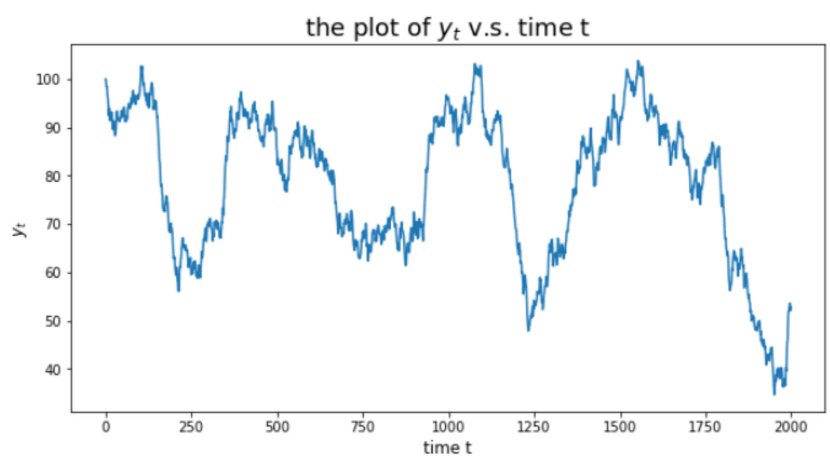
Fitting & Evaluating Three Basic Strategies in Synthetic Time Series

Jinxing Xie



1.Introduction:

Traditionally, technical analysis is always suspected by the academic community, since under the market efficiency theorem, no technical trading system can generate excessive profit from the market, and most of the investment books only contribute few pages for the technical analysis (Dahlquist, 2005). However, due to the development of computation, more and more mathematic parameters or correlations among underlying assets or the market sentiment can be observed, calculated then used in the technical analysis.

In this paper, we will mainly discuss the fittings and performances of three basic but popular technical strategies in a synthetic stochastic price series and also evaluate each of them by three indicators: Sharpe Ratio, Calmar Ratio and Value at Risk (VaR). Since Sharpe Ratio can be transformed into a t-statistic (Andrew, 2002), single and multiple tests can both be used to test the profitability of strategy in single-period and multi-periods (Harvey & Liu, 2015). Please also bear in mind that, all the strategy fittings in this paper, including the coefficients and trailing-stop criteria, are done in the training set.

2.Methodology:

2.1. Time Series Generation & Analysis:

The time series used in this report is generated by the formula:

$$\Delta y_t - d = \phi(\Delta y_{t-1} - d) + \varepsilon_t + \theta\varepsilon_{t\text{-}1}$$

where t ranges 0 to 2000, $y_0 = y_1 = 100$, $\varphi = 0.6$, $d = 0.025$, $\theta = -0.4$, y_t represents the (close) price in day t, $\Delta y_t = y_t - y_{t-1}$ and ε_t is a sequence of white noise with zero mean and unit variance.

The formula given above is an ARIMA(1,1,1) model, however, in the following sections of strategy making, the synthetic sequences of returns and prices will be considered without any prior knowledge about this generating function. For the need of analysis and comparing in-sample and out-of-sample performances of the strategies, the time series is split into two subsets, named training set and test set, representing the 70% and 30% of the original data, and all the strategy fittings later are done by using training set only.

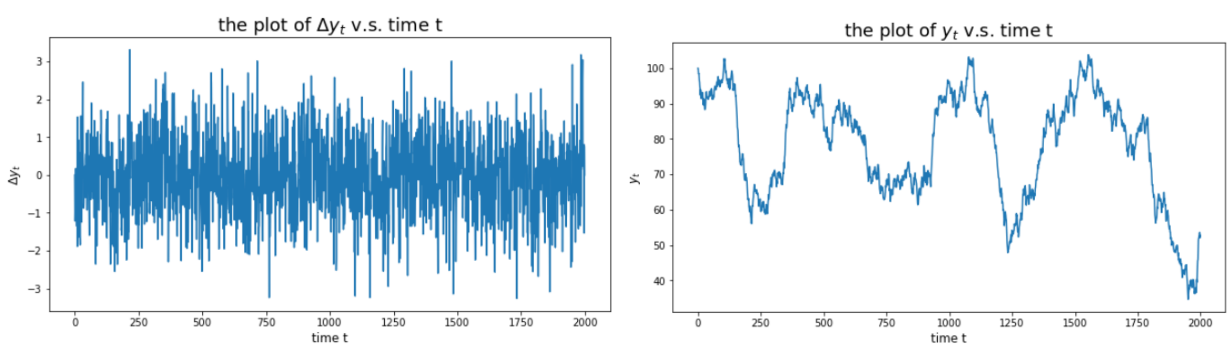


Figure 1: The synthetic returns from day 0 to day 2000 (Left) and corresponding price plot (Right).

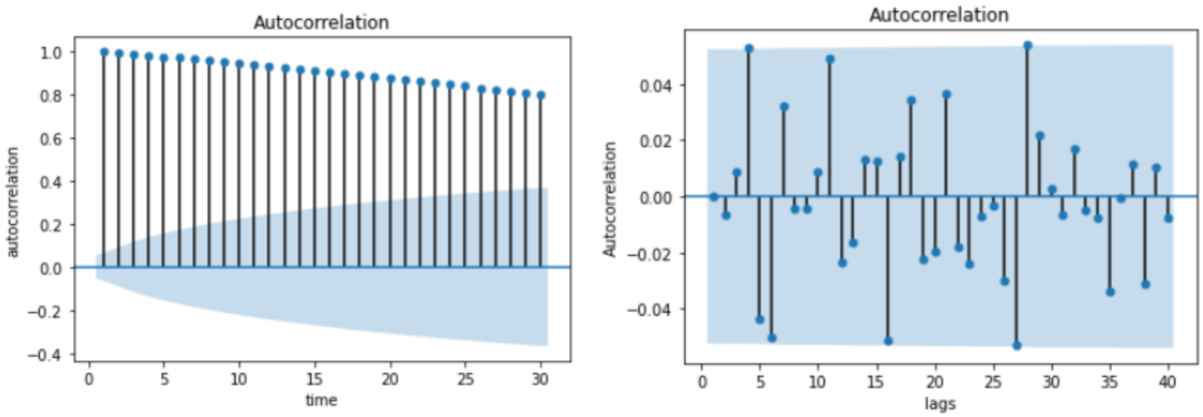


Figure 2: The ACF plots of prices (Left) and Residual ACF plot of ARMA(2,1,1) of prices (Right).

In the plot of returns Δy shown in Figure 1, no significant seasonality or trend can be observed, however, in Figure 2, the Autocorrelation Function (ACF) plot of price series implies a potential linear relation between autocorrelation and time t. Therefore, there may exist an ARIMA model for price time series.

By trying automatic ARIMA fitting function in python, the best ARIMA model for price series in training set with the lowest AIC criterion can be represented as:

$$\Delta y_t = 1.661 * \Delta y_{t-1} - 0.1906 * \Delta y_{t-2} - 0.9511 * \epsilon_{t-1} + \epsilon_t$$

This result will be used to fit an ARMA model for return Δy_t, which will be introduced in Section 2.2.2.

2.2. Trading Strategies:

2.2.1 TF Strategy:

Motivated by the paper from Dahlquist (2005), the Trend Following Strategies, named TF, uses the crossover of the 21-day Moving Average and 1-day Moving average as the signal for buying or selling, that signals buying as long as the price breaches the 21-day Moving Average, and selling once the 21-day Moving Average rises above the price. Comparing with the moving averages with longer look-back periods, Dahlquist (2005) argued, 21-day Moving Average has more latest price information which is relatively more important. Furthermore, since in the ARIMA strategy using model which will be introduced later, trades happen almost every day, a TF strategy with 21-day Moving Average may have a significant difference from it and bring us more mid-term information, so 21-day Moving Average is selected.

In this strategy, two parameters will be considered:

p: The percentage of cash used to purchase stocks after the purchase signal is generated.

q: The proportion of position which is closed once the sell signal is generated.

The final total value of TF strategy with different p, q is shown in Figure 4, and the optimal p and q which will bring the highest final total value are 0.91 and 0.11 respectively, which implies that in this TF strategy, Fast-in-but-slow-out leads to the highest profit in the training set. The optimal p and q will be used in testing.

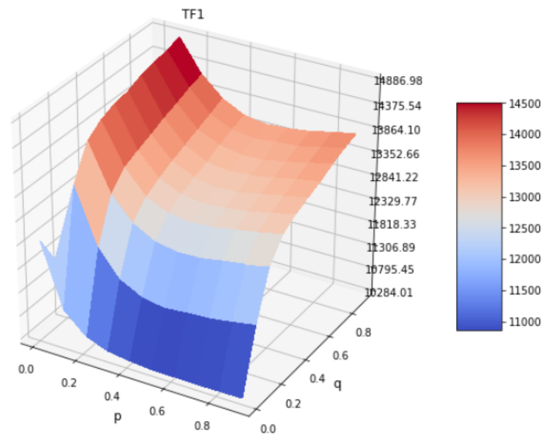


Figure 4: The final total value over training set versus corresponding *p* and *q* under TF strategy.

2.2.2 ARMA Strategy:

In this section, we will develop a simple on-line high-frequency trading strategy based on the ARMA (2,1,1) model discovered above. The basic idea behind this strategy is that by using the equation of the ARIMA model, when the historic prices data is given, we can forecast the change of price in the next period, and based on the signs of these changes (positive or negative), we can add position if the sign is positive, otherwise, reduce or clear the position. Since intuitively, investing more cash in the asset can amplify the profit when we have a potential positive return prediction, and reducing the position can decrease the uncertain loss if negative forecasts are generated. Besides, since there won't be too many zero forecasts, transactions happen almost every day.

Similar to the TF Strategy, the proportions of position opening and closing, designated as *p* and *q* respectively, will be considered. But different from the result of the TF strategy, in the ARMA strategy, slow-in-and-slow-out can generate the highest profit, where optimal *p* and *q* are both 0.11. More details are shown in Figure 5 below.

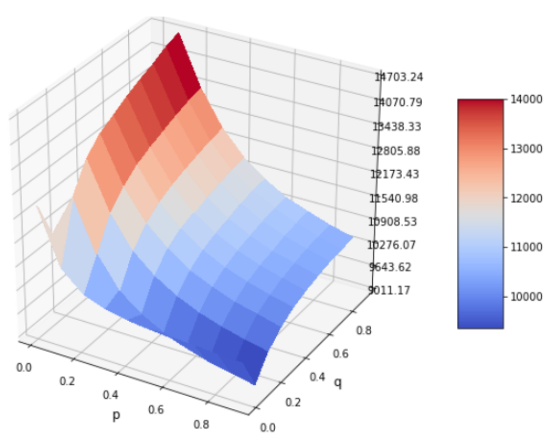


Figure 5: The final TV over training set versus corresponding p and q in ARMA strategy.

2.2.3 MR Strategy:

Defined in the paper (Temnov, 2017), Mean-Reverting (MR) Strategy believes in the reverting driving forces that revert the price to the long-term mean and gains profits from this price reversion. According to Temonov's suggestions in that paper, here, the long-term mean is represented by 27-day Moving Average MA and three parameters of the strategy need to consider:

D_1 - the predetermined level of price to MA for position opening;

D_2 – the predetermined level for position closing when the price reverts back to mean successfully;

TS - The predetermined level for trailing stop;

By MR strategy, we will open a long position at time k as long as the price $p_\mathrm{k} < D_1 * MA_k$, and close the position at time t if our $TV_t < TS * TV_k$ for trailing stopping or when the price p_t rises back to the MA_t, namely $p_t = D_2 * MA_t$. Also, he suggested that every after a fixed period of time, these parameters should be updated such that they could maximize the profit in the last period, and then be used in next period. Therefore, the length of the time window is also a parameter in strategy making. I simply prepare 3 time-windows of 25-day, 50-day and 100-day, and their performances in the training set are shown in Figure 6. Obviously, the MR strategy with a 50-day time-window has an outstanding profit and will be tested in the test set.

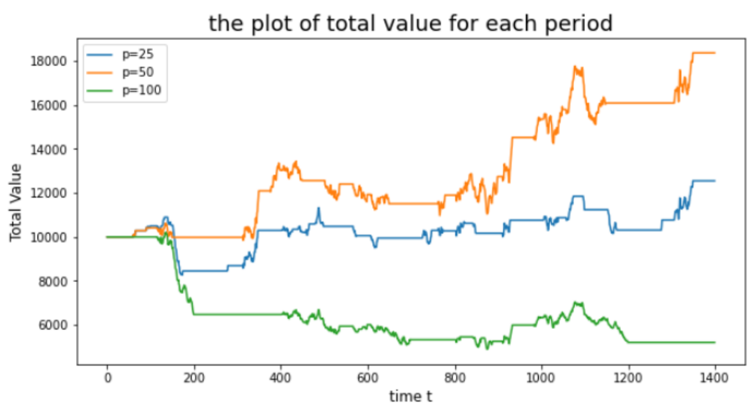


Figure 6: The performances of MR strategies of time-windows of 25, 50 and 100.

2.3 Performance Indicators:

2.3.1 Sharpe Ratio (SR):

Sharpe Ratio is a popular indicator used in the industry for performance evaluation, depicting the mean returns per unit of volatility, written as:

$$SR = \frac{\mu}{\sigma}$$

Where μ is the mean (excess) return and σ is the volatility. Since we only perform our strategies in a synthetic price series without consideration of risk-free rate, μ is just the mean of logarithmic returns.

2.3.2 Value-at-Risk (VaR):

VaR indicator represents the maximum loss which can happen for a given confidence level α, written as:

$$VaR_R(\alpha) = inf\{\, r: F_{\text{-}R}(r) < \alpha\}$$

Where R is the random variable of the daily log returns of a strategy and $F_{\text{-}R}$ is the cumulative distribution function of -R.

2.3.3 Calmar Ratio (CR):

Calmar Ratio represents a risk-adjusted average return measured on the maximum drawdowns in the history, written as:

$$\mathrm{C}R = \frac{\mu}{\mathrm{MDD}}$$

Similarly, μ is the mean of the strategic daily returns, and MDD is the maximum drawdown written as:

$$MDD = \frac{\text{Peak Value - MTV}}{\text{Peak Value}}$$

Where Peak Value is the maximum strategic total value in the history and MTV is the abbreviation for Minimum Through Value, representing the minimum total value after reaching the peak value.

2.4 Hypothesis Test of profitability:

2.4.1 Single Hypothesis:

Recall the Sharpe Ratio formula:

$$SR = \frac{\mu}{\sigma}$$

In Andrew Lo's paper (2002), student t-distribution of Sharpe Ratio was first documented. Suppose we have an annualized Sharpe Ratio over T days, which is annualized from the daily Sharpe Ratio, then the t-statistic will be calculated by:

$$t = SR\sqrt{\frac{T}{252}}$$

With the t-statistic and corresponding degree of freedom df = T-1, we can make a hypothesis test following the suggestion given in the research (Harvey & Liu, 2015) for testing whether the strategy has a true non-zero profit even with a non-zero Sharp Ratio in back-testing with:

$$H_0: The\ ture\ Sharpe\ Ratio\ is\ zero\ (namely, true\ \mu\ is\ zero)$$

$$H_1: The\ ture\ Sharpe\ Ratio\ is\ non-zero\ (namely, true\ \mu\ is\ \text{non-}zero)$$

If confidence level $\alpha = 0.5$, the null hypothesis will be rejected if the corresponding p-value is less than 0.05. Since in this paper, all the strategies are long-only and conducted in a synthetic price process, the correlation which may cause misinformation won't be considered.

2.4.2 Multi-test Hypothesis:

To guarantee the non-zero profitability of strategy can still be significant at the confidence level $\alpha = 0.05$ even in the out-of-sample test, multiple tests are needed, and we can measure them from a single test by Family Wise Error Rate (FWER) Control. In the paper from Harvey & Liu (2015), they also argued that, Bonferroni Correction can be used to achieve this purpose, where suppose p^s is the p-value of a single test and the number of tests is M, and by Bonferroni Correction the adjusted p-value p^m:

$$p^m = \min(\mathrm{M}^* p^s, 1)$$

The hypothesis test has:

$$H_0\text{: } No\ strategies\ can\ generate\ non - zero\ profit$$

$$H_1\text{: } At\ least\ one\ strategy\ can\ generate\ non - zero\ profit$$

If the only strategy under consideration has *p-value* $p^s < 0.05$ which rejects the null hypothesis shown in section 2.4.1, and prices in different periods are assumed mutually independent, therefore, its Sharpe Ratio in each period can be seen as a single test then p^m can be calculated to evaluate its profitability in M periods (years) by Bonferroni Correction. If p^m is not low enough to reject the null hypothesis, then in the aggregation of M periods, no evidence to reject the hypothesis of the non-profitability of the strategy.

For example, there are 600 days in our test set, approximately 2.5 years, so N = 2. Suppose one of my strategies has a low enough *p-value* p^s in training set to reject the null hypothesis of single test, now to ensure its profitability is non-zero in an unseen sample (*e.g.* the test set), we need multi tests with N = 2. By Bonferroni Adjustment, $p^m = 2p^s$, if $p^m > 0.05$, we have no evidence to reject the null hypothesis, otherwise, its profitability can be guaranteed in two years at a 5% confidence level.

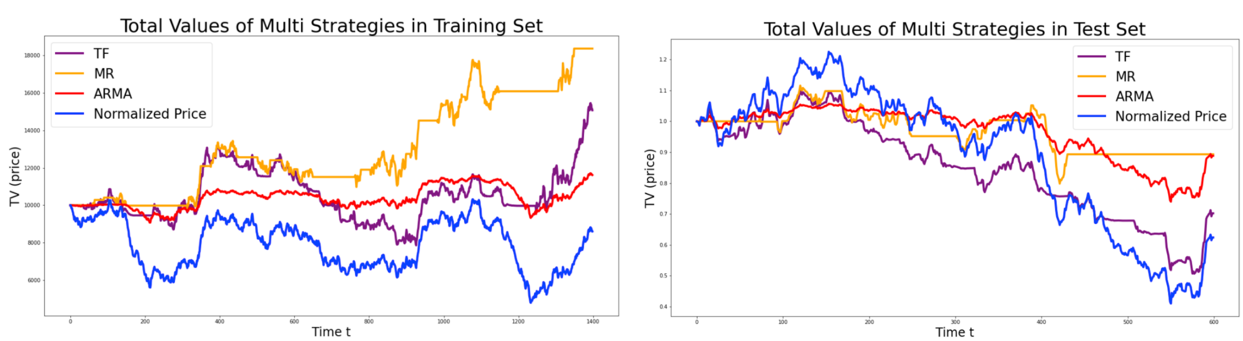


Figure 7: Total value of three Strategies in both Training Set and Test Set.

3. Results:

All the strategies introduced above have been conducted in both training set and test set, the results are shown in Figure 7. It is clear that, among the three strategies, MR strategy has the best performance in both the training set and test set, especially in the training set, its final return is almost 80%. The TF strategy also has an excellent performance in the training set with a 40% final return, but a huge loss in the test set. For the ARMA strategy, it is quite stable in both sets, but cannot bring any considerable profit for investors, three indicators may bring us a more comprehensive perspective than just with the figure.

The returns normalized by Sharpe Ratio, Calmar Ratio and VaR of each strategy are shown in Table 1. Coordinated with the result shown in Figure 7, MR strategy has fabulous returns even after being adjusted by volatility and the drawdown. In comparison with the TF strategy, it can bring more profit but with less loss and has better performance in the "bear market". Recalling Figure 7, the strategic total value of MR reaches the maximum at the end of the training set, so its MDD (Maximum Drawdown) in the training set is zero which brings about the null value in its CR calculation. Furthermore, since in this paper, strategies are all long-only and there is a consecutive drop of prices in the test set, all the things that three strategies can do are avoiding losses rather than gaining any profits, the MR strategy with a trailing stop signal generator has an excellent performance in even bear market, but TF does not. Although all strategies failed in gaining profit in the test set, the negative values of SR and CR in the test set still offer us some useful information in the anti-risk capacity: the ARMA strategy remain its stability in the test set with the highest SR, highest CR and lowest VaR among three strategies. MR strategy also has an excellent performance in this period with similar values in these indicators to ARMA strategy. According to Figure 7, the TF suffers a huge loss in the test set, and makes many awful operations during the 'recession' in the test set, so its values of three indicators are worst in comparison with the other two strategies.

To guarantee the profitability of these strategies in multi-period, their annualized Sharpe Ratios in the training set are all transformed into t-statistic by multiplying $\sqrt{1400/252}$, then their *p-value* are calculated. But only best strategy MR has significant evidence to reject the null hypothesis at the confidence level $\alpha = 0.05$, as 0.0456. When we repeat the calculation in the paper (Harvey & Liu, 2015) with M = 2, the *p-value* of MR becomes 0.09, without enough evidence to reject the null hypothesis. Therefore, the annualized Sharpe Ratio of MR strategy in the training set cannot guarantee the rejection of non-profitability in the time horizon of more than 1 year. This result and the high *p-value* of TR and ARMA strategies can be proven in the test set where all the *p-value* of single tests are insignificant to reject the null hypothesis.

In summary, MR has the best performance among three strategies because of both outstanding profit-gaining ability and distinguished anti-recession capacity. TF surely has a good performance in a bull market but its poor trailing-stop ability will shrink the profit accumulated from the stock price increment. ARMA model cannot prove its profitability in both the training set and test set.

	Training Set			**Test Set**		
	SR	**CR**	**VaR**	**SR**	**CR**	**VaR**
TF	0.4782	0.0117	0.0160	-0.8463	-0.0011	0.0183
ARMA	0.4200	0.0131	0.0064	-0.5490	-0.0006	0.0087
MR	0.8490	Null	0.0124	-0.5831	-0.0008	0.0102

Table 1: The indicators Annualized Sharpe Ratio (SR), Calmar Ratio (CR) and Value at Risk (VaR) of three strategies.

4.Discussion:

In this paper, we have introduced three self-finance long-only strategies of trend-following, mean-reverting and ARIMA model, and performed them in both the training set and test set of a synthetic price time series. However, several limitations and questions also come up in my programming and report writing. Firstly, limited by the size of synthetic time series, the ARIMA (2,1,1) model that is fitted by the training set doesn't match the time series generating function of ARIMA (1,1,1) model given in the Section 2.1. The difference will cause errors accumulating in daily return forecast. However, when conducting an ARIMA model fitting for the entire time series, the new model is similar to the initial generating function. Secondly, also because of the data size limitation, when running the hypothesis tests, none of the strategies can guarantee non-zero returns for more than a year. Since the 1400 observations in the training set are not large enough to ensure the strategy profitability even with 80% final strategic returns, it implies that when evaluating a strategy, a long enough time series is necessary, unless you really find a cash-machine-like algorithm to generate a fabulous return to pass the multi-tests. Lastly, in the test set, the price series looks like during a recession, three long-only strategies all fail to protect the principal, therefore, if the shorting behavior can be considered, better strategies can be refitted with these methods then outperform even in a bear market.

5.Bibliography:

[1] Dahlquist, Julie R, 2005. Trading currencies using moving averages - trend following versus contrarian strategies. Managerial finance, 31(5), pp.60–66.

[2] Lubnau, Thorben & Todorova, Neda, 2015. Trading on mean-reversion in energy futures markets. Energy economics, 51, pp.312–319.

[3] Temnov, Grigory, 2017. A strategy based on mean reverting property of markets and applications to foreign exchange trading with trailing stops. Applied stochastic models in business and industry, pp.Applied stochastic models in business and industry, 2017.

[4] Harvey, Campbell R. and Liu, Yan, Backtesting (July 28, 2015). Available at SSRN: https://ssrn.com/abstract=2345489 or http://dx.doi.org/10.2139/ssrn.2345489

[5] Andrew W. Lo, 2002. The Statistics of Sharpe Ratios. Financial analysts journal, 58(4), pp.36–52.