

Asset Allocation Assignment

Planning steps:

- Exploratory data analysis (null numbers, distributions, data type, missing values, correlation between features and index)
- Feature Selection
- Feature Engineering
- Machine Learning Index Forecasting (train/test time series split, training & testing, models evaluation)
- Asset Allocation Strategy
- Strategy Backtesting and portfolio performance evaluation
- Conclusions
- Further steps

Assumptions:

- MAPE is used as a Machine Learning Index Forecasting evaluation metrics
- Recursive Feature Elimination and correlation coefficients are used for feature selection
- The features based on an index like moving average or previous month price are not used for prediction
- RandomForest used for forecasting of the index based on literature review
- The prediction of the index is made one month ahead. The model training period is a sliding window of 12- 24 months prior to the prediction
- Strategy is based on moving averages crossovers
- Fees, dividends, transaction costs, taxes are not taken into account

Methodology

Data exploration

Data exploration demonstrated 0 missing values and 0 null numbers, all of the float type. Correlation between the features shows a high correlation between for example US Real Personal Income and US Real Personal Income excl Trans, which also have a high correlation with the index. The highly correlated (correlation coefficient > 0.97) features that contain similar macroeconomic/sentiment information are excluded from the features.

Lagged, aggregated, and averaged parameters can be considered as an additional feature. Percent changes in features do not demonstrate a noticeable correlation with the index.

Highly correlated are removed and RFE and Boruta Algorithm are used for Feature Selection

Machine Learning Index Forecasting

To prevent look-ahead bias split train/test data for time series is done by sliding 12 or 24 months period for training and one month ahead test.

Train	Test		
	Train		Test

Model

Random Forest is used based on the results of the literature review (prediction of indexes from macroeconomic data). The number of training months used for model training was obtained by optimization. It is not a straightforward value because the economic regimes were different and monetary policies as well. MAPE for the time frame is used as a model evaluation metric.

Asset Allocation Strategy, Backtesting, and Evaluation

As an asset allocation strategy moving average crossover is implemented. The predicted index value is used for calculating the moving averages for the next month and create buy/hold/sell signal depending on the moving averages (long moving average < short moving average = sell signal, short < long = buy signal, the rest is hold).

Backtesting of buy/hold/sell signal using historical index value, calculating returns and portfolio metrics are done and reported to Html file.

Conclusions

- Information in the features can be used for one-month ahead prediction with MAPE = 4.1% for the data timeframe.
- Highly correlated features like Fed Balance Sheet/Fed Excess Reserves can be replaced with one feature
- Features that are important for index prediction are different for different periods (for example, Fed Balance Sheet after 2008 crisis), that is why the sliding training data is used to retrain the model.
- Moving average crossover asset allocation model was implemented using index prediction one month ahead to predict moving averages crossover
- Portfolio performance evaluation shows that model based on predicted values outperforms the model based on real historical index values. Overall the portfolio performance is poor in terms of total return (7.6%), sharpe ratio (0.21), and drawdown (-52%). The benchmarking with buy and hold strategy should be done for more precise evaluation.

Further steps:

- Feature Engineering (lagged data, aggregated and averaged data, difference economic periods/cycles, previous index price information)
- ML model tuning and optimization (historical index data used for training, recent data has higher weights than older data)
- Testing other algorithms like QRF, ANN and in combination with Time Series Algorithms
- Strategy optimization in terms of long and short windows based on portfolio performance metrics
- Testing and benchmarking different buy/sell strategies

Files included:

S_Main.py	python script
Technical_requirements.txt	All the necessary libraries
html_assignment_report.html	html_assignment_report
figures	figures required for the report
Assessment	jupyter notebook