ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

Research on Cold Chain Logistics Demand under the Context of New Retail

Haodong Yu*

Department of Statistics, Guangxi University, Guilin, 541006, China
**Corresponding author: 764167438@qq.com*

Abstract: *This article uses the relevant indicators of cold chain logistics and socio-economic development indicators in Zhejiang Province from 2000 to 2021. First, it screens the indicators through correlation analysis and multiple linear regression, and then solves multicollinearity through principal component analysis to obtain the prediction equation for the demand of cold chain logistics. Finally, a grey prediction model is established to predict the development trend of the indicators for the next four years, thereby obtaining the dynamic forecast values of cold chain logistics demand under the new retail background. Lastly, based on the development trend of cold chain logistics nationwide over the next four years, this study provides feasible suggestions for the development of China's cold chain logistics industry.*

Keywords: *New Retail; Cold Chain Logistics; Principal Component Analysis; Multiple Linear Regression; Grey Prediction*

1. Introduction

New Retail, proposed by Jack Ma in 2016, is a new retail model that deeply integrates online and offline service experiences with modern logistics. Cold chain logistics is a form of low-temperature logistics mainly targeting perishable food.

Currently, research on cold chain logistics mainly falls into the following three categories:

1) Analyzing the importance of its development. Wang Yuxia (2016) analyzed the development characteristics, necessity, and current status of cold chain logistics and drew conclusions [1]. Zhu Kunping (2016) proposed specific suggestions from the perspectives of modernizing agricultural cold chain logistics hardware facilities, standardizing management, intensifying resources, humanizing policies, and professionalizing talents [2].

2) Proposing improvements and innovations based on practicality. Du Jianping et al. (2020) conducted research on the cooling performance of portable boxes with thermal energy storage suitable for cold chain applications [3]. Zhao Jiao Jiao (2020) et al. conducted research on reusable low-temperature phase change microcapsules in short-term cold chain transportation [4].

3) Predicting cold chain logistics demand based on mathematical models. Liang Yan et al. (2018) conducted a predictive analysis of the demand for agricultural cold chain logistics in Tianjin based on multiple linear regression [5]. Zeng Hao (2022) predicted China's cold chain logistics demand based on the grey prediction model [6]. Xu Wen (2023) predicted China's cold chain logistics demand based on the grey prediction model and BP neural network [7].

This article belongs to the third category, which predicts cold chain logistics demand based on mathematical models. By reviewing the literature, we found that other scholars' predictions of the demand for fresh agricultural cold chain logistics mainly include: time series prediction methods, such as exponential smoothing, growth rate analysis, and ARIMA methods; BP neural network prediction methods; grey prediction methods, and regression analysis methods.

In this study, we used Spearman correlation analysis and multiple linear regression to select six suitable indicators and then established a multiple linear regression model. We discovered that multicollinearity exists in the multiple linear equations, so we decided to use principal component analysis to reduce dimensions. We verified the feasibility of principal component analysis through KMO and Bartlett tests, and then selected the principal components from the variance explanation table to obtain the prediction equation for cold chain logistics demand. The accuracy of the prediction equation

ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

was tested using residual analysis. We then conducted grey predictions for each individual indicator, and the fitting effect was found to be satisfactory. The model was tested using residual analysis and level ratio deviation tests, and good results were obtained. We summarized all the indicators in the grey prediction table, and then combined the prediction results with the multiple linear regression after principal component analysis for dynamic prediction, obtaining the forecast results of cold chain logistics demand for 2023-2026. Finally, based on the development trends, we offer some feasible suggestions for the government and enterprises

2. The basic fundamental of Model

2.1 Principle of Multiple Linear Regression Model

Multiple Linear Regression is a statistical method used to study the linear relationship between multiple independent variables (also known as explanatory variables or features) and a dependent variable (also known as the response variable). The multiple linear regression model assumes that the relationship between the dependent variable and the independent variables can be represented by a linear equation, and there is no strict collinearity among the independent variables. The multiple linear regression model can be expressed as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ip} + \varepsilon_i, i = 1,2,\cdots,n$$

In matrix form:

$$y = X\beta + \varepsilon, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

If there is multicollinearity in the multiple linear regression model, principal component analysis can be used for dimensionality reduction.

2.2 Principle of Principal Component Analysis Model

Principal Component Analysis (PCA) is a statistical method used to reduce the dimensionality of high-dimensional original data through linear transformation. The goal of PCA is to retain the main information in the original data while reducing the data dimension for easier analysis, visualization, and storage.

The Principal Component Analysis method transforms p original variables x_i (after standardization) that have correlation into another set of uncorrelated variables F_i through coordinate transformation. Mathematical model of principal component analysis as:

$$\begin{cases} F_i = \mu_{11}x_1 + \mu_{12}x_2 + \ldots + \mu_{1p}x_p \\ F_2 = \mu_{21}x_1 + \mu_{22}x_2 + \ldots + \mu_{1p}x_p \\ \vdots \\ F_p = \mu_{p1}x_1 + \mu_{p2}x_2 + \ldots + \mu_{pp}x_p \end{cases}, \quad {\mu_{i1}}^2 + {\mu_{i2}}^2 + \ldots + {\mu_{ip}}^2 = 1 (i = 1,2,3\ldots,p)$$

2.3 Principle of the $GM(1,1)$ Model

The Grey Model is a prediction and analysis method based on grey system theory, mainly used for dealing with systems with uncertainty and incomplete information. The core idea of the grey model is to predict the future by constructing a simple differential equation model.

The most commonly used model in grey models is the $GM(1,1)$ model. $GM(1,1)$ is a first-order, single-variable grey model suitable for single-variable, single-factor time series data. The basic form of the $GM(1,1)$ model is:

$$x^{(0)}(k) + \hat{a}z^{(1)}(k) = \hat{b}, k = 2,3,\ldots,n,$$

where $b$represents the grey action quantity, and$-a$represents the development coefficient.

The advantage of the grey model is its low data requirements, making it capable of making predictions even with limited and incomplete information. Considering that the concept of new retail has only been

ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

proposed for 6 years, the related data is scarce and imperfect, making the grey model a suitable choice for prediction. However, the applicability of the grey model is limited, as it struggles with multivariable and multifactor complex systems. Therefore, in this paper, the grey model is used in conjunction with the multiple linear regression model to improve accuracy.

3. Establishing Prediction Model

3.1 Determine the prediction indicators

The data involved in this paper are all from the statistical yearbook of Zhejiang Province from 2000 to 2021. The selected indicators include: total provincial GDP, the primary, secondary, and tertiary industry GDPs, total consumer price index, retail price index, total freight volume, cargo turnover, total retail sales of consumer goods, total import and export value, and cold chain logistics demand. Among them, cold chain logistics demand = vegetable and edible fungus production + aquatic product output + meat output + fruit output.

Through the Spearman correlation analysis, it was found that the total provincial GDP and the primary, secondary, and tertiary industry GDPs are completely positively correlated. Since the total GDP is composed of the primary, secondary, and tertiary industry GDPs, the total provincial GDP will be removed in the following multiple linear regression model. The results are as follows Table 1:

Table 1: Summary of multiple linear regression model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.67958	0.97611	-0.696	0.49956
GDP of primary industry (100 million yuan)	0.10299	0.25001	0.412	0.68764
GDP of secondary industry (100 million yuan)	0.07141	0.11614	0.615	0.55013
GDP of the tertiary industry (100 million yuan)	-0.09012	0.1831	-0.492	0.63147
Total consumption index of residents	-0.55354	0.47161	-1.174	0.26328
Retail price index	0.78959	0.34992	2.256	0.04349
Total freight volume (10000 tons)	0.11009	0.09789	1.125	0.28272
Total cargo turnover (100 million ton kilometers)	0.1434	0.06292	2.279	0.04175
Total retail sales of consumer goods	0.43025	0.10819	3.977	0.00184
Total value of import and export (10000 yuan)	0.26139	0.01131	23.12	2.55E-11

Since the p-values of the production values of the first, second, and third industries are too large, all greater than 0.05, indicating that they are not significant at the 95% confidence level, and they also contain the subsequent variables in reality, we do not choose the production values of the first, second, and third industries as indicators. Apparently, this is due to multicollinearity among the indicators, and because the first variable contains the other three variables. Moreover, considering the practical significance, the first four indicators all include the later indicators to some extent, and the first variable can also be linearly represented by all other variables. In order to ensure the rationality of the principal component regression, we choose to abandon the first four indicators.

So, we set the demand for cold chain logistics as Y, and the six indicators are: the total consumption index of residents as x_1, the retail price index of goods as x_2, the total freight volume as x_3, the total cargo turnover as x_4, the total retail sales of social consumer goods as x_5, and the total import and export value as x_6.

3.2 Establishing a Principal Component Regression Model

Based on the determined indicators, we initially establish a multiple linear regression model for Zhejiang Province's cold chain logistics demand as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

According to the calculation results of the statistical software R, the adjusted R-squared of the model is 0.9054, indicating that the goodness of fit is very good. The model's p-value is 5.989e-08, which indicates that the model is significant at the 95% confidence level. However, there are still a few coefficients with p-values greater than 0.05, indicating that they are not significant at the 95% confidence level. This is clearly due to the multicollinearity among the various indicators. In this case, we use principal component analysis (PCA) to shrink the variables and eliminate multicollinearity.

ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

To determine whether principal component analysis can be performed, we first conduct the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test , as shown in Table 2 and Table 3:

Table 2: Summary of multiple linear regression model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.46E+02	6.64E+02	1.275	0.2219
x_1	7.71E+00	2.66E+00	2.901	0.011
x_2	-8.31E+00	2.44E+00	-3.407	0.0039
x_3	-1.60E-03	8.68E-04	-1.849	0.0843
x_4	4.99E-02	2.61E-02	1.916	0.0747
x_5	-9.93E-03	1.16E-02	-0.857	0.4049
x_6	-3.69E-03	3.47E-03	-1.065	0.3037

Table 3: KMO test and Bartlett test

KMO test and Bartlett test		
KMO value		0.842
Bartlett sphericity test	Approximate chi square	348.898
	df	15
	P	0.000***
Note: * * *, * * and * represent the significance levels of 1%, 5% and 10% respectively		

The KMO test result shows that the KMO value is 0.842. Meanwhile, the result of the Bartlett's sphericity test shows that the significance P-value is 0.000***, indicating a significant level, rejecting the null hypothesis. The variables are correlated, and the principal component analysis is valid, with a suitable degree.

Next, we determine the number of principal components by analyzing the variance explanation table. The variance explanation table mainly looks at the contribution rate of the principal components to the explanation of variables. If it's too low (e.g., below 60%), the principal component data needs to be adjusted. The evaluation criterion is that the cumulative variance contribution rate of the principal components should generally reach over 90% before it is acceptable; otherwise, the factor data needs to be adjusted, as shown in Table 4.

Table 4: Variance Interpretation Table

component			
	Characteristic root	Variance interpretation rate (%)	Cumulative variance interpretation rate (%)
x_1	5.705	95.079	95.079
x_2	0.241	4.012	99.091
x_3	0.039	0.657	99.748
x_4	0.007	0.122	99.87
x_5	0.006	0.097	99.967
x_6	0.002	0.033	100

The results show that the variance contribution rate of the first two principal components is about 99%, which means they contain about 99.1% of the information of the six variables. Therefore, we choose the first two principal components. In addition, according to the component matrix table, we save the linear expression coefficients of the first two principal components with respect to x_1-x_6:

Table 5: Composition matrix

name		
	Component 1	Component 2
x_1	0.174	-0.345
x_2	0.172	-0.69
x_3	0.173	0.091
x_4	0.172	-0.801
x_5	0.174	0.168
x_6	0.16	1.697

$$F1 = 0.174x_1 + 0.172x_2 + 0.173x_3 + 0.172x_4 + 0.174x_5 + 0.16x_6$$

$$F2 = -0.345x_1 - 0.69x_2 + 0.091x_3 - 0.801x_4 + 0.168x_5 + 1.697x_6$$

ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

$$F = (0.951/0.991) \times F1 + (0.04/0.991) \times F2$$

Using the two principal components for fitting the multiple linear regression model:

Table 6: Summary of regression analysis model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	898.3717	52.05756	17.257	4.57e-13 ***
F1	0.015837	0.002501	6.332	4.47e-06 ***
F2	-0.00568	0.001493	-3.804	0.0012 **

The results show that the adjusted $R^2 = 0.8$, indicating a good fit that meets the expected standard. The model's p-value is 8.661e-08, indicating that the model is significant at the 95% confidence level. At the same time, the p-value of the principal component F1 is 4.47e-06, and the p-value of the principal component F2 is 0.0012, both significant at the 95% confidence level , as shown in Table 5 and Table 6.

Finally, we use the statistical software R to test whether there is heteroscedasticity in the model , as shown in Figure 1.

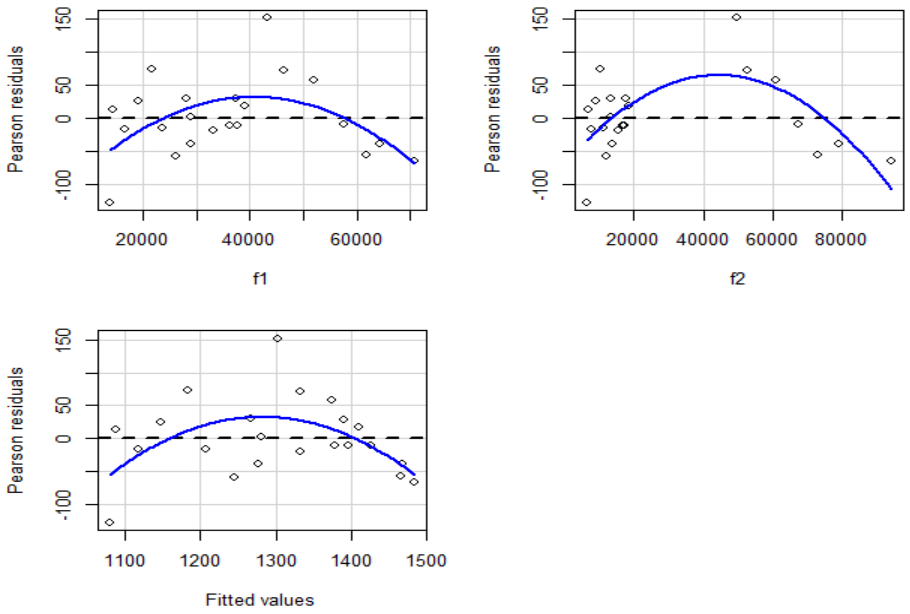

Figure 1: Residual plot

From the residual plot, it can be seen that the model residuals are evenly distributed around the mean value of 0, with no heteroscedasticity. Thus, the model passes various tests, and the prediction equation for the cold chain logistics demand is obtained:

$$Y = 898.37 + 0.016\mathrm{F}_1 - 0.0057\mathrm{F}_2$$

3.3 Establishing a Principal Grey Prediction Model

To obtain the predicted values of cold chain logistics demand in China for the next five years, we predict the values of each indicator and then input them into the principal component regression model. Compared to directly using the grey model for dynamic prediction, this approach offers higher accuracy. Taking the total consumer expenditure index x_1 as an example, based on the $GM(1,1)$ model introduced earlier, the basic form is $x^{(0)}(k) + \hat{a}z^{(1)}(k) = \hat{b}$. After running the model, we can obtain the following results:

$$a = -0.022,\ b = 512.92,\ C = 0.009$$

The development coefficient a represents the development pattern and trend of the data series, while the grey action quantity b reflects the changing relationship within the series. The posterior difference ratio C is used to verify the accuracy of grey prediction; a smaller ratio indicates higher accuracy. Generally, if the C value is less than 0.35, the model has high accuracy. For x_1, the C value is 0.009, indicating that the model's accuracy is very high.

ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

The grey prediction results for the x_1 indicator, representing consumer price index, are presented below , as shown in Figure 2:

Figure 2: Grey prediction plot

The graph demonstrates a good fit between 2017 and 2021. This suggests that the grey model is accurate in characterizing and predicting consumption data, effectively capturing consumer behavior patterns.

3.4 Model Validation

Two methods are commonly used for validating the $GM(1,1)$ model: residual testing and ratio deviation testing. Upon successful validation, the model can be employed for prediction. Results are obtained as presented below:

Table 7: Residual table

Index entry	Original value	predicted value	residual	Relative error (%)
2017	516.3	516.3	0	0
2018	528.1	529.928	-1.828	0.346
2019	543.3	541.606	1.694	0.312
2020	555.6	553.541	2.059	0.371
2021	563.9	565.739	-1.839	0.326

Under general circumstances, a relative error of less than 20% indicates that the fitting degree meets general requirements, while a relative error of less than 10% suggests a very good fitting degree. The model's relative error remains stably between 0.3% and 0.4%, implying that the model's fitting effect is very good. Therefore, it is believed that the $GM(1,1)$ model has an excellent fitting degree for the original data.

Table 8: Deviation of grade ratio

Index entry	Original value	Grade ratio
2017	516.3	-
2018	528.1	0.978
2019	543.3	0.972
2020	555.6	0.978
2021	563.9	0.985

If all the level ratio values are within the interval $(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$, then the model construction is suitable. All the level ratio values of the original sequence are within the interval $(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ i.e. (0.717, 1.396), indicating that the original sequence is suitable for constructing a gray prediction model , as shown in Table 7 and Table 8.

3.5 Prediction Results

By repeating the above model solving and verification steps, the predicted values for each indicator for the next four years are obtained as follows:

ISSN 2616-5902 Vol. 5, Issue 14: 142-148, DOI: 10.25236/AJBM.2023.051423

By substituting the predicted values of the above indicators into the principal component regression model, we obtain the cold chain logistics demand for the next four years, as shown in Table 9 and Table 10:

Table 9: Gray Prediction Table for All Indicators

	x_1	x_2	x_3	x_4	x_5	x_6
2023	590.947	435.809	369040.917	13812.439	31370.610	51935.881
2024	603.969	444.083	393336.905	14281.934	32726.722	59002.444
2025	617.278	452.514	419232.431	14767.387	34141.457	67030.508
2026	630.881	461.104	446832.801	15269.340	35617.350	76150.895

Table 10: Cold Chain Logistics Demand

year	2023	2024	2025	2026
Cold chain logistics demand	1523.13	1533.545	1540.945	1544.645

4. Conclusion

Cold chain logistics demand is influenced by various factors, including logistics demand and socio-economic environment. Time series prediction methods only consider the impact of time factors, so they are not used. Due to the late start of new retail development in China and limited historical data. The development of China's new retail industry started relatively late, resulting in limited historical data availability. Therefore, using the BP neural network method for prediction is not suitable. Moreover, the single grey model is not suitable for multi-factor and multi-variable analysis, and multiple linear regression may produce unreliable predictions due to multicollinearity. To improve accuracy, the article combines the grey model with multiple linear regression model for prediction. Principal component analysis is used to reduce the dimensionality of the multiple linear regression model, addressing the issue of multicollinearity. Model results demonstrate that the above issues have been successfully resolved, thereby improving prediction accuracy.

References

[1] Wang Yuxia. Problems and countermeasures of cold chain logistics of agricultural products in China [J]. Logistics engineering and management. 2016(2): 80-84.
[2] Zhu Kunping, Jiang Linlin and Wang Henan. Market analysis and countermeasures of cold chain logistics of agricultural products in Hebei Province [J]. Price Monthly, 2016 (12): 64-68
[3] Du J. et al., Cooling performance of a thermal energy storage-based portable box for cold chain applications [J]. Journal of Energy Storage, 2020. 28: p. 101238.
[4] Zhao J. et al., Recyclable low-temperature phase change microcapsules for cold storage [J]. Journal of Colloid and Interface Science, 2020. 564: p. 286-295.
[5] Liang Yan, Yang Huihui and Su Huihui. Forecast analysis of Tianjin agricultural products cold chain logistics demand based on multiple linear regression [J]. Southern Agricultural Machinery, 2018 49 (18): 230-231.
[6] Zeng Hao, Zhu Wenjuan. Forecast analysis of cold chain logistics demand of fresh agricultural products in Hunan Province based on grey GM (1, 1) model [J]. Journal of Xinyang Agriculture and Forestry University, 2022, 32 (04): 40-46
[7] Xu Wen, Wen Jialin. Research on Cold Chain Logistics Demand Forecast and Influencing Factors of Aquatic Products in Zhejiang Province [J]. Logistics Engineering and Management, 2023, 45 (02): 41-45.