Algothon 2019 – Phase 2

Phase 2 Objectives:

1. *Work on implementing a proof-of-concept of your model from your initial hypothesis.*
2. *Perform a simple backtest of your model using real data. Discuss the results and plot the cumulative out-of-sample performance.*
3. *Produce the comparative score – this score will be used to compare your model to baselines. The score can be produced by following the notebook and using your backtest results. We are looking to see evidence that your model has some predictive power; this score will enable us to evaluate this.*

A key challenge in Phase 2 is to overcome look-ahead bias. This means your model can only use information that was available before each trading period. This can be challenging because when training a model, the historic data often has different forms of data leakage that will introduce some future information, such as revisions, averaging or normalisation that includes future periods. Often when these leakages are genuinely corrected, the initial predictive power disappears. Therefore, we also want to encourage models based on sound economic, market or psychological concepts that may reveal some form of inefficiency in the market that could be exploited to generate alpha (and improve market efficiency).

Don't worry if your model is not producing alpha yet. Interesting ideas and concepts are valuable entries as well. Throughout this programme you will learn how to improve your model and you will be able to resubmit when your results improve.

Also note that at this stage you don't need to implement a full trading algorithm, just a predictive score for the beginning of each period. Your model should produce a positive number where it is predicting a higher probability that the variable will go up over the next period, and negative where it is predicting it will go down over the next period.

A model with some predictive power is the basis of most alpha gen models, but it is also only the first step. In later phases, you will be tasked to refine your model to show the risk adjusted return, the post-transaction cost return, and overcome various other challenges such as overfitting, regime change, or alpha decay.

Steps for phase 2

1. *Build and refine a model that implements the idea and produces a forecast.*
2. *The model forecast should be a score, for each period, and for each asset in your target asset universe; the score should be a prediction of the level or movement of a price, return, or some other unknown asset variable.*
3. *Run your model over multiple periods and multiple assets to produce a backtest which should show the forecast performance versus the actual movement; you can use the notebook provided to produce this*

Recommendations:

- Divide the data your model uses into training, validation and test sets.
- Produce a backtest over several hundred periods, e.g. over several years for a daily model.

Notes for Trading Algo Models

You will only need to submit the predictive model results in phase 2. You can start separating the predictive aspect of your trading model by providing 0 in periods with no trades, 1 when you want to buy, and -1 when you want to sell. You can also adjust these scores based on the scale and confidence/probability of the prediction.

Example – Phase 2

Model Name:

Crossover Triggered Momentum Reversal

Model process and implementation

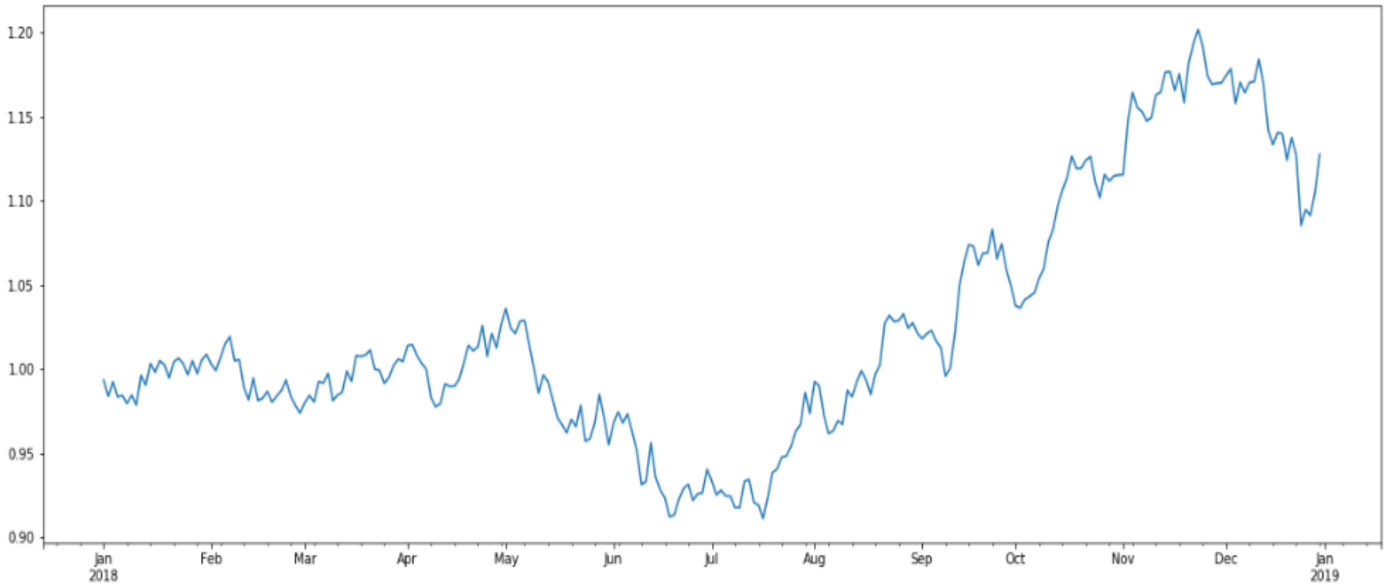
We are using a moving average of a period of 200 data points and 20 data points. Our back test has shown that these are the best look back periods for the asset. A buy signal is generated when the short-term average crosses above the long-term average, while a sell signal is triggered by a short-term average crossing below a long-term average. The cross over identifies a shift in momentum and forms the basis of our strategy.

Model Testing Details

- Asset Universe: FTSE 100
- Training Period: 2015-2016
- Validation Period: 2017
- Test Period: 2018

Model Back Test Performance

Show the cumulated returns only on the test period



Discussion of results and planned enhancements

There are several ways in which our model can be refined:

- Using EMAs which react quicker to market changes, allowing for faster entry and exits to trades when prices change.
- Use Bollinger Bands in addition to the MA, as these reflect volatility in conjunction with the trend portrayed by the MA.
- Create MA ribbon, that uses an array of MAs of different lengths to indicate the trend strength. This can be used with both SMAs and EMAs

Evaluation

We are using a unitised performance measure of the lagged correlation between the model forecasts and realised variables such as returns. The example in the notebook uses returns and implements the following methodology to create a uniform measure of forecasting ability:

- **Data setup**: identify universe of assets, get returns for each period per asset;
- **Forecast model**: run model to get model/forecast score for the start of each period;
- **Forecast evaluation**:
  - **Scale forecasts**: unitize forecast vector by taking scores, dot-prod with its transpose, divide forecast by sqrt of ratio;
  - **Calculate periodic returns**: for each period: take scaled forecast at beginning of period, multiply by end of period asset-return to give asset-level 'model-return'; sum to get portfolio-level 'model-return';
  - **Calculate cumulative returns**: multiply portfolio-level 'model-return' return across all periods to get total 'model-return' across backtest timeline.

Note: This is an **unconstrained** measure of performance. We do not consider risk adjustments and trading costs at this point.

Please see separate attachment for notebook implementation, *backtest-v8{.ipynb,.html}*.

Learning Resources

Further Reading

If you need more help to implement your idea, please start with the following guides and references:

- https://www.quantstart.com/articles/Successful-Backtesting-of-Algorithmic-Trading-Strategies-Part-I
- https://www.quantstart.com/articles/Successful-Backtesting-of-Algorithmic-Trading-Strategies-Part-II
- https://www.quantstart.com/articles/Research-Backtesting-Environments-in-Python-with-pandas

If you are eager to understand more and anticipate the later phases, we also recommend the following:

- Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk (McGraw-Hill Library of Investment & Finance) by Richard C. Grinold, Ronald N. Kahn. Book available in the Central Library of Imperial College
- Arnott, Robert D. and Harvey, Campbell R. and Markowitz, Harry, A Backtesting Protocol in the Era of Machine Learning (November 21, 2018). Available at SSRN: https://ssrn.com/abstract=3275654 or http://dx.doi.org/10.2139/ssrn.3275654
- López de Prado, Marcos, What to Look for in a Backtest (August 11, 2013). Available at SSRN: https://ssrn.com/abstract=2308682 or http://dx.doi.org/10.2139/ssrn.2308682
- Lo, Andrew. (2003). The Statistics of Sharpe Ratios. Financial Analysts Journal. 58. 10.2469/faj.v58.n4.2453. Available at: https://www.researchgate.net/publication/228139699_The_Statistics_of_Sharpe_Ratios
- Garleanu and Pedersen (2013). "Dynamic trading with predictable returns and transaction costs," Journal of Finance. Available at: http://www.lhpedersen.com/research