

ISyE 6420

Bayesian Statistics

Junqing Ma

April 28, 2015

# **Time Series Forecast with Bayesian Approach Case Study of Apple Inc. Sales**

# 1. Introduction

Time series data that involve sequential ordering and correlation occur everywhere around us. Analyzing time series is the problem of discovering the pattern in the sequential data that develop in a logical way as the series continue. In the past few decades, research has been conducted on Bayesian method application in time series analysis field. In general, Bayesian inference has a distinct advantage over classical statistics in non-standard problems where concepts such as sufficiency or completeness do not apply. Autoregressive-moving average (ARMA) time series models are quite non-standard, even if the usual assumption of normality is retained. For example, classical analysis of ARMA models must rely on asymptotic behavior such as consistency, asymptotic normality, and efficiency. But in many cases, as few as 50 observations are quite common that is not a big enough size for classical data analysis. However, Bayesian statistics are little affected by sample size. Therefore, through my project, I want to explore how Bayesian methods perform in the case of time series data and how their performance is compared to classical time series analysis.

I use Apple Inc.'s quarterly revenue data from 2002 to 2012 which count to 50 data points. Then I will identify the most suitable parameter  $p$  and  $q$  for ARMA model by fitting ARMA models with different parameters and examining the accuracy. To test the accuracy, I take out 10% as the test data that is about 5 quarterly numbers. I will predict these 10% data with the models I train under Bayesian ARMA model and classical ARMA model and compare the forecasted values to the real values that I leave out. Here

Mean Squared Error (MSE), defined as 
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2.$$
 will serve as the measurement of prediction accuracy.

## 2. Data Investigation

As we know, autoregressive moving average models provide a parsimonious description of a weakly stationary stochastic process in terms of two polynomials, one for the auto regression part and the second for the moving average part. So before I apply ARMA models to the data, I need to examine the data to make sure the data are stationary so that ARMA models are suitable in such case. The weak stationarity is defined as: 1) the mean is constant 2) the covariance only depends on the distance. If we take a look at the time series plot of the data (figure 1), we can see there is clearly a trend going up

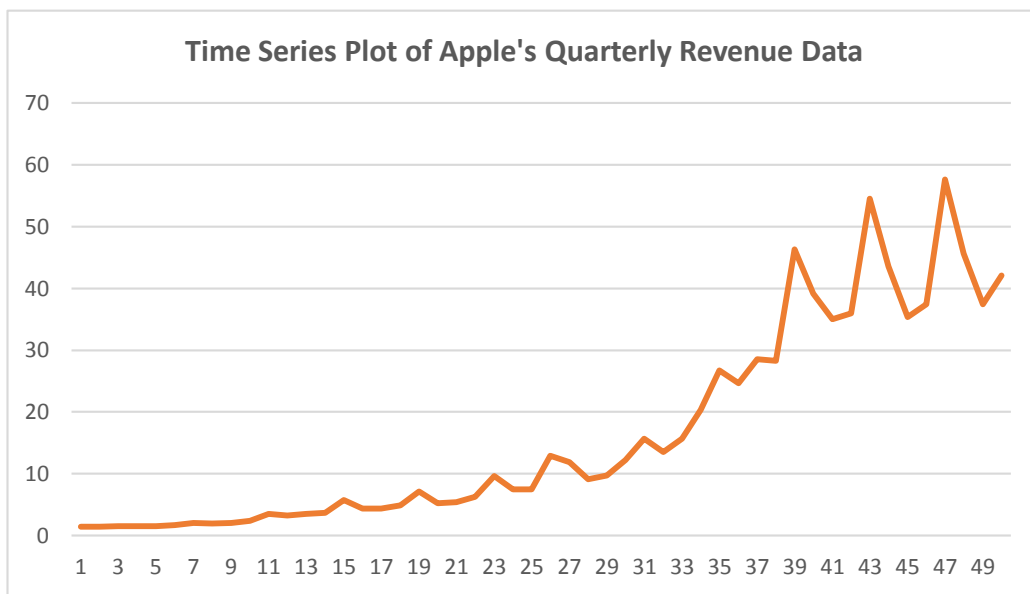


Figure 1

over time, which violates the constant mean rule. The typical technique to deal with non-stationary data is to take differencing (study the difference between two consecutive data points instead of studying themselves). Looking at the time series plot of new data, we can see now the data are bouncing around zero and thus the mean is about zero. While the variance is getting greater over time, the ACF and PACF graphs show the data can still be considered stationary. In addition, taking second differencing does not help too much. So I will proceed using data with first differencing.

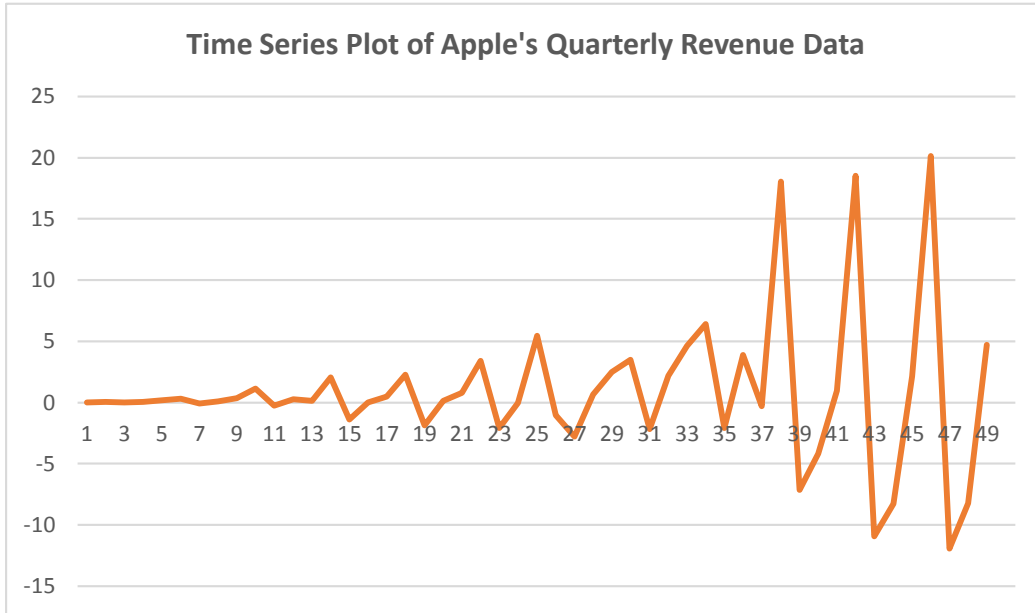


Figure 2

### 3. Model Identification & Results

ARMA models are presented in the form of:

$$X_t = \mu + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

where  $X_{t-1}, \dots, X_{t-p}$  stand for  $p$  autoregression terms and  $a_{t-1}, \dots, a_{t-q}$  stand for  $q$  moving average terms. If we go back to the time series plot (Figure 2), we can detect a strong seasonality – the data go up and down every four periods. Therefore we also need to take into account the seasonality in our ARMA models.

$$X_t = \mu + \phi_1 X_{t-k} + \dots + \phi_p X_{t-nk} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

where  $k$  is the interval of seasonality. In our case,  $k$  is 4.

After we determine the seasonality, we start to identify  $p$  and  $q$  in the models. By rule of thumb, the sum of  $p$  and  $q$  is usually no greater than 3 especially in not complex case. Thus there are three ARMA models that I want to test:

(1) ARMA(1,1):  $X_t = \mu + \phi_1 X_{t-4} + a_t - \theta_1 a_{t-1}$

$$(2) \text{ ARMA (2,1)} \quad X_t = \mu + \phi_1 X_{t-4} + \phi_2 X_{t-8} + a_t - \theta_1 a_{t-1}$$

$$(3) \text{ ARMA (1,2)} \quad X_t = \mu + \phi_1 X_{t-4} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Identifying three possible ARMA models, I implement them in WinBUGS and generate three different groups of forecasting. I use common priors such normal and gamma conjugates and beta prior for parameter  $p$  and  $q$ . Roughly comparing the results to real data, I conclude ARMA (2,1) and ARMA (1,2) are slightly better. So I run the same model in Minitab where classical data analysis is performed. The final results are summarized in Table 1 as follows:

	ARMA(1,1) WinBUGS	ARMA(2,1) WinBUGS	ARMA(2,1) Minitab	ARMA(1,2) WinBUGS	ARMA(1,2) Minitab	Real Data
46	16.39	16.43	17.67074	18.36	22.09094	20.12
47	-10.5	-17.06	-13.4881	-11.26	-11.7631	-11.94
48	-6.431	-9.433	-12.699	-7.614	-9.45646	-8.22
49	1.668	2.424	3.350647	2.193	3.586423	4.69
50	16.79	31.32	11.30552	16.81	21.73788	32.48
MSE:	54.899	9.5565	95.7219	51.14224	57.3922	----

It's not hard to see ARMA (2,1) generated very good forecasting that has MSE only at 9.5565, significantly smaller than that of forecasting from others models. Five data points forecasted under this model are fairly closed to the real data. Two ARMA models trained under classical data analysis in Minitab perform far worse.

Therefore, we conclude the best ARMA model for our case study is ARMA (2,1)

Bayesian model, presented as below:

$$X_t = 0.1358 + 0.7792X_{t-4} + 0.03092X_{t-8} + a_t - 0.1845a_{t-1}$$

We can see  $X_t$  is heavily dependent on the its previous term, a seasonality of 4 periods before, and slightly dependent on the next previous term, two seasonalities of 8 periods before.

## 4. Conclusion

In my project, I have shown Bayesian ARMA model performs better and forecasts more accurate than classical ARMA model under the dataset of Apple Inc.'s quarterly sales. While recognizing the advantages of Bayesian ARMA models, my conclusion does have a few major limitations.

First of all, the data size is small for classical ARMA model and this may negatively affect its performance. Second, my conclusion can only be interpreted in this sample of data. It may not be true if same analysis is conducted under other datasets, such as IBM sales or Microsoft. Apple Inc.'s sales data are relatively stable and predictable than other kinds of data.

In the future study, I want to explore how different choices of prior distribution affects the performance of Bayesian models. I especially want to experiment with informative prior distribution with a larger dataset so that a more reliable conclusion could be reached. On the other hand, it is always interesting to include other time series factors in the time series regression. In the example of Apple Inc.'s revenue, I would like to try other important indicators such as inventory data and demand data.