Multi-asset Prediction Model Description

Assumptions

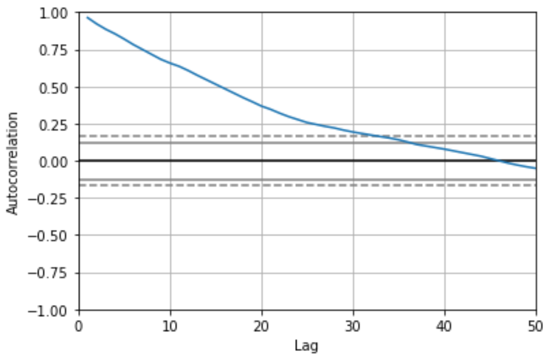
- Encoded integer value for each variable has ordinal relation.
- The markets are dynamic correlated, many indicators (economic, technical, sentiment etc.) can give impact on the MSCI index price.
- The trading decision critical value is created according to last 12 months average return with standard deviation, due to the assumption that equity market drifting by time.

Data treatment

1. Index price evaluations

evaluation metrics	value
max monthly gain	0.1222
max monthly loss	(0.2015)
maximum drawdown	0.6517
annualized Sharpe Ratio	0.1645
average rolling 3years SR	0.6740

2. Months lagging features
   Generate 15 months price lagging features from MSCI index data according to autocorrelation plot. Plot below shows up to 15 months lags, the correlation is above 0.5.



3. Shift index price one month up
   The date of feature data and index data are matched, but to forecast future one month index price according to current feature, we need to shift the index data one month back.
4. Data truncation
   Because of the lagging features and shifted index price, first 15 rows feature data contains NA, so that we need to cut them off. Now there are **237** months in total instead of 252 months.
   So far, we have 237 months and (59+15) features data and 236 index price data.
5. No standardization
   XGBoost's tree structure decides that data standardization is not necessary.

Methodology

1. Train, test, forecast data sets splitting - KFold CV
   Keep last 2 months data as validation & forecast data, then split rest of them into train & test sets with adjustable ratio (currently 25% test data).

Use K-fold cross validation resampling method to evaluate the model performance on train dataset.

2. Forecast model – XGBoost
XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. With the default XGBregressor, the model is already overperformed comparing with ARIMA & LSTM model:

	ARIMA(benchmark)	LSTM	XGBoost
MAE	31.32	18.48	11.12
RMSE	33.29	26.54	13.71

Here we do not do parameter tuning for XGB model, neither LSTM model, since after tuning, the model performed worse than the default one. But there are space to improve the model performance from the configuration perspective.

3. Feature importance - SHAP

Choose XGB model over LSTM model is mainly because NN model can not be explained by features, while XGB can achieve that easily by feature importance functions. Here we use Shapley value, which is the average of the marginal contributions across all permutations.

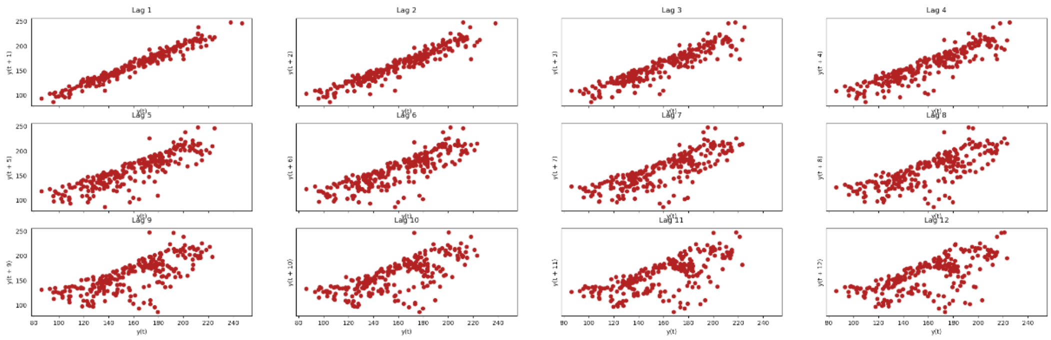
4. Trading decision

Instead of predicting last month index price, we predict last 2 months index prices to calculate the return rate. Then comparing the last return rate with last 12 months average absolute return rate to make decisions.

Model predicted return rate is 0.003, and real price mean absolute return rate (L12M) is 0.06 with 0.043 standard deviation. Predicted return rate does not surpass 0.06±0.043.

Conclusion

- Neutral on equity markets over the next month.
- Top 3 important features are index price lag_1, lag_3 and lag_4. From the scatter plot below we can tell lag_1, lag_2, lag_3, lag_4 are all have strong correlation with dependent variable, while Lag_12 is already disperse a bit.



- Besides lags, **EU HY spreads**, **Eurozone Retail Sales**, **US Credit Standards**, **US Indus Prod Index** are the most influential features for the model.