Group 17: Rohit, Ajay, Rishi, Sudhansh, Gurmeet

Multi-Asset Portfolio Construction and Return Prediction with Dynamic Features

Applied Finance Project

Masters of Financial Engineering, UC Berkeley

Goals

I Robust Multi-Asset Return & Variance Prediction

II Economic Intuition: Understanding Key Drivers

III Utility Maximization

Roadmap

I Data & Factor Construction

Macro, Technical, Pattern, CSA & Regime

II Prediction Models

Returns: Methodology, Results

Variance: Methodology, Results

III Portfolio Optimization

Mean-Variance, Sharpe Maximisation

Trading Cost, Frequency Assumptions

I Data & Factor Construction

i) Data & Factor Construction

Asset Space

SPX Equity Index

Gold

Silver

Corn

Oil

Commodity Index

10 Yr Treasury

GBPUSD

USDJPY

EURUSD

REITS

Soy

Daily Data from January, 2000 to December 2023

In Sample: Till 2016

Out of Sample: 2017 to 2023*

What will we try to do?

Weekly Prediction: Returns and Variances

*Actual Out of Sample period is last ~330 weeks

i) Data & Factor Construction

Macro

- Inflation
- Industrial Production
- Unemployment
- Economic Policy Uncertainty
- BBB-AAA yield spread
- Consumption-Wealth ratio
- Fama-French 5 factors
- Risk Free rate
- M1/M2 money supply

Technical

- Lagged Returns
- Denoised Returns
- Return Momentum
- Return Reversals
- Return Skew/ Kurtosis/ Volatility
- RSI/MACD/ADX

Pattern

- Doji Star
- Breakaway
- Shooting star
- Harami cross
- Short Line Candle

Cross Section

- Returns (R) &
- lagged returns (L . R) of other assets

Regime

- Macro/ Technical/CSA all features used in determination
- 3 state and 5 state models were decided

i) Data & Factor Construction

Regime

- Macro/ Technical/CSA all features used in determination
- 3 state and 5 state models were decided

Two Approaches:

1. K-Means
2. Hidden Markov Models

Both models were trained on historical data, using our base feature universe and lagged features as our state-observations. Once trained, they were then used to predict the states for the training data that we will use in return prediction.

1. K-Means

Advantages:
- Simplicity
- Quick and Scalable

Disadvantages:
- Static
- Sphericity of Clusters

Optimal number of clusters, k found using elbow method

2. Hidden Markov Models

Advantages:
- Temporal Dynamics
- Flexibility in Model Distributions
- Predictive Insights

Disadvantages:
- Complexity
- Markov Assumption

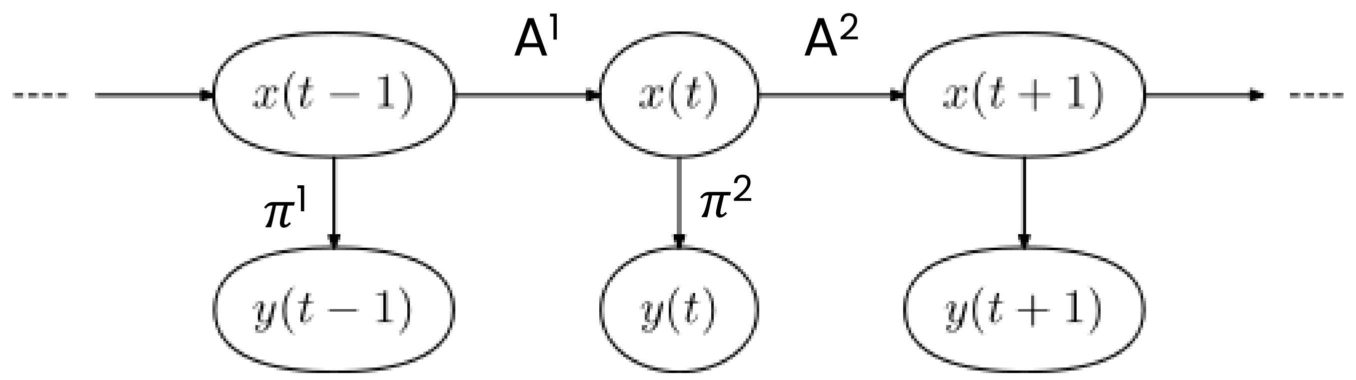
Optimal number of clusters found using AIC, BIC

i) Data & Factor Construction

Regime

- Macro/ Technical/CSA all features used in determination
- 3 state and 5 state models were decided

Hidden Markov Models



A is transition probability

π is emission probability

Since x cannot be directly observed, one makes educated guesses based on observations y.

In our case, each y is our entire feature universe and x are the market regime states

With HMMs, you can answer three fundamental questions:

1. Evaluation: Given Y and $\lambda = (A, \pi)$, what is $P(Y|\,\lambda)$?
2. Training: Given Y, how do we learn the model parameters $\lambda = (A, \pi)$ to maximize $P(Y|\,\lambda)$?
3. Recognition: Given Y and $\lambda = (A, \pi)$, what is the best hidden state sequence X that best explains Y?

These problems are solved using a Foward-Backward algorithm, Baum-Welch algorithm, and Viterbi algorithm respectively.

i) Data & Factor Construction

Hidden Markov Models: Gaussian

Regime

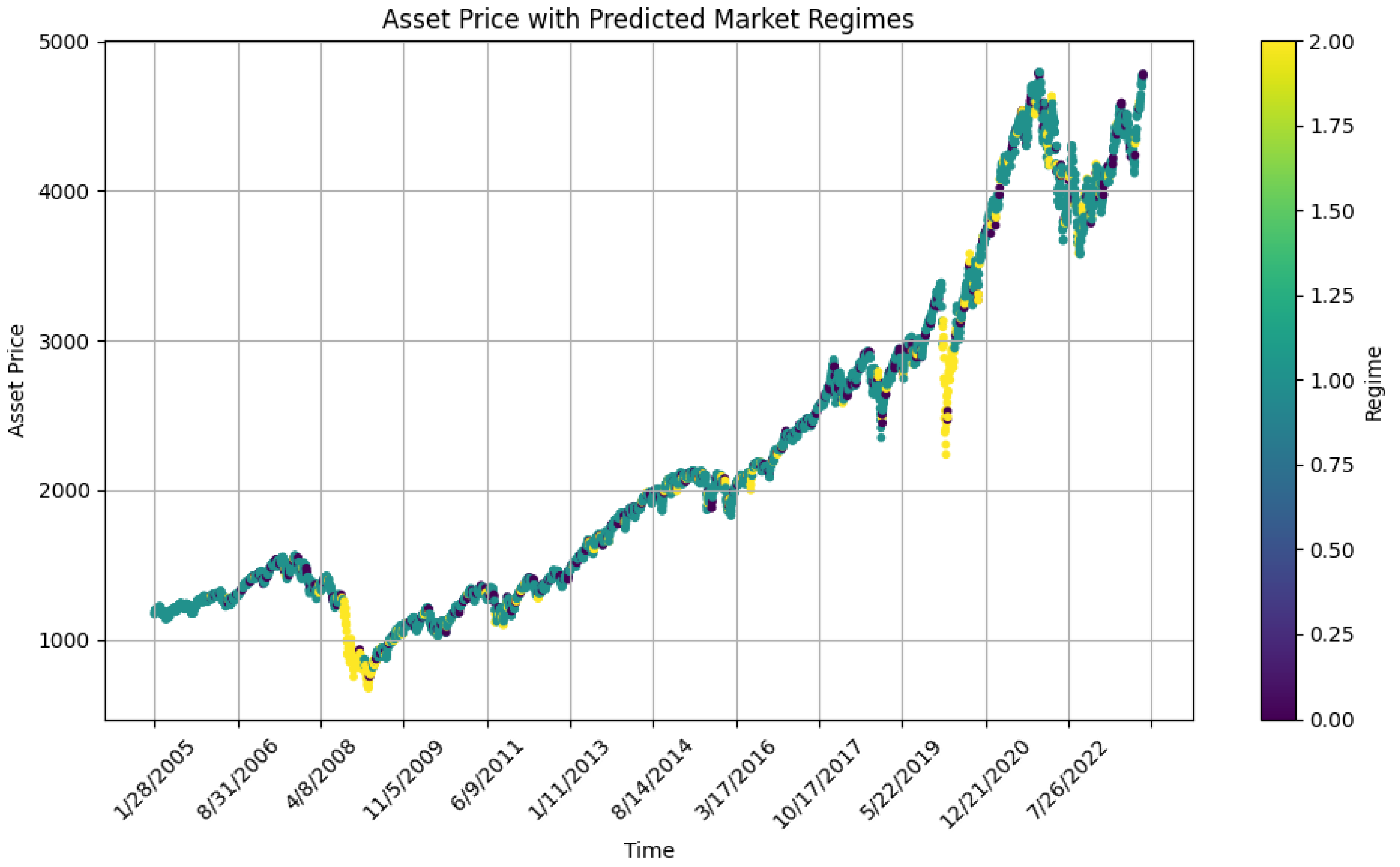
- Macro/ Technical/CSA all features used in determination
- 3 state and 5 state models were decided

$$\lambda = (A, \pi)$$

$$\lambda = (A, \theta, \pi)$$

Same as before but with an added parameter for the distribution of our observation/emission probabilities, which here are assumed Gaussian.

- $A = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$
- $\theta = [\mathcal{N}(1,1), \mathcal{N}(5,1)]$
- $\pi = [0.6, 0.4]$



i) Data & Factor Construction

The Curse of Dimensionality

Macro

75 features:

I. Levels **(15)**

II. Monthly differences **(15)**

III. 1/2/3-month lags **(15 * 3)**

Technical

28 features for each assets:

I. Lagged Daily Differences **(15)**

II. RSI, MACD, ADX **(3)**

III. Return Momentums for past 7, 15, 30, 45-day cumulative returns **(4)**

IV. Return Reversals using 7 & 30-day **(2)**

V. Skewness, Kurtosis, Volatility using 22-day lookback window **(3)**

VI. Denoised return **(1)**

Pattern

61 features:

Categorical labels to convey presence/absence of certain patterns in the price/return historical time series

Cross Section

~1000 features

Cross sectional features, including lagged returns, patterns and technical features of other assets

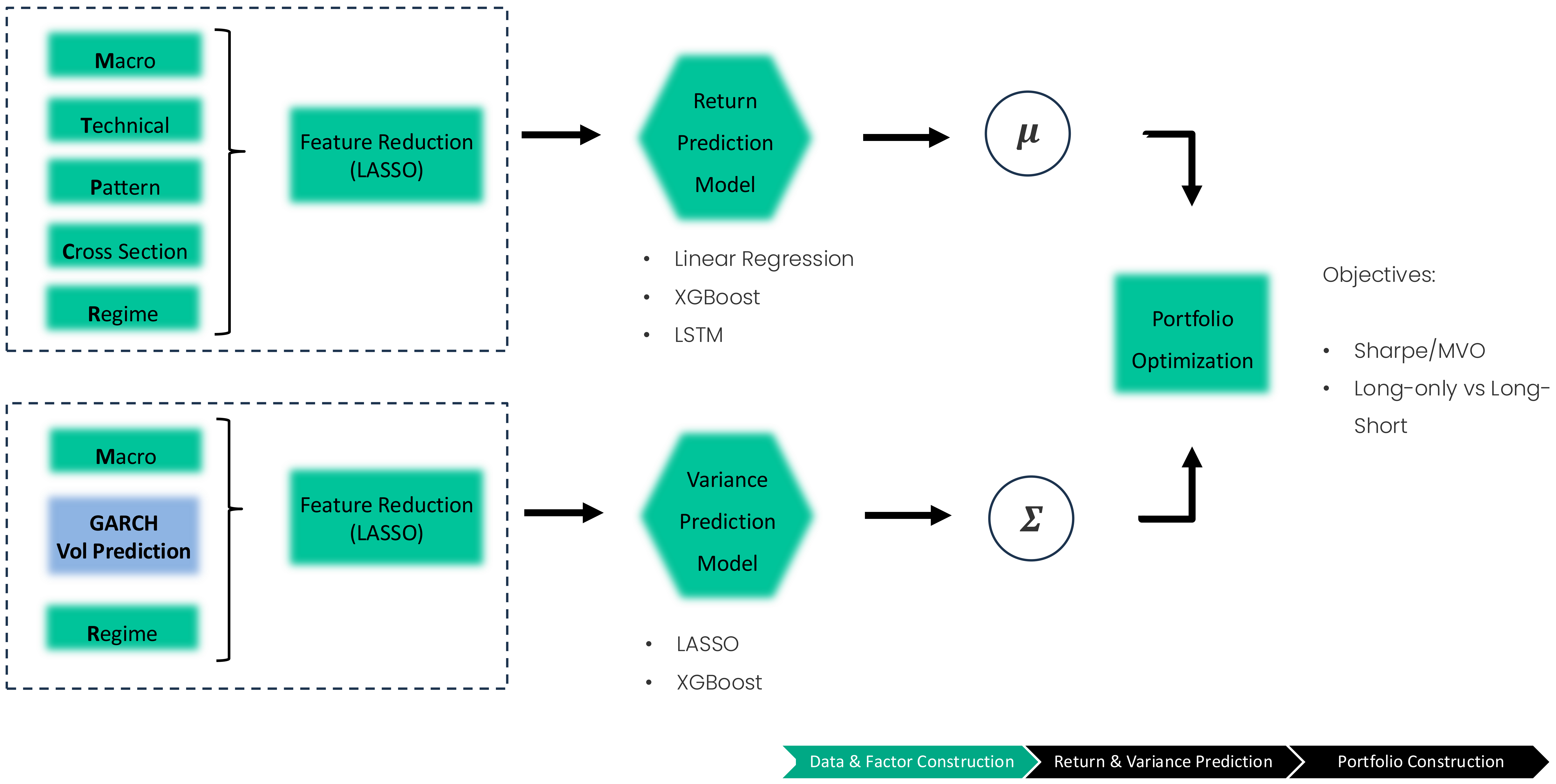
Regime

2 features

We considered 3 state and 5 state models

Hidden Markov Models and k-Means Clustering utilized for determination of regimes

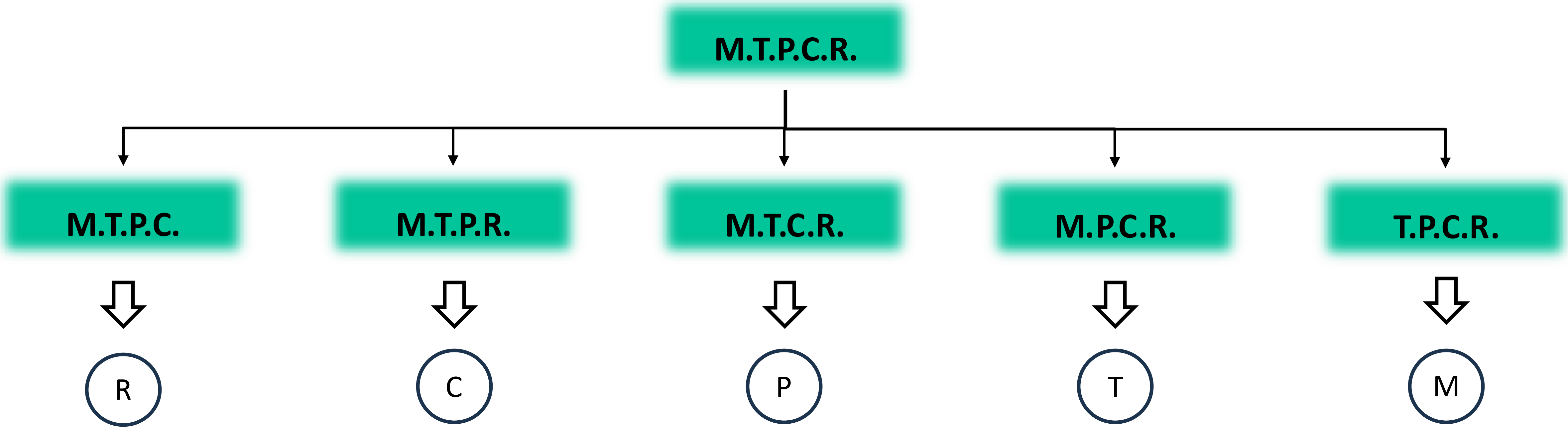
Framework Overview



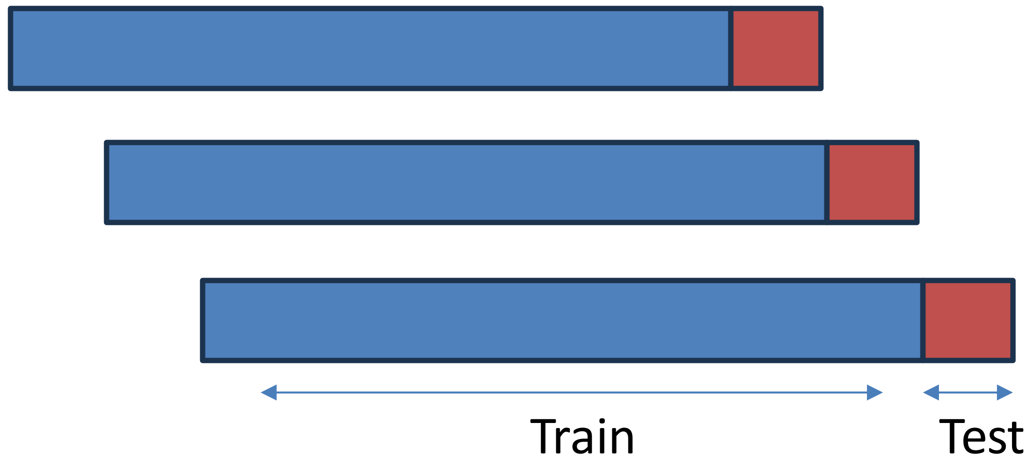
II Predicting Returns

Prediction Models: **Return Prediction**

Approach: Kitchen-sink OOS explained variance loss



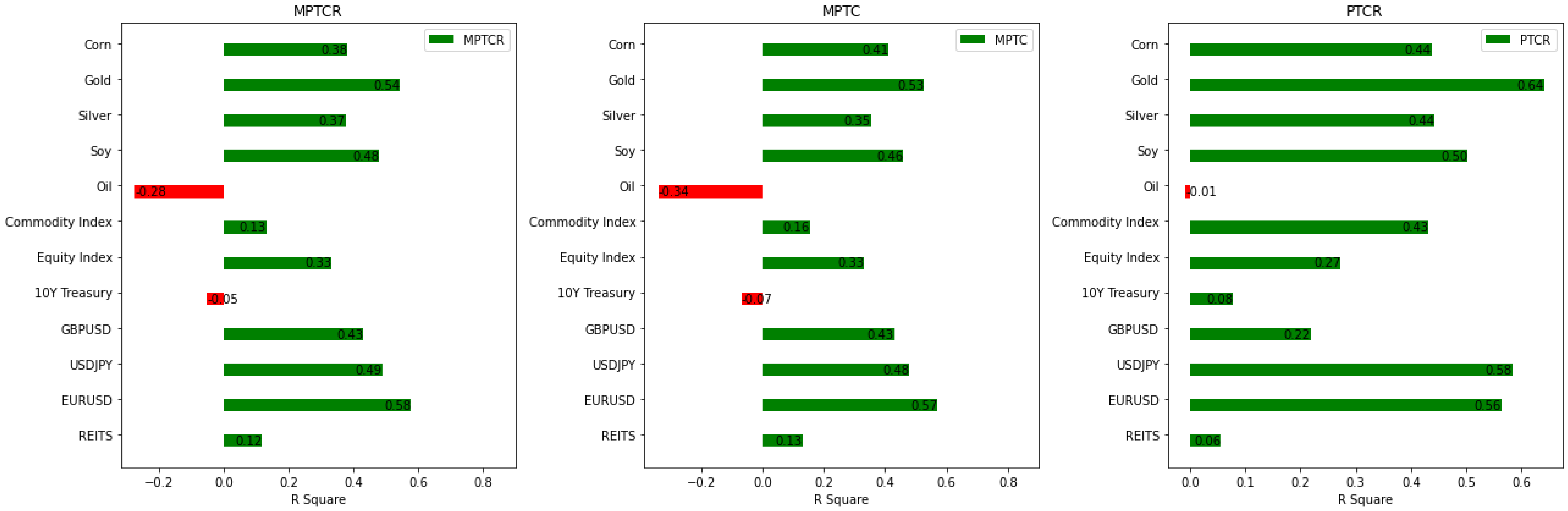
- Rolling: 600:1 weeks train-test window (OOS 2017 to 2023)
- Feature Reduction: PCA vs Lasso(OOS vs IS)
- Stability with number of features
- Linear vs Non-Linear
- Why weekly?
- What else did we try?



- Feature importance was determined using a step-by-step feature-removed regression/model.

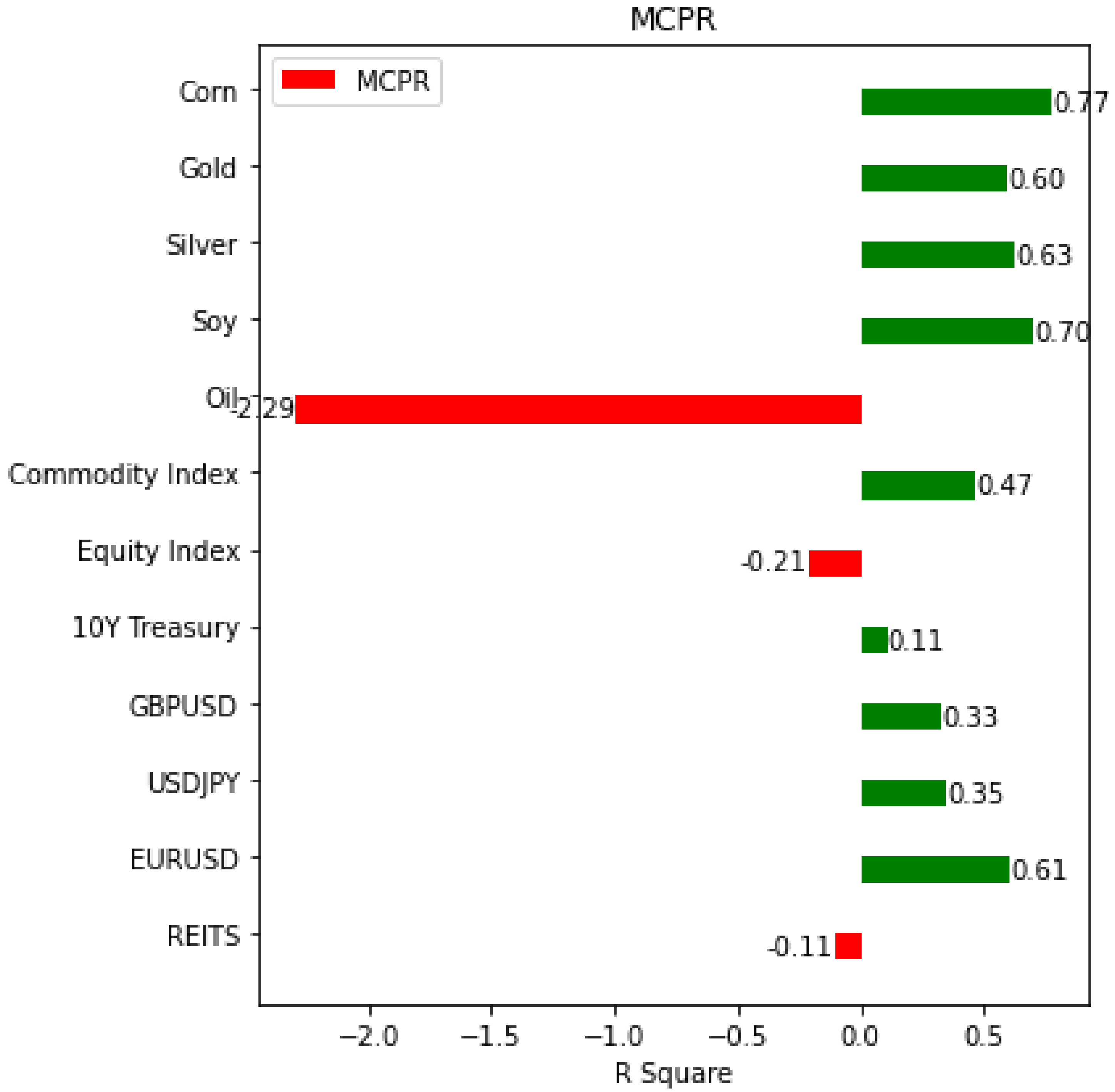
*M = Macro, T = Technical, P = Pattern, C = Cross-sectional Returns, R = Regimes

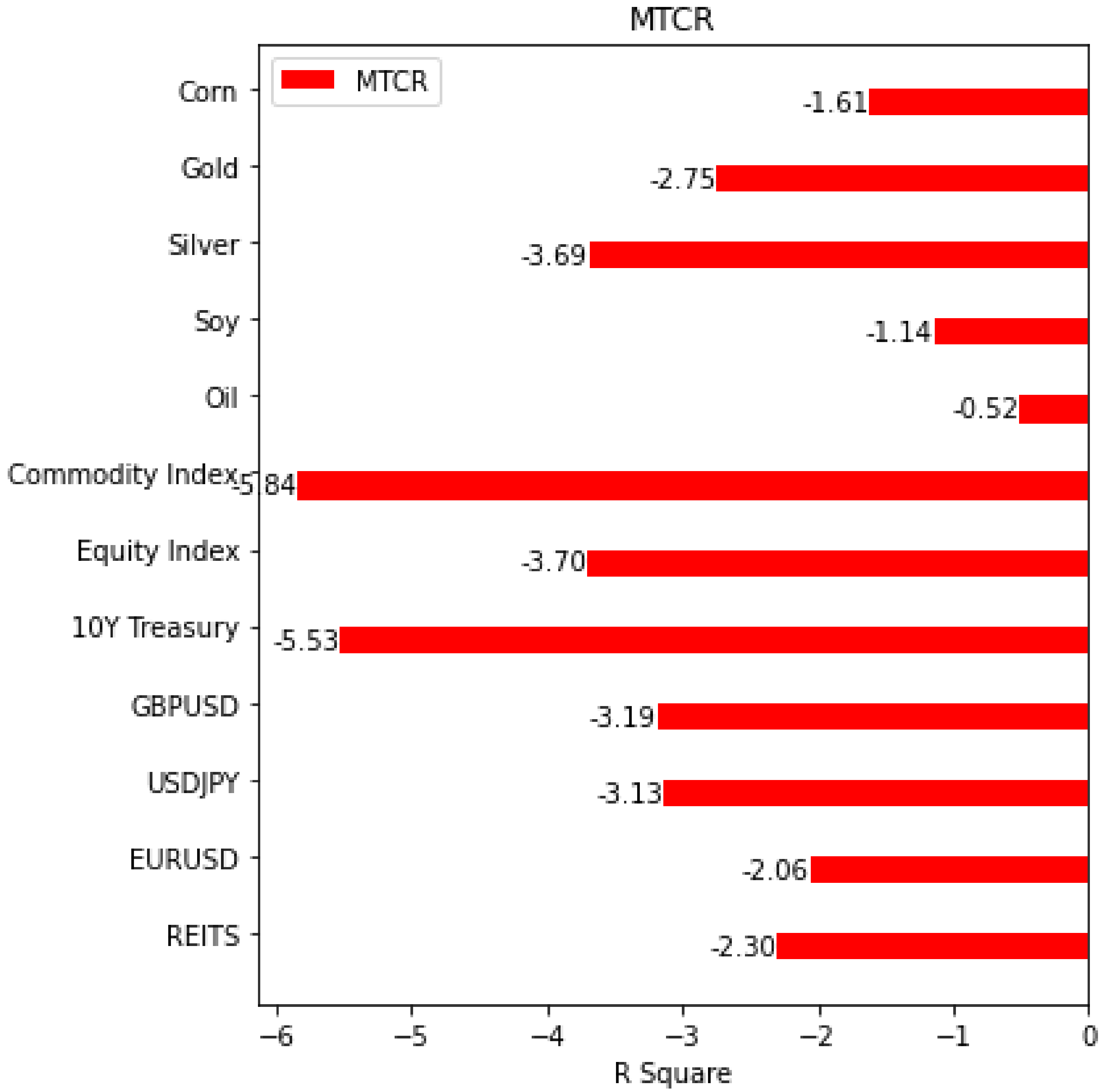
OOS R^2



*M = Macro, T = Technical, P = Pattern, R = Regimes, C = Cross-sectional Returns

OOS R^2





*M = Macro, T = Technical, P = Pattern, R = Regimes, C = Cross-sectional Returns ii Prediction Models: Return Prediction

Results

Gain in R-squared picture

Highest Average Gains

	Technical	Pattern	Regime	Macro	CSA
REITS	0.2241	2.4192	-0.0136	0.0609	0.1062
EURUSD	-0.0325	2.6333	0.0059	0.0144	0.1199
USDJPY	0.1385	3.6194	0.0112	-0.0936	0.1572
GBPUSD	0.0985	3.6151	0.0007	0.2112	0.4292
10Y Treasury	-0.1590	5.4776	0.0144	-0.1310	0.0289
Equity Index	0.5429	4.0322	0.0000	0.0586	0.2569
Commodity Index	-0.3359	5.9671	-0.0263	-0.3012	-0.0065
Oil	2.0187	0.2402	0.0627	-0.2677	0.8619
Soy	-0.2196	1.6228	0.0215	-0.0223	-0.0713
Silver	-0.2522	4.0671	0.0209	-0.0683	-0.0205
Gold	-0.0547	3.2897	0.0171	-0.0980	-0.0081
Corn	-0.3933	1.9952	-0.0294	-0.0562	-0.0216

Least Gains

REITS and GBPUSD most impacted for Macro

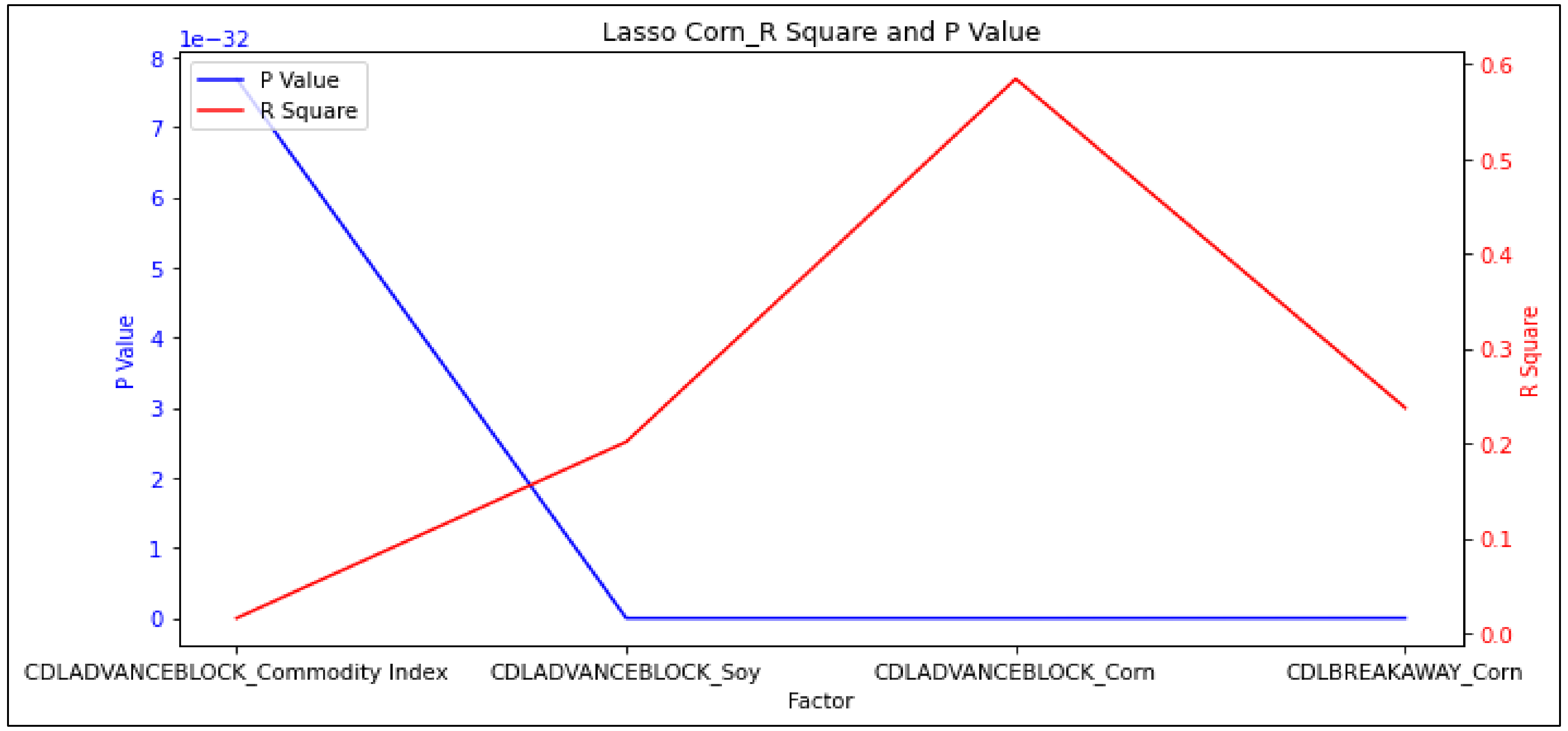
Small impact for multiple assets

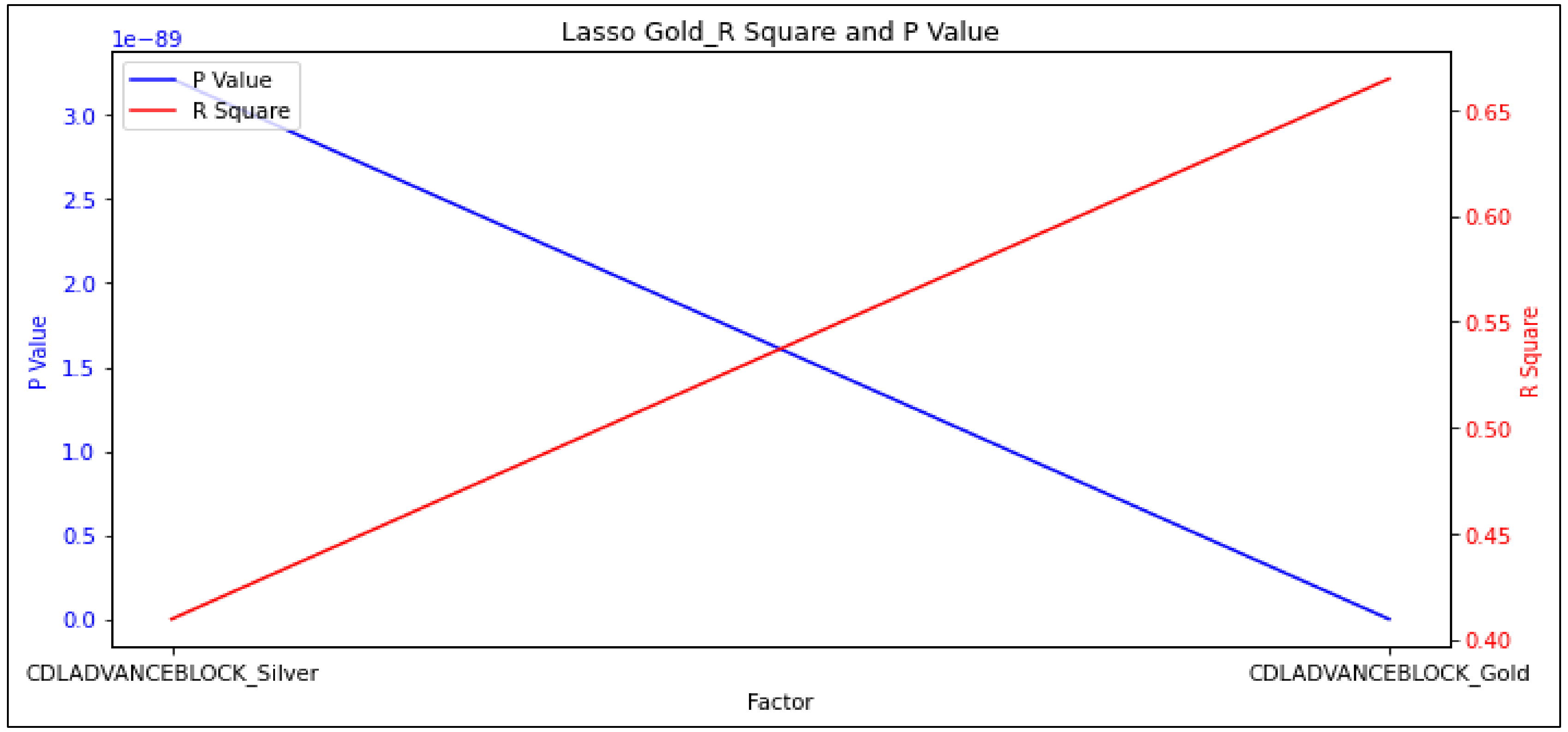
Patterns are positive for all assets

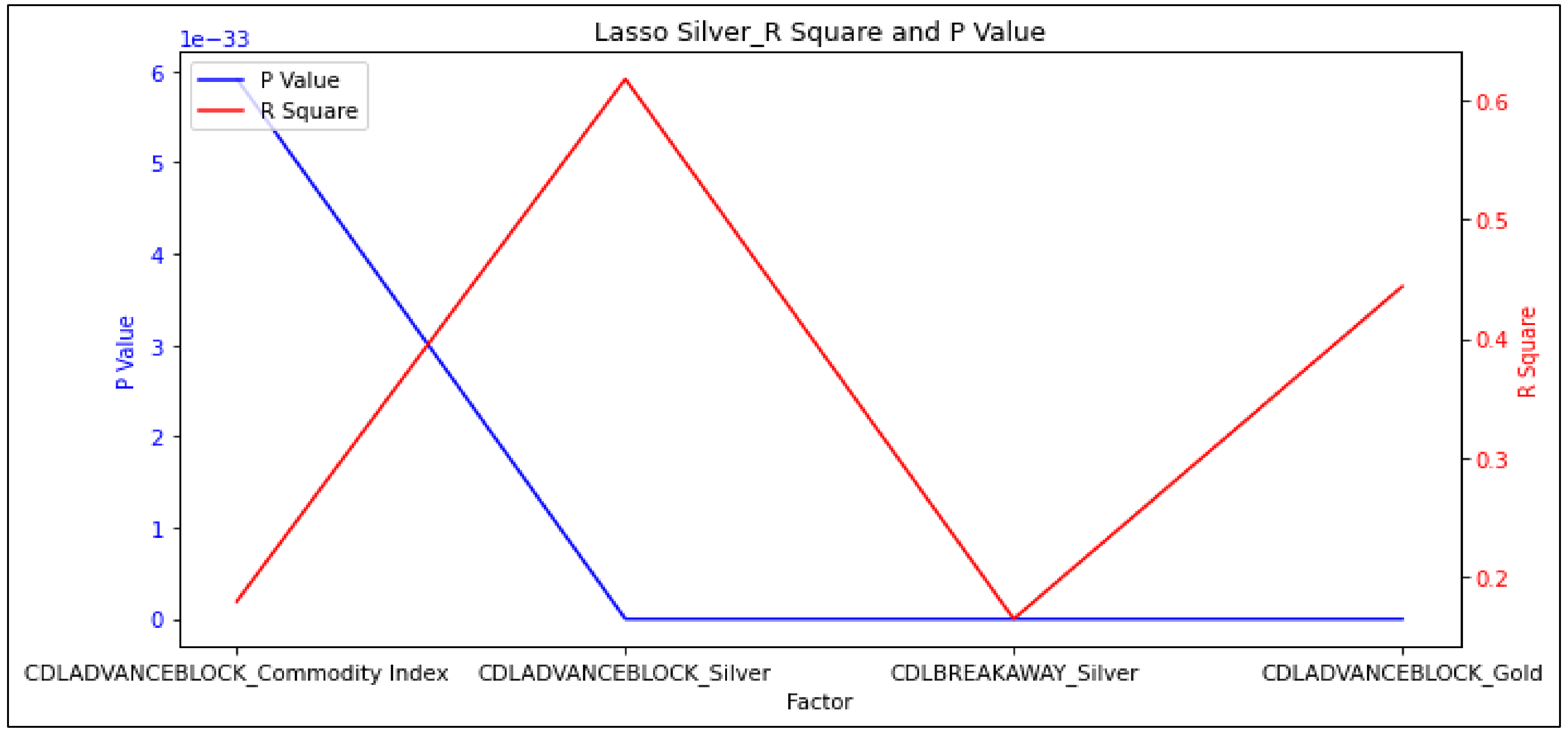
Technical has a big positive impact on a few assets

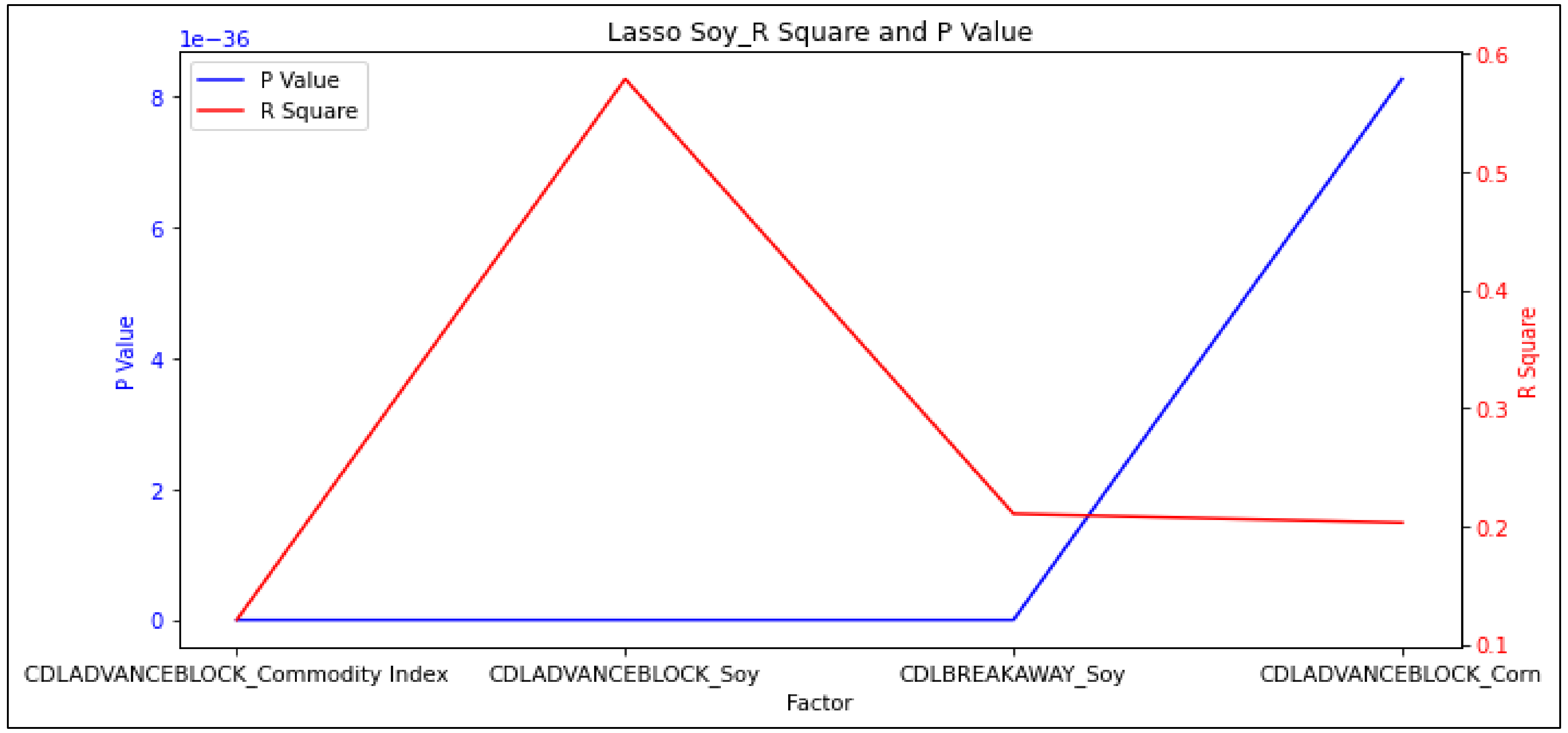
Adding variables reduces OOS R squared?

Univariate Lasso of Top Factors

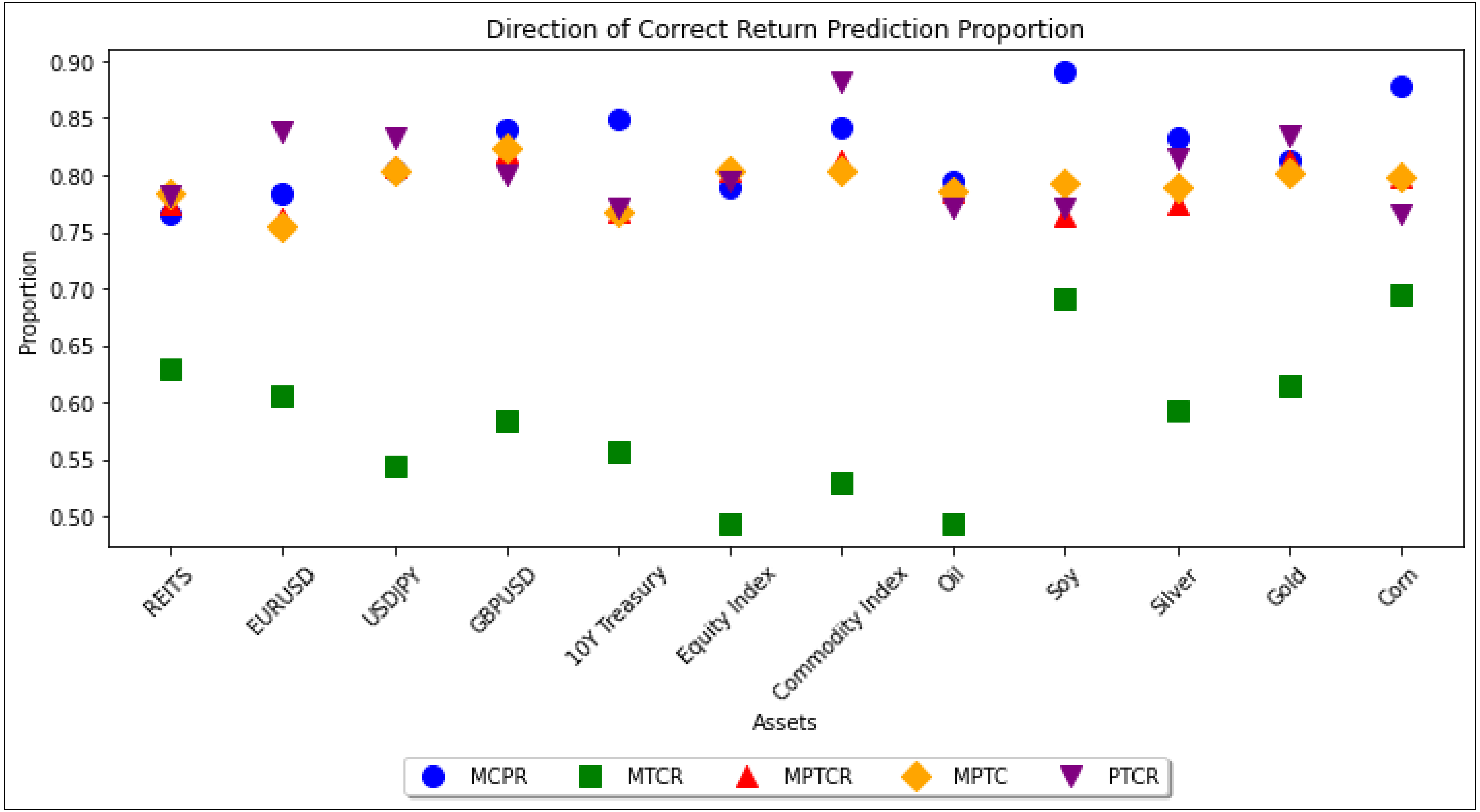








ii Prediction Models: Return Prediction



II Predicting Variance

ii

Prediction Models: Variance Prediction

Model 1: GARCH(1,1)

Returns → GARCH

- GARCH gives us Conditional volatility
- Estimate based on daily returns
- Realized is 1 day variance calculated over 5 days

Model 2: Linear Regression

GARCH → 1 Day Realized Volatility

- Realized Volatility is 1 day variance calculated over 5 days
- Regress GARCH on Realized Volatility, over 22 days and 200 days

Model 3: Linear Regression

GARCH, REGIME → 1 Day Realized Volatility

- Regime is encoded as dummy variable
- Predict Realized Volatility

Model 4: Lasso Regression

Macro, GARCH, REGIME, Lag Variances → 1 Day Realized Volatility

- Use Lasso for Feature Selection

ii Prediction Models: Variance Prediction

Results

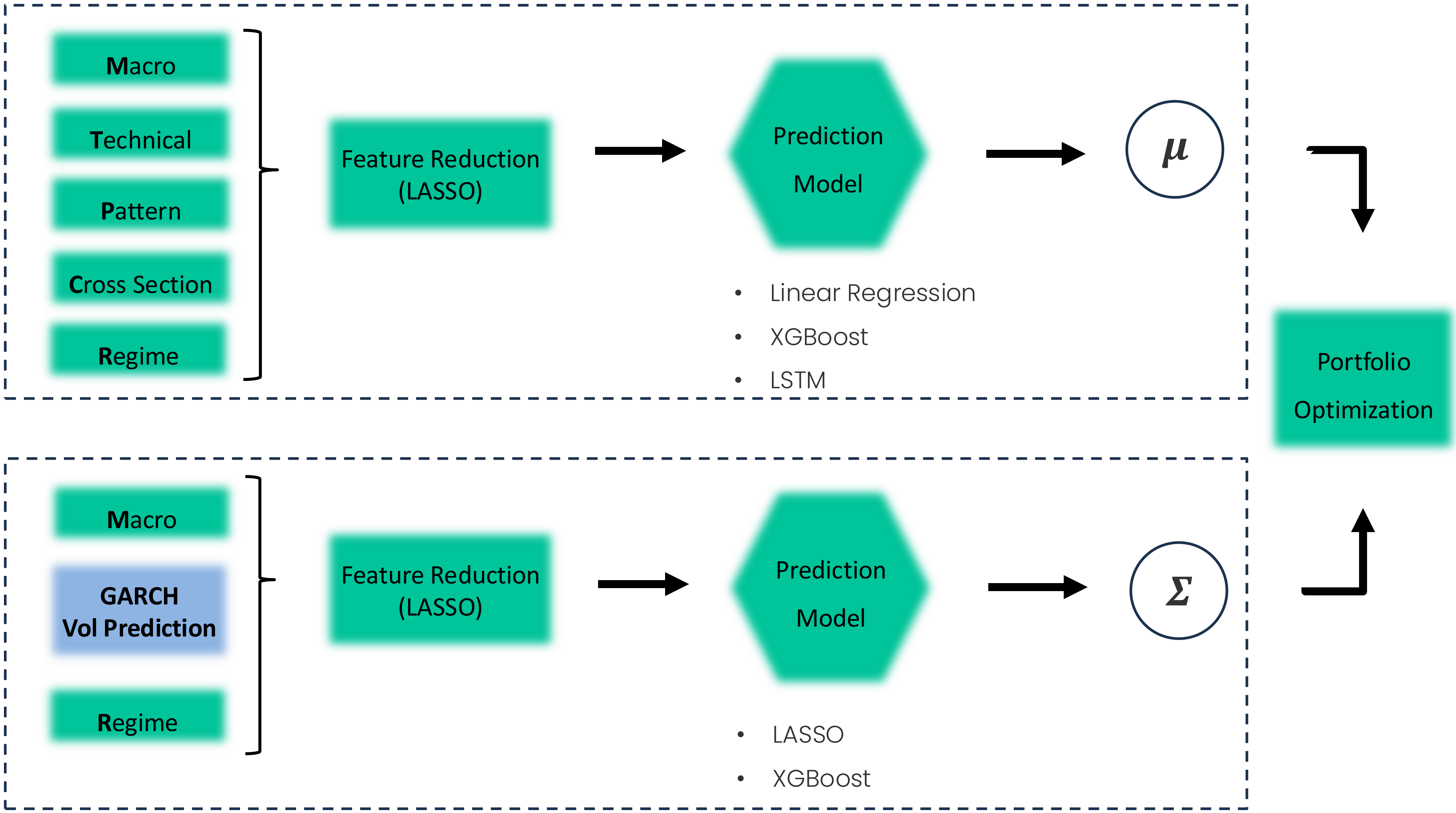
22 days training: R Square				
Asset	GARCH	Linear Regression on Garch	Linear Regression on GARCH and Regime	Linear Regression with feature selection
REITS	0.171	0.491	0.434	0.330
EURUSD	-5.698	0.261	0.103	0.268
USDJPY	-2.054	-0.141	-0.253	0.080
GBPUSD	-2.150	-3.536	-5.108	0.073
10Y Treasury	0.027	-1.171	-1.571	-0.006
Equity Index	0.201	0.341	-9.37E+23	0.081
Commodity Index	-0.362	0.166	0.038	0.221
Oil	0.000	-0.122	-0.136	-0.049
Soy	0.120	0.033	-4.30E+26	0.077
Silver	0.090	-0.151	-1.92E+24	0.056
Gold	-0.301	-0.397	-1.395	0.137
Corn	0.048	-0.028	-8.19E+24	-0.010

200 day training: R Square				
	GARCH	Linear Regression on Garch	Linear Regression on GARCH and Regime	Linear Regression with feature selection
REITS	0.171	-0.010	-0.047	-0.212
EURUSD	-5.698	0.128	0.111	-0.162
USDJPY	-2.054	0.120	0.103	-0.031
GBPUSD	-2.150	0.033	0.019	-0.005
10Y Treasury	0.027	0.026	0.028	-0.006
Equity Index	0.201	0.192	0.200	-0.032
Commodity Index	-0.362	0.215	0.214	-0.137
Oil	0.000	-0.032	-0.040	-0.003
Soy	0.120	0.159	0.155	-0.112
Silver	0.090	0.113	0.101	-0.107
Gold	-0.301	0.091	0.059	-0.327
Corn	0.048	0.056	0.017	-0.062

Figure 6: R Square Gain for Volatilities

III Portfolio Optimization

Framework Overview



Objectives:

- Sharpe/MVO
- Long-only vs Long-Short

Data & Factor Construction › Return & Variance Prediction › Portfolio Construction iii

Portfolio Optimization

Methodology

- Mean Variance Optimization
- Sharpe Maximization

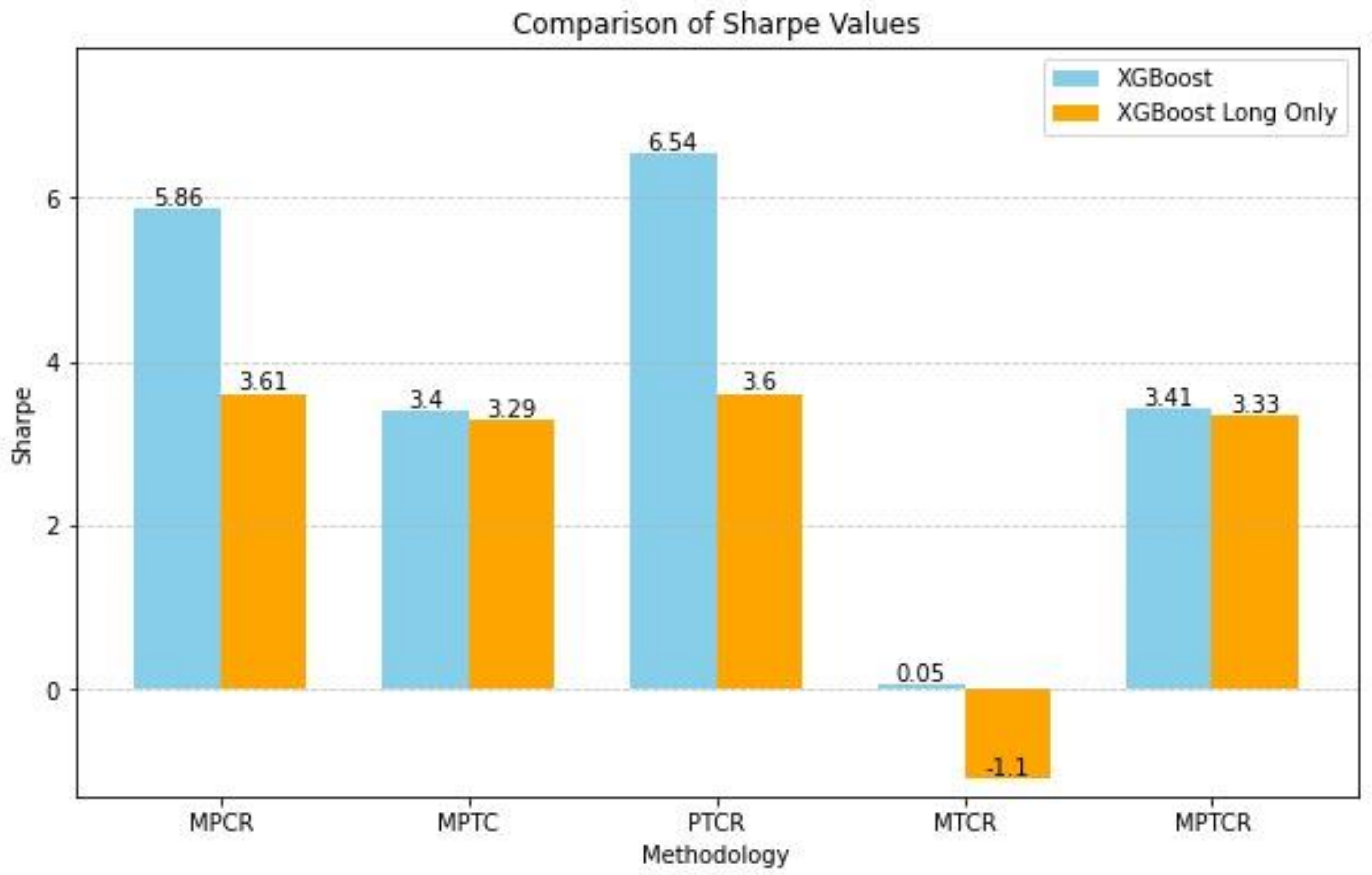
Weight constraints:
-Long-only: [0, 0.2]
-Long-short [-0.2, 0.2]

Transaction costs:
-60 bips per transaction (to account for futures margin holding, short selling and transactions)

iii

Portfolio Optimization

Results



Conclusion

1. Patterns have good predictive power and significantly improve OOS R Square

2. Model is sensitive to training. If Lasso is not able to reduce the features sufficiently, the OOS Prediction is not good

3. Non-Linear movement of features can be better captured with CNN (short term) and LSTM(for inter-temporal relationship). We need to further explore that aspect.

4. GARCH, when added with other features, adds value in Variance Prediction.

Thank You