A Method for Predicting the Average Selling Price of Cigarette Per Carton Based on Multiple Linear Regression

Xin Qu[1], Huahua Wu[2], Yan Liang[2], Wei Yu[1], Jianbo Xue[1], Xiaoxia Shen[1], Qiao Liang[3,*], Changhua Chen[3]

[1]Sichuan Provincial Tobacco Company Liangshan Prefecture Company, Xichang, 610039, China
[2]China National Tobacco Corporation Sichuan Company, Chengdu, 610039, China
[3]School of Management, Xihua University, Chengdu, 610039, China

1306760628@qq.com

Abstract. As the cigarette industry shifted from the dual-growth model of "sales + structure" to the single-channel development model of "structure", the sustainable growth of the single-carton sales structure has become an important hand for the industry to realize high-quality development. However, the current prediction methods for the average price of cigarette sales per unit rely heavily on empirical forecasting, lacking in theoretical and scientific basis, and fall short of practical needs. In order to further improve the prediction accuracy of the average price of cigarette sales per unit, this paper combines macroeconomic indicators, constructs a multiple linear regression model, and screens the economic indicators and tests the model through the unit root testing, difference handing and autocorrelation test. Focusing on the cigarette business budget of Liangshan Yi Autonomous Prefecture, this study finds that the structure of cigarette sales per unit directly affects the average sales price, which is closely tied to consumer demand. It further reveals that cigarette consumption is influenced by seven macroeconomic factors: the growth rate of tax benefits, urban disposable income growth rate, national per-unit sales, GDP, the added value of the tertiary sector, CPI index, and household food expenditure. Consequently, a multivariate linear regression model is constructed with these seven factors as independent variables to predict the average price of cigarette sales per unit. The results demonstrate that this model can effectively explain 95.6% of the variance in the average price per unit, which can significantly enhance the accuracy and scientific rigor of measurable and controllable cigarette marketing with an average prediction accuracy of 99.58%.

Keywords: Multivariate Linear Regression, Operational budget, Average selling price of cigarettes per carton

F. Cao et al. (eds.), *Proceedings of the 2023 5th International Conference on Economic Management and Cultural Industry (ICEMCI 2023)*, Advances in Economics, Business and Management Research 276,
https://doi.org/10.2991/978-94-6463-368-9_15

1 Introduction

Under the situation of approaching the demand turning point, increasing the average sales price per carton of cigarettes is an important way to fully tap the market resources and promote the income of the cigarette supply chain. Optimizing the cigarette structure is an important way to increase the average sales price per carton of cigarettes, so it is also called "improving the cigarette structure". Under the new normal of economic development, improving the cigarette structure has been placed in an important position of the economic operation regulation of the tobacco industry. It is crucial to accurately predict the trend of the average sales price per carton of cigarettes based on the level of economic and social development and the consumption ability of residents.

In terms of forecasting methods, existing studies mainly use time series forecasting methods and causal analysis methods. Time series methods mainly include moving average method, exponential smoothing method, and BOX-Jenkins method; causal analysis methods are divided into linear regression forecasting, nonlinear regression forecasting, gray system forecasting model, and Markov forecasting model. For example, the methods used by Ding Zi Xiang in his research on the application of combination-based forecasting methods in cigarette sales include quadratic curve regression, quadratic curve regression, gray prediction, and stepwise regression analysis [1]. Hua Yong proposed a model and algorithm for forecasting the monthly delivery volume of cigarettes, which can coordinate the industrial and commercial sides, and combine the market and the plan. That is, using the time series method to determine the commercial sales forecasting model. At the same time, an innovative application algorithm of seasonal variation forecasting method was designed to use two quantitative regulation forecasting tools and based on the management of stock-to-sale ratio to finally calculate the monthly cigarette placement volume [2]. In addition, existing studies use Markov models, gray models, regression models, etc. to make predictions in some years with small sample sizes. For example, Yang Wan and Xia Zhengwei both used the Markov chain model in the prediction of future cigarette demand, and obtained accurate and effective analysis [3]. Zhao Hui constructs a quantitative forecasting model to calculate cigarette demand, analyzes the three methods of trend extrapolation, time series decomposition, and multiple regression by comparing the fit and efficacy of the three methods, and comes to the conclusion that the model of time series data can effectively use historical sales data to calculate a more accurate forecast [4].

In terms of the object of forecasting, the forecasting studies of total annual cigarette sales and monthly sales accounted for a larger proportion of the studies than the studies in terms of the average price of sales per carton. Xiong Li established a monthly demand forecasting model for Guiyang city area based on the principle of control charts, using two time series forecasting methods, namely, moving average method and exponential smoothing method [5]. And Yang Lei also used regression model, exponential smoothing model, time series decomposition model, seasonal ARIMA model and X13-A-S method to forecast the monthly sales of Yuxi (soft) cigarettes from 2015 to 2020 in the study of time series forecasting method of monthly sales of cigarette single specification [6]. However, there are fewer studies on the prediction of cigarette unit carton value. Sheng Juan predicts the average price of cigarette unit carton in Inner Mongolia

Autonomous Region through correlation analysis and regression analysis, but it uses fewer indicators and the results of the study are less applicable [7].

In summary, the literature on annual and monthly cigarette sales forecasts is abundant and the forecasting methods are relatively mature, but there are fewer studies on the forecasting of the average price of a single carton sold. As a special industry in China, the change of its average sales price per carton has a strong linear correlation with the regional macroeconomic level. Therefore, based on the literature combing and reality analysis, this paper concludes that the regression prediction model is applicable to the prediction of the average price of cigarette single carton sales. This paper combines macroeconomic indicators and constructs a multiple linear regression model to predict the future average price of cigarette unit carton sales in Liangshan Yi Autonomous Prefecture as an example. Furthermore, referring to the trends of socio-economic development and the requirements of the higher-level work, the average selling price per carton was forecasted and planned with the aim of increasing the sales revenue per carton and focusing on the structural improvement. This would assist the economic policy regulation of Liangshan Prefecture and strengthen the cigarette marketing management [8]. First of all, this paper creates a set of predictive indicators including economic level, product factors and residents' living standard on the basis of literature combing, and screens out 16 usable indicators such as tax profit, single case growth rate and consumer price index through unit root test and difference handing. Secondly, because the lag effect of economic variables, time series non-stationarity may trigger autocorrelation leading to inaccurate parameter estimation, this paper further screens the time series data through autocorrelation test. Finally, a multiple linear regression model is established to predict the average price of cigarette single carton sales, which is expected to provide a more meaningful reference for the production and operation decisions of tobacco enterprises as well as the regulation of regional economic policies.

2 Selection of Classification Indicators

According to the principle of "the population determines the sales volume, the economy determines the unit carton", we build the internal and external influence indicators combined with the relevant literature. In the literature on forecasting price models, internal influencing factors are usually set up as corporate profitability, sales price of related products or observed value of previous years' prices [9]. Therefore, among the internal influencing factors, this paper adopts the factors that are linearly related to the rate of change of the average price per case sold, mainly the average price per carton sold at all levels and the rate of growth of per carton. Analyzing the external environment, the growth rate of disposable income per capita, GDP and its growth rate, and the value of social production have an important influence on the forecasting of economic prices [10]. And there is a significant positive correlation between the average price of cigarettes sold per carton and the above factors [7]. Therefore, the main conditions for considering external influences are the growth rate of disposable income per capita, GDP and its growth rate, and the value of social production and other relevant economic factors. The impulse response function is used to analyze the VAR model constructed

by the CPI and the hog slaughter price index, and it is clear that there is a dynamic relationship between the hog slaughter price and the CPI [11]. And the correlation between residents' food consumption expenditure and residents' annual disposable income is strong [12], so residential food consumption expenditure and CPI are taken into account in the initial indicators. Applying tax big data for analysis and research can reflect the level of economic development from the bottom up [13]. Therefore, the tax and tax growth rate will be taken into account in the forecasting model of this paper, which is expected to have a linear correlation with the average price of cigarette single carton sales, which is more accurate for forecasting the average price of cigarette single carton sales.

In summary, to determine the external economic factors, the standard of living of the population, internal factors and other factors required to predict the single carton, the above factors are specifically divided into 17 indicators, which can be screened according to the actual situation and calculations. The specific indicators are shown in Table 1.

Table 1. Initial correlates and indicators for single-carton forecasting

Factor	Composition of indicators	Data sources	Data unit
External economic factors	Urban disposable income per capita growth rate	Liangshan Statistical Yearbook	%
	Rural disposable income per capita growth rate	Liangshan Statistical Yearbook	%
	Citywide GDP	Liangshan Statistical Yearbook	Ten thousand yuan
	Citywide GDP growth rate	Liangshan Statistical Yearbook	%
	Resident food consumption expenditure	Liangshan Statistical Yearbook	Ten thousand yuan
	Total retail sales of social consumer goods	Liangshan Statistical Yearbook	Ten thousand yuan
	Value added of the tertiary industry	Liangshan Statistical Yearbook	Ten thousand yuan
	Tax profit	Liangshan Statistical Yearbook	Ten thousand yuan
	Tax profit growth rate	Liangshan Statistical Yearbook	%
	CPI index	Liangshan Statistical Yearbook	/
Internal factors	Cigarette tax and profit	Liangshan Prefecture Tobacco Bureau	Ten thousand yuan
	Cigarette tax and profit growth rate	Liangshan Prefecture Tobacco Bureau	%
	National average selling price per carton	Liangshan Prefecture Tobacco Bureau	yuan
	National average price growth rate of single carton sales	Liangshan Prefecture Tobacco Bureau	%
	Average selling price per carton in the province	Liangshan Prefecture Tobacco Bureau	yuan
	Average price growth rate of single carton sales in the province	Liangshan Prefecture Tobacco Bureau	%
	Average price growth rate of single carton sales in the city	Liangshan Prefecture Tobacco Bureau	%

3 Model Construction

Multiple linear regression method is a statistical method used to research the linear relationship between a dependent variable (or response variable) and multiple independent variables (or explanatory variables). Multiple linear regression methods can be used to describe, predict, control and analyze multivariate data. Using multiple regression can see if two or more variables change together and quantify the relationship between them [14]. On this basis, the test substitutes the screened indicators into the

model, and the historical values are used to predict the future values. Finally, the feasibility of the indicators and models is judged by calculating the goodness of fit between the real and predicted values [15]. The factors that affect the average selling price per carton were identified through literature review. Data were collected annually from relevant databases, and a dataset was constructed and cleaned. Then, all possible combinations of indicators were tested, and the models trained by each combination were evaluated. The best-performing model was selected as the final output model.

3.1 Data Processing and Verification

Linear Interpolation Method to Complete Data.

Linear interpolation is a common numerical analysis method used to estimate the value of an unknown data point between two known data points. It is based on the simple assumption that values vary uniformly between two known data points [16]. The calculation method is:

Firstly, calculate the slope k between two known data points, and then calculate the value of the unknown data point based on this slope. The slope k is obtained by calculating the difference between the known data points $(y_2 - y_1)$ and the corresponding x-coordinate $(x_2 - x_1)$. Using the slope k and the distance between the unknown stronghold and the first known data point, use the following formula to calculate the value of the unknown stronghold:

$$y = y_1 + k * (x - x_1)$$

Where y is the value of the unknown stronghold, y_1 is the value of the first known data point, slope k is the slope between the two known data points, x is the horizontal coordinate of the unknown stronghold, and x_1 is the horizontal coordinate of the first known data point.

Data Smoothness Test.

In this paper, the ADF unit root test is used to test the smoothness of the time series [17]. The inspection process is shown below. Take the AR(p) model as an example:

$$x_t = \varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p} + \varepsilon_t$$

Assume

$$\rho = \varphi_1 + \cdots + \varphi_p - 1$$

In the ADF unit root test, the original and alternative hypotheses are:

$$H0: \rho=0$$

$$H1: \rho<0$$

The original hypothesis indicates that the series is non-stationary and the alternative hypothesis indicates that the series is stationary.

The ADF statistic was constructed as follows:

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$$

Where $S(\hat{\rho})$ is the sample standard deviation of ρ. Based on the table of critical values and the P-value corresponding to τ, it is possible to determine whether the ADF value accepts the original hypothesis or rejects it.

3.2 Construction of Multiple Linear Regression Model

Regression analysis is the process of processing and analyzing the historical data of the independent and dependent variables and training them to find out the quantitative relationship between the independent and dependent variables, and then build a predictive model of the dependent variable [18].

Suppose the linear regression model between the explanatory variable y and multiple explanatory variables $x_1\ x_2 \ldots\ x_k$ is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \mu$$

Where y is the explanatory variable $x_j (j = 1\ 2\ \ldots\ k)$ is the k explanatory variables $\beta_j (j = 1\ 2\ \ldots\ k)$ is the $k+1$ unknown parameter, and μ is the random error term [19].

For n sets of observations $y_i\ \ x_{1i}\ \ x_{2i} \ldots\ \ x_{ki} (i = 1\ 2\ \ldots\ n)$ the system of equations has the form

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \mu_i\ (i = 1\ 2\ \ldots \mathrm{n})$$

so

$$y = x\beta + \mu$$

The parameters $(\beta_0, \beta_1, \beta_2 \ldots \beta_k)$ were estimated by least squares with the aim of minimizing the residual sum of squares.

$$Q = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$$

According to the principle of finding the limit in calculus, find the partial quotient for $\beta_0, \beta_1, \beta_2 \ldots \beta_k$, respectively, such that it satisfies

$$\frac{\partial Q}{\partial \beta_k} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik}) x_{ik} = 0$$

The system of equations is obtained by replacing β with $\hat{\beta}$: $X^T\left(y - X\hat{\beta}\right) = 0$ so: $X^T X\hat{\beta} = X^T Y$

The least squares solution of the regression parameters after shifting the terms is given by$\hat{\beta} = (X^T X)^{-1} X^T y$, This results in a multivariate linear transcendental regression equation:

$$y = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \cdots + \widehat{\beta_k} x_k$$

3.3 Model Verification

Test Data Correlation.

Correlation analysis is the study of whether there is some kind of dependence between phenomena, testing for independence between the multiple observations that make up the sample [20].

Autocorrelation refers to the existence of interrelationships between the random error terms, undermining the basic assumption of the linear regression model that the random disturbance terms are independent or uncorrelated with each other [21]. The goal of the autocorrelation test is to determine whether there is a statistically significant autocorrelation structure in the time series, i.e., whether the values of past observations are correlated with current observations. Since the autocorrelation problem leads to an increase in the up and down swing of the regression line estimated from the sample data, resulting in inaccurate parameter estimation, an autocorrelation test is performed on the indicator combinations while calculating. For small samples of spherical perturbations, no autocorrelation is required. The test can be used to assess randomness and trends in time series data.

Goodness of fit.

The method used to statistically test the effect of the regression model is the goodness-of-fit test. Goodness-of-fit is whether, how close, and to what extent the modeled trajectory matches the observed actual situation, and is usually measured statistically by the constructive statistic R^2. Calculated from the sample data, if the modeling is closer to reality, R^2 is closer to 1 [22]. The R^2 formula is as follows:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Accuracy Inspection.

To verify the accuracy of the model, the test results were subjected to a relative error test [23]. The formula is as follows:

$$W = \frac{|P_0 - P_1|}{P_0}$$

Where W is the error rate, P_0 is the actual value and P_1 is the predicted value.

4 Case Study

4.1 Data Collection

Based on data availability, this study's dataset spans a total of 13 years from 2010-2022 and involves the following 17 indicators as well as the historical annual average price of single carton sales in Liangshan Yi Autonomous Prefecture. Due to the early years, the Liangshan Yi Autonomous Prefecture Bureau of Statistics did not show the economic indicators for 2010, and the Liangshan Yi Autonomous Prefecture Tobacco Bureau only provided some relevant data for 2011-2022. Therefore, the missing part of the data in 2010 was supplemented using the formula of the linear interpolation method, and the supplemented indicators are: tax profit, tax profit growth rate, cigarette tax profit, cigarette tax profit growth rate, the national single carton, the national single carton growth rate, the province's single carton, the province's single carton growth rate, as well as the state's single carton growth rate. The final partial year data are shown in Table 2.

Table 2. Forecast Dataset

Year	2010	2011	2012	…	2020	2021	2022
Average selling price of single carton cigarettes in Liangshan City	16515.93	18525.99	20536.05	…	34177.09	36336.77	37217
Tax profit	150151.32	213913.48	277675.6	…	393385.2	430575.4	415130.3
Citywide GDP growth rate	0.25	0.152	0.138	…	0.039	0.072	0.06
Urban disposable income per capita growth rate	0.134	0.157	0.152	…	0.048	0.081	0.051
National average price growth rate of single carton sales	0.1973	0.156	0.1147	…	0.0358	0.0521	0.0537
Average price growth rate of single carton sales in the province	0.215	0.17	0.125	…	0.035	0.0433	-0.0283
Average price growth rate of single carton sales in the city	0.1033	0.1085	0.1137	…	0.0658	0.0632	0.0242
CPI index	102.9	105.2	102.6	…	103.8	99.8	100.2
Cigarette tax and profit	53328.45	58422.77	63517.09	…	135544.3	142868.7	154150.1
Tax profit growth rate	-0.1021	0.098	0.2981	…	-0.0381	0.0945	-0.0359
Rural disposable income per capita growth rate	0.153	0.213	0.159	…	0.095	0.103	0.068
National average selling price per carton	18571	20929	23287	…	33987	35758	37678
Average selling price per carton in the province	16507	19888	23269	…	37329	38945	40045
Citywide GDP	7841900	10001275	11226693	…	17331465	19011773	20813591
Value added of the tertiary industry	2410818	2820021	3160070	…	7668555	8186443	8773542
Total retail sales of social consumer goods	2448765	2879371	3471364	…	6782290	7616599	7829510
Cigarette tax and profit growth rate	0.5616	0.3244	0.0872	…	0.1017	0.054	0.789
Resident food consumption expenditure	204327	214953	259019	…	256505	255992	256503

As can be seen from the scatter plot, the average price of a single carton of sales in Liangshan Yi Autonomous Prefecture has been increasing with the trend of years. However, if only eight years of data from 2015-2022 are used for the regression, the sample is too small to take into account more relevant macroeconomic factors. Therefore, data from 2010-2020 was chosen to train the model, and data from 2021 and 2022 were used for forecasting. Figure 1 shows the trend of the average price of a single carton from 2010 to 2020.

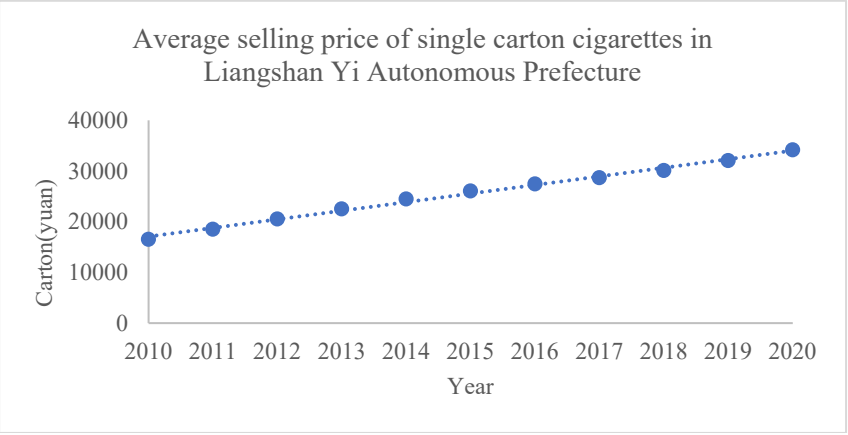

Fig. 1. 2010-2020 Annual trend of average single carton sales price

4.2 Data Processing

Before estimating the long-term relationship between the above time series data, a unit root test must be done [24]. The presence of a unit root process in the sequence is not smooth and can cause pseudo-regression in the regression analysis. After the unit root test, the partial unsteady series are respectively processed with each median difference, and the results as in Table 3 are obtained. The results show that the unit root test of the data of cigarette tax profit growth rate and residents' food consumption expenditure is not ideal, and it cannot be optimized after differential processing, so 16 indicators out of 17 indicators passed the test.

Table 3. Unit Root Test of Variables

Unit Root Test / Variable	0th-order difference	1th-order difference	2th-order difference	3th-order difference	Final difference order
Tax profit	0.6534	0.1714	0.0000	/	2th-order
Tax profit growth rate	0.0187	/	/	/	0th-order
Cigarette tax and profit	0.5485	0.3496	0.0724	0.0117	3th-order
Citywide GDP growth rate	0.0166	/	/	/	0th-order
Urban disposable income per capita growth rate	0.3301	0.0008	/	/	1th-order

Unit Root Test / Variable	0th-order difference	1th-order difference	2th-order difference	3th-order difference	Final difference order
Rural disposable income per capita growth rate	0.1587	0.0000	/	/	1th-order
National average selling price per carton	0.1680	0.7846	0.0002	/	2th-order
National average price growth rate of single carton sales	0.6909	0.0033	/	/	1th-order
Average selling price per carton in the province	0.0571	0.7188	0.0174	/	2th-order
Average price growth rate of single carton sales in the province	0.3263	0.6818	0.0203	/	2th-order
Average price growth rate of single carton sales in the city	0.8048	0.8976	0.4289	0.0012	3th-order
Citywide GDP	0.6683	0.1720	0.0000	/	2th-order
Value added of the tertiary industry	0.7382	0.0170	/	/	1th-order
Total retail sales of social consumer goods	0.4660	0.0000	/	/	1th-order
CPI index	0.2967	0.0182	/	/	1th-order
Resident food consumption expenditure	0.8998	0.4305	0.0933	0.0122	3th-order
Cigarette tax and profit growth rate	0.9820	0.6045	0.7248	0.3975	/

4.3 Multiple Regression Model

The metrics are ranked and combined, the models trained by testing all possible combinations are tested, and the model with the best performance is selected as the final output model.

The 16 pending indicators are substituted into STATA software for multiple regression analysis for prediction by layer stacking, with a stepwise method of adding indicators from 2 indicators to n indicators. The results show that the optimal model indicators include tax and profit growth rate, urban disposable income growth rate, national single carton, GDP, value added of the tertiary industry, CPI index, and residents' food consumption expenditures a total of seven indicators. The results are shown in Table 4.

Table 4. 2022 Forecast Results and Autocorrelation Test Results

Using Variables	Projected results for 2022	Autocorrelation test
Tax profit growth rate, National average selling price per carton	32402.113	0.0638
Tax profit growth rate, National average selling price per carton, Value added of the tertiary industry	32405.006	0.0643
Tax profit growth rate, National average selling price per carton, Value added of the tertiary industry, Resident food consumption expenditure	33257.93	0.1311

Using Variables	Projected results for 2022	Autocorrelation test
Tax profit growth rate, National average selling price per carton, Value added of the tertiary industry, Resident food consumption expenditure, Citywide GDP	34228.211	0.3904
Tax profit growth rate, National average selling price per carton, Value added of the tertiary industry, Resident food consumption expenditure, Citywide GDP, Urban disposable income per capita growth rate	37986.461	0.0759
Tax profit growth rate, National average selling price per carton, Value added of the tertiary industry, Resident food consumption expenditure, Citywide GDP, Urban disposable income per capita growth rate, CPI index	37295.613	0.0849

Autocorrelation was tested for the above seven indicators using the Ljung-CARTON Q test. The results show the probability of Prob>chi2 of 0.0849, so the original hypothesis that there is no serial autocorrelation. Therefore, the original hypothesis of "no serial autocorrelation" can be accepted at the 5% level of significance.

Table 5. Results of autocorrelation test

Wntestq e1,lags(1) Portmanteau test for white noise		
Portmanteau(Q) statistic	=	2.9685
Prob>chi2(1)	=	0.0849

The result of the model calculations is shown in Table 5. where the goodness of fit, R^2, is the parameter estimate using least squares and is the ratio of the regression sum of squares to the total sum of squares of deviation. It represents the proportion of the total sum of squared deviations that can be explained by the regression sum of squares. The result of the calculation is 0.9562, which indicates that the model regression fits well. Table 6 shows the optimal calculation results of the model.

Table 6. The optimal calculation results of the model

Variable	Symbol	Ratio
Tax profit growth rate	x_1	-55130.87
Urban disposable income per capita growth rate	x_2	-112968.8
National average selling price per carton	x_3	7.165451
Citywide GDP	x_4	0.0047438
Value added of the tertiary industry	x_5	-0.006655
CPI index	x_6	-1541.965
Resident food consumption expenditure	x_7	0.0351485
Constant	c	34765.54
Goodness of fit	R^2	0.9562

Converting this to a mathematical equation is:

$$Y = 34765.54 - 55130.87\,x_1 - 112968.8x_2 + 7.165451x_3 + 0.0047438x_4 - 0.006655x_5 - 1541.965x_6 + 0.0351485x_7$$

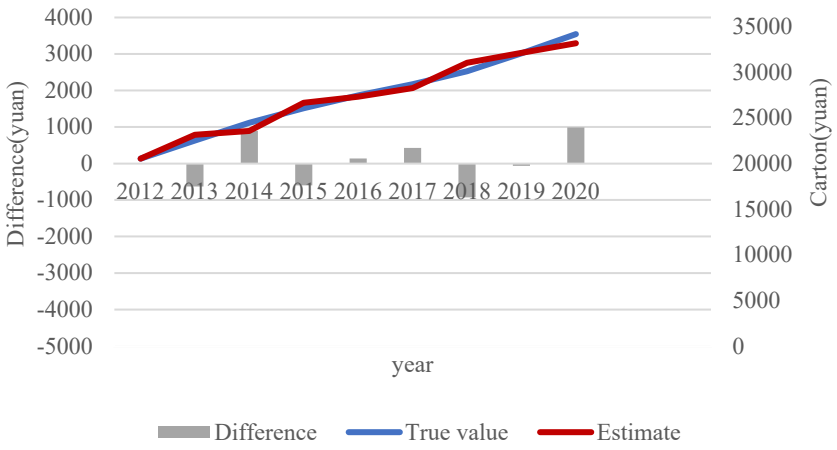

Fig. 2. Comparison of real and projected values, 2012-2020

The comparison between the true and predicted values can be visualized from Figure 2, and the results show that the model predicts better.

4.4 Model Application Analysis

Cigarette industry belongs to the indicator sales, so most of the prediction value of cigarettes by previous researchers used empirical prediction method, due to the lack of empirical data. On this basis, this paper forecasts the single carton of cigarettes in Liangshan Yi Autonomous Prefecture, and establishes a model through linear regression method, which gives substantial help in forecasting the average sales price of single case of cigarettes in Liangshan Yi Autonomous Prefecture. Among them, the average price of single case sales in 2021 is forecasted to be 36565.367 yuan, and the actual value is 36336.77 yuan, with a difference of 228.597 yuan; The projected average price per case sold in 2022 is 37,295.613 yuan and the actual value is 37,217 yuan, with a difference of 78.613 yuan. This model is based on multivariate linear regression, and the predicted average price per case sold for the next year is obtained by inputting the relevant data available for the previous year This model was based on multivariate linear regression, and the predicted average price per case sold for the next year was obtained by inputting the relevant data available for the previous year. The average prediction accuracy is 99.58%. The projections for 2021 and 2022 are shown in Figure 3.

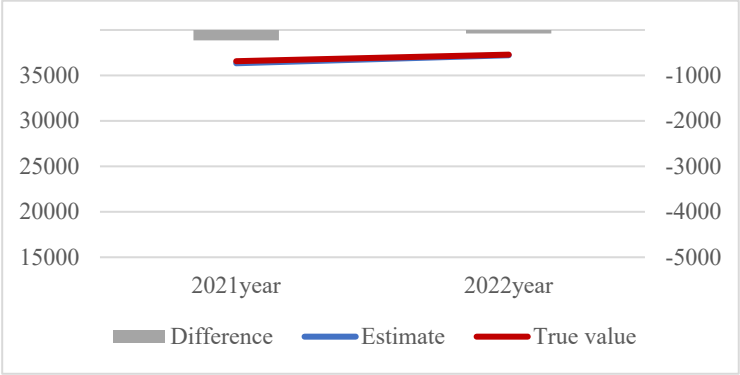

Fig. 3. Errors in 2021 and 2022

5 Conclusion

Using the unit root test, difference handing, autocorrelation test and other methods, this paper constructs a multiple linear regression model to optimize and improve the prediction method of the average price of a single carton of cigarettes. Taking Liangshan Tobacco Commercial Company's single average price forecast as an example, based on the principle of "population determines sales volume and economy determines structure" of cigarette sales. In this paper, seven indicators, including tax and profit growth rate, urban disposable income growth rate, national single case, GDP, value added of the tertiary industry, CPI index, and residents' food consumption expenditures, were selected as explanatory variables, and a forecasting model for the average price of a single case of cigarettes in the cigarette business in Liangshan Yi Autonomous Prefecture was established. This changed the original empirical prediction mode, improved the scientific nature of the business budget, and could provide a more meaningful reference for the production and operation decision-making of tobacco enterprises. However, there are still shortcomings in this study, such as the model in this paper may have regional limitations and may require some adjustments if it is used in other regional budget studies. Moreover, changes in cigarette consumption perceptions and the overall supply and demand in the cigarette market due to health needs can reduce the accuracy of the model. Based on this, there is still a lot of room for improvement in this study, such as the introduction of cutting-edge methods like machine learning, to continuously improve the scientific and applicability of the study.

References

1. Ding, Zixiang. (2023). Research on the application of combined forecasting methods in cigarette sales Doctoral dissertation, Anhui University of Science and Technology]. DOI: 10.26918/d.cnki.ghngc.2021.000445.
2. Hua, Y. (2015). A forecasting model for monthly cigarette distribution volume based on industrial and commercial coordination. Journal of China Tobacco, 21(4), 99-106. DOI: 10.16472/j.chinatobacco.2014.207.
3. Xia, Zhengwei. (2013). Cigarette demand forecasting based on grey Markov model. Journal of Shanghai Second Polytechnic University, 30(2), 101-106. DOI: 10. 19570/ j. cnki. jsspu. 2013.02.004.
4. Zhao, Hui, & Wang, Hui. (2011). Comparison of cigarette demand forecasts in a city of Hebei Province based on various time series models. Economic Research Guide, (08), 142-145.
5. Xiong, Li. (2009). Application of CV-control charts and time series analysis in monthly cigarette demand forecasting for customer manager areas in Guiyang City. Modern Market, (07), 221-223.
6. Yang, Lei, He, Xuefeng, Zhang, Tao et al. (2023). Research on time series forecasting methods for monthly sales volume of single-specification cigarettes. China Market, (20), 137-143. DOI: 10.13939/j.cnki.zgsc.2023.20.137.
7. Sheng, Juan. (2020). Analysis of the trend of single-carton value changes of cigarettes based on national economic data. Modern Market, (04), 123-127. DOI: 10. 14013/ j. cnki. scxdh. 2020.04.055.
8. Li, Kunzhong. (2011, February 15). Timely and comprehensive deployment of tobacco production and operation in Liangshan Prefecture. Zhonghua Cooperative Times, p. D03.
9. Xu, Jie. (2009). Analysis of financial indicators in stock price prediction. Chinese Chief Accountant, (10), 96-97.
10. Ma, Chengwen, Xue, Qianyu, & Wang, Yufei. (2019). Factors influencing the market price of commercial residential buildings and price prediction. Journal of Changchun University of Technology, 40(3), 282-288. DOI: 10.15923/j.cnki.cn22-1382/t.2019.3.14.
11. National Bureau of Statistics of Guangxi Investigation Corps Research Group, & Liao, Jinchang. (2022). Study on the characteristics of pork price fluctuations and its dynamic relationship with CPI: An empirical analysis based on the VAR model. Market Forum, (11), 1-13.
12. Deng, Lebin, Li, Junhua, & Wei, Yifeng. (2008). Forecasting model of urban residents' consumption expenditure and income: An empirical analysis based on Shiyan City. Modern Market, (36), 144-145.
13. Fan, Yong, & Du, Han. (2021). Big data in taxation: Theory, application, and limitations. Taxation Research, (09), 57-62. DOI: 10.19376/j.cnki.cn11-1011/f.2021.09.009.
14. Kiss I, Cioată G V, Ratiu A S, et al. (2018). Statistical experiments using the multiple regression research for prediction of proper hardness in areas of phosphorus cast–iron brake shoes manufacturing[J]. IOP Conference Series: Materials Science and Engineering, 2018, 294(1).
15. Wang, Huiwen, & Meng, Jie. (2007). Forecasting modeling method of multiple linear regression. Journal of Beijing University of Aeronautics and Astronautics, 2007(04), 500-504. DOI: 10.13700/j.bh.1001-5965.2007.04.028.
16. McElroy, S. T, Politis, et al. (2022). Optimal linear interpolation of multiple missing values[J]. Statistical Inference for Stochastic Processes,2022,25(3).

17. Guan, Haotian. (2022). Research and application of time series forecasting methods based on autoregressive models [Doctoral dissertation, Yunnan University]. DOI: 10. 27456/ d. cnki.gyndu.2022.000384.
18. Zhan, Haisong, & Gao, Xiangzhou. (2023). Establishing a prediction model for rebar prices based on regression analysis. Metallurgical Management, (16), 37-40.
19. Qiu, Mei, Shi, Longqing, Teng, Chao, et al. (2013). Multivariate linear regression prediction model for the extension length of small faults in mines. Coal Geology & Exploration, 41(6), 7-10.
20. S. W L. (2007). A Test for Serial Correlation in Multivariate Data[J]. The Annals of Statistics,2007,5(2).
21. Wang, Zengmin, & Li, Tong. (2022). Study on the test power of autocorrelation test methods. Quantitative Economics and Technical Economics Research, 39(1), 174-189.
22. Song, Yunxue, Zhang, Kexing, & Shi, Yongsheng. (2009). Prediction of engine performance parameters based on multiple linear regression. Journal of Aeronautical Power, 24(2), 427-431. DOI: 10.13224/j.cnki.jasp.2009.02.019.
23. Zhang Jianlong, He Zheng Siyu, Jiang Zhihui. (2022). Research on the Prediction of Live Chicken Market Prices in China in the Post-pandemic Era. China Market, 2022(10): 126-128. DOI: 10.13939/j.cnki.zgsc.2022.10.126.
24. Cui Daoling. (2005). Unit Root Test and Error Correction Model: Principle and Application. Journal of Nanjing Audit University, 2005(03): 15-18.