International Review of Social Sciences and Humanities
Vol. 10, No. 1 (2015), pp. 46-55
www.irssh.com
ISSN 2248-9010 (Online), ISSN 2250-0715 (Print)

ARMA Forecasting Model and Its Application Based on Sales Stock of Lead Refining Enterprise under the Background of Strategic Emerging Industries

BU Hua-bai
(Corresponding Author)
Department of Economics and Management
Hengyang Normal University, Hengyang, P.R. China, 421008
E-mail: buhuabai@163.com

BU Shi-zhen
Department of Economics and Management
Hengyang Normal University, Hengyang, P.R. China, 421008
E-mail: bushizhen@163.com

(Received: 3-7-15 / Accepted: 26-9-15)

Abstract

This paper analyzes the mathematical methods in building the ARMA model, and builds an ARMA Forecasting Model of Lead Refining Enterprise Production and Sales Stock under Low-Carbon Economy, The test results show, ARMA (4, 0) model is a better short-term forecasting model of forecasting Lead Refining Enterprise Production and Sales Stock. Finally, the article also uses the model to forecast Chinese refined lead marketing stock in the next five years ,and predicted provide three policy recommendations according to the testing results, Seriously implement the guidelines of lead resource protection, require the lead industrial enterprises to meet the needs of domestic lead resource as the development goals. And Eliminate backward production within limited time, avoid over-exploitation of lead resources, and improve the efficiency of the lead resource use, in order to promote low-carbon sustainable development of China's refined lead market.

Keywords: Application, ARMA model, Chinese lead refineries, Sales forecasting.

1. Introduction

Because China's lead industry is in the upstream of much other industry, its development status and level will directly influence other industries, such as the construction industry, automobile industry, transportation industry, military industry and chemical industry and so on, the study on the development of lead industry has been a key research of theoretical and practical industry. At the same time, the lead industry belongs to the traditional "high energy consumption, high emission, high pollution" industry, in the face of the impact of the tide of the of low carbon economic development, research of China's lead industry's low carbon development is still an urgent problem to be solved in order to develop the new low carbon lead industry. As everyone knows, under the background of the development of low carbon economy, the market production and sales of lead products (production and consumption difference) is not only directly influenced by lead industry's structural adjustment and lead products' market fluctuation, but also indirectly influenced by other countries' laws and regulations and the development level of the social economy. Quantitative prediction of China's refined lead market and sales stock is an important decision- making tool for the scientific development of China's lead industry. There are many prediction methods and tools, such as the grey forecasting method, support vector machine forecast method, neural network forecasting method, etc. This paper chooses the ARMA model which is based on time series to forecast. Therefore, this article has first analyzed the mathematical methods in the construction of ARMA model, and then examines the feasibility of ARMA model which predict market and sales stock based on the sequences of the historical data, finally established the ARMA model and predicted the market and sales stock of the Chinese lead refining industry in the coming 5 years.

2. Mathematical Method of the ARMA Model

2.1 Mathematical Expression of the ARMA Model

Refer to relevant literature, we can see that ARMA (Auto-Regressive and Moving Average Model) is an important method to study time order, which is first brought forward by an American statistician G.E.P.Box and G.M.Jenkins in 1970, It was a mathematically mature random time series forecasting model, which was initially only for stationary time series prediction, after the further improvement by Box and Jenkins (1976), finally obtained no stationary time series ARMA model which can be converted to a smooth process through one or more difference. AR (auto-regressive model) and MA (moving average model) are two special forms of ARMA (mixed model), and its mathematics expression can be derived according to the following path.

Suppose $x_1, x_2, \ldots, x_k$ are time series that is composed by extrinsic factors that affect a certain observation variable Y, and Y is the observed values of the prediction object, and e is error, then the regression equation is:

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k + e$$

Meantime, the prediction variable Yt can also be influenced by its own internal factors, the variation rule is:

$$Yt = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + e_t$$

Due to the error term et has different dependencies relationship with xt-n (n = 1, 2, p) in different periods, this dependency is expressed by:

$$e_t = \beta_0 + \beta_1 e_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_q x_{t-q} + \mu_t$$

Sorting out the above equation, we can get the mathematical expression of ARMA (p, q) model:

$$Yt = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + \beta_0 + \beta_1 e_{t-1} + \beta_2 x_{t-2} + \ldots + \beta_q x_{t-q} + \mu_t$$

Among them, p is the auto-regressive, q is moving average, index $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_k$ are auto-regressive number, $\beta_0, \beta_1, \beta_2, \ldots, \beta_q$ are moving average number, is the unvalued index of this model, and μ_t is a random item.

2.2 Parameter Estimation Method in the Construction of ARMA Model

ARMA model parameter estimation can be divided into the AR (p) model and MA model. AR (p) model parameter estimation method includes Yule-Walker estimation, least squares estimation and maximum likelihood estimation. MA model parameter estimation method includes the inverse correlation function method, square estimation and the new interest rate estimation method. In this paper, the AR (p) model's parameters were estimated by the Yule-Walker method, MA model's parameters were estimated by the inverse correlation function method. The specific methods are discussed as follows:

2.2.1 Use Yule-Walker to Estimate the Parameter of AR (p) Model

Regard α_p as auto-regression parameter. $\{\gamma_k\}$ As auto-covariance function, and α_p is only decided by $\{\gamma_k\}$, their identified relationship can be expressed as follows:

σ^2, the variance of the white noise, is decided by $\sigma^2 = \gamma_0 - \gamma_p^T \alpha_p$ .because sequence $\{x_1, x_2, \cdots, x_N\}$ is composed by historical observed data, it is given value. So when $N > p$, we can generate function $y_t = x_t - \overline{x}_N, t = 1 \sim N$ and $\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-k} y_j y_{j+k},\ k = 0,1,\cdots,p$ .If two or more value are not equal, then it can be proved that $\hat{\Gamma}_p$ is positive definite. Solve it; we can get the unique solution.

2.2.2 The Estimation of the Inverse Correlation Function of MA (q) Model's Parameter

Suppose $\hat{X}_1 = 0$, $\hat{X}_{k+1} = L(X_{k+1} | X_k, \cdots, X_1)$, then sample information is

$\hat{\varepsilon}_{k+1} = X_{k+1} - L(X_{k+1} | X_k, \cdots, X_1)$, we can forecast variance $\nu_k = \mathrm{E}\hat{\varepsilon}_{k+1}^2$.

Because $\hat{X}_{m+1} = \sum_{j=1}^{q} \theta_{m,j} \hat{\varepsilon}_{m+1-j},\ m \ge q$ and $\theta_{m,j}$ can be obtained by recurrence, then if m is relatively big, we can get the estimated value ε_m of the new information by $\hat{\varepsilon}_m = X_m - \hat{X}_m$.

Based on the above calculated result, we can get the bigger t 's approximate MA (q) model, its specific expression is as follow:

$$X_t \approx \hat{\varepsilon}_t + \sum_{j=1}^{q} b_j \hat{\varepsilon}_{t-j} = X_t - \hat{X}_t + \sum_{j=1}^{q} b_j \hat{\varepsilon}_{t-j} .$$

Based on the proximate MA (q) model, we can get:

$$\hat{X}_t \approx \sum_{j=1}^{q} b_j \hat{\varepsilon}_{t-j}\ ,\ \hat{X}_{m+1} = \sum_{j=1}^{q} \theta_{m,j} \hat{\varepsilon}_{m-j}$$

After calculating, we can get a reasonable parameter estimation of the MA model:

$\hat{b}_j = \theta_{m,j}$, $j = 1 \sim q$, $\hat{\sigma}^2 = v_m$

2.3 The Determination Method of the Order Number in the Construction of ARMA Model

The order number p and q in the construction of ARMA model can be determined. Use AR (p) model's order determination to determinate p.Use MA (q) model's order determination to determinate q.

2.3.1 The Order Determination Method of AR (p) Model

The order determination method of AR (p) model is built on the basis of theorem 1 and theorem 2.As for the verification of these two theorems; we can refer to other relevant document.

Theorem 1: *If* $\varepsilon_t \sim \mathrm{WN}(0,\sigma^2)$ *is independent identical distributed in the AR (p), then as long as k>p,* $\lim_{N\to\infty}\hat{\alpha}_{k,j} = \begin{cases}\alpha_j, & j \le p \\ 0, & j > p\end{cases}$*.In order to check* $H_0 : \alpha_{k,k} = 0$*, we can use limiting distribution* $\hat{\alpha}_{k,k} - \alpha_{k,k}$.

Theorem 2: *If* $\varepsilon_t \sim \mathrm{WN}(0,\sigma^2)$ *is independent identical distributed in the AR (p), and* $\mathrm{E}\varepsilon_t^4 < \infty$ *,and if it is certain that k>p, then*

$$\sqrt{N}(\hat{\alpha}_{k,1} - \alpha_{k,1}, \cdots, \hat{\alpha}_{k,k} - \alpha_{k,k})^T \xrightarrow{依分布} N(0, \sigma^2 \boldsymbol{\Gamma}_k^{-1})$$

$$\sqrt{N}(\hat{\alpha}_{k,1} - \alpha_{k,1}, \cdots, \hat{\alpha}_{k,k} - \alpha_{k,k})^T \xrightarrow{依分布} N(0, \sigma^2 \boldsymbol{\Gamma}_k^{-1})$$

To k>p, it can prove that $\sqrt{N}\,\hat{\alpha}_{k,k} \xrightarrow{依分布} N(0,1)$, so there is 95% probability that $\hat{\alpha}_{k,k}$ is in $(-1.96/\sqrt{N}, 1.96/\sqrt{N})$, so there is a higher probability to choose p' s estimation $\hat{p} = \sup\{ j : |\hat{\alpha}_{j,j}| > \frac{1.96}{\sqrt{N}}, 1 \le j \le k \approx 10\}$

If partial correlation coefficient $\alpha_{k,k} \overset{k=\hat{p}}{\neq} 0$, $\alpha_{k,k} \overset{k>\hat{p}}{\approx} 0$, then the model's order number is $p = \hat{p}$, its calculating formula is as follow:

$$\begin{bmatrix}\hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \vdots \\ \hat{\gamma}_k\end{bmatrix} = \begin{bmatrix}\hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0\end{bmatrix}\begin{bmatrix}\alpha_{k,1} \\ \alpha_{k,2} \\ \vdots \\ \alpha_{k,k}\end{bmatrix}$$

Among them, $[\alpha_{k,1} \quad \alpha_{k,2} \quad \cdots \quad \alpha_{k,k}]^T = [\alpha_1 \quad \ldots \quad \alpha_p \quad 0 \quad \ldots \quad 0]^T$, $\{\gamma_k\}$ is auto-covariance function.

2.3.2 The Determination Method of the Order Number of MA (q) Model

The determination MA(q)model's order number can be achieved by the following steps: first, calculate $\hat{q}$,its value can be got by the following method: $\hat{q} = \{\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0\}$;second, use AIC to determine MA(q)model's order number, its calculating procedure is as follows: first suppose the upper limit of q, Q_0 has already been given, then calculate respectively the $\hat{\sigma}_m^2$ of MA(m) and $\mathrm{AIC}(m) = \ln(\hat{\sigma}_m^2) + 2m / N$ when m=0,1,2,3,…,Q_0, . At last get the smallest m of the minimal value by comparison. This smallest m is the estimation of q.

3. The Construction of Chinese Lead Refinery Enterprises' Production and Sales Stock Prediction ARMA

The construction of Chinese lead refinery enterprises' production and sales stock prediction ARMA model. This paper is guided by the method of Box－Jenkins（1976）to construct ARMA model, dividing the construction of ARMA model into four procedures: Data preprocessing, model recognition, parameter estimation, model testing. Combine the observed historical data into time series; get the best ARMA model after running and testing them. Then, predictive variable can be predicted.

3.1 Data Resource and Preprocessing

The data regard the different value of 2005--2014 Chinese lead industry's lead refinery market stock as the analytic target, using software Reviews to fit the model. Table 1 has listed 2005--2014 Chinese refined lead stock.

Table 1: Change of Chinese refined lead stock in 2005-2014

		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Refined lead	**Production**	1187	1323	1555	1941	2389	2714	2755	3141	3679	4191
	Consumption	721	954	1181	1509	1981	2214	2533	3075	3661	4199
	Stock	477	365	359	425	416	501	234	66	334	-107

(Data resource: H.ZSG, WRMS, southwestern security research center, the information of the author)

Analyze the data that use 2005～2014 Chinese lead industry's lead refinery production and sales stock as a model. First, get the natural logarithm from the lead refinery production and sales stock series, then eliminate the heteroskedasticity, Next carry out first-order difference to the new series, eliminate the tendency, making the data resource steady.

3.2 The Recognition and Selection of the Model

Table 2 is the relevant testing results of the Chinese lead refinery enterprises’ production and sales stock series. According to the analysis result, we can find that the first-order lag, second-order lag, and third-order lag's partial autocorrelation coefficient is twice more than the standard. And others are all trailing and in the double standard deviation. So it is proper to use this series to build ARMA (p, q) model.

Table 2: Relevant Testing of the Chinese Lead Refinery Enterprises Productions and Sales Stock Series in 2005-2014

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |**** |*	. |**** |*	1	0.479	0.4791	7.4019	0.006
. | . |	. ***| . |*	2	0.006	-0.271	7.4008	0.023
. | . |	. |***. |*	3	0.029	0.218	7.4071	0.057
. **| . |*	.**| . |*	4	-0.102	-0.317	7.8004	0.095
. ***| . |*	. *| . |*	5	-0.227	0.009	10.532	0.079
. **| . |*	. **| . |*	6	-0.149	-0.068	10.695	0.101
. | . |	. *| . |*	7	-0.041	0.026	10.891	0.150
. **| . |*	. **| . |*	8	-0.112	-0.158	10.988	0.174

Sample: 2005- 2014; included observations: 10

Based on the relevant data in the above table, through a gradual attempt from low level to high level, we can judge Chinese lead refinery enterprises' production and sales stock as fourth-order auto-regressive process. The best form of the ARMA model is verified to be ARMA (4, 0).

In addition, by calculating the value of AIC and SIC (as table3), we can also confirm the exponent number of the model. The value of AIC and SIC of ARMA (4, 0) is minimal; it also shows that ARMA (4, 0) can well be similar to disturbance term.

Table 3: Change of AIC and SIC

	ARMA (p, q) order					
(p,q)	4,0	4,1	4,2	4,3	4,4	5,1
AIC	-1.46	-1.37	-1.45	-1.42	-1.36	-1.41
SIC	-1.38	-1.36	-1.39	-1.25	-1.21	-1.35

3.3 The Estimation of the Parameter of the Model and the Determination of its Specific Formula

Use Yule-Walke method to predict the relevant parameter, and have a test over the significance of the parameter and the randomness of the residual sequence. The results are shown in table 4:

Table 4: Parameter Prediction and Testing Result of ARMA (4, 0) Model of the Chinese Lead Refinery Enterprises Productions and Sales Stock

	Coefficient	**Std. Error**	**t-Statistic**	**Prob.**
C	377.2536	24.16398	16.00845	0.0000
AR(1)	1.501009	0.154901	9.001679	0.0000
AR(2)	-0.708199	0.170339	-4.172665	0.0004
AR(3)	0.360079	0.142983	2.497049	0.0209
AR(4)	-0.392005	0.119201	-3.279778	0.0037
R-squared	0.812977	Mean dependent var		379.7900
Adjusted R-squared	0.800191	S.D. dependent var		87.05333
S.E. of regression	40.01755	Akaike info criterion		10.40031
Sum squared resid	32978.25	Schwarz criterion		11.00132
Log likelihood	-130.0029	Hannan-Quinn criter.		10.45001
F-statistic	24.49884	Durbin-Watson stat		2.001319
Prob(F-statistic)	0.000000			

At last, get the concrete form of ARMA (4, 0) model:

$Y_t = 377.255+1.501Y_{t-1}-0.708Y_{t-2}+0.36Y_{t-3}-0.392Y_{t-4}+u_t$, $R^2 = 0.79$

3.4 The Adaptive Test of the Model

By $\{x_1, x_2, \cdots, x_N\}$, get $\hat{p}$, $(\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_p)$, $\hat{\sigma}^2$, use white noise to test residuals:

$$\hat{\varepsilon}_t = y_t - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{t-j}, \ t = \hat{p}+1 \sim N$$

If the test is accordance, then accept the Model, and apply it in predicting. And if the test is not accordance, it means there is still some useful information remained in the residual series, then re-predicts parameters, and uses other ARMA (p, q) model.

Have a white noise residual test on the random error of the obtained model The testing results are as table 4.From the testing results, we can see that the value of Q are all less than the distribution threshold of X2's, whose test value is 0.05.Because the probability values in the figure are all greater than 0.05, and the minimum is 0.092, the maximum is 285, so the random error of the model is a white noise sequence, so the model can be accepted and can be used to predict.

T**able 5:** The White Noise Testing Results of Random Error Item

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. | . |	. | . |	1	0.038	0.038	0.0517	
. | . |	. | . |	2	0.014	0.014	0.0578	
. ***| . |	. ***| . |	3	-0.208	-0.214	1.5002	
. | . |	. | . |	4	0.031	0.051	1.4947	
. **| . |	. **| . |	5	-0.201	-0.201	2.8115	0.091
. |* *. |	. |* *. |	6	0.152	0.138	3.5923	0.165
. | . |	. | . |	7	-0.052	-0.064	3.7090	0.279
. |* *. |	. | . |	8	0.105	0.031	3.9046	0.408

Sample: 2005- 2014; included observations: 10

4. The Application of the Prediction Model ARMA in Chinese Lead Refinery Enterprises' Productions and Sales Stock

Since the prediction model is appropriate, then we can use the obtained model to predict, its prediction results are as table 6:

Table 6: Prediction of Chinese Lead Productions and Sales Stock (Unit: one million tons)

Year	2015	2016	2017	2018	2019
Predictive value	93	-508	-455	-447	-103

The automobile industry has always been the major market of lead and zinc industry. Lead is mainly used in producing lea-acid battery, so the development of the future automobile market will inevitably influence the development of the lead market. With the increasing demand of the automobile market, lead and zinc industry will get a good prospect. Not only the automobile industry but also communication facilities and the IT networking products has a demand for lead, so the demand for lead-acid battery increase a lot, but at the same time the requirements for the lead-acid will be higher and higher. Therefore, from the definite aspect of the prediction analysis, it is apparent that the lead products have a wide space for the development of the market. From the quantitive prediction of Chinese lead refinery enterprises' production and sales stock prediction model ARMA,the overall trend of the lead resource is that the energy power is surplus, and the stock is relative large.

5. Suggestion for the Policy

China lead smelting capacity caused by the rapid increase of the reasons may be many, but mainly of lead smelting in the continuous upgrading of technology, industrial policy, optimize ceaselessly adjust, market competition and the growing Chinese local governments continue to promote wait for a reason. The current lead enterprises smelting capacity increases quickly continuously, is conducive to the promotion of Chinese lead industry “energy saving, emission reduction and pollution reducing”, so as to facilitate the Chinese lead smelting enterprises and low carbon development. However, this kind of smelting capacity of rapid growth if the lead yields substantial excess, would be prejudicial to lead the healthy development of the industry. In view of our country present lead the development trend of the market, in order to ensure that Chinese refined lead market low carbon development.

To seriously implement the guidelines of lead resource protection, require the lead industrial enterprises to meet the needs of domestic lead resource as the development goals. Improve the comprehensive utilization level and the efficiency level of lead resources through rational planning and utilization of domestic lead resources, not only ensure a moderate lead storage, but also avoid excessive storage which will lead to vicious competition in the market.

To achieve the low-carbon sustainable development of the lead industry, ensure the security of lead resources. As for the government, the key is to strengthen macro-control and supervisory management over the lead resource, put an end to the disordered mining and excessive mining to achieve the goal of protecting and saving the lead resources. As for the downstream enterprises of the lead industry chain, the key is to achieve the recycling use of lead resources based on network perspective, maintain the stability of lead market.

To eliminate backward production within limited time, avoid over-exploitation of lead resources, and improve the efficiency of the lead resource use. Eliminate backward production capacity of smelting lead Within the limited, develop scientific environmental protection, quality, and scale standards, higher the access threshold of lead smelting production, prevent excessive investment of lead which will induce overcapacity, inhibit the increase of lead reserve by reducing the lead production from the source, promote the upgrading and low carbon sustainable development of the lead industry.

Acknowledgements

The research of paper is supported by Philosophy and Social Science Foundation of China (12BGL005); Education Ministry Philosophy and Social Science Foundation of China ---“Embedding, the consequences of strategic positioning and growth—based on intelligent network of logistics companies” (12YJA630004).

References

[1] Box, George and Pierce, Distribution of autocorrelation in autoregressive moving average time series models, *The Journal of the American statistician*, 6(5) (2007), 23-30.
[2] C. Hualiu, The ARMA consumption model, *The Journal of China the Health Statistics*, 13(7) (2011), 322-324.
[3] J. Zongli, *Analysis Based on the ARMA Model: The Total Retail Sales in Jiangsu Province (5rd edition)*, (2002), Beijing: Higher Education Press.
[4] D. Dickey, W. Bell and R. Miller, Unit roots in time series models: Tests implications, *The Journal of American Statistician*, 40(3) (2010), 98-102.
[5] H. Fu, The ARMA model of industry growth forecasts in the life insurance, *The Journal of Statistical Sinica*, 126(5) (2009), 79-94.
[6] J. Liu, On stationary and asymptotic inference of bilinear time series models, *The Journal of Statistical Sinica*, 56(2) (1992), 479-494.
[7] Z.J. Xin and H. Dawei, Inverse autocorrelation function and its role in the ARMA model identification, *The Journal of China Health Statistics*, 16 (6) (1999), 322-324.
[8] G. Xue, Evolutionary optimization of ARMA model based polymerization /

depolymerization method and the power system load forecasting application, *Ph. D Thesis*, Beijing University of Technology, (2014).

[9] B. Huabai, Lead and zinc industry for a low carbon economy, *Ph. D Thesis*, Central South University, (2011).

[10] X. Jian, Graph model test method and ARMA model, *The Journal of Guangzhou University*, 125(4) (2013), 32-38.

[11] Z. Caiping, ARMA model and residual model application, *Ph. D Thesis*, Yanshan University, (2014).