An empirical analysis about forecasting Tmall air-conditioning sales using time series model

Yan Xia

Department of Mathematics, Ocean University of China, China

Abstract— *Time series model is a hotspot in the research of statistics. On November 11, 2015, Tmall platform's turnover was more than $91.2 billion which caused the attention of scholars both at home and abroad. So this paper aims to forecast sales of Tmall, which is helpful to the enterprises. Research methods are ARIMA model and VAR model. The first model is single-variable model and the later is multi-variable model. In the study, ARIMA model makes the sequence smooth by using two difference operation. In VAR model, five explanatory variables are transformed into one main component. By contrast, VAR model does not give detailed accurate prediction, but ARIMA model does. Therefore, single-variable time series model is more suitable for sales forecast than multi-variable model.*

Keywords— ARIMA model, VAR model, Sales forecast.

I. INTRODUCTION

In recent years, the prosperity of electronic commerce makes the enterprises pay more and more attention to marketing strategy, especially accurate sales forecast. Time series model is that using historical data related to past behavior to infer the future behavior of time sequence. So this paper aims to do an empirical analysis about sales forecast using time series model. Industry data such as sales, price, online active stores, online stores clinching a deal, online products and products clinching a deal need collecting. Use MATLAB and EVIEWS software to process the data.

The body of this paper includes four parts. The first part is introduction. The second part is model definition about ARIMA model and VAR model. The third part is modeling and forecasting sales. The last part is conclusion.

II. MODEL DEFINITION

Generally, AR(p) formula is as follow:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots \phi_p X_{t-p} + \mu_t \tag{1}$$

If the error term is white noise sequence, then the process is described as pure AR(p). On the contrary, μ_t is pure MA(q) process as follow:

$$\mu_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{2}$$

ARMA model[1] process has the model structure as follow:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{3}$$

Autoregressive integrated moving average model, ARIMA model for short, its structure is as follow:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ Ex_s \varepsilon_t = 0, \forall s < t \end{cases} \tag{4}$$

Its basic idea is that let non-stationary time series smooth by difference, so that it can use ARMA model to establish stationary time series model.

Vector autoregressive model [2], VAR model for short, its structure is as follow:

$$X_t = \delta + A_1 X_{t-1} + A_2 X_{t-2} + \dots A_p X_{t-p} + U_t \quad (5)$$

A_i and X_t are m dimension matrix. U_t and δ are m dimension vector. Let $E(U_t) = 0, E(U_t U_t^T) = \sum_{uu}$, and $E(U_t U_s^T) = 0, s \neq t$

Above two models, when the reciprocal of characteristic roots are inside the unit circle, they are smooth.

III. MODELING AND FORECAST

Monthly data are used covering the period from January 2010 to December 2014. It was provided by a enterprises.

3.1 ARIMA modeling

We need to do pure random inspection. Because only the non-white noise sequence modeling makes sense. LB[3] test result is as follows.

TABLE 1
LB TEST

Delayed order	LB statistic	P value
3	55.339	0
6	67.568	0
12	87.188	0

Table 1 show that P value is less than the significance level of 0.05. Therefore, the alternative hypothesis may be accepted. This sequence is not white noise sequence.

Then let us do stationarity test. ADF test shows P value approximates 1, so this sequence is non-stationary sequence.

Modeling process is as follows:

To eliminate seasonal fluctuations, do first-order difference operation by step 12. To eliminate trend changes, do first-order difference operation by step 1. Differenced sequence is smooth. Considering sample size, let p and q change between 0 to 3. When AIC function reaches minimum value, we get a row vector including p and q [4].

TABLE 2
AIC FUNCTION VALUE

p	q	AIC	p	q	AIC
			2	0	43.5429
0	1	39.1627	2	1	45.5429
0	2	40.44	2	2	46.3551
0	3	41.6112	**2**	**3**	**38.3967**
1	0	51.4209	3	0	45.5425
1	1	40.2587	3	1	47.5215
1	2	42.1878	3	2	42.6222
1	3	43.4502	3	3	40.1274

Table 2 shows that the structure of this model should be confirmed as ARMA(2,3).

Next, let us estimate the parameters and forecast sales. Its formula is as follow:

$$X_t = 0.9763 X_{t-1} + 0.7508 X_{t-2} - 0.0607 \varepsilon_{t-1} - 0.0085 \varepsilon_{t-2} - 0.9308 \varepsilon_{t-3} \quad (6)$$

Forecast can be seen from Fig.1

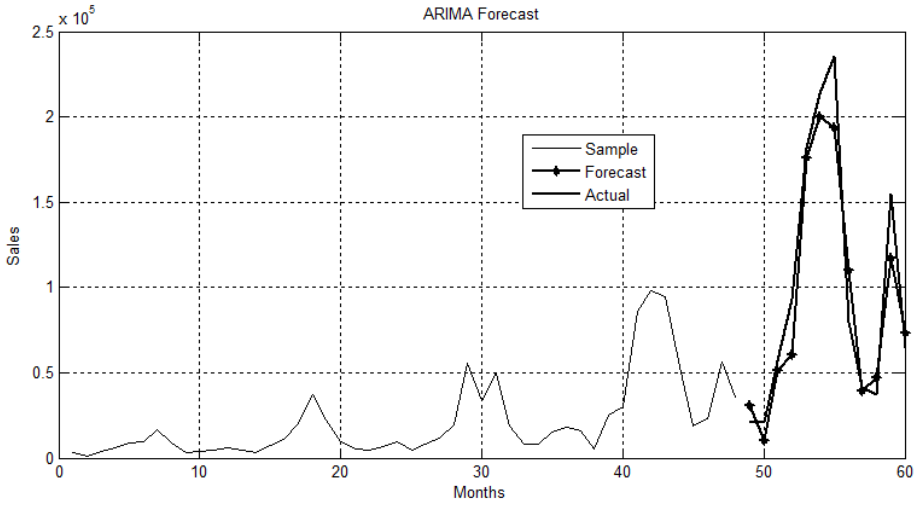

FIG.1 SALES FORECAST OF 2014 USING ARIMA MODEL

The average error rate calculated is 0.2313, accurate for sales. The residual sequence is white noise.

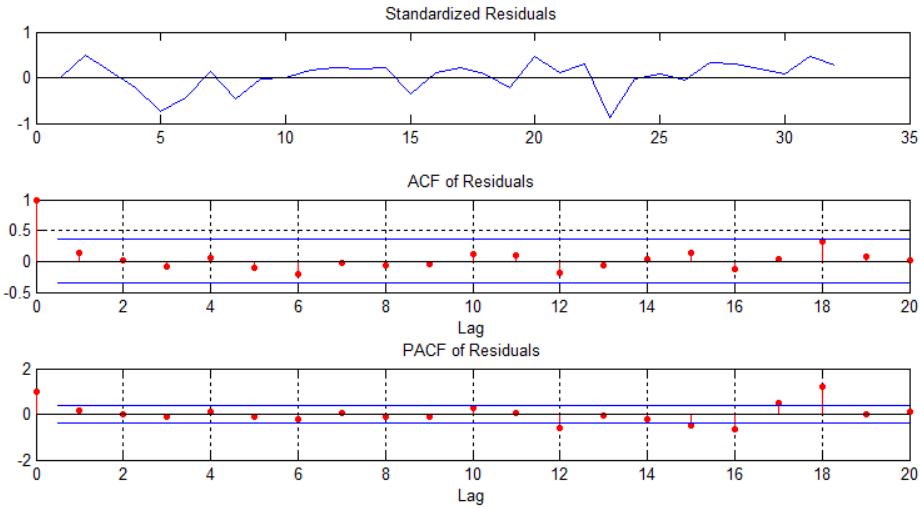

FIG.2 THE RESIDUALS OF ARMA MODEL AND ITS ACF AND PACF

3.2 VAR modeling

Factors that affect sales are five variables, so it is suitable to use principal component analysis or PCA [5] for short to reduce dimensions before establishing the VAR model.

TABLE 3
RESULTS OF PCA

Number	Principal Component	Contribution Rate	Cumulative Contribution Rate
1	z1	0.9654	0.9654
2	z2	0.0343	0.9996
3	z3	0.0003	1.0000
4	z4	0.0000	1.0000
5	z5	0.0000	1.0000

The contribution rate of the first principal component is above 90%.Its formula as follow:

$$z_1 = -0.0039x_1 + 0.0288x_2 + 0.0047x_3 + 0.9559x_4 + 0.2922x_5 \quad (7)$$

The VAR model uses two variables, namely z1 and sales. Firstly, stationarity test shows that when $p \leq 6$, that is, the reciprocal of characteristic roots are inside the unit circle, model is smooth.

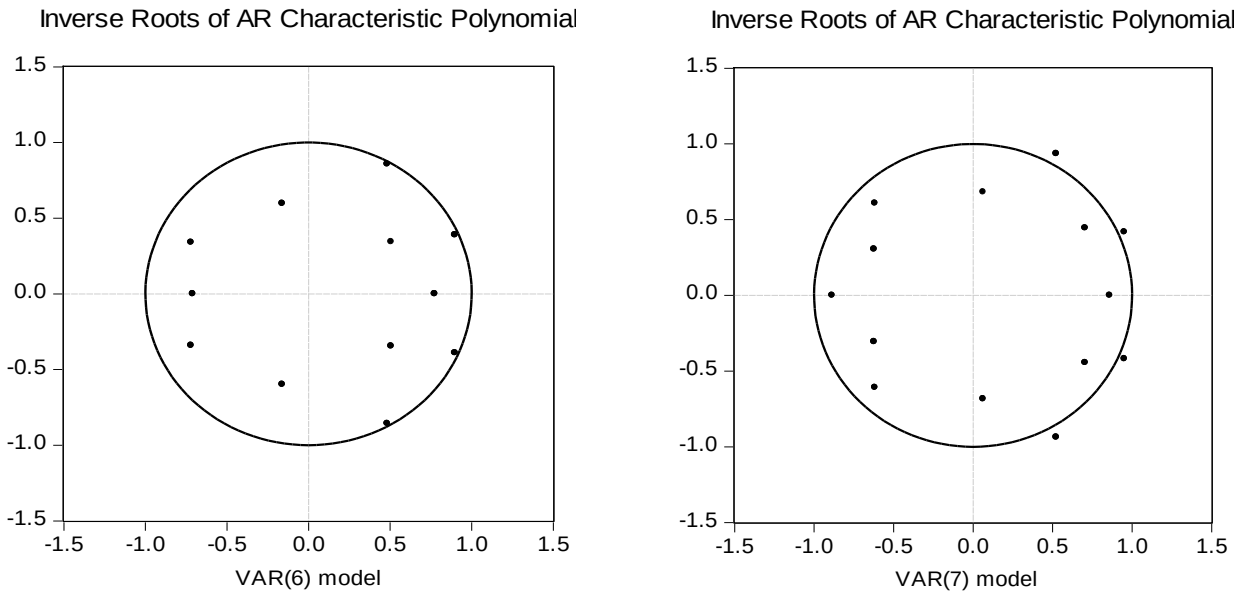

FIG.3 THE RECIPROCAL OF CHARACTERISTIC ROOTS DISTRIBUTION ABOUT VAR(6) AND VAR(7)

Similar with ARIMA model, using MATLAB software to calculate AIC function. The structure of this model is eventually determined as VAR (6). Its parameters calculated can be seen on equation (8):

$$\begin{aligned} sales &= 0.56 \cdot sales(-1) - 0.04 \cdot sales(-2) - 0.58 \cdot sales(-3) \\ &+0.28 \cdot sales(-4) + 0.22 \cdot sales(-5) - 0.45 \cdot sales(-6) \\ &+0.85 \cdot z1(-1) - 0.32 \cdot z1(-2) - 0.24 \cdot z1(-3) \\ &+0.09 \cdot z1(-4) - 0.05 \cdot z1(-5) + 0.16 \cdot z1(-6) \\ &+19701.23 \end{aligned} \tag{8}$$

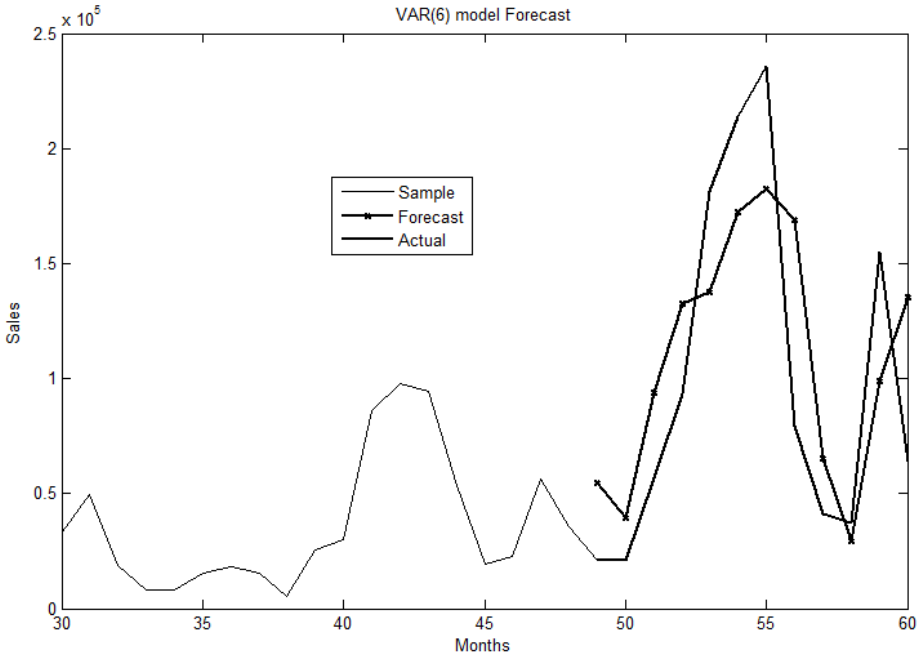

FIG.4 VAR MODEL FORECAST SALES OF 2014

VAR model can grasp the change of time series from the macroscopic aspect. But its precision is not high.

IV. CONCLUSION

As single variable model, ARIMA model establishes equation by calculating ACF and PACF, the residuals included. According to akaike information criterion, the model structure is identified as ARMA (2,3). It is very precise and credible. With regard to multi-variable time series model, it did principal component analysis to reduce dimensions. Five explanatory variables are transformed to one main component. VAR model has two variables. On the contrary, prediction accuracy of VAR model is not high. It is suitable for macro-economic analysis.

This paper studies the issues of sales forecast, which belongs to the category of microeconomic. It comes to a conclusion that single-variable time series model is more suitable for this problem, especially the stochastic time series models.

Although we find an ideal time series model, sales forecast in November may have too big error. This is a problem of structural breaks, which need further research.

REFERENCES

[1] Shuyuan He. Random process [M]. Peking University Press, 2008.

[2] Tiemei Gao. Econometric analysis method and modeling: application and instance EViews [M]. Beijing: Tsinghua University press, 2006:50-51.

[3] Jianhong Wu, Lixing Zhu. Multivariate time series GARCH type model of diagnostic test [J]. Journal of mathematical physics, 2010, 30 (3) : 630-638.

[4] Sifan Yang. Time series modeling and forecasting [D]. Tsinghua university, 2014.

[5] Shoukui Si. Mathematical modeling algorithm and application [M]. National defense industry press, 2011