ARMA model for large cap stocks returns

Kunal Rajani (kr386@cornell.edu)*

27th Aug 2012

Abstract

Time series is a vast topic that is extensively used in the supply chain and operations industry to forecast future values of randomly varying but correlated to past quantities. Although its a science to model a time varying series using various models such as AR, MA, GARCH, etc. it is an art to make it work and deduce the right conclusions out of the result. The purpose of this paper is to predict the returns of large cap stocks using the ARMA time series approach and create a portfolio of stocks based on such prediction that can beat the SNP 500.

1 Introduction

Stock returns are one of the most volatile forms of time series as they can simply vary based on perception without any fundamental underlying cause. Unlike a product demand that can be studied and forecasted based on the industry and the market and hence that tends to be stable or rather predictive, stock prices can take a nose dive due to one negative report or surge beyond bounds based on a surprise announcement. This can be attributed to the fact that stocks are highly liquid instruments and hence their prices tend to have a considerably higher random variation.

It is a challenege to model such a high degree of random variation and there is a lot of research that goes into coming up with customized algorithms to predict stock returns for different industries. In this paper I will try to fit ARMA model to a portfolio of large cap stocks and study its predictive power to obtain an SNP 500 beating portfolio.

1.1 The ARMA model

The ARMA(p,q) model is stated as:

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + ... + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q}$$

The ϕ are the coeff for the AR model that take into account the impact of previous values of the series while θ are the coeff of the MA model that consider the impact of the previous random values of the series. In our case the series is the returns on a family of stocks.

*Please reach out to me if you need any additional details on this analysis

1.2 The Data

This analysis was done on companies that had a market capitalization of between \$5B and \$50B at some point after 2008, i.e. post the financial crisis. For those companies I obtained daily stock quotes that were adjusted for splits. Finally, I fit an ARMA model on these stock quotes and tried to create a portfolio that could beat the SNP 500 over the analysis period.

2 The Analysis

2.1 The Procedure

After obtaining the data as described above I began creating my portfolio from May '11 to Aug '12. I rebalanced my portfolio weekly (every 5 trading days) and looked at the previous 100 days as an input to the ARMA model construction. Note that as simple as it sounds when these analyses are done from scratch there are numerous obstacles that can bring you to a deadlock and it is best to be aware of them before hand. I am listing a few of those practical challenges below:

- Each stock can have its own stationarity characteristic and hence can have different values of p and q and also a non-zero drift term. This causes inconsistency in developing a generic model. I dealt with that by using an *auto.arima* function in R.
- The data can be in different formats and computations can have different results on different types. This helped me get used to different data types in MATLAB and R.
- Data can be missing or a company can be too new/old to be active on the exchange for the entire analysis period.

During the rebalancing I took the top 10% of the stocks based on their predicted returns and standard errors, i.e. those stocks having the maximum sharpe ratio. Because of unavailable data throughout the analysis period I filtered the list to work with 700 large cap stocks. These stocks were held for 5 days after which the ARMA model was again fit to all the stocks and the top 10% were taken from them. This procedure continued until the end of the time period.

2.2 The observations

The following plot Fig 1 shows the evolution of \$1 invested in three portfolios:

- ARMA - the portfolio obtained by computing the expected return of the stocks based on an ARMA model and the selecting the top 10% of the stocks, rebalancing every holding period.
- SNP 500 - the portolio obtained by investing only in SNP 500

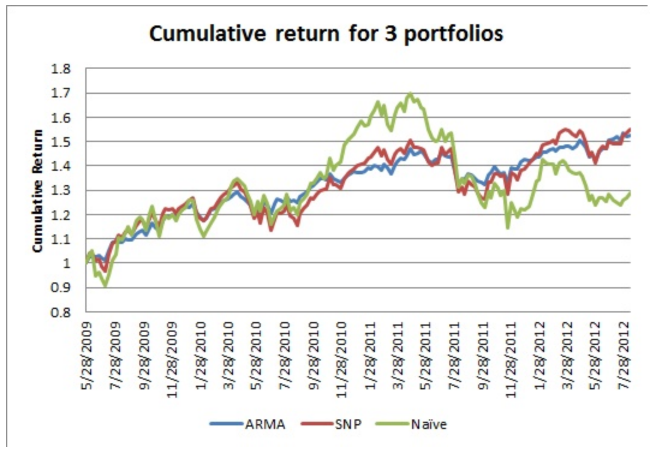

Figure 1: Cumulative return of three portfolios from 28 May '09 to 28 Jul '12

- Naive - the portfolio obtained by investing in the top 10% of the stocks based on the previous 20 days of return. This would help in understanding if there is any value add of considering the previous 100 days of history and fitting it to a time series model.

Apart from comparing the portfolio thus obtained it would be beneficial to see if our predicted returns were indeed able to rank order the stocks based on their actual returns. To see that we can take a look at the graph (Fig 2 on the following page) below that tells the %age of selected stocks that were actually one of the top 25 stocks of that period.

3 Conclusions

We see from Fig 1 that our portfolio doesn't seem to provide a clear evidence of the value add of using time series modeling alone to predict future stock returns. We see that our portfolio pretty much sticks with SNP throughout the analysis period but manages to do so with a lower volatility of 1.5% as compared to SNP's 2.3% that yields a Sharpe Ratio of 66 to SNP and 99 to our portfolio. This helps conclude that SNP is a good representation of the market index. We see that the naive portfolio, on the other hand, is quite reactive and does well when the market was recovering in 2010 and tolled back during the euro crisis of 2011.

As further evidence of the above conclusion, when the same analysis was carried out with a farther horizon of a month, i.e. a holding period of one

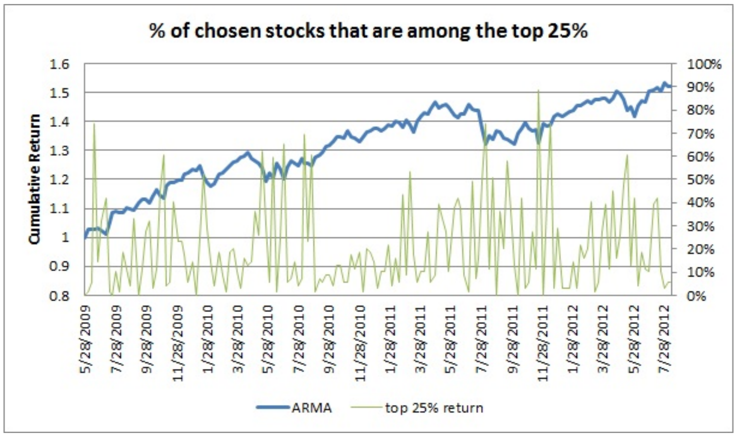

Figure 2: %age of selected stocks that were in the actual top 25% of the stocks in the period

month, we observe the below evolution in Fig. 3 . We again see that the naive portolfio is quite reactive although it seems to do better in 2011 and sustains the eurocrisis better as compared to short horizons of a week. Even though the naive portfolio ends up higher its high volatility causes it to have a lower Sharpe Ratio of 27.6 as compared to 36 for SNP and 50 for our portfolio.

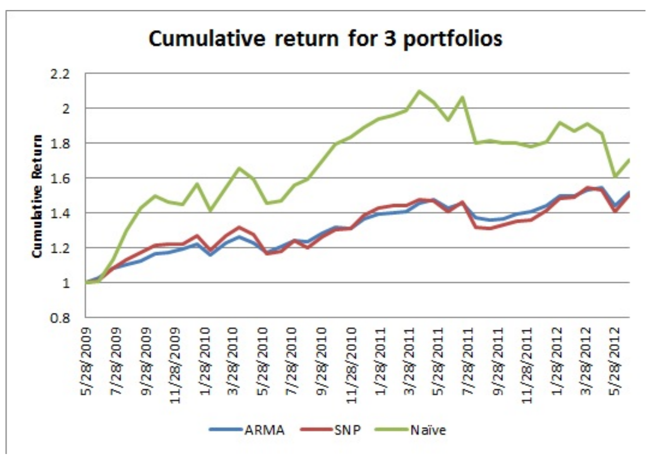

Figure 3: One month holding period

From Fig. 2 we can see that ARMA modeling wasn't able to consistently select a majority of top 25% stocks. It actually ends up picking 20% of the top 25% stocks on an average but with very high swings as can be seen in the Fig.

Hence we can say that ARMA modeling alone is not able to provide a competitive edge to predict the future short term stock returns of large cap stocks.

This was quite expected since, as was discussed before, stock returns are highly volatile and past performances are not very good indicators of the future.

4 Further Work

This was a preliminary analysis to study the application of time series modeling to predict short term stock returns. Though such returns tend to be more volatile, long term returns are more stable and it would be interesting to see if such models can provide an edge in that horizon although we would need to be careful about seasonality effects. Furthermore it would be insightful to see if such a methodology works for any particular segment (like an industry) more so for others.

The next phase of this project will explore these areas together with an analysis on a different time horizon and short term stocks to paint a better picture. I will also study vector ARMA models that consider the impact of one stock over the future of another although the magnitude and volatility of the problem might make that unfeasible and futile.