Comparison of AR(1) Model and Garch(1, 1) Model for Fitting Amazon Stock Price Data

Helena Li, Jin Yan, Yi Han

Abstract

Introduction & Background

Stock prices data give investors insight into the company's current operating status and prospect of the company. Time series modeling gives a prediction on the stock prices and for investors interested in the company it is crucial for them to make wise decisions.

The time series dataset we are using in this experiment is the daily Amazon Stock Price dataset recorded from the beginning of 2017 to around the end of March in 2019. This data is downloaded from Yahoo Finance. The measurements are recorded daily and do not include weekends.

Stocks prices in the dataset refer to daily adjusted closing prices. The closing price is the raw price, which is just the cash value of the last transacted price before the market closes. The adjusted closing price factors in anything that might affect the stock price after the market closes. (https://www.investopedia.com/terms/a/adjusted_closing_price.asp) .

On the left is the plot of this dataset. It shows that the time series is not stationary and it has a increasing trend from 2017 to 2018.5. Besides, there is an increased variability with time.

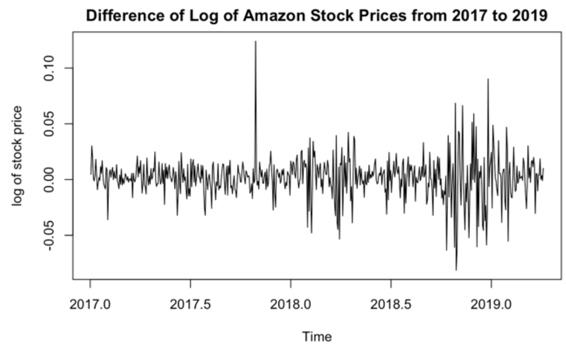
In this experiment, we fit two models using the Amazon Stock Price Data: the AR(1) model and GARCH(1, 1) model. Before comparing the two model outcomes, we will first introducing this two models.

The AR model, or the autoregressive model, is the model specifices that the output variable depends linearly on its own pervious values and on a stochastics term (normally written as the white noise term), thus forming a stochastic difference equation (引用 1).

When the autoregressive model is combined with moveing average (MA) model, it becomes an autoregressive moveing average (ARMA) model, and if the ARMA model is assumed for the error variance, the model is considered as a generalized autorgressive conditional heteroskedasticity (GARCH) model (引用 2). The GARCH model is always used to estimate volatility in financial markets (引用 3), that is why we choose it to analyze our stock price data here.

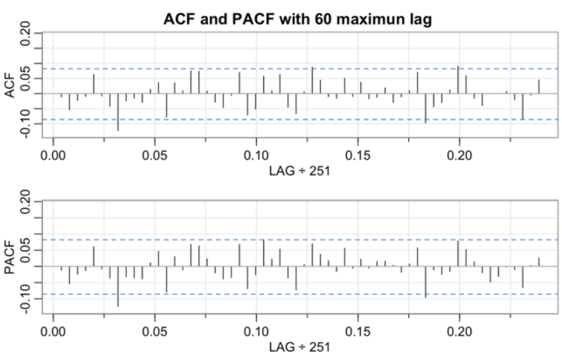
ARIMA Model Fitting

From the plot above, there is a linear trend in different periods in the original time series and log transformation is usually applied to financial time series.



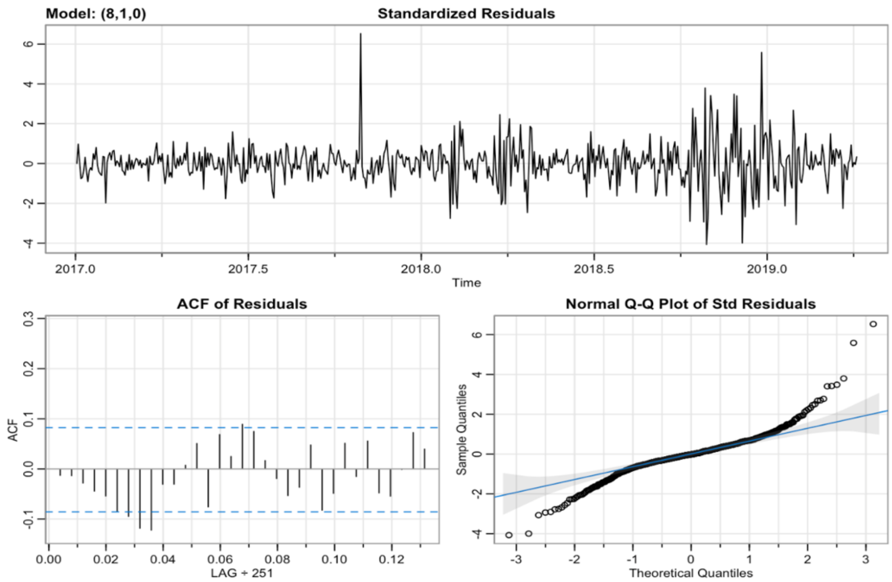
After log transoformation, the time series is still not stationary. As is common with financial time series, we'll look instead at the percent change/growth rate of the series. This is well approximated by the difference of a log transformation. Now we get a stationary time series with an approximate zero mean. Consider differencing was part of our transformation. We'll model log of stock price after differencing as an ARMA(p,q) process for some values of p and q, which means that log(stock price) is an ARIMA(p,1,q) process.

To choose p and q, we look at the sample ACF and PACF of the series. The sample ACF and PACF look consistent with x being either an MA(8) process or an AR(8) process. We also incorporate a ARMA(8,8) model for model selection.



MODEL	AIC	AICC	BIC
ARIMA(0,1,8)	-2822.171	-2821.774	-2778.785
ARIMA(8,1,0)	-3078.464	-3078.067	-3035.078
ARIMA(8,1,8)	-2889.949	-2888.699	-2811.855

To conduct model selection, we compare AIC, AICc, and BIC value of the three models. The ARIMA(8,1,0) model achieves the lowest information criteria.

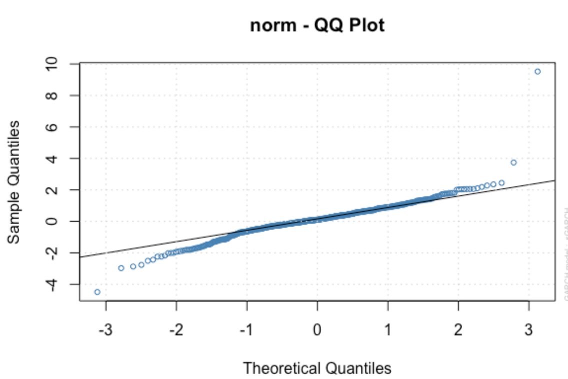


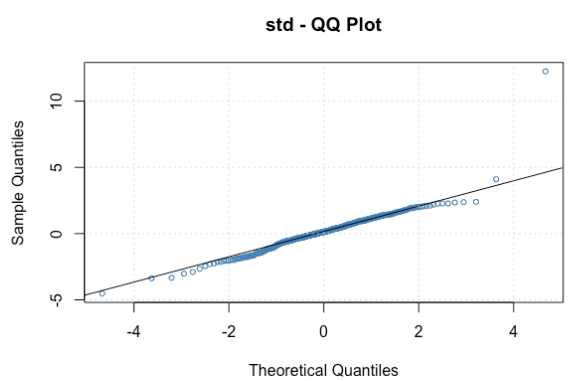
GARCH Model Fitting

In this section, we fit the stock price data with AR(1) as the mean model and GARCH(1,1) as the variance model to specify the burst of volatility among the periods of stability.

In the GARCH model specification, we compare two models. The two models use almost the same information, the only difference lies in their model distribution: one uses normal model, while the other uses a t-distriution model. We especially focused on the QQ-plot of the standardized residuals, and found the model basing on t-distribution does somehow better than the normal one. Here are the plots:

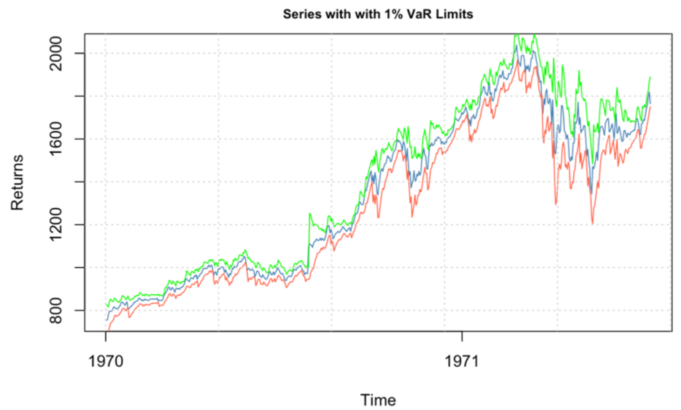
While both model performed quite well for the regions around 0, normal model shows quite a deviation when theoretical quantiles smaller than -1, while the same area for std normal





still work quite well. The std model also shows some deviation when theoretical quantiles greater than 2, but there are few data points at that place, and the deviation at the same place for normal model does not seem better.

The comparison of information criteria also gives the same result. For normal mode, AIC=9.0421, BIC=9.0809; while for the std model, AIC=8.8332, BIC=8.8797. Because we should choose model with the least information criteria, here both AIC and BIC prefer std model.



Then we use the std model to do forecast. In the left plot below, the blue line is the actual time series and the green and the red ones represent 1% variance limits of the std model. It shows that almost all the actual time series fall into the 1% variance limits which means the std model is a good fit to the original time series.

Then, we try to use the std model to do forecast. In a rolling forecast, we make a 1-step-ahead forecast, then observe the next value in the series, then we make another 1-step ahead forecast taking this information account, and so on. In the right plot below, we do 10 times of rolling forecast.