Model Card for Investment Strategy Final Project

**Model Description**

## Model inputs

The model takes as its input 19 different sets of data covering financial, macro-economic and behavioural features across 14 different equity markets.

The model also takes the excess returns (returns over cash) for each market.

## Model architecture

The model is a random forest classification model. Its main hyperparameters are set in the in-sample (or training/validation phase) using Bayesian optimisation and then used in the out-of-sample or testing phase.

## Model output

The output of the model, arrived at in the testing phases, is both the accuracy score (53% in the testing phase) and a set of important features identified through permutation analysis.

**Model Performance**

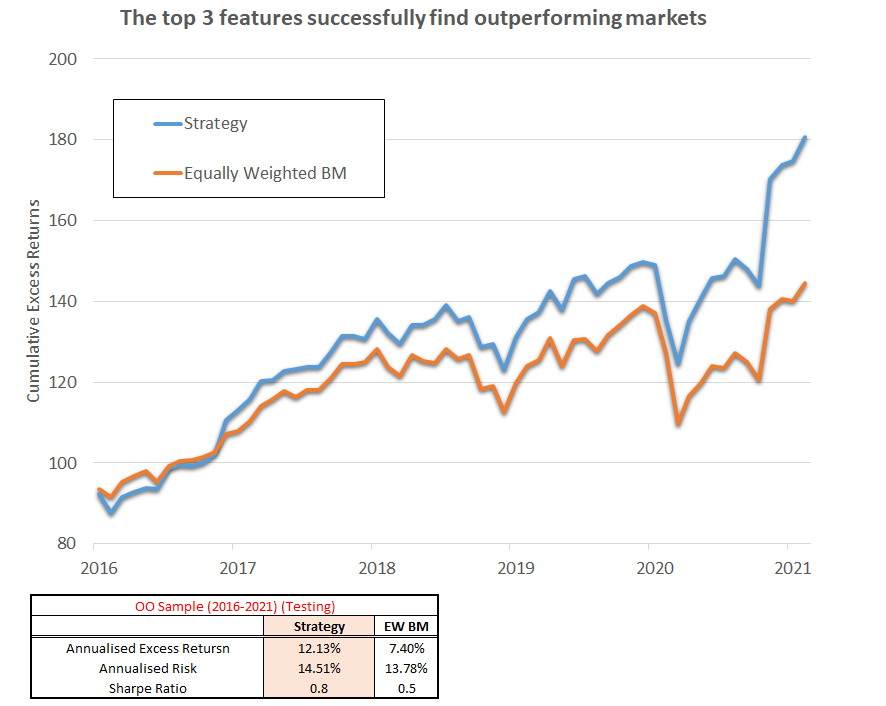
For financial models, the accuracy score noted above is actually less useful than with non-financial models. The noisy nature of financial data means that most of the metrics beloved by ML practitioners will have very low outcomes.

For example, a standard monthly linear regression using financial data to predict excess returns will typically have an < 1%. Similarly, ML models applied to financial data will have accuracy scores in the 50% – 60% range since financial data is notoriously noisy.

To assess the model’s performance we look at the three most important features that it identifies as predicting excess returns and then use these to select markets out-of-sample. In our case the most important features are VRP, MOM1 and EQSTYLE.

As noted in the graph below, these three features that the random forest identifies as the most important for selecting outperforming markets, do indeed generate significant outperformance over the testing period.

Even with an accuracy of 53%, the model strategy outperforms an equally weighted benchmark. The Sharpe ratio (the risk adjusted return) – an important finance metric – is also more attractive for the strategy than for the benchmark.



To me this indicates that the model is performing well and that the random forest approach has indeed identified features that are very important for predicting outperformance.

**Model Limitations and Trade Offs**

Before we rush off to set up a hedge fund to run this model and perhaps buy a small island with the profits (say Indonesia!), there are a few limitations we should consider:

1. The model does not consider trading costs but assumes that you can rebalance the portfolio of country positions costlessly. This is not the case and frequently perfectly good models fail to work in practice since the cost of running them eats away any alpha.
2. While this model works fine, it would be useful to run the same model in a completely different universe – say the Emerging Markets – to see if it is still effective. If it is, this would certainly boosts the case for this model truly predicting market returns.
3. While the model works over the testing period, a more dynamic approach that adapts to changing conditions might be useful to explore. This could be achieved by using an expanding window to re-run the random trees against and observe how the features change in importance.
4. We have, by necessity, only selected a small number of features to allow the random forest to iterate over. In reality we would want a far larger universe with sentiment, volatility and other asset allocation factors to consider.
5. Our portfolio construction is also very basic being a simple, equally weighted approach to creating our strategy portfolio. A more complex approach that takes into account the correlations between different countries might prove useful here.

Despite these limitations, I think the random forest approach and the strategy it produces work well in the testing phase.