B. Chen, and E. Veglianti. Platform service supply chain network equilibrium model with data empowerment. *Sustainability*, 14(9):5419, 2022. <https://doi.org/10.3390/su14095419>.
- [16] L. Rubio, A. Palacio Pinedo, A. Mejía Castaño, and F. Ramos. Forecasting volatility by using wavelet transform, arima and garch models. *Eurasian Economic Review*, 13(3):803–830, 2023. <https://doi.org/10.1007/s40822-023-00243-x>.
- [17] S. Tiwari, H.-M. Wee, and Y. Daryanto. Big data analytics in supply chain management between 2010 and 2016: insights to industries. *Computers & Industrial Engineering*, 115:319–330, 2018. <https://doi.org/10.1016/j.cie.2017.11.017>.
- [18] P. William O’Brien PhD and P. Scott Bucking PhD. From static to dynamic: improving scope 2 emission factor predictions with arima data-driven models. *ASHRAE Transactions*, 130:182–191, 2024.
- [19] Y. Xue and L. Ge. Cost optimization control of logistics service supply chain based on cloud genetic algorithm. *Wireless Personal Communications*, 102:3171–3186, 2018. <https://doi.org/10.1007/s11277-018-5335-z>.
- [20] V. S. Yadav, A. R. Singh, R. D. Raut, S. K. Mangla, S. Luthra, and A. Kumar. Exploring the application of industry 4.0 technologies in the agricultural food supply chain: a systematic literature review. *Computers & Industrial Engineering*, 169:108304, 2022. <https://doi.org/10.1016/j.cie.2022.108304>.
- [21] S. M. Zahraee, N. Shiwakoti, and P. Stasinopoulos. Biomass supply chain environmental and socio-economic analysis: 40-years comprehensive review of methods, decision issues, sustainability challenges, and the way forward. *Biomass and Bioenergy*, 142:105777, 2020. <https://doi.org/10.1016/j.biombioe.2020.105777>.
- [22] Y. Zhang and L. Wang. A dynamic scheduling method for logistics supply chain based on adaptive ant colony algorithm. *International Journal of Computational Intelligence Systems*, 17(1):198, 2024. <https://doi.org/10.1007/s44196-024-00606-5>.
- [23] C. Zhao, D. Wang, A. Younas, and B. Zhang. Coordination of closed-loop supply chain considering loss-aversion and remanufactured products quality control. *Annals of Operations Research*:1–24, 2022. <https://doi.org/10.1007/s10479-022-04619-1>.