Prediction to Industrial Added Value Based on Holt-Winters Model and ARIMA Model

Lei Kuang, Chengyu Lin, Wenwen Wang and Xi Fang[*]
School of Science Wuhan University of Technology Wuhan, China 430070
[*]Corresponding author

Abstract—**Industrial added value (IAV) is an important indicator to measure a country's industrial development level, at the same time, is also a core indicator of national economy accounting system. The paper respectively using ARIMA and Holt - Winters, the two time series forecasting model, to fit the monthly data of industrial added value of Hubei province in 2009-2014 and on the basis of it we forecast for next year. For ARIMA model, this paper introduces the smooth sequence processing, model identification, parameter estimation and model diagnosis and prediction, such as the key to predict; For Holt-Winters model, this paper introduces the selection of smoothing coefficient alpha, beta, gamma and initial smoothing factor, recursive calculation process, such as the key to predict, and the corresponding algorithm and the prediction model is designed. Finally use example to compare the two methods and analyze the pros and cons of two kinds of models to predict.**

Keywords-ARIMA model; holt-winters model; industrial production; the optimal smoothing coefficient

I. INTRODUCTION

Industry, is the acquisition of the raw materials and put them in the factory production into a work of the product and process. In our country, the industry as the main material production departments, occupies the largest share of GDP. Especially since the Chinese economic reform and opening up, the revitalization of industry is an important guarantee of promoting the sustainable and healthy development. In today's rapid development of global economy, industrial economy is showing the tremendous power to change the face of the whole country and society. There is no doubt that the industrial development is strategic focus in the development of national economy. Hubei province is one of the largest and most populous province in the "five" period project layout, it has a good industrial base. History repeatedly shows that Hubei province strengthen as its industry strengthen. At present, the level of Hubei province industrialization is still in mid-term stage. Since the international financial crisis, the domestic and foreign economic situation is complicated, also in the fluctuation of economic growth in our province. Grasp the regulation of economic operation requires proactive, predictive, pertinence, forecast the situation become a indispensable support for steady and healthy development of the industrial economy. How to grasp the macro level of industrial development of Hubei to ensure the stable and rapid development of industrial economy in our province, is a problem we must and necessary to solve. Therefore, in order to promote the healthy and smooth development of the province's industrial economy, to accurately predict the province IAV is particularly important

On the basis of the industrial production data from Hubei Economic and Information Technology Committee of each month from 2009 to 2014. According to the trend and seasonal characteristics of the output sequence, this paper forecast the Hubei province IAV based on Holt - Winters model and ARIMA model, and further verify its effectiveness and the pros and cons of the two models.

II. MODEL

A. The Basic Idea of ARIMA Model

Autoregressive Integrated Moving Average Model, referred to as ARIMA, is by the Mr. Box and Mr. Jenkins in the early 70 s put forward a famous time series prediction method, so they are called Box-Jenkins Model method. In fact, the expansion of the autoregressive moving average model, referred to ARMA. The basic idea is to did not meet the requirements of stability after the d order by phase difference the time sequence of smooth, then using the ARMA (p, q) modeling of the stationary series, then through the inverse transformation to get the original sequence. Expression of model are as follows:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + L + \varphi_p X_{t-p} + \varepsilon - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - L \ \theta_q \varepsilon_{t-q}$$

Among then, $\varphi_1, \varphi_2, \ldots \varphi_p$ is autoregressive coefficient; p is autoregressive order; $\theta_1, \theta_2, \ldots, \theta_p$ is moving average coefficients; q is moving average order; $\{\varepsilon\}$ is s white noise sequence, usually the model can be expressed as ARIMA (p, d, q), d for the difference among them.

B. The Basic Idea of Holt-Winters Model

The basic idea of Holt-Winters model is making a decomposition research of the linear trend, seasonal fluctuation and random fluctuation time series. And combine with exponential smoothing method, for the long-term trend (Ut), the trend of increment (Bt) and the seasonal change (Ft) make an estimate respectively. Then set up the forecast model

to extrapolation. The method is a forecast method which takes linear trend, seasonal changes and irregular changes of time series break down, combining with exponential smoothing. It can be dealt with trend and seasonal change at the same time. As well, the influence of random fluctuations can be appropriately filtered out [1]. So this model is particulaly suitable for the trend and seasonal variation of time series prediction research. Holt - Winters model includes none-season model, additive model and multiplication model. In this paper, we use Holt - Winters multiplication model to forecast.

Holt - Winters multiplication model consists of the following three basic formula [2].

$$U_t = \alpha \frac{d_t}{F_{t-L}} + (1-\alpha)(U_{t-1} + b_{t-1}) \tag{1}$$

$$b_t = \beta(U_t - U_{t-1}) + (1-\beta)b_{t-1} \tag{2}$$

$$F_t = \gamma \frac{d_t}{U_t} + (1-\gamma)F_{t-L} \tag{3}$$

Among them: alpha, beta, gamma for smoothing coefficient, the value between the (0, 1);

Ut is stable composition, refers to remove the seasonal variation influences the time series of exponential smoothing average;

Bt is linear composition, refers to the time sequence change trend of the exponential smoothing average;

Ft expressed seasonal ingredients, refers to the exponential smoothing average of the seasonal factor;

Dt is the actual value of the current time;

L represents the length of season length or time period;

In the calculation of the first part of the exponential smoothing average Ut, we eliminate the seasonal variation effects in the dt through to the actual value of the current time dt divided by the season factor Ft - L; For the second part of Ut, add the trends of average bt-1 on the average number of Ut-1 in the first time to Further improve the accuracy of the exponential smoothing results; At the same time for both alpha and 1 - alpha weighted.

Bt is used to describe a smooth time series trends. Here using difference "Ut - Ut - 1"said the trend of increment, at the same time as a result of the existence of random disturbance, we should also have difference smoothly correction. In other words, we put this difference into the trend of incremental weighted average of the previous period, as the trend incremental estimation. Using the coefficient of beta (0<beta<1) weighted; at the same time, for previous trend of weighted value of bt-1 to weighted with 1 –beta.

Ft which equal to the seasonal index, is the ratio of the current moment of time series observations (dt) and time sequence of the exponential smoothing average (ut) . Ut includes the change of the trend, but excluding the influence of the seasonal change, by giving seasonal factor multiplied by the coefficient gamma 0<gamma< 1) ; At the same time season index Ft - L multiplied by 1 – gamma for last same season, both weighted. Based on the Holt - Winters, seasonal exponential smoothing model to predict the formula is as follows

$$f_{t+m} = (U_t + mb_t)F_{t-L+m} \tag{4}$$

Among them: m is now time to predict time distance time interval number.

C. Determine the Value of Smoothing Coefficient

Generally speaking, the smoothing coefficient alpha, beta, gamma value is set according to practical experience. They can take the same value or different values. Selecting principle is to make the predicted values and the actual value is close to or equal. This will inevitabley lead to the unreliability of the forecast due to different value of smoothing coefficient predicated difference results. This article quotes the thought of "the optimal smoothing factor"[3], which is to choose a set of smoothing coefficient value that can make recently predicted data over to minimize the sum of the squares of the value of the relative error. In the related economic situation without a dramatic change, we can think of to take the group smoothing coefficient value can make optimal prediction of this year. In software design, through exhaustive any combination of three smoothing coefficient (from 0.01 to 0.99) , IT MADE THE CORRESPONDING FORECAST, RESPECTIVELY. And through the calculation results to calculate the sum of the squared relative error of the last two years, choosing the minimum relative error value of the squares of the smoothing coefficient as "the optimal smoothing factor".

D. Determine the Initial Smoothing Factor

To determine the initial smoothing factor using the following formula:

$$\bar{d} = \sum_{i=1}^{n} d_i \tag{5}$$

$$U_0 = \sum_{i=1}^{L} d_i \tag{6}$$

$$b_0 = d_2 - d_1 \tag{7}$$

$$F_t = \frac{(\sum_{i=0}^{n/L-1} d_{i \times L+t})/(n/L-1)}{\overline{d}}$$

(8)

E. Recursive Calculation

Calculated the initial smoothing factor into the formula (1) ~ (5), and successive recursion repeatedly, each index smoothing factor are calculated respectively U_t, b_t, F_t and predicted $F_t + m$ [4].
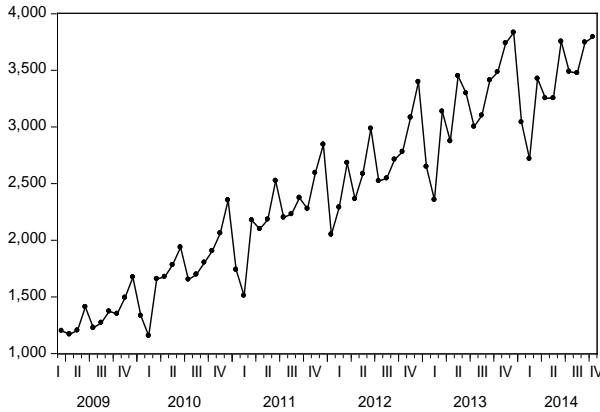
III. MODEL APPLICATION

A. ARIMA Model

FIGURE I. MARCH 2009 TO OCTOBER 2014 IN HUBEI PROVINCE INDUSTRIAL OUTPUT SEQUENCE

Figure 1 is in Hubei province in March 2009 to October 2014 IAV of historical time series y (t) graphics, roughly as you can see, the sequence is not only a trend, also has the period length of 12 seasonal effect. In order to eliminate the trend component, we first to y(t) for the order by phase difference, namely

$$y1(t) = \nabla y(t) = y(t) - y(t-1)$$

Through EViews software available difference sequence of linear graph (see figure 2)
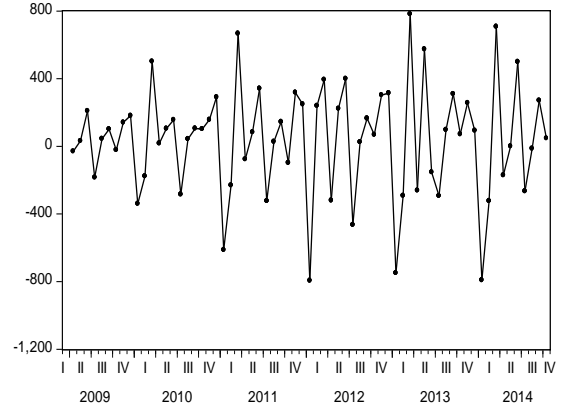

FIGURE II. DIFFERENCE SEQUENCE OF THE LINEAR GRAPH

Can be seen in figure 2, the difference after the sequence $y1(t)$ of the mean is steady near zero, namely through the first order difference $y(t)$, the original sequence trend basically eliminated.

The next season to the first-order difference $y(t)$ sequence difference, there is

$$y2(t) = y1(t) - y1(t-12)$$

In the SPSS runs test was carried out on the season difference sequence $y2(t)$ (in the average of data division),

Smooth sequences $y2(t)$ have been found. Sequence $y2(t)$ of the sample mean m $\approx$ -2.6987, so to zero mean seasonal difference sequence $y2(t)$, namely

$$y3(t) = y2(t) + 2.6987$$

Through EViews available sequence of sample autocorrelation coefficient of the graphics as shown in figure 3:

Sample: 2010M04 2014M10
Included observations: 55

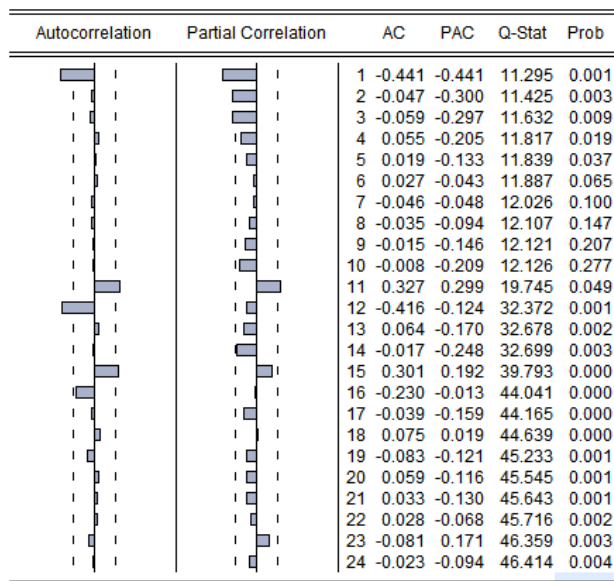| Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.441	-0.441	11.295	0.001
		2	-0.047	-0.300	11.425	0.003
		3	-0.059	-0.297	11.632	0.009
		4	0.055	-0.205	11.817	0.019
		5	0.019	-0.133	11.839	0.037
		6	0.027	-0.043	11.887	0.065
		7	-0.046	-0.048	12.026	0.100
		8	-0.035	-0.094	12.107	0.147
		9	-0.015	-0.146	12.121	0.207
		10	-0.008	-0.209	12.126	0.277
		11	0.327	0.299	19.745	0.049
		12	-0.416	-0.124	32.372	0.001
		13	0.064	-0.170	32.678	0.002
		14	-0.017	-0.248	32.699	0.003
		15	0.301	0.192	39.793	0.000
		16	-0.230	-0.013	44.041	0.000
		17	-0.039	-0.159	44.165	0.000
		18	0.075	0.019	44.639	0.000
		19	-0.083	-0.121	45.233	0.001
		20	0.059	-0.116	45.545	0.001
		21	0.033	-0.130	45.643	0.001
		22	0.028	-0.068	45.716	0.002
		23	-0.081	0.171	46.359	0.003
		24	-0.023	-0.094	46.414	0.004

FIGURE III. SEQUENCE Y3(T) OF SAMPLE AUTOCORRELATION FUNCTION DIAGRAM

Can be seen from figure 3, the sample autocorrelation coefficient after step 1 is truncated, partial autocorrelation coefficient is trailing, therefore available MA (1) the sequence y3(t). Use EViews software on sequence y3(t), MA (1) model parameters test of concrete numerical value as shown in figure 4.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.975505	0.026834	-36.35298	0.0000
R-squared	0.428538	Mean dependent var		-2.73E-05
Adjusted R-squared	0.428538	S.D. dependent var		214.7997
S.E. of regression	162.3780	Akaike info criterion		13.03575
Sum squared resid	1423798.	Schwarz criterion		13.07224
Log likelihood	-357.4830	Hannan-Quinn criter.		13.04986
Durbin-Watson stat	1.791334			
Inverted MA Roots	.98			

FIGURE IV. SEQUENCE Y3 (T) MA (1) MODEL PARAMETERS OF THE TEST RESULTS

Can be seen from the associated probability and unit root, the use of MA (1) of sample time series y3(t) model fitting is appropriate.

In conclusion, in Hubei province in March 2009 to October 2014 monthly data y(t) model for industrial production

$$(1-B^{12})(1-B)y_t = -2.6987 + \varepsilon_t + 0.9755\varepsilon_{t-1}$$

In this model, the months in 2014 in Hubei province industrial added value of prediction results compared with the actual values are shown in table 1.

TABLE I. 2014 INDUSTRIAL OUTPUT PREDICTION RESULTS

Month	Actual value /billion	Predicted value/billion		Forcast error/%	
		H-W	ARIMA	H-W	ARIMA
Jan	3041.60	2994.55	2954.27	1.55	2.87
Feb	2718.50	2754.23	2840.23	1.31	4.48
Mar	3425.60	3696.8	3702.47	7.92	8.08
Apr	3254.20	3456.57	3407.57	6.22	4.71
May	3254.20	3690.66	3684.32	13.41	13.22
Jun	3754.10	3996.8	4002.61	6.46	6.62
July	3488.50	3425.85	3588.37	1.8	2.86
Aug	3476.00	3474.7	3626.53	0.04	4.33
Sept	3747.00	3696.7	3857.92	1.34	2.96
Oct	3794.00	3698.05	3452.34	2.53	9.01

B. Holt-Winters Model

This paper selects Hubei province industrial output data from March 2009 to December 2013 as sample data set. We built up a IAV prediction model based on Holt-Winters multiplication model. EVIEWS 8.0 as a platform to predict the IAV of Hubei province and verify model prediction effect.

By the original sequence diagram (figure 1) analysis, industrial production of Hubei province presents clear trend and seasonal change. Thus suitable for Holt - Winters multiplication model to forecast.

When determining the smoothing coefficient alpha, beta, gamma, we adopted the thought of "the optimal smoothing factor" in literature[5], using the function of EVIEWS that can estimate minimize the error sum of squares of parameter value. When alpha is equal to 0.06, the beta is equal to 0.08, gamma is equal to 0.01, we get the sum of squared residuals is 876834.7. Under this parameter, Holt - Winters prediction model for Hubei province in2014 IAV of the prediction results compared with the actual values are shown in table 1.

IV. CONCLUSION AND SUGGESTION

By table 1 , analysis of ARIMA model prediction error is 13.22%, the biggest of the smallest prediction error is 2.86%, average error is 5.91%; Holt - Winters model's biggest prediction error is 13.41%, the smallest prediction error is 0.04%, average error is 4.258%, so the Holt - Winters model is better than that of ARIMA model.

By the contrastive analysis of the predicted results with the actual industrial production, the biggest prediction error is 13.41%, the smallest prediction error is 0.04%, and most of the prediction error is less than 5%, indicating that the prediction model has certain scientific.

Due to the above data is adopted by the Hubei province monthly total industrial output data, and is affected by the financial crisis, China's slowing economic growth in 2009. On the other hand, prediction methods used in this paper is based on sequence itself, regardless of the other economic and social factors, so the prediction of the maximum error is slightly large.in order to improve the accuracy of the predication, when departments of provincial industrial added value do short-term prediction, should be appropriate to consider the factors that influence the industrial added value fluctuation.

REFERENCES

[1] WEI Cun-ping,"A Discussion on the Forecast of the Air Traffic Performed, " CIUIL aviation economics & technology, vol.8, pp.41-43, 2000

[2] ZHAO Yi-fei, WANG Hong-yong, "A Model for Short-term Prediction of Air Traffic Flow with Holt-Winters Model," Journal of Transportation Engineering and Information,vol.4,pp.23-28,May 2007.

[3] ZHAO Yi-fei, WANG Hong-yong, ZHANG Liang,"Using Holt-Winters Model to Forecast Air Transportation Turnover Volume," journal of civil aviation university of china,vol.25,pp.1-3,April 2007.

[4] WAN Jie,"ARIMA model and Holt - Winters model In the railway passenger turnover forecast,"Railway Transport and Economy,vol.32,pp.91-94,2010.

[5] TONG Ming-rong,XUE Heng-xin,LIN Lin,"Study on the Forecast of Railway Freight Traffic Volume Based on HoIt-Winter Model,"railway transport and economy,pp.81-86,2007.

[6] ZHANG Xiao-qing,"The discussion about linear season addition trend prediction model," Statistics and Decision,pp.22-23,2005.

[7] GAYNOC P E, KIRKPATRICK R C. Introducion to time-series modeling and forecasting in business and economics[M]. New York:Mc Graw-Hill, 1994.

[8] WANG Qing-rong,"Combination for ecast based on neural network with Holt - Winters of railway freight volume, " Journal of Lanzhou Jiaotong University,vol.29,pp.122-125,August 2010.