Model Performance Report

A baseline measure of model performance was done to evaluate the performance of both linear and non-linear models to classify the market regimes using a Logistic Regression model and a random Forest model. The Logistic Regression model was only slightly better than random guessing in three classes, which is around thirty-six to thirty-seven percent. Analysis of the confusion matrix showed that there was a strong bias to predicting the sideway regime and much fewer correct predictions of uptrend and downtrend classes. The trend shows that the model had a challenge to disentangle this overlap of regimes and instead fell to the most frequent or less uncertain class.

The weaknesses of the baseline model can be explained by the fact it assumes that the decision boundary is linear in a feature space where highly correlated indicators dominate. High interdependence between trend-relevant attributes lowered the predictive capacity of the model to draw significant class boundaries leading to a breakdown in the predictive trust into one prevailing category. That is why, Logistic Regression model was only used as a diagnostic benchmark as it proved that linear approaches were not sufficient to solve this issue.

The Random Forest model indicated more balanced prediction behavior although this produced the same overall accuracy. It had a more balanced distribution of the predictions between all of the three classes with better rates of recognition of the uptrend regime and downtrend regime. This mean is an indication of the ability of the model to manage non-linear interactions between volatility, trend as well as momentum features. Even though the accuracy of headlines did not improve significantly, the Random Forest model generated the results which were more understandable and had a more practical use in regard to the identification of the regimes.

The important features analysis also explained the learning behavior of the model. The measures of volatility and moving averages proved to be the most significant predictors, ATR, and MACD came next. Momentum indicators and RSI played a mediocre role, whereas volume-

based features always occupied the last place in the list of importance. These are in line with the search analysis and financial intuition, and the results are that the model learns mostly the market structure and the dynamics of regime, and not the short-term noise.

Even though there are these strong points, there are weaknesses. The two models were characterized by difficulty when it comes to the issue of transitioning regimes where the indicators signals are ambiguous and the borders of the classes are blurred. Also, the usual metrics of accuracy are not effective enough to assess the model performance in this sense, which indicates the necessity to use some other metrics of evaluation, like the macro-averaged scores.