

Research and Practice of Tobacco Industrial Companies' Sales Forecast Based on ARMA Model, Seasonal Fluctuation and Policy Factor

Qiao Shi¹, Yong Cen^{1,*}, Liping Xu² and Lina Zhang²

¹Information Center of China Tobacco Zhejiang Industrial Co., Ltd., Hangzhou, 310009, China

²Technology Center of China Tobacco Zhejiang Industrial Co., Ltd., Hangzhou, 310024, China

*Corresponding author

Abstract—In the wave of marketization reform, the tobacco industry has been gradually transforming to the market-driven planning model from fully planned economic model. In the tobacco market demand forecast, multiple reasons have been showing impact on each other like industry's own growth, seasonal fluctuations, policy impact etc. It's an industry challenge for many tobacco companies to provide rational, reliable and accurate prediction of sales trends. This essay, based on Zhejiang Tobacco's actual needs, established a hybrid tobacco sales forecast model based on Autoregressive Moving Average (ARMA) algorithm as well as taking seasonal fluctuations and policy influence into account. This model is validated in the simulation process.

Keywords—ARMA; autoregressive; moving average; policy influence; seasonal forecasting; tobacco; sales forecast

I. INTRODUCTION

Since Joining WTO, China State Tobacco Monopoly Administration has required "By straightening out the relationship of property rights, reforming industrial and commercial management system to separate the management of those two, therefore establishing a nation unified yet hierarchically managed tobacco sales network, enabling all industrial companies to compete equally" [1] to reinforce the competence of China tobacco industry from 2003. Tremendous changes have happened in the manufacturing and selling process of tobacco industry. Provincial and municipal entities have gradually formed the new tobacco circulation model with separated industrial and commercial companies, the former is responsible for manufacturing according to orders from different tobacco monopoly bureaus and the delivering process while the latter one is responsible for organizing retail stores to making orders based on their order plans as well as the delivering process to those retailers.

Under the industrial and commercial separation system, commercial companies are in charge of retail channels and requesting supplies from different industrial companies, which, as the main brand owner and manufacturing entity, need to comprehensively co-ordinate the requirements of domestic provinces as well as export needs. Therefore, industrial companies should be capable of advanced procurement of raw materials, arrangement of manufacturing and allocation of cigarettes packages, effective control of inventory level and

cost as well as fully satisfying the needs of customers and retailers in order to improve the overall efficiency of tobacco supply chain including procurement, manufacture and delivery.

Under the new model of "organizing supply by order", ordering requirements are usually very urgent or with rapid change, thus calling for higher capability of industrial companies in balancing order requirement and their own production capability, raw material supply and inventory level. To be competitive in the market, to improve brand awareness and market share, it's not only a challenge for commercial companies but also for industrial companies to accurately grasp market change and continuously improve sales forecast.

Compared to other traditional industries, due to the existence of a certain degree of planning and monopoly system and a relatively stable demand of smokers, there's less disorderly competition in tobacco industry. [2] In view of such features, the general time series method can be applied to simply and effectively predict sales demand from the long-term perspective.

Cigarette sales forecast will also be influenced by local market's population size, demographic structure, income level, local policies, brand tendency etc. [3] This essay focuses on the application of total tobacco sales forecast, thus local factors are excluded here.

In view of this, this essay proposes average tobacco sales forecast based on Autoregressive Moving Average (ARMA) algorithm while taking major festivals into account by comparing the historical data to determine the seasonal coefficient of each festival and embodying those coefficients in the model. Compared with the traditional ARMA model which requires manual setting of parameters and could not automatically adjust itself when deviation occurs, Zhejiang Tobacco conducted the comparison analysis of forecasting data and real sales data to adjust the parameters and optimized the model automatically to obtain more accurate sales forecast. According to Zhejiang Tobacco's real business situation, the automatic balancing of the production planning between joint factories and its own factories based on annual agreement with each one could make good use of manufacture capability, lower inventory level and increase delivery accuracy. The feasibility and validity of the model are proved by the comparison of the simulation data from April to July in 2015.

II. TOBACCO SALES FORECAST MODEL

Tobacco market not only aligns with general growth trends due to economic growth but also in line with certain degree of seasonal feature. Therefore, long-term trends and short-term seasonal impact have both been taken into consideration.

In the process of calculation, the moving average method is used to process the historical data to calculate the moving average M_t of sales by eliminating the influence of seasonal factors. Then we should determine the long-term growth trend factor Y_t of each month, which could be obtained by calculating the weighted average of the historical data by autoregressive method and determining the parameters according to historical data. Trend factor for each month in the future will then be obtained. Finally, we use the exponential smoothing method to determine the seasonal factor T_t , which is used to modify the effect of major festivals on tobacco sales.

By comparing the data before and after the policies with a significant impact on tobacco and excluding the growth trend factor, we could calculate and obtain the impact of the policy factor P_t in percent.

Forecast Formula:

$$X_t = Y_t * T_t * P_t \quad (1)$$

A. Moving Average Method

To eliminate the influence of seasonal factors, we use the natural annual (12-month) period to calculate the average demand for each month, that is, the monthly sales volume is calculated by the moving average of this month and previous 11 months which will then be used as trend analysis factor.

$$M_t = (Y_{t-11} + Y_{t-10} + \dots + Y_t) / 12 \quad (2)$$

B. Weighted Historical Data of ARMA Model

ARMA model is a commonly used stochastic time series analysis model, founded by Box and Jenkins, also known as B-J method. The basic idea is that some time series are one time series variables dependent on time. Although the individual sequence values of the time series have uncertainties, the changes of the whole sequence do have certain regularity which could be described by relevant mathematical model. Through the analysis and study of the mathematical model, we can understand the structure and characteristics of time series more precisely, and reach the optimal prediction under the minimum variance. [4]

Time series analysis is usually not based on the basis of economic relations reflected in economic theory, but "let the data speak for themselves." Y_t can be explained by its own hysteresis and random error terms. [5]

ARMA model is a prediction model based on historical data, so in the modeling stage, the pre-processing of historical data is the basis of the model's reliability. In ARMA model, hysteresis is infinite in infinite distribution hysteresis model while the

sample observations are always limited, so it is impossible to directly conduct the estimation. To proceed smoothly, some constraints or assumptions must be applied to transform the structure of the model. The Koyck transform is a representative method.

A time series P-order regression model, Y_t is a linear combination of the first P series and a function of the error terms, see as follows: [6]

$$Y_t = \beta_0 + \beta_1 * M_{t-1} + \beta_2 * M_{t-2} + \beta_3 * M_{t-3} + \dots + \beta_P * M_{t-P} + e_t \quad (3)$$

β_0 is a constant term, $\beta_1 \dots \beta_P$ is parameter of the regression model, e_t is a white noise with mean 0 and variance σ .

The lag-explanatory variables can be assumed in the Kuhnian transformation, and the effects on the interpreted variables are attenuated by the geometric progression as the lag time increases. That is, the attenuation of the lag coefficient obeys a geometric series whose common ratio is less than one [6].

$$\beta_i = \beta_0 * \lambda^i, 0 < \lambda < 1, i=0,1,2,\dots \quad (4)$$

β_i is usually referred to as the distribution delay decay rate, the value closer to zero, the faster the decay rate (Figure I)

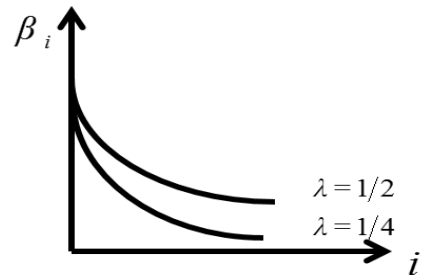


FIGURE I. HYSTERESIS STRUCTURE DECAYING BY GEOMETRIC PROGRESSION

In this essay, the older the sales data, the weaker the reference for the future. Compared to the traditional "generalizations", data transformed through Kuyke are more valuable and referable.

C. Seasonal Exponential Smoothing

The seasonal variation prediction method is a special method for forecasting the time series including seasonal variation. The first thing is to study the seasonal variation pattern of time series. In the study of seasonal changes, long-term trends are normally contained. Thus, the predicted data of the time series can be seen as a function of the long-term trend and the seasonal index.

Certain seasonal model could be chosen based on seasonal change pattern, along with long-term trend, which could be applied in economic forecasting. Trend exclusion method is to calculate the seasonal trend by calculating the trend factor of the sequence and then calculating the seasonal index after

removing the trend factor. In contrast to direct averaging, this approach eliminates the possible inclusion of trend factors for all time periods and for the mean of the same period to find pure seasonal factors and normalize them as seasonal indices in the seasonal prediction model.

$$T_t = \frac{Y_t}{M_t} \quad (5)$$

D. Policy Factor

Policy impact is also too important to be ignored in tobacco market. After 6 years, in May 10, 2015, the state once again raised the tobacco industry consumption tax from 5% to 11%. Affected by this tax policy, localities began to raise the wholesale and retail price of cigarettes. Several months' cigarette shipments declined rapidly compared to the same period last year under such situation. If these factors are not to be considered, our prediction clearly would have a great deviation.

In the beginning phase of the policy release, it's hard to use quantitative methods to predict and measure the impact, thus qualitative prediction method, which means experts' prediction of future general nature and degree based on personal business knowledge, rich experience and comprehensive analytical capability, existing historical materials as well as various advice, is used. Commonly used qualitative prediction methods include market research, expert forecasting, Delphi and so on.

Market research: an approach that uses random sampling or typical survey to select a certain number of consumers from the candidates, conducts comprehensive analysis of consumer buying intention through publication or interview and then predicts the future demand for the goods.

Expert forecasting method: a method that takes experts as the target to obtain required information. It organizes experts in various fields to comprehensively analyze past and current problems based on professional experience and knowledge to find the pattern and predict the future growth.

Delphi method: a method that uses back-to-back communication to consult the expert group members of forecast advice and makes the final market trends predictions when the expert group forecasts tend to focus after several rounds of consultation.

Different methods should be chosen based on different policies. A synthesized result from a comprehensive analysis based on several forecasting methods could also be used.

The policy factor will be shown as a positive or negative percentage (100% positive, <100% negative) in the final market analysis. The factor has both long-term impact and short-term impact which could be divided into before and after the incident and will be reflected eventually in the forecast result.

III. SOLUTION AND APPLICATION OF MODEL

Taking the commercial cigarettes allocating (selling) data

of Zhejiang Tobacco from 2011 to 2015 as the sample and conducting the calculation according to different cigarettes classes in order to validate the model and solve the parameters.

- Sample data: Figure II and Figure III are monthly sales volume curve for first and second class cigarettes during 2012 and 2014. Obvious seasonal change could be noticed here.

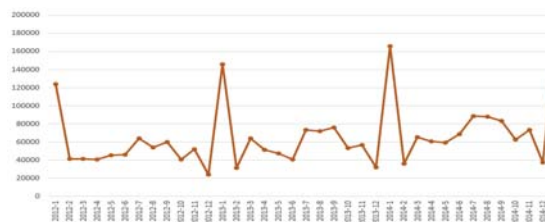


FIGURE II. FIRST CLASS CIGARETTES 2012-2014 SALES VOLUME CURVE



FIGURE III. SECOND CLASS CIGARETTES 2012-2014 SALES VOLUME CURVE

- Overall Trend Analysis: Figure IV and Figure V are monthly sales volume trend line for first and second class cigarettes during 2012 and 2014.

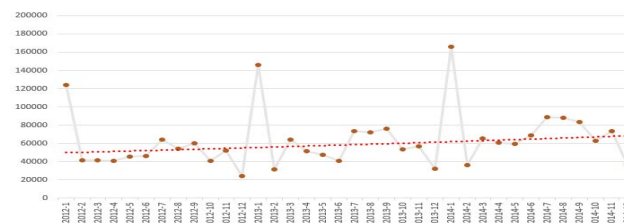


FIGURE IV. FIRST CLASS CIGARETTES 2012-2014 SALES VOLUME TREND LINE

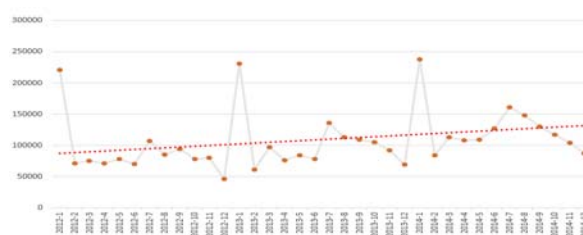


FIGURE V. SECOND CLASS CIGARETTES 2012-2014 SALES VOLUME TREND LINE

- Moving Average analysis: To eliminate the impact of seasonal fluctuations and better analyze the long term trend, monthly sales is calculated by previous 12 months' moving average number. Figure VI and VII below are monthly sales volume moving average curve for first and second class cigarettes during 2012 and 2014.

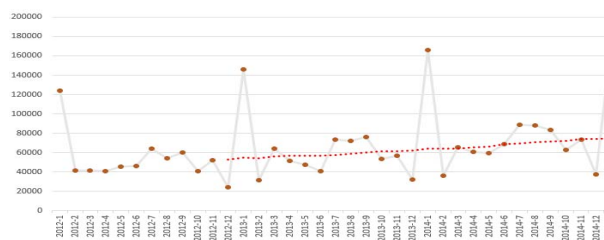


FIGURE VI. FIRST CLASS CIGARETTES 2012-2014 TREND LINE OF SALES VOLUME MOVING AVERAGE EFFICIENCY

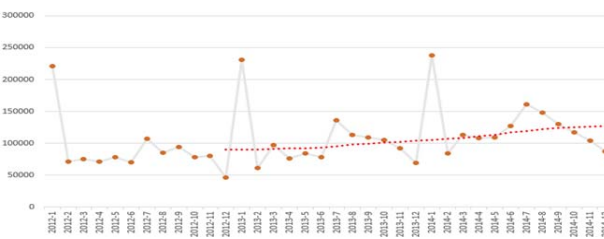


FIGURE VII. SECOND CLASS CIGARETTES 2012-2014 TREND LINE OF SALES VOLUME MOVING AVERAGE EFFICIENCY

By using the ARMA model, the sales data from 2012 to 2014 are taken as samples and the lagged term Y_{t-1} as the dependent variable and the confidence level is 98%. EXCEL is used to solve the Y_t autoregressive parameters. The regression statistics for first class cigarettes are shown in Figure VIII.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.996583078							
R Square	0.99317783							
Adjusted R Square	0.992977178							
Standard Error	698.8602989							
Observations	36							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	2417487109	2417487109	4949.751861	2.04689E-38			
Residual	34	16605794.39	488405.7174					
Total	35	2434092903						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 98.0%	Upper 98.0%
Intercept	521.1813489	896.7252599	0.622882293	0.537517762	-1179.248966	2221.631664	-1521.390208	2563.752906
X Variable 1	1.005433515	0.014290968	70.35447293	2.04689E-38	0.976390774	1.034476256	0.970547124	1.040319906

FIGURE VIII. REGRESSION STATISTICS ANALYSIS RESULT FOR FIRST CLASS CIGARETTES

According to the regression parameter table, autoregressive function with the fitting degree over 99.3% is shown below:

$$Y_t = 521.181348865888 + 1.00543351467545 * Y_{t-1} \quad (6)$$

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.997750174								
R Square	0.99550541								
Adjusted R Square	0.995373216								
Standard Error	993.5135								
Observations	36								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	1	7433271179	7433271179	7930.649444	1.69693E-41				
Residual	34	33560348.54	987069.0747						
Total	35	7466831527							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 98.0%	Upper 98.0%	
Intercept	-985.1902015	1165.395606	-0.845369758	0.403814504	-3353.559022	1383.178619	-3830.095251	1839.714848	
X Variable 1	1.024240392	0.011802817	86.72931461	1.69693E-41	1.000234181	1.048226602	0.995427949	1.053052835	

FIGURE IX. REGRESSION STATISTICS ANALYSIS RESULT FOR SECOND CLASS CIGARETTES

The regression statistics for second class cigarettes are shown in Figure IX.

According to the regression parameter table, autoregressive function with the fitting degree over 99.3% is shown below:

$$Y_t = -985.190201517704 + 1.02424039185279 * Y_{t-1} \quad (7)$$

- Seasonal variation Curve: In order to eliminate the impact of policy factors in the seasonal variation curve, we select the monthly sales volume change curve from 2011 to 2014, and analyze the seasonal fluctuation trend with the year as the analysis theme and the natural month as the dimension. Figures X and XI below indicate an obvious and regular seasonal fluctuation in the sales of both first and second class cigarettes in each month of the year.

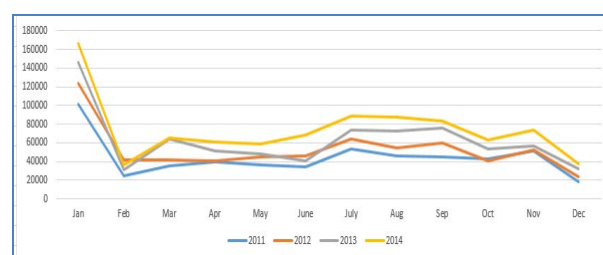


FIGURE X. FIRST CLASS CIGARETTES 2011-2014 SALES VOLUME EFFICIENCY SEASONAL VARIATION CURVE

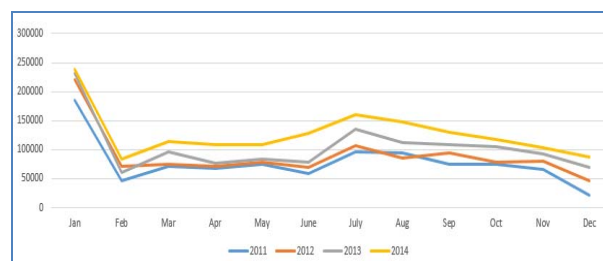


FIGURE XI. SECOND CLASS CIGARETTES 2011-2014 SALES VOLUME EFFICIENCY SEASONAL VARIATION CURVE

To solve seasonal parameters:

$$T_t = \frac{Y_t}{M_t} \quad (8)$$

According to data between 2012 and 2014, monthly index for first and second class cigarettes are listed below.

- Policy impact analysis: with the increase of tobacco consumption tax, a basic agreement of different experts is proposed in a production scheduling meeting of Zhejiang Tobacco in early May 2015. The increase of consumption tax from 5% to 11% is a substantial rise which will have a direct impact on retail price. The amount of increase is especially high for first class cigarettes which are sold with higher price.

It had caused huge impact on short term sales: before the incident, many retailers would store more stock which caused higher demand than normal situation but within a few months

after the occurrence, a sharp decline in demand for low-priced cigarettes would be seen because the price increase is not significant, and consumption amount in a short time is difficult to change greatly, besides, part of first class cigarette customers may lower the standard and choose the second class. From a long-term perspective, the overall impact of the market is neutral. Long-term impact index: 90% of first class cigarettes, 105% of second-class. Before the policy, May 2015, 150% of first class and after the policy, the index changes from June 70%, July 75%, August 80% and September 90%; short-term impact on second class cigarettes could be ignored while only long term effect needs to be considered.

TABLE I. MONTHLY INDEX FOR INDEX FOR FIRST CLASS CIGARETTES BETWEEN 2012 AND 2014

	2012	2013	2014	Average
Jan	269.81%	267.07%	260.64%	265.84%
Feb	87.21%	58.41%	56.20%	67.27%
Mar	86.63%	114.39%	101.71%	100.91%
Apr	84.27%	91.18%	93.38%	89.61%
May	93.27%	83.84%	89.48%	88.86%
June	92.80%	72.03%	100.65%	88.50%
July	126.48%	128.33%	126.90%	127.24%
Aug	105.95%	123.04%	123.87%	117.62%
Sep	114.95%	126.67%	116.89%	119.50%
Oct	78.15%	87.94%	86.67%	84.25%
Nov	99.63%	92.74%	99.75%	97.38%
Dec	45.41%	51.44%	50.57%	49.14%

TABLE II. MONTHLY INDEX FOR INDEX FOR SECOND CLASS CIGARETTES BETWEEN 2012 AND 2014

	2012	2013	2014	Average
Jan	275.53%	255.14%	227.24%	252.64%
Feb	85.81%	68.17%	78.54%	77.50%
Mar	90.11%	105.50%	104.68%	100.09%
Apr	85.66%	82.71%	97.34%	88.57%
May	94.08%	91.33%	96.89%	94.10%
June	82.82%	83.45%	108.97%	91.74%
July	125.87%	142.56%	135.22%	134.55%
Aug	100.39%	115.46%	121.11%	112.32%
Sep	108.75%	110.43%	104.72%	107.96%
Oct	90.38%	103.90%	94.00%	96.09%
Nov	90.81%	89.75%	82.67%	87.74%
Dec	51.49%	65.99%	68.09%	61.86%

IV. SIMULATION ANALYSIS

Measuring back the actual value of January to December of 2015 based on the forecasting model.

First of all, analyzing the policy factors by comparing 2015 actual sales volatility with that of 2011 to 2014. See Figure XII and XIII.



FIGURE XII. FIRST CLASS CIGARETTES SALES VOLUME COMPARISON OF 2015 AND PREVIOUS FOUR YEARS

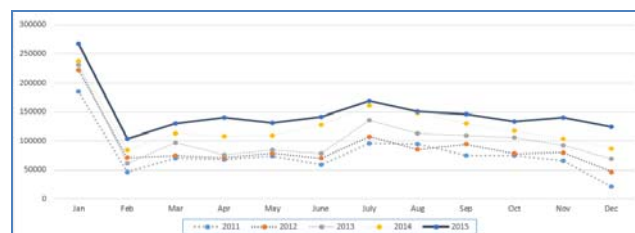


FIGURE XIII. SECOND CLASS CIGARETTES SALES VOLUME COMPARISON OF 2015 AND PREVIOUS FOUR YEARS

The result is mostly compatible with the qualitative analysis of group experts. Sales volume of first class cigarettes in May increased a lot due to stocking up before the price increase of retailers and smokers in order to reduce costs. However, there's no obvious decrease in June while the subsequent July and August plummeted, much lower than the same period in 2014. Sales in following September began to stabilize and returned to normal. There's not much impact on total second class cigarettes' sales performance. Analysis is down below in Table III and Table IV. Modeling based on data from 2012 to 2014, Y trend value, T season index and P policy index could be calculated respectively and parameters would be solved too. Then we could predict the sales number of January to December in 2015 and compare it with the actual sales data.

The average annual prediction accuracy of first and second class cigarette is 97.7% and 98.7% respectively. Because the qualitative prediction of the market psychology of the study is not enough, we mistakenly believed that first class of cigarette sales in June will be significantly reduced, so June sales forecast and actual data are quite different. The accuracy of the second class of smoke prediction will be stable. Cigarette industry is not only affected by market but also government plans, local annual tax requirements, the volume of shipments increased abnormally compared to previous years and predicted numbers.

V. CONCLUSIONS

Due to the impact of market trend, seasonality and policy on tobacco industry, a hybrid forecasting model is proposed to divide sales data into trend factor, seasonal index and policy index. The trend factors are regressed by the ARMA autoregressive moving average model. Seasonal fluctuations are balanced by seasonal factors. In addition to the quantitative forecasts, market research, expert forecasting, and Delphi methods should be used for qualitative analysis of policies to quantify the policy factors. Finally, sales forecast is completed through the combination of three factors and model parameters

are automatically optimized in the rolling prediction process through the test of actual data. In the sales forecast project of Zhejiang Tobacco, the overall forecast accuracy of the model is higher than 97% and over 80% averagely within 3 months, with a total average of 90%, representing a certain degree of reference. This model is not comprehensive enough for the policy forecasting, lacking research on market psychological analysis.

REFERENCES

- [1] *Organizing Supply by Order Business Operation Specification. State Tobacco Monopoly Administration, March, 2007.*
- [2] Wang Sen, Wu Chunming, "Discussion of Tobacco Demand Forecast Model", Modern Mechanical Press, 2007(2), pp.4-6.
- [3] Zhao Hong, "Study on and design of tobacco requirement forecast system for commercial companies", Shandong University Master Essay, March 2009.
- [4] Pan Hongyu, "Time Series Analysis", University of International Business and Economics Press, January, 2006.
- [5] Jeffrey Woodridge, "Introduction to Econometrics", Tsinghua University Press, July, 2009.
- [6] Third ed. R. H. Shumway and D.S.Stoffer, "Time Series Analysis and Its Applications with R Examples", Springer, 2011.
- [7] Stoke Watson, Introduction to Econometrics (Version 3). Gezhi Press, April, 2012.

TABLE III. COMPARISON OF PREDICTIVE AND ACTUAL SALES DATA OF 2015 FIRST CLASS CIGARETTE

MONTH	Y trend value	T season index	P policy index	PREDICT ED VALUE	ACTUALVAL UE	Residual	Error (%)
FORMER VALUE	127166.75						
Jan	128378.8938	252.64%	100%	324330.4	268117.4	-56213	-17.33%
Feb	129597.6237	77.50%	100%	100444.5	103071.2	2626.716	2.62%
Mar	130822.9756	100.09%	100%	130946.1	129565.2	-1380.88	-1.05%
Apr	132054.9855	88.57%	100%	116959.8	139881.8	22922.03	19.60%
May	133293.6896	94.10%	100%	125429.5	131403.2	5973.743	4.76%
June	134539.1242	91.74%	105%	129603.3	141872.2	12268.92	9.47%
July	135791.3258	134.55%	105%	191839.6	169491.2	-22348.4	-11.65%
Aug	137050.3313	112.32%	105%	161633.4	150812.6	-10820.8	-6.69%
Sep	138316.1777	107.96%	105%	156799.2	145648	-11151.2	-7.11%
Oct	139588.902	96.09%	105%	140841.8	133146.2	-7695.59	-5.46%
Nov	140868.5417	87.74%	105%	129783.8	140605.2	10821.35	8.34%
Dec	142155.1343	61.86%	105%	92327.94	124352.8	32024.86	34.69%
Total	1622457.705			1800939	1777967	-22972.3	-1.28%

TABLE IV. COMPARISON OF PREDICTIVE AND ACTUAL SALES DATA OF 2015 SECOND CLASS CIGARETTE

MONTH	Y trend value	T season index	P policy index	PREDICTED VALUE	ACTUALV ALUE	Residual	Error (%)
FORMER VALUE	74157.71667						
Jan	75081.835	265.84%	100%	199599.6	173802	-25798	-12.92%
Feb	76010.975	67.27%	100%	51135.53	59630.8	8495.27	16.61%
Mar	76945.163	100.91%	100%	77646.18	83170.2	5524.02	7.11%
Apr	77884.427	89.61%	100%	69793.17	62631.6	-7161.6	-10.26%
May	78828.794	88.86%	150%	105075.4	96534.6	-8540.8	-8.13%
June	79778.293	88.50%	70%	49420.3	74372.6	24952.3	50.49%
July	80732.951	127.24%	75%	77041.8	67764.4	-9277.4	-12.04%
Aug	81692.796	117.62%	80%	76869.38	71688.8	-5180.6	-6.74%
Sep	82657.856	119.50%	90%	88899.6	83972.2	-4927.4	-5.54%
Oct	83628.16	84.25%	90%	63413.77	58054.2	-5359.6	-8.45%
Nov	84603.737	97.38%	90%	74145.27	63510.6	-10635	-14.34%
Dec	84603.737	49.14%	90%	37419.33	53687.8	16268.5	43.48%
Total	962448.72			970459.3	948820	-21640	-2.23%