Predicting Bull & Bear Markets with Machine Learning

Kelly, Jonathan, Kai Lin, Shu Ting, Jessica

Table of contents

01
Introduction

Project Background & Aims

Research Aim

This study examines whether machine learning (ML) models can enhance stock market state predictions (bull/bear markets) compared to traditional econometric models.

Reference Study

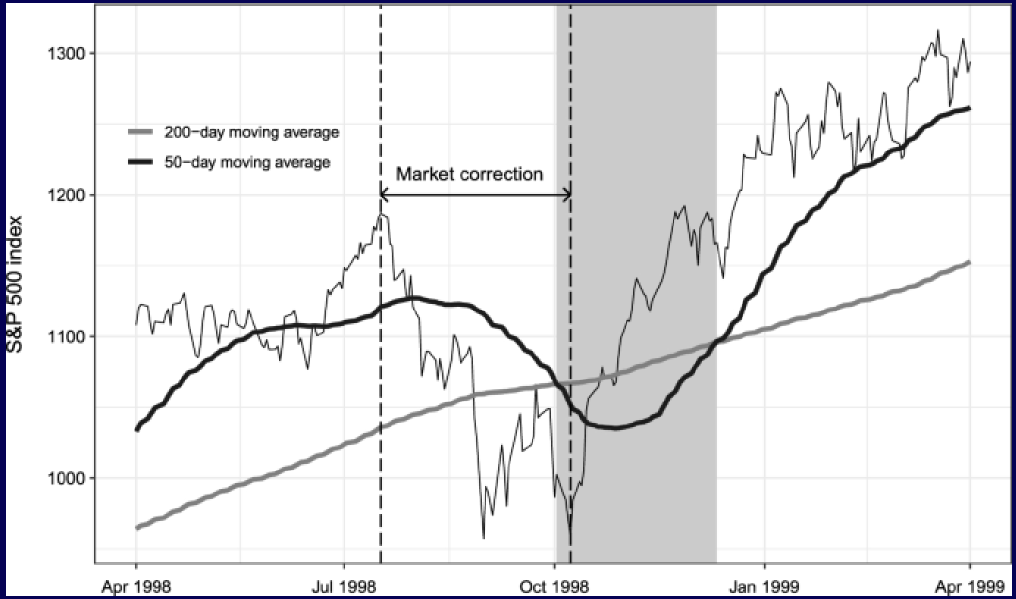
Nyberg, 2013, examined the effectiveness of various types of static and dynamic binary probit models in predicting bull/bear markets

How does our project value-add?

Applying ML classification detects any non-linear interactions that may be present (Nyberg's econometrics models assume linear relationships)

Definition of bull/bear market

Bry & Boschan Algorithm



- 2 sided moving average (using the S&P500 return in the last 6 months & the next 6 months to locate peaks and troughs)
- Minimum length of bull/bear market = 6 months
- 1 complete cycle = at least 15 months

02
Data

Data Source

Overall Dataset

Robert Shiller's website

- US Stock Price, Earnings and Dividends as well as Interest Rates and Cyclically Adjusted Price Earnings Ratio (CAPE)

Amit Goyal's website

- Variables: Useful variables for predicting equity premiums – stock-related predictors useful for predicting bull-bear market states.

FRED-MD

- Macroeconomic indicators (useful for predicting bear markets): Unemployment rate, federal funds rate and industrial production growth rate

Data Transformation*

Variable	Transformation
Quarterly and yearly variables, variables with all NAs, real prices (used nominal prices**), CAPE (TR_CAPE: more comprehensive measure of market returns (Siegel, 2016)***)	Removed variables
Ratio variables	Took the natural logarithm of the variable before first differencing it.
Index variables	Took the first difference of the variable

- Ensured stationarity of variables pre- and post-transformation by conducting unit root tests.

*Followed FRED-MD Guidelines

**Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments* (10th ed.). McGraw-Hill Education.

***Siegel, J. J. (2016). *Stocks for the long run: The definitive guide to financial market returns and long-term investment strategies* (5th ed.). McGraw-Hill Education.

Data Transformation (Lags)

Forecast Horizons

- Selected Horizons: 1, 3, and 6 months - investor interest in short- to medium-term returns.
- Excluded Horizon: 12 months - difficulty in predicting longer-term market trends

Lags for Predictors

Up to 6 Lags for Each Predictor: Captures potential delayed effects on market state while avoiding model complexity.

Lag Selection by Horizon:

- 1-Month Horizon (h=1): Lags 1 to 6
- 3-Month Horizon (h=3): Lags 3 to 8
- 6-Month Horizon (h=6): Lags 6 to 11

Lags for Dependent Variable

Dependent Variable Lags as Predictors:

- Used lags 7 to 12 for h=1, lags 9-14 for h=3, and lags 12-17 for h=6.

Market State Classification:

- Bry and Boschan Algorithm (R's bbdetection package)
- Peaks/troughs identified with a two-sided moving average - next and last 6 months of S&P500 nominal prices

Data Transformation (Final Cleaning)

01

Initial Observations:
594 after aligning variables' start and end dates and accounting for missing data.

02

Market State Adjustment:

- Removed last 6 observations to account for Bry and Boschan algorithm accuracy requirements (future S&P500 data needed).

Lag Adjustment:

- Removed first 17 observations to remove any NAs from lags and standardise data for train-test split for all models.

Final Dataset: 571 observations.

Train-Test Split

Total observations: 571

Training Set: 421

Test Set: 150

Validation set required only for boosting

Training Set: 321

Validation Set: 100

Rolling Window

Time-series data:

- Order of preservation matters
- Ensures consistency of forecast horizon
- Recursive estimation: maintain time sequence while evaluating model performance

03
Models

Models Used

- Logit
  - Standard Logit (Benchmark)
  - Lasso Logit
- Tree-based models
  - Random Forest
    - Random Selection
    - Bagging
  - Boosting
    - Gradient Boosting (GBM)
    - Extreme Gradient Boosting (XGboost)

New Models

Lasso Logit

- Logit regression with L1 regularisation
- Chosen for feature selection ability

XGBoost

- Idea of the GBM algorithm is kept, but incorporates more mechanism
  - Uses both first and second order loss function
- More optimised and efficient
- Able to get better performance than GBM

Tuning Models

Lasso Logit

- The model was ran using rlassologit from the hdm package
- The optimal lambda provided by the function was chosen as the lambda for prediction, no additional tuning
- Lambda updated dynamically in rolling window

Random Forest

- The number of predictors chosen for each tree was the default choice:

$$m = \sqrt{P} = \sqrt{276} \approx 16$$

- 500 trees with a minimum node size of 10 was used.

Tuning Models

Boosting

- Cross validation was used for tuning the tree size and finding the simplest model with the lowest misclassification rate.
- Across all forecast horizons, boosting with depth = 5 generally outperformed depth = 2
- Depth = 5 was used for all models for ease of comparison

Model Type	Forecast Horizon	Best Tree Size	Misclassification
GBM	1	510	0.07
	3	496	
	6	488	
GBM (Sample Mean)	1	952	
	3	1058	
	6	1090	
XGB	1	174	
	3	179	
	6	145	

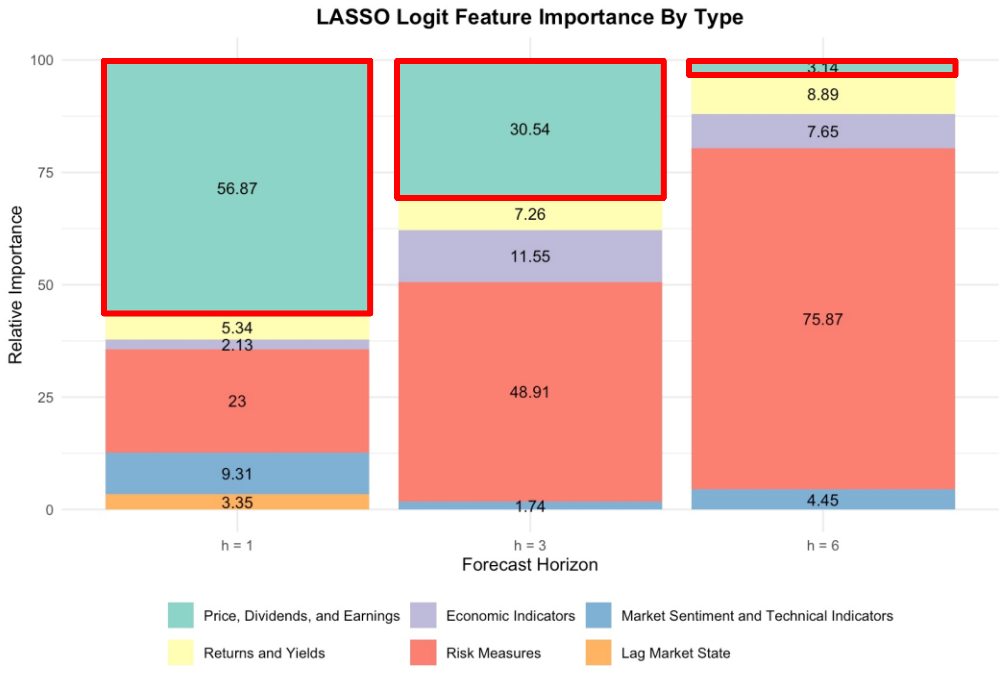
04
Results

Feature Importance

For the machine learning models, the predictors were grouped into 6 groups:

- Price, Dividends and Earnings
- Economic Indicators
- Returns and Yields
- Risk Measures
- Market Sentiment and Technical Indicators
- Lags of Market State (dependent variable)

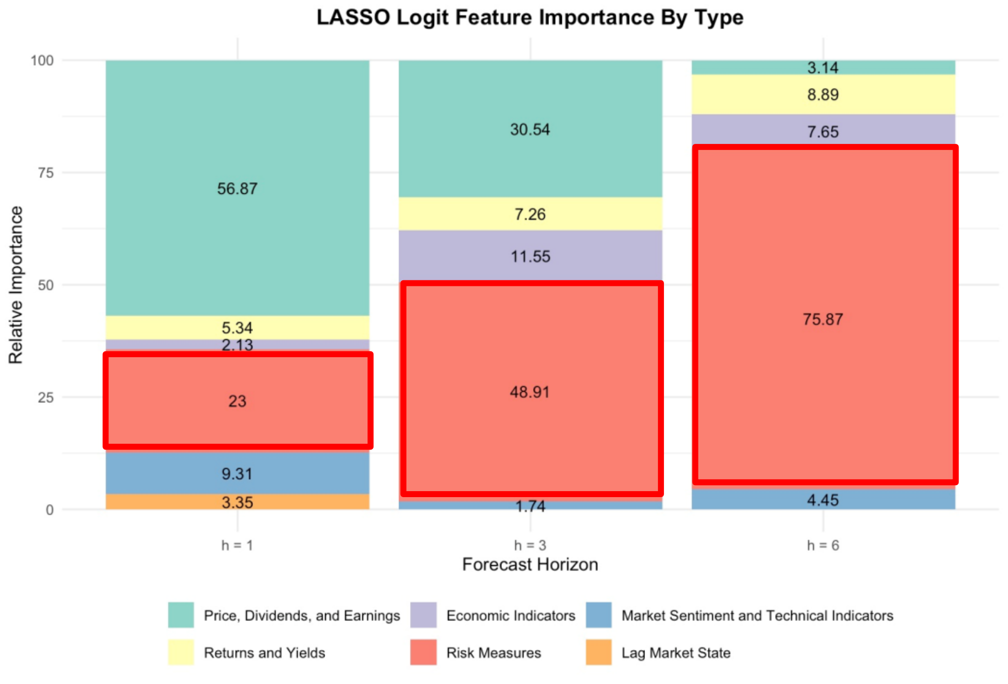
Feature Importance – Lasso Logit



<u>As forecast horizon increases</u>

Importance of short-term indicators on recent market conditions drops

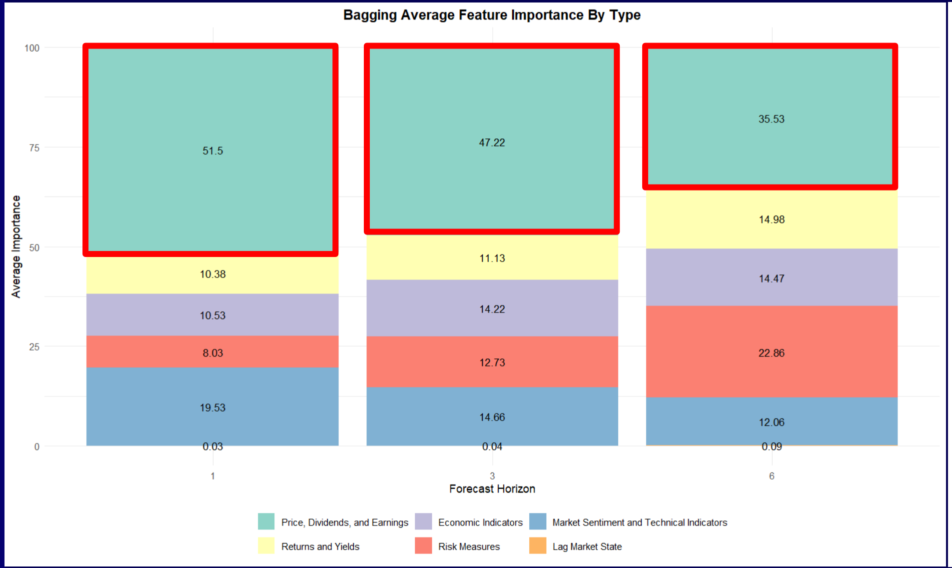
Feature Importance – Lasso Logit

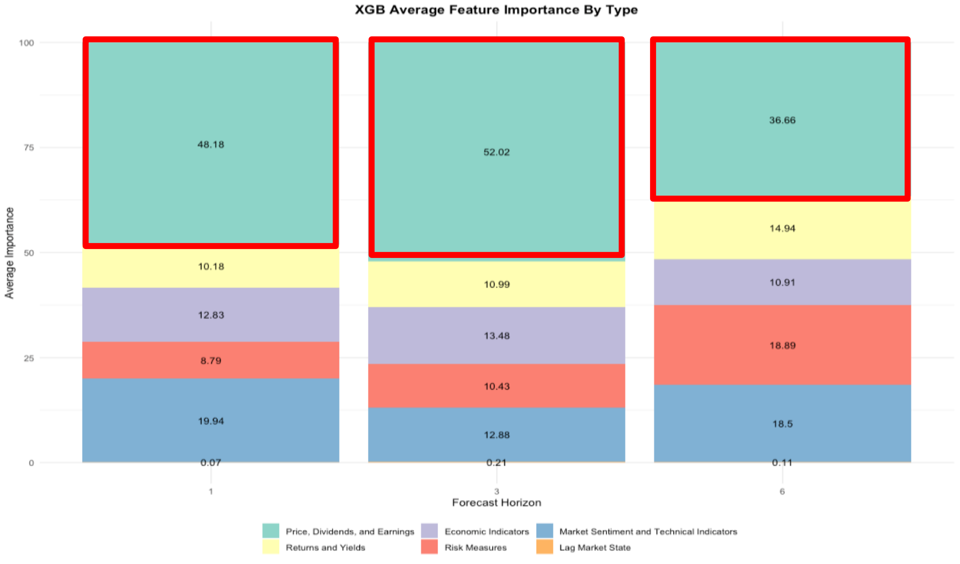


<u>As forecast horizon increases</u>

Signals that provide a more macro view become more important, capturing long-term uncertainty and systematic risks

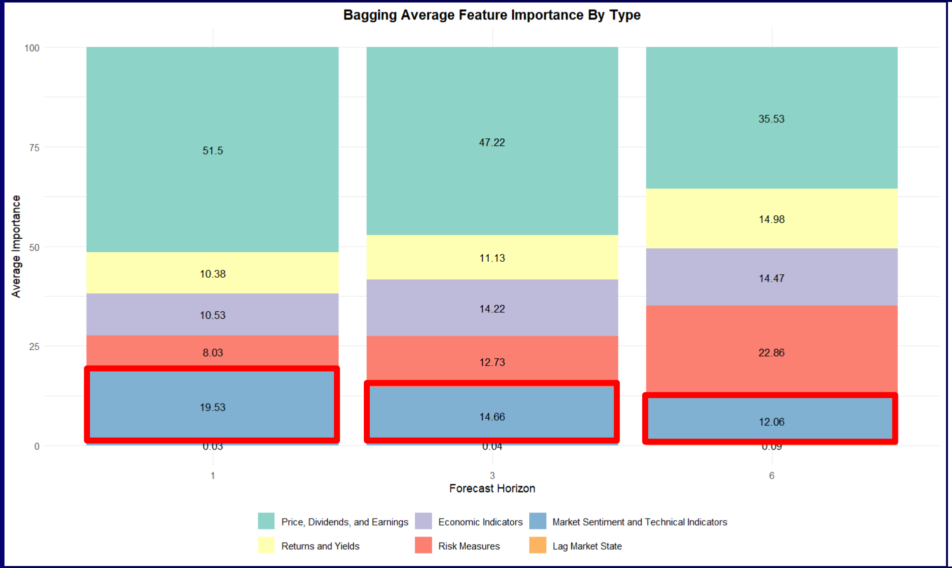
Feature Importance – Tree-based

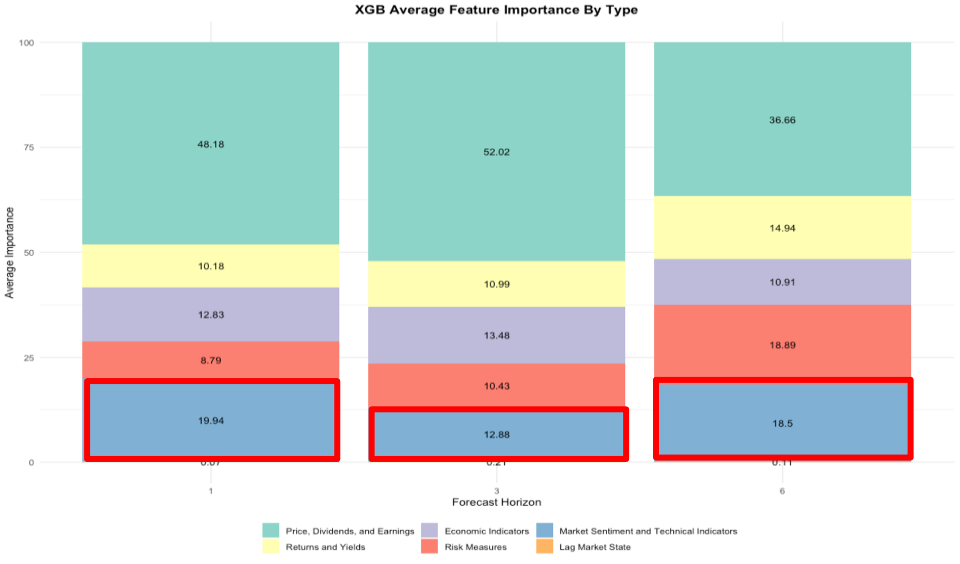




- Price, Dividends, and Earnings have the highest importance across all forecast horizons

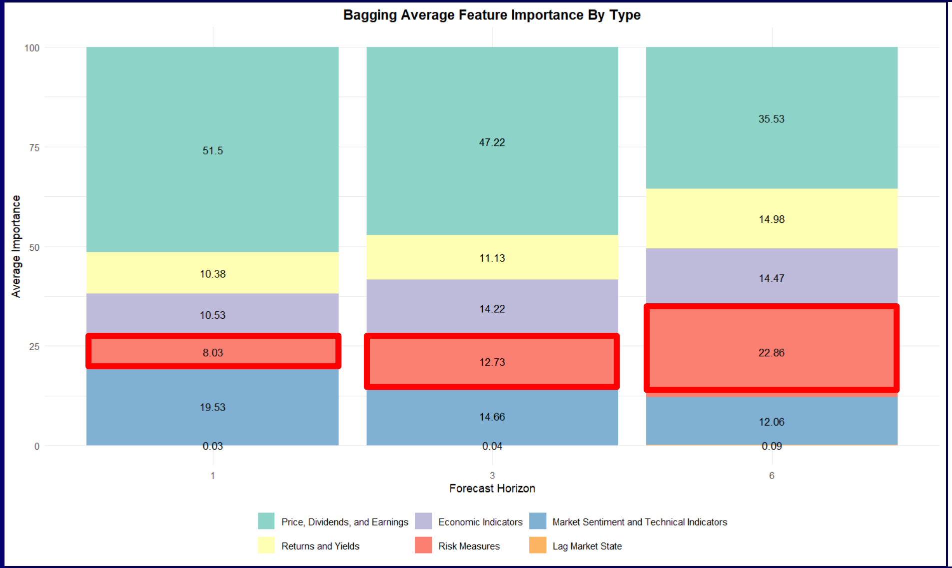
Feature Importance – Tree-based

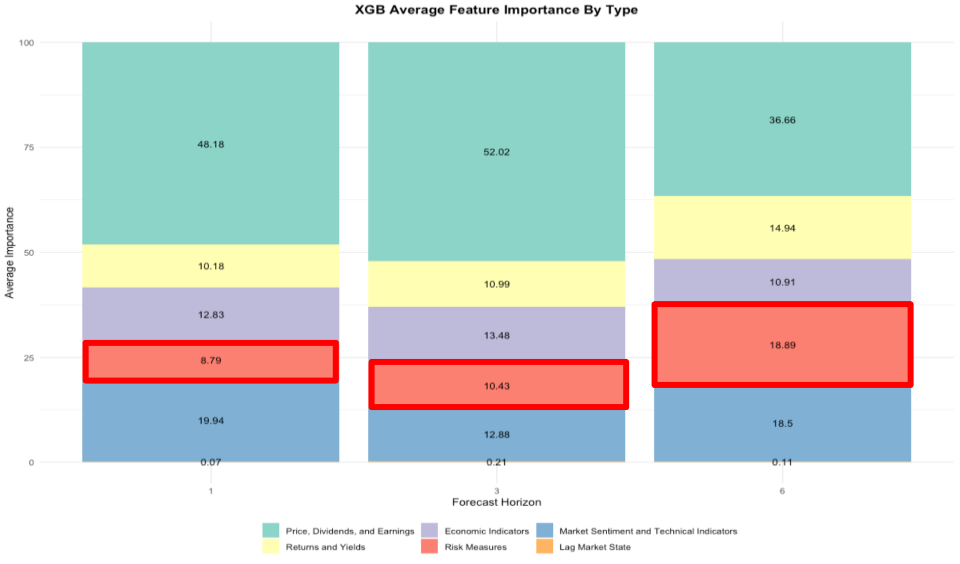




- Importance of Market Sentiment and Technical Indicators generally decreases
- Reflect general attitudes of investors, affecting how they act in the short term

Feature Importance – Tree-based





- Risk Measure group shows a significant rise in importance
- Aligns with lasso logit

F1 score

	Forecast Horizon		
Model	h = 1	h = 3	h = 6
Logit	0.955	0.955	0.955
Lasso Logit	0.912	0.917	0.917
Bagging	0.959	0.952	0.949
Random Forest	0.956	0.960	0.957
Boosting GBM	0.985	0.985	0.982
Boosting GBM (sample mean)	0.985	0.985	0.985
Boosting XGB	0.982	0.985	0.985

- High F1 scores might be misleading due to the imbalance of market states
- More comprehensive metric is needed

05

Profit-based Portfolio Strategy

How the strategy works (threshold = 0.50)

1. Classify bear/bull market for each observation.
2. If probability of bear market > 0.5, invest in risk-free assets e.g. treasury bill.
3. Otherwise, invest in stocks.
4. Cumulative return is calculated by compounding the monthly returns over time

vs

How the strategy works (threshold = sample average)

1. Classify bear/bull market for each observation.
2. If probability of bull market > average of bear market months of the training set, invest in stocks.
3. Otherwise, invest in risk-free assets e.g. treasury bills.
4. Cumulative return is calculated by compounding the monthly returns over time

vs

Market timing experiment

- Time period: July 2009 to December 2021
- Method 1: threshold = 0.50
- Method 2: threshold = average of bear market probability of the training set ≈ 0.26

1-step ahead forecast

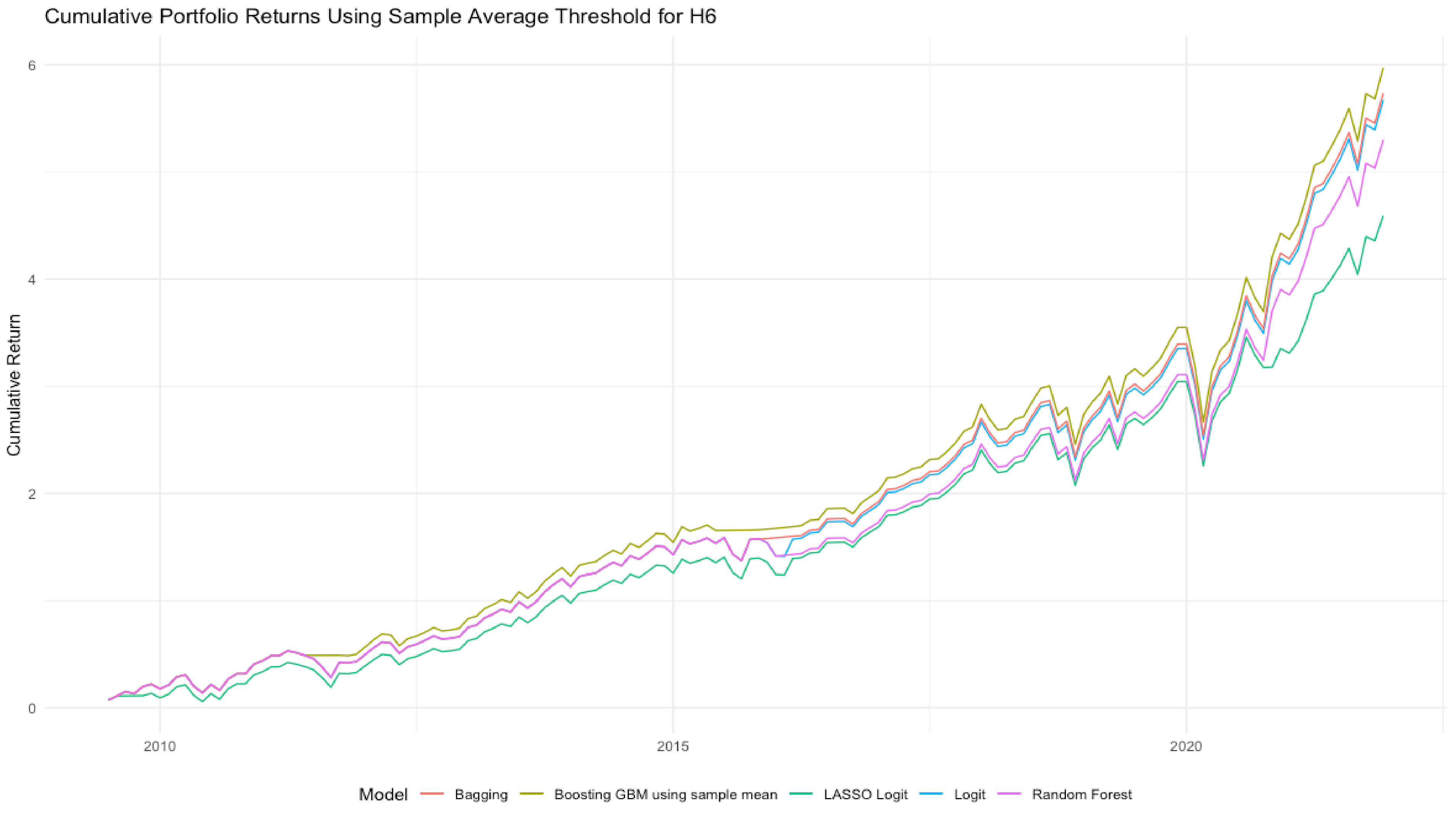
Model	Cumulative returns when threshold = 0.50	Cumulative returns when threshold = sample average
Logit	5.67031	5.67031
Lasso Logit	3.86681	4.87568
Bagging	5.41376	$\boxed{6.72304}$
Random Forest	3.88215	6.06708
Boosting GBM	5.97120	4.97808
Boosting GBM (sample mean)	5.97120	5.97120
Boosting XGB	5.83025	5.12975

3-steps ahead forecast

Model	Cumulative returns when threshold = 0.50	Cumulative returns when threshold = sample average
Logit	5.67031	5.67031
Lasso Logit	3.60500	3.61037
Bagging	3.69823	5.22590
Random Forest	4.68467	5.29800
Boosting GBM	5.97120	5.69690
Boosting GBM (sample mean)	5.97120	5.97120
Boosting XGB	5.97120	5.12975

6-steps ahead forecast

Model	Cumulative returns when threshold = 0.50	Cumulative returns when threshold = sample average
Logit	5.67031	5.67031
Lasso Logit	3.83951	4.59110
Bagging	4.21127	5.73455
Random Forest	5.08841	5.29800
Boosting GBM	6.11753	5.03982
Boosting GBM (sample mean)	5.97120	5.97120
Boosting XGB	5.97120	4.75748



Cumulative Portfolio Returns Using Sample Average Threshold for H6
Cumulative Return
0
2
4
6
2010
2015
2020
Model
Bagging
Boosting GBM using sample mean
LASSO Logit
Logit
Random Forest

06
Conclusion

Aim:

Find which model gives the best predictive performance, and if ML models can beat the benchmark logit model

Our Findings

- Gradient Boosting Machines tuned using sample mean performed consistently well for all forecast horizons (h = 1, 3, 6)
    - GBM generally out-performed XGBoost
- Robustness of GBM model makes it the most consistent and reliable model for market state prediction
- LASSO Logit generally performed the worst across all forecast horizons
- Bagging and Random Forest performed especially well for h = 1, and reasonably well for longer forecast horizons
    - Bagging generally out-performed Random Forest

Compared to logit?

- Logit model had a very stable performance
  - Generated same cumulative returns of 567% for all forecast horizons
- Classified all months in test set (Jul 2009 – Dec 2021) as bull markets
  - Might not perform well for periods that have more bear market months

- Only GBM model consistently out-performed the logit model

Limitations

- Performance of models highly dependent on proportion of bear and bull months in the market
- Used historical data from most recent vintage instead of real-time vintages

Future Work

- Can look into expanding our analysis to include periods with difference economic conditions
- Use real-time vintages to evaluate and test models
- Explore prediction of additional market states, eg recession

Thank You!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**