1 Introduction

In this report we generate a price time series using an ARIMA(1,1,1) model. We then define 3 self-financing dynamic long-only trading strategies and a static simple buy and hold strategy which is a typical benchmark against which the performance of the dynamic strategies will be evaluated. In total we analyse the performance of four trading strategies: i) Buy and hold; ii) Trend following based on EMA crossover; iii) Mean reversion based on ARIMA model and iv) Mean reversion based on ARIMA-GARCH model.

2 Methodology

2.1 Time Series Generation

In this section we will generate a price time series, that will be the basis for the trading strategies defined in Section II of this report.

- To ensure the stability of results each time we run the model, we will initialise the random number generator in Python using a seed equal to the unique student number.
- We will then generate a price time series based on the equation:

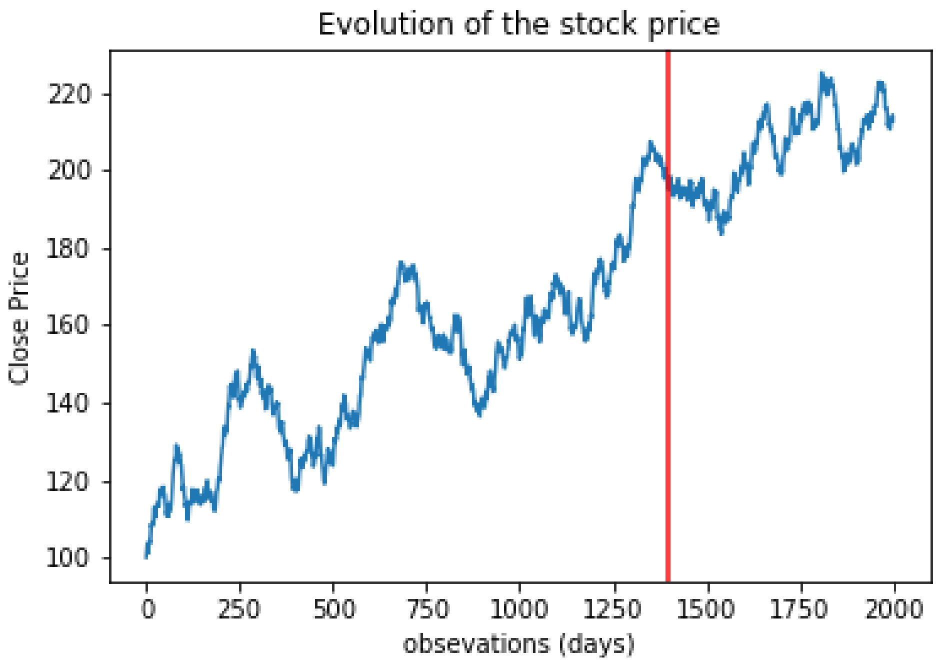
$$\Delta y_t - d = \phi\left(\Delta y_{t-1} - d\right) + \varepsilon_t + \theta\varepsilon_{t-1} \tag{1}$$

  The equation above depicts an ARIMA model of order (1,1,1), which is a generalization of an autoregressive moving average (ARMA) model that considers the observed time series as the integration(i.e. the difference) of the fundamental time series. [1]

- Then we divide the time series into a training set and a test set, representing 70% and 30% of the data respectively. The train set consists of the most recent 600 observations.

The application of the methodology described above for time series generation, results in the sequence of prices depicted in Figure 1. We can visually observe that the series of prices is non-stationary as it displays a strong upward trend.

Figure 1: Evolution of the stock price generated by the ARIMA(1,1,1) model



The red line indicates the split between the training set and the test set.

2.2 Trading Strategies

In this section we will define 3 self-financing long-only trading strategies with initial cash C_0 = 10,000. The self-financing condition for the update of cash and volume at each time step is given by:

$$TV(t) = C(t) + p(t)V(t) = C(t+1) + p(t)V(t+1) \tag{2}$$

for all time steps t. The long-only condition is given by $V(t) \geq 0$ for all time steps. No borrowing is considered and $C(t) \geq 0$ for all time steps. We will estimate the log returns of the strategy a at time t as follows:

$$r_a(t) = \log\left(\frac{TV(t)}{TV(t-1)}\right) \tag{3}$$

In order to benchmark the performance of the trading strategies defined in this section, we will initially build a simple buy and hold strategy with a total value given by $TV_t = X_t p_t + C_t$ at each time step t. The buy and hold strategy is very simple and it is a typical benchmark against which the dynamic strategies defined below will be evaluated.

2.2.1 Trend following based on EMA crossover

The trend following strategy follows market winners needs to recognise when the markets change trend, in order to identify potential trading opportunities.We are using a strategy based on the Exponential Moving Averages (EMA) indicators, because it gives more weight to the most recent periods, thus being a better representation of the recent performance of the asset.
Two separate EMA filters with varying look-back periods are created based on the price time series generated above. A crossover occurs when a faster moving average (i.e. the shorter period moving average) crosses a slower moving average (i.e. the longer period moving average). This provides an indication of an upward trend in price behaviour which is interpreted as a signal to buy.[2]

The following trading signals are generated from the EMA crossover strategy:

- Purchase the asset when the faster EMA exceeds the slower EMA.
- If the longer average subsequently exceeds the shorter average, the asset is sold back.(i.e. we liquidate our position in the asset and hold only cash in our portfolio).
- The strategy works well when a time series enters a period of strong trend and then slowly reverses the trend.

Mathematical formulation
The EMA indicator applies weighting factors which decrease exponentially for older observations. It is mathematically defined as follows[1]:

$$s_{t-1} = c(\lambda) \sum_{k=1}^{\infty} \lambda^k r_{t-k} \tag{4}$$

where λ, represents the degree of weighting decrease, a constant smoothing factor between 0 and 1. A higher coefficient discounts older observations faster.Considering two EMAs, where EMA 1 is the short term moving average while EMA 2 is the long term moving average, the crossover strategy would generate a trading signal defined as follows:

$$s_t = c(\lambda_1) \sum_{k=1}^{\infty} \lambda_1^k r_{t-k} - c(\lambda_2) \sum_{k=1}^{\infty} \lambda_2^k r_{t-k} \tag{5}$$

As explained above, if $s_{t-1} > 0$ then the faster EMA is above the slower EMA and trading signal is a buy, whereas if $s_t < 0$ the asset is sold back.

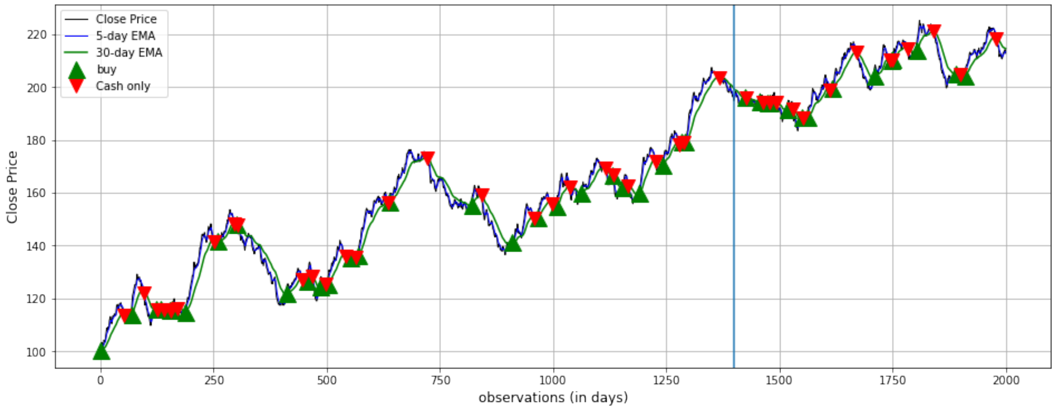
Implementation

- In the existing price series dataframe, we create a new column ‘Signal’ such that if the short term EMA is greater than the long term EMA then we set Signal value as 1 else its value is 0.

- From these ‘Signal’ values, the position orders can be generated to represent trading signals. When ‘Position’ = 1, it implies that the Signal has changed from 0 to 1 meaning that the short-term EMA has crossed above the long-term EMA, thereby triggering a buy signal. When ‘Position’ = -1, it implies that the Signal has changed from 1 to 0, thereby triggering a signal to liquidate our position in the asset.[3]
- We then calculate the strategy returns using Equation (3).
- To achieve better results, we want to optimise the window lengths of the faster and slower EMA. We create an array of values that represent the different lengths of short moving average window and long moving average window that we wish to run the tests over. Cumulative returns for different combinations of window lengths will be calculated by utilising the method described in the previous three steps. We select the optimal window lengths that achieve the best cumulative returns for this strategy.

The optimal window length that we obtain for the faster and slower EMA is **5 days** and **30 days** respectively. We use the selected optimal window lengths to generate trading signals from the EMA crossover strategy. As you can see in the plot below, a signal to buy (as represented by the green up-triangle) is triggered when the fast EMA crosses above the slow EMA. This shows a shift in trend i.e. the average price over the last 5 days has risen above the average price of the past 30 days. If the reverse happens, we liquidate our position in the asset and hold only cash in our portfolio.

Figure 2: Generating trading signals based on the EMA crossover strategy



The vertical line shows the split between the training set and the test set.

2.2.2 Mean reversion strategy: ARIMA model

It is widely accepted in finance that stock prices tend to follow a random walk model.If we think of the Random Walk as an AR(1) model then the coefficient of ϕ_{t-1} is 1. A conventional approach to handle unit root non-stationarity is by differencing.

The non-stationarity property justifies the application of ARIMA in modelling price-time series. Mathematically, an ARIMA(p,d,q) model requires three parameters: i) p: the order of the autoregressive process; ii) d: the degree of differentiation (number of times it was differenced); iii) q: the order of the moving average process

Given that returns are estimated as the difference of log prices, if a log price series is ARIMA (p,1,q) then log returns follow a stationary and invertible ARMA(p,q) model.[4]

Our motivation in using a mean reversion strategy based on ARIMA model, is that in general financial time series tend to exhibit mean reversion over very short time scale. Therefore, we will inverse the logic of the trend following strategy to apply mean-reversion, relying on the signals generated by the ARIMA model.

Therefore, we will predict log returns using an ARMA(p,q) model. We will then derive the predicted prices, and we will implement our trading strategy based on the following criteria:

- If the actual price at time t is greater than the predicted price, then we liquidate our position in the asset and hold only cash in our portfolio.
- If the actual price at time t is greater than the predicted price, then we buy the asset.
- The strategy works well over short time scales when mean reversion is more likely.

Mathematical formulation

As mentioned above we would expected for log returns to follow an ARMA (p,q) model which is equivalent to log prices following an ARIMA(p,1,q) model. The predicted log returns generated from an ARMA(p,q) model are defined as follows:

$$\widehat{r_t} = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots \phi_p r_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots \theta_q \varepsilon_{t-q} \tag{6}$$

We can derive the predicted prices for each time step t based on the series of predicted returns, as follows:

$$\widehat{y_t} = \exp\left(\widehat{r_t}\right) \cdot y_{t-1} \tag{7}$$

Trading signals can be generated by considering the difference between the predicted price and the actual price at time t:

$$s_t = y_t - \widehat{y_t} \tag{8}$$

The logic of mean reversion strategy dictates that: If $s_t > 0$, we liquidate our position in the asset, and if $s_t < 0$, we buy the asset.

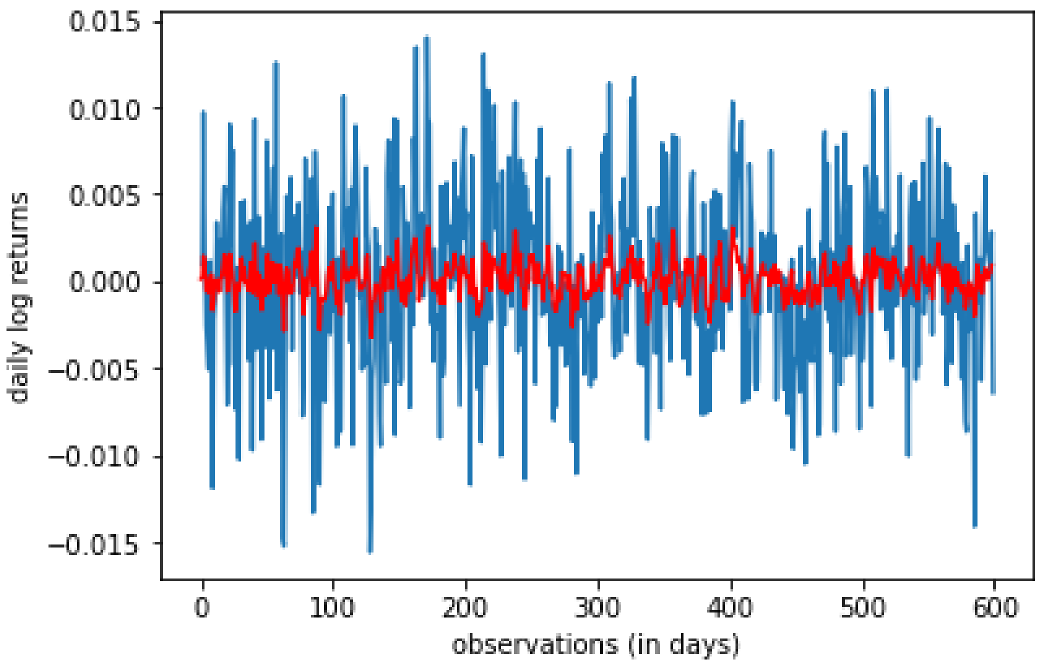
Implementation

- First, I take differences for log transformation of the price series to obtain the series of returns.
- Second, I identify the optimal order of the AR and MA process using the AIC criteria for model selection. The fitting process will test each combination of p, d, and q. The Akaike information criterion (AIC) is an estimator of prediction error and provides a means for model selection.AIC is given by:

$$AIC = 2k - 2\ln\left(\widehat{L}\right) \tag{9}$$

  where k is the number of estimated parameters in the model and L is the maximum value of the likelihood function for the model. Given a set of candidate models for the data, the preferred model is **the one with the minimum AIC value**. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting.[5] Based on the AIC criteria the best selected order for our series of returns is **ARMA(4,3).**. The optimisation of the lag parameters of the ARMA model is done on the train set. The actual vs. predicted returns estimated on the test set are presented in the figure below:

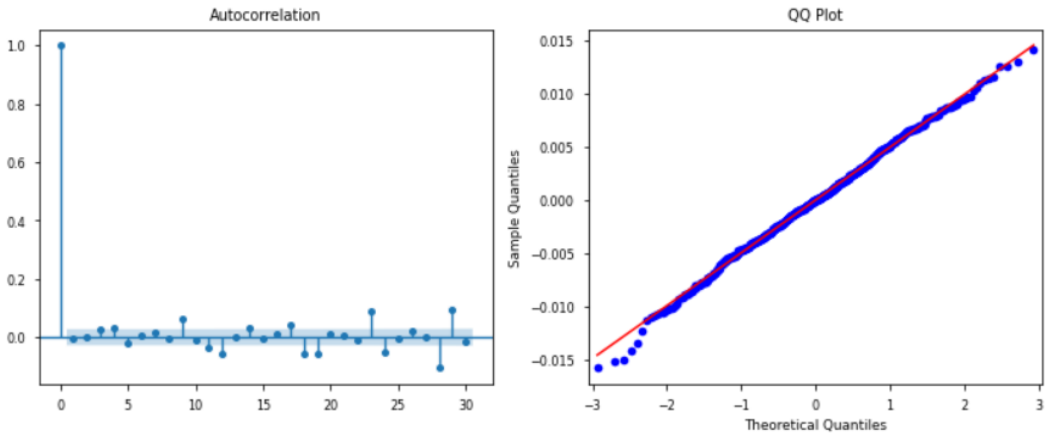
Figure 3: Actual returns vs. predicted returns based on the ARIMA(4,1,3) model



This figure illustrates the out-of-sample predictions of the ARIMA model.

- Finally, we want to understand whether the selected model is a good fit to our data. The QQ-plot indicates a slight deviation from the normal distribution and the correlograms show that there is still some significant pattern left in the autocorrelation function of the residuals. This means that the residuals of the selected model are not considered white noise and normally distributed, which undermines the goodness of fit of the ARMA (4,3) for log returns on the test set.

Figure 4: Residuals of the fitted model on the test set (ARIMA model)



2.2.3 ARIMA-GARCH model

The ARIMA (p,1,q) model addresses the non-stationarity of price time series and effectively removes the serial dependence in the expected component of realised returns. However, it fails to capture a significant property of financial time series, which is volatility clustering. Heteroskedasticity describes the irregular pattern of variation of an error term. Essentially, where there is heteroskedasticity, observations do not conform to a linear pattern. Instead, they tend to cluster.The result is that the conclusions and predictive values drawn from the ARIMA model will not be entirely reliable. [6]

How can we incorporate volatility clustering in our model? To do that, first we need to investigate whether there is any serial correlation in the series of squared error terms ε_t^2. If we observe auto-correlation in the ε_t^2, this is an indication that the variance of the original series of returns also exhibits auto-correlation.

One way to address this could be to create an AR model for the variance itself — a model that actually accounts for the changes in the variance over time using past values of the variance. This is the basis of ARCH model. Just like ARCH(p) is AR(p) applied to the variance of a time series, GARCH(p, q) is an ARMA(p,q) model applied to the variance of a time series. The AR(p) models the variance of the residuals (squared errors). The MA(q) portion models the variance of the process.[7]

$$\sigma_t^2 = \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 = \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{10}$$

And the series of residuals is given by:

$$\varepsilon_t = w_t \sigma_t \tag{11}$$

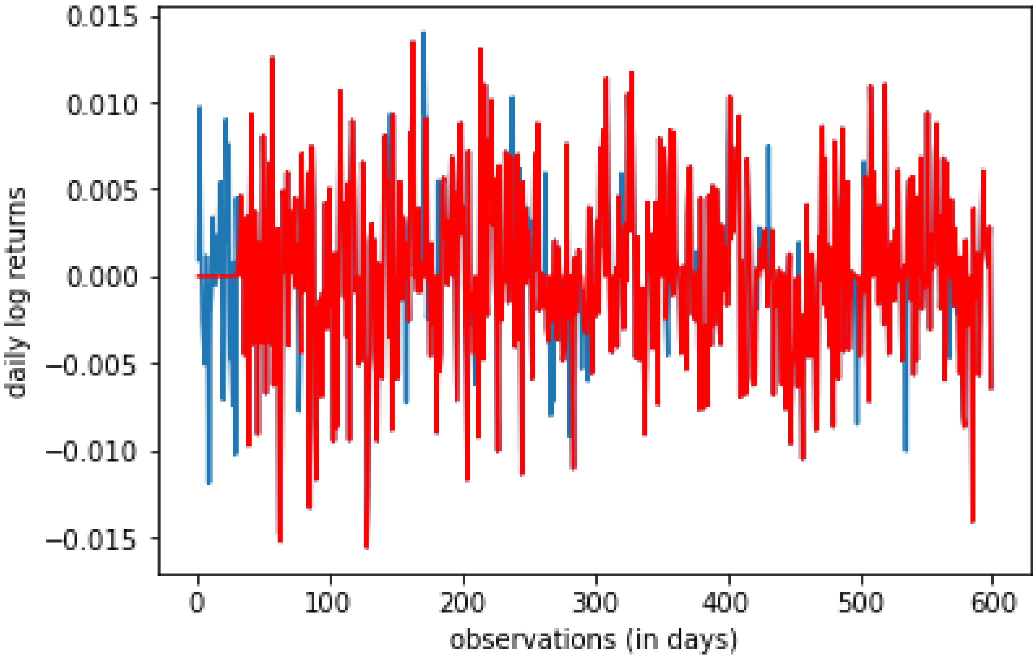
where w_t is white noise.

Implementation

The general process for a GARCH model involves the following steps:

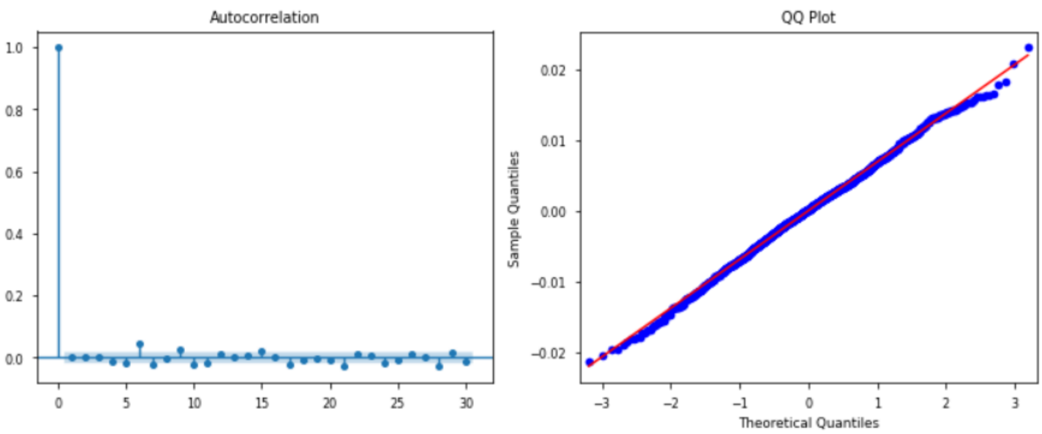
- First,we use the same procedure as before to find a good fit of the ARMA (p,q) model to our series of log returns. The looping procedure will provide us with the "best" fitting ARMA model, in terms of the AIC, which we can then use to determine the appropriate order of the GARCH model.
- After the fitting process we observe that the residuals of the ARMA model for returns closely resemble white noise. However when we look at the squared residual terms, we observe that there is substantial evidence of a conditionally heteroskedastic process via the decay of successive lags. The significance of the lags in both the ACF and PACF indicate that we need both AR and MA components for our model. Thus the GARCH model is considered appropriate.
- The residuals of the ARMA model, then will be added to the GARCH (p,q) model to predict volatility.
- The generation of trading signals is done on a "rolling" basis. For each day t, the previous 30 days of the logarithmic returns are used as a window for fitting an optimal ARIMA and GARCH model. The combined model is used to make a prediction for the next day returns.The following series of predicted vs. actual returns is obtained:

Figure 5: Actual returns vs. predicted returns based on the ARIMA-GARCH model



- Finally, the QQ-plot indicates a strong resemblance of residuals to the normal distribution and correlograms show that there is no significant pattern left in the autocorrelation function of the residuals. This implies that the ARIMA-GARCH model is a better fit to our test data.

Figure 6: Residuals of the fitted model on the test set (ARIMA-GARCH model)



- Based on the series of predicted returns we can retrieve the series of predicted prices using Equation (7). Trading signals can be generated by considering the difference between the predicted price and the actual price at time t:

$$s_t = y_t - \widehat{y}_t \tag{12}$$

Similar to the ARIMA strategy above, if $s_t > 0$, we liquidate our position in the asset, and if $s_t < 0$, we buy the asset.

2.3 Performance Indicators

In this section we will compute 3 representative performance indicators to evaluate the trading strategies defined above:

- **1. Sharpe Ratio**: a measure of risk-adjusted returns. The ratio compares the mean average of the excess returns of the strategy with the standard deviation of those returns. The benchmark for the performance of different strategies is usually the annualised Sharpe. For a strategy based on daily returns (as in our case), the annualised Sharpe Ratio is mathematically defined as:

$$SR(annualised) = \sqrt{252}\frac{\mu}{\sigma} \tag{13}$$

  where μ is the average of daily returns (we have assumed that the risk-free rate is zero), and σ is the volatility of daily returns.

- **2. Conditional Sharpe Ratio**:a measure of lower partial risk-adjusted excess returns of a strategy. CSR replaces volatility with Conditional VaR (i.e expected shortfall) in the denominator of the Sharpe ratio. Clearly if the downside risk (represented by the expected shortfall) is the major concern for the investor then the CSR is demonstrably favourable to evaluate the performance of trading strategies.[8] It is mathematically defined as:

$$CSR(annualised) = \sqrt{252}\frac{\mu}{ES(\alpha)} \tag{14}$$

  where $ES(\alpha)$ is the expected shortfall for a given confidence level $(alpha)$.

- **3. Calmar Ratio**: Calmar ratio is a Sharpe type measure that uses maximum drawdown rather than standard deviation to reflect the investor's risk. It is mathematically defined as:

$$Calmar = \frac{\mu}{MDD} \tag{15}$$

  where MDD is the maximum drawdown and it is given by: $\frac{MinThroughValue - MaxPeakValue}{MaxPeakValue}$

For each of these performance metrics we will provide two independent measures: i) within the training set(i.e. the first 1400 observations) and ii) within the test set (i.e. the last 600 observations). See Table below:

Figure 7: Summary of performance indicators

Train set	Buy &Hold	EMA crossover	ARIMA	ARIMA-GARCCH
Annualised Sharpe	1.05	2.46	2.91	2.46
Conditional Sharpe Ratio	1.43	3.27	3.96	3.30
Calmar Ratio	0.51	2.96	3.18	2.43

Test set	Buy &Hold	EMA crossover	ARIMA	ARIMA-GARCCH
Annualised Sharpe	0.46	2.28	0.51	0.55
Conditional Sharpe Ratio	0.61	2.98	0.67	0.70
Calmar Ratio	0.33	2.52	0.35	0.35

2.4 Statistical Tests

When we try many strategies, we need to adjust our evaluation method for these multiple tests because Sharpe ratios and other statistics will be overstated. Why is the adjustment for multiple testing important?
Consider a study that finds genetic linkages for Parkinson's disease. About a half a million genetic sequences are tested for the potential association with the disease. Given this large number of tests, many genetic sequences will appear to affect the disease under conventional standards. Given the large number of tests, more restrictive standards must be applied to lower the possibility of false discoveries. In fact, the identified genetic linkages from the tests have t-statistics that exceed 5.3.[9]

Based on the definition of the $t-stat = \frac{\mu}{\sigma\sqrt{T}}$ and the Sharpe Ratio (i.e. $SR = \frac{\mu}{\sigma}$, we obtain the following relationship:

$$t-ratio = SR * \sqrt{T} \tag{16}$$

Therefore, for a fixed T (where T denotes the number of returns in our sample), a higher Sharpe ratio implies a higher t-statistic, which in turn implies a higher significance level (lower p-value) for the investment strategy.

For the hypotheses that the strategies defined in this report have a non-zero Sharpe ratio, we should use a statistical test that controls for the multiple testing problem.There are two main approaches to the multiple testing problem: i) family-wise error rate (FWER) and ii) the false discovery rate (FDR). In the family-wise error rate, it is unacceptable to make a single false discovery. FWER calculates the probability of making at least one false discovery. In contrast, the FDR focuses on the proportion of false rejections. Here we will use two statistical tests to control for the FWER at a confidence level of 5%:

- **Bonferroni test:** Given the chance that one test could randomly show up as significant, the Bonferroni requires the confidence level to increase. The p-value adjustment is according to the following rule:

  $$p_i^B = \min\left[Np_i, 1\right] i = 1, 2...N \tag{17}$$

  where N is the number of tests.One important issue with the Bonferroni test is that it omits important information in the individual collection of test statistics.

- **Holm test:** provides a way to deal with the information in the test statistics, using a sequential testing method. Starting from the first test, we sequentially compare the p-values with their hurdles. When we first come across the test such that its p-value fails to meet the hurdle, we reject this test and all others with higher p-values.The p-value adjustment based on the Holm test is as follows:

  $$p_i^H = \min\left[\max_{j \le i}\left[(N-j+1)\, p_j\right], 1\right] i = 1, 2...N \tag{18}$$

  The Holm test captures the information in the distribution of the test statistics and it is less stringent than the Bonferroni because the hurdles are relaxed after the first test.

The Bonferroni and Holm adjustments to the p-values for our trading strategies (in ascending orders) are summarised below:

Figure 8: Adjustments to p-values to account for multiple testing

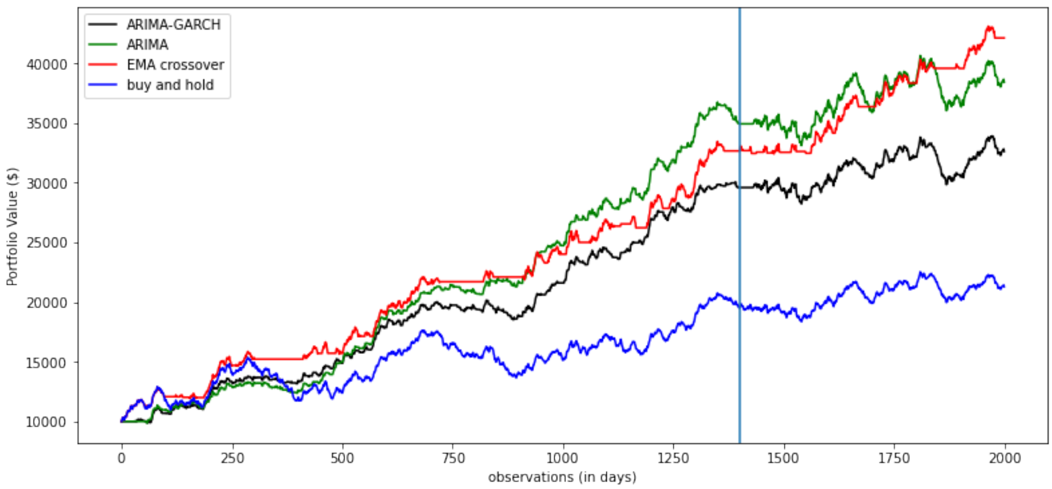
Train set	ARIMA	EMA crossover	ARIMA-GARCH	Buy & Hold
Unadjusted p-value	4.76E-12	3.98E-09	3.90E-09	0.01
Bonferroni adjustment	1.90E-11	1.59E-08	1.56E-08	0.03
Holm adjustment	1.90E-11	1.19E-08	7.80E-09	0.01

Test set	EMA crossover	ARIMA-GARCH	ARIMA	Buy & Hold
Unadjusted p-value	7.79E-11	0.20	0.22	0.24
Bonferroni adjustment	3.12E-10	0.79	0.86	0.96
Holm adjustment	3.12E-10	0.59	0.43	0.24

3 Results and Discussion

- We observe that the ARIMA mean-reversion strategy **considerably outperforms** the other strategies in the train set. However, **the EMA crossover strategy performs better on the test set**. The cumulative performance of different trading strategies is summarised in the figure below:

Figure 9: Cumulative performance of different trading strategies



- The over-performance of ARIMA mean reversion strategy in the train set, is also supported by the higher values it achieved across several performance indicators such as: Annualised Sharpe Ratio, Conditional Sharpe Ratio and Calmar Ratio. However, in the test set the best-performing strategy is the EMA crossover, while the other strategies achieve sub-optimal results in terms of the excess returns per unit of risk. We also see that ARIMA-GARCH strategy peforms relatively better than ARIMA strategy in the test set. This is consistent with the results of our residual analysis which suggested that ARIMA-GARCH was a better fit to our test set data. Nonetheless, its performance still lags significantly behind the EMA crossover.

- We also observe that after accounting for multiple testing in the train set, all four trading strategies have a non-zero Sharpe ratio, which is significant at a confidence level of 5%. **While in the test set, the only statistically significant non-zero Sharpe Ratio is the one generated by the EMA crossover strategy.**

- Overall, we conclude that one could expect a statically significant higher level of excess returns per unit of risk under the EMA crossover strategy compared to the other strategies reviewed in this report. This is because EMA crossover strategy is more effective in capturing the upward trending behaviour displayed by our generated price time series.

References

[1] Algorithmic Trading Course Notes. 2021.

[2] Fidelity. "How to Trade With Exponential Moving Average Strategy". 2021.

[3] Pratik Nabriya. "Generating Trade Signals using Moving Average(MA) Crossover Strategy". 2020.

[4] H. Dong. "Predictive power of ARIMA models in forecasting equity returns: a sliding window method". 2020.

[5] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. "The Elements of Statistical Learning,Second Edition".

[6] Dominic Roe. "Developing a Hybrid ARIMA-GARCH Model for Long-Term Forecasting of Financial Time Series". 2018.

[7] Piotr Fryzlewicz. "Lecture notes: Financial time series, ARCH and GARCH models".

[8] Carl Bacon. "How sharp is the Sharpe-ratio? - Risk-adjusted Performance Measures ".

[9] Campbell Harvey and Yan Liu. "Evaluating Trading Strategies ".