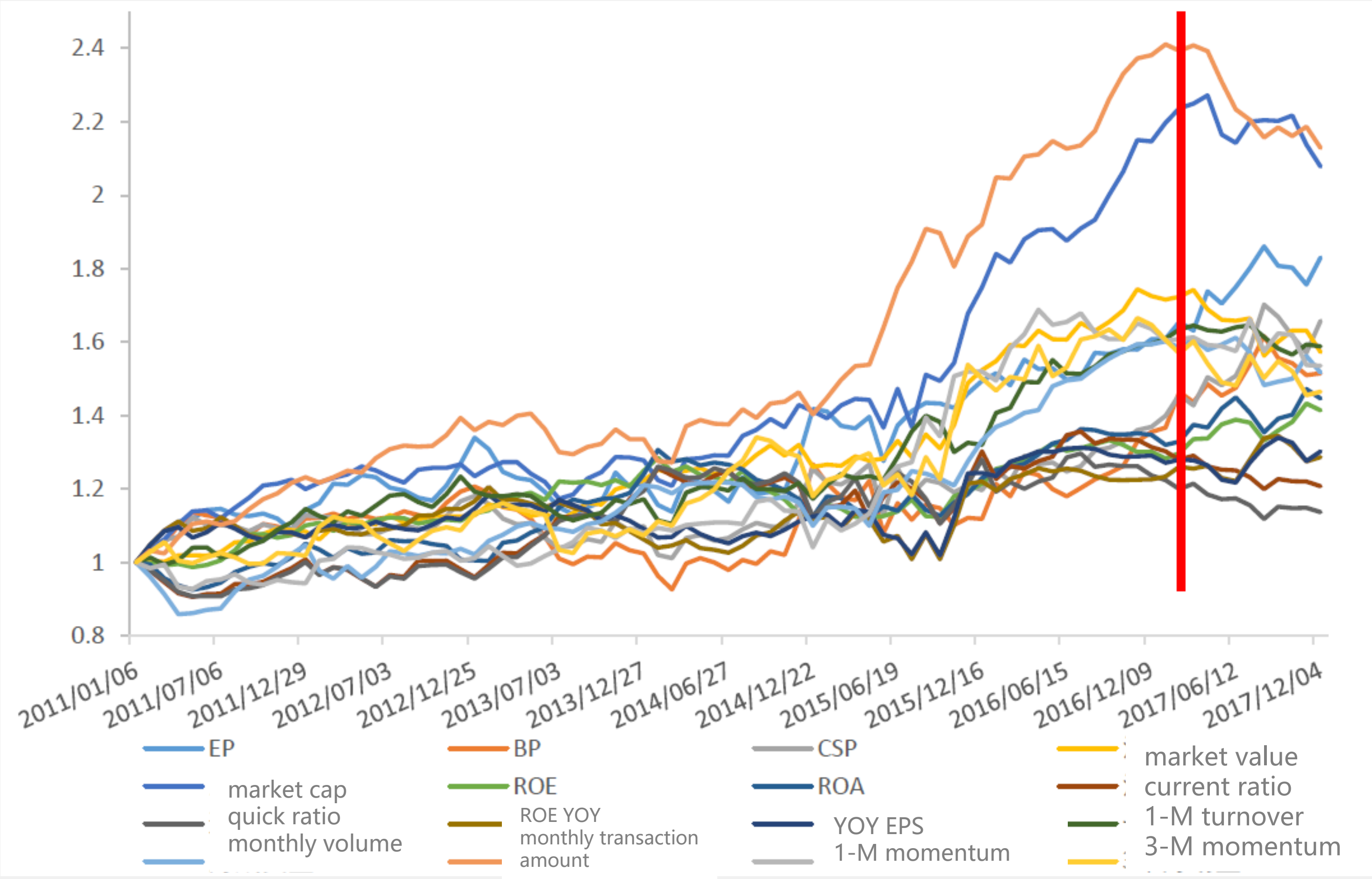
the Timing Method of Multi-factor Stock Selection Model Based on XGBoost Model

Cindy (Zixin) Zhou
Washington University in St. Louis

CONTENTS

01 Background

Since 2017, traditional factors performs poor



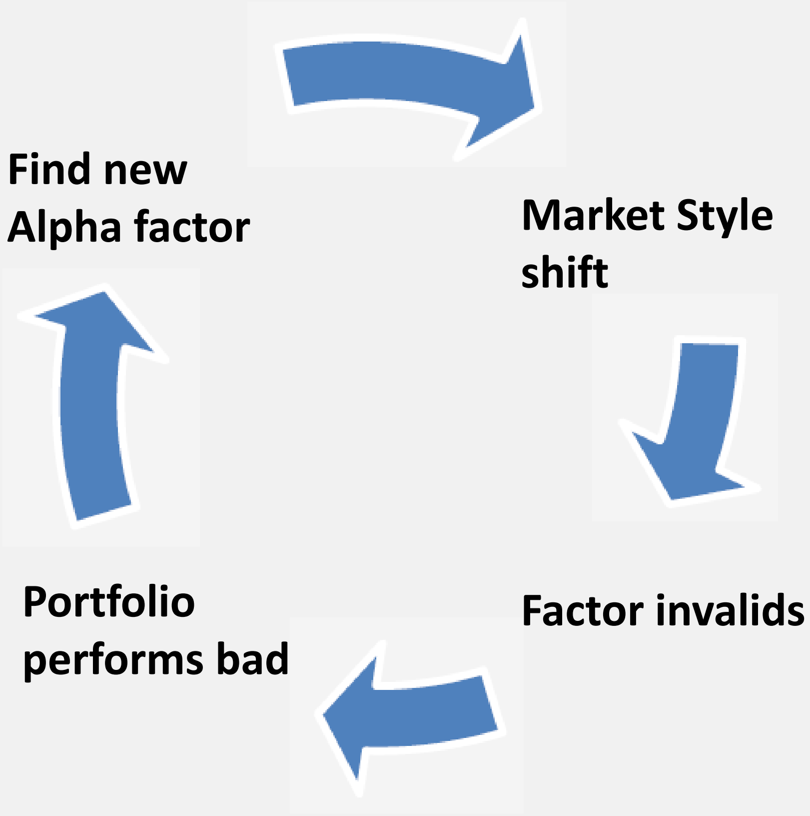
The performances of utilizing widely-used factors to select stocks from 2011.01 to 2017.12

Since 2017, the styles of factors change significantly

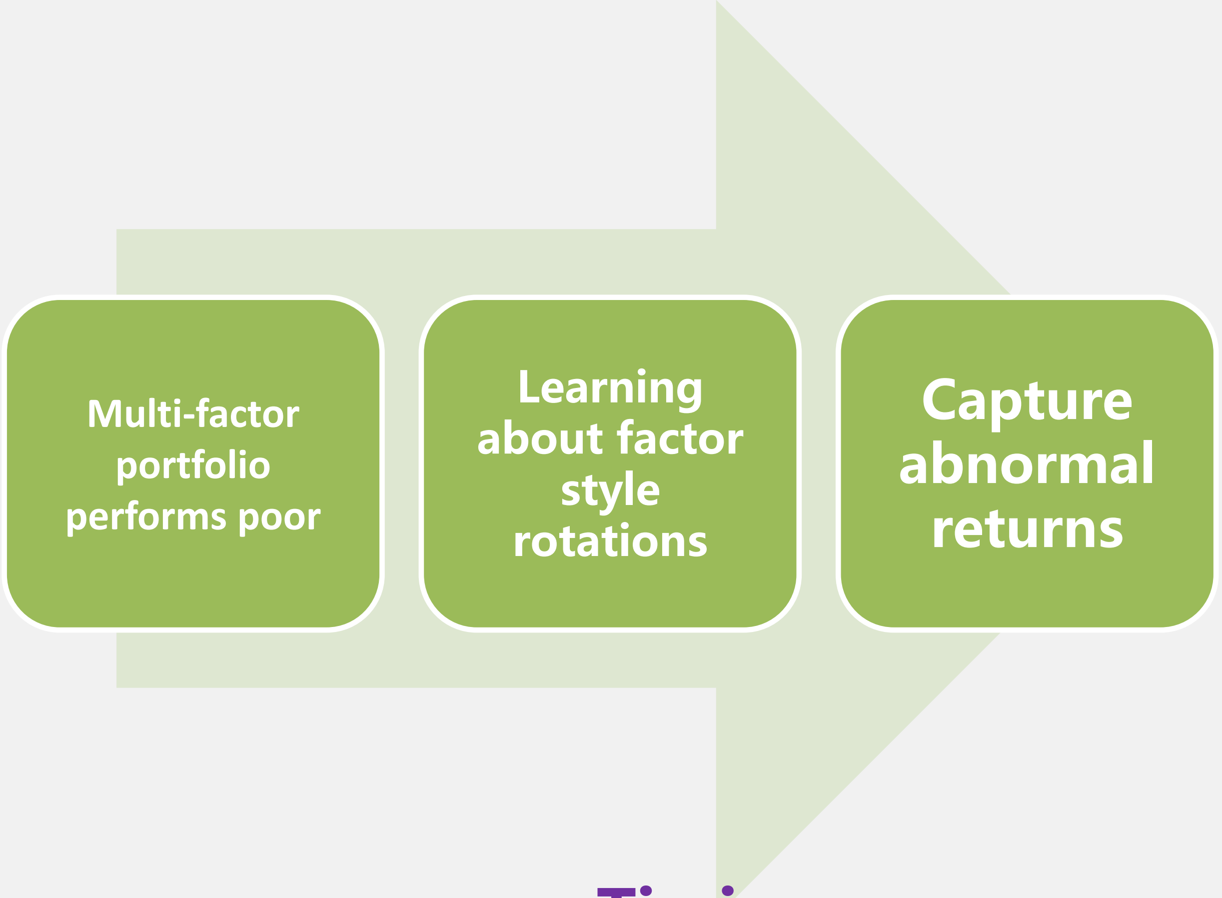
Factor Name	Correlation	Past Year Avg IC	Past 2-year Avg IC	Past 3-year Avg IC	Past 11-year Avg IC	Past 11-year IC Std	Past 11-year ICIR	Past 1-year IC – Past 11-year IC
ROE	Positive	3.83	2.56	2.15	0.72	12.42	0.06	3.10
Quick Ratio	Positive	-0.62	-0.39	0.66	0.68	11.22	0.06	-1.30
ROE YOY	Positive	3.28	2.02	1.84	0.88	7.62	0.12	2.39
EP	Positive	4.95	4.84	4.06	3.01	12.14	0.25	1.94
Market Value	Negative	-2.83	0.94	3.18	2.87	9.91	0.29	-5.71
Monthly Transaction Amount	Negative	2.41	5.98	6.55	5.87	13.55	0.43	-3.46
One-month Momentum	Negative	1.93	4.69	6.63	6.44	15.94	0.40	-4.52

2011.01-2017.12 Widely-used Factors Historical IC Values

It’s hard to find new Alpha factors → Develop good timing methods

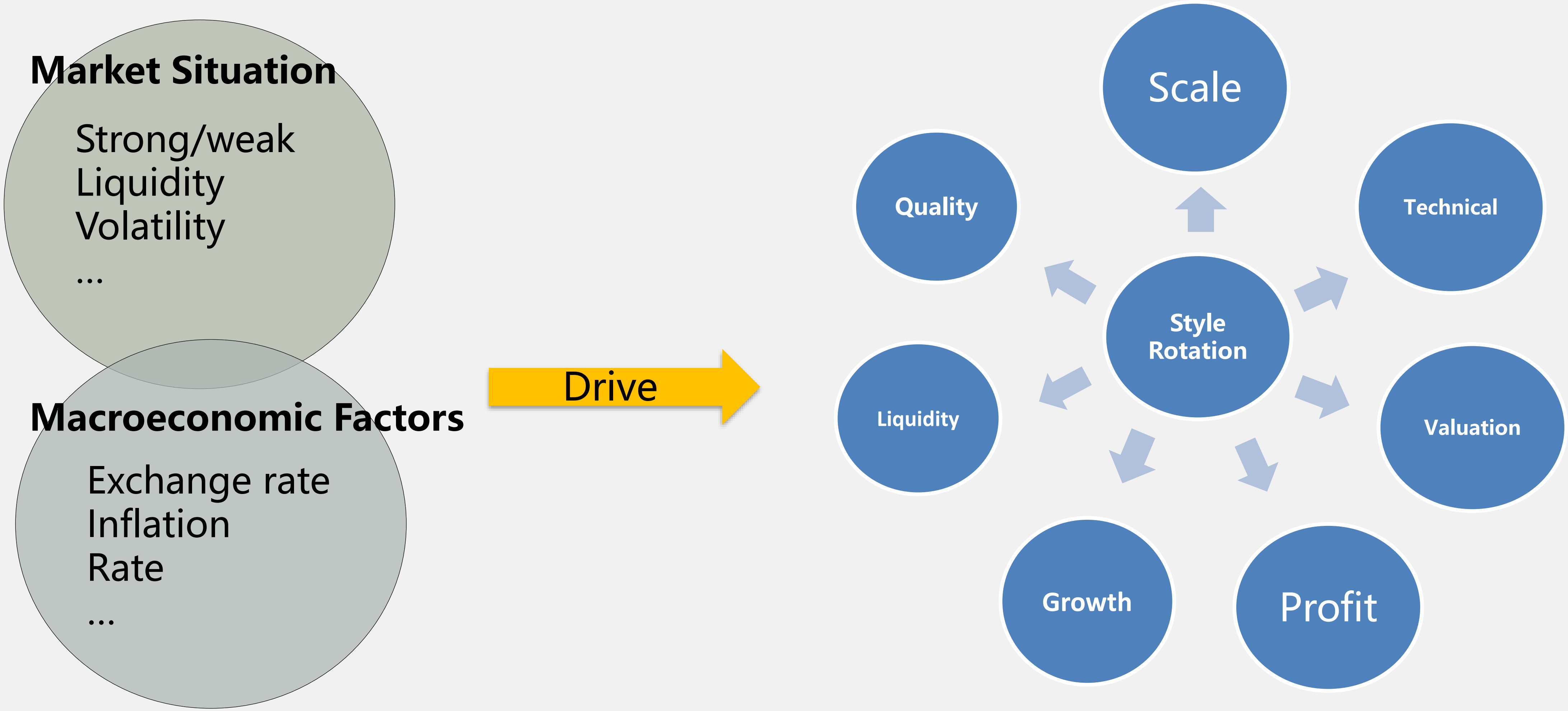


Dig out Alpha factors



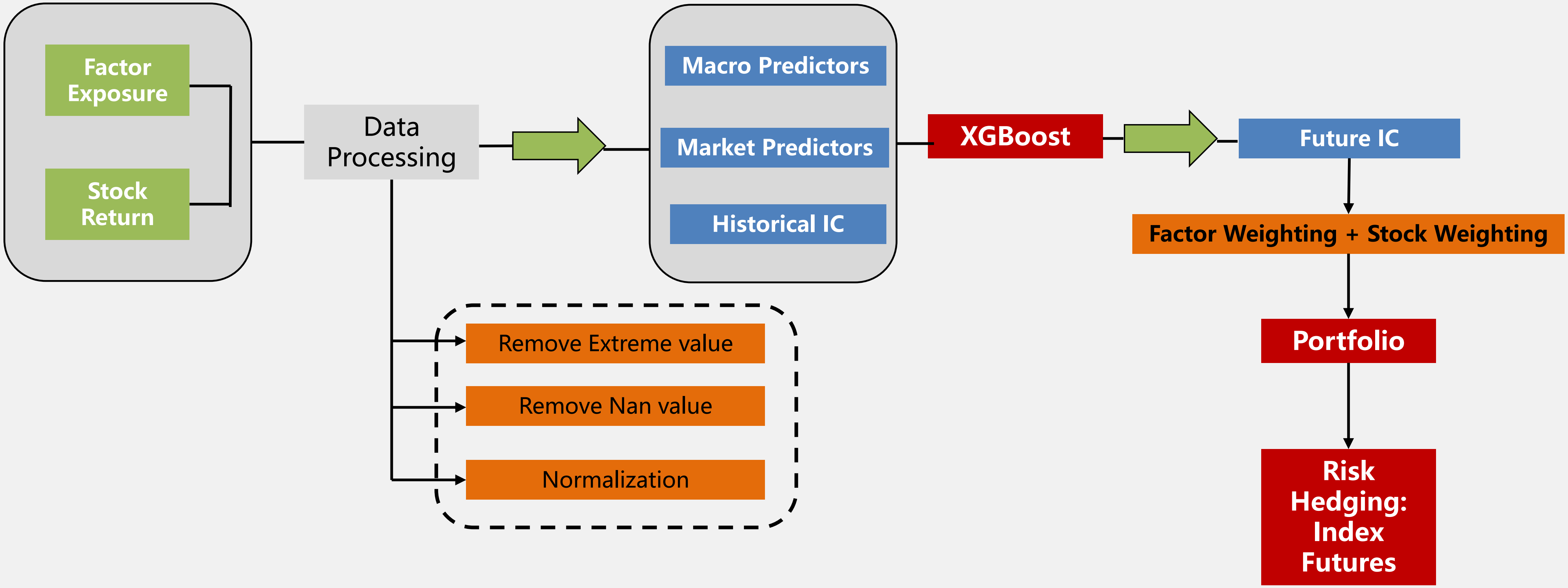
Timing

The elements to effect factor performance: Market, Macro Economy, Factors’ Style Rotation



02 Timing Method Framework

Framework

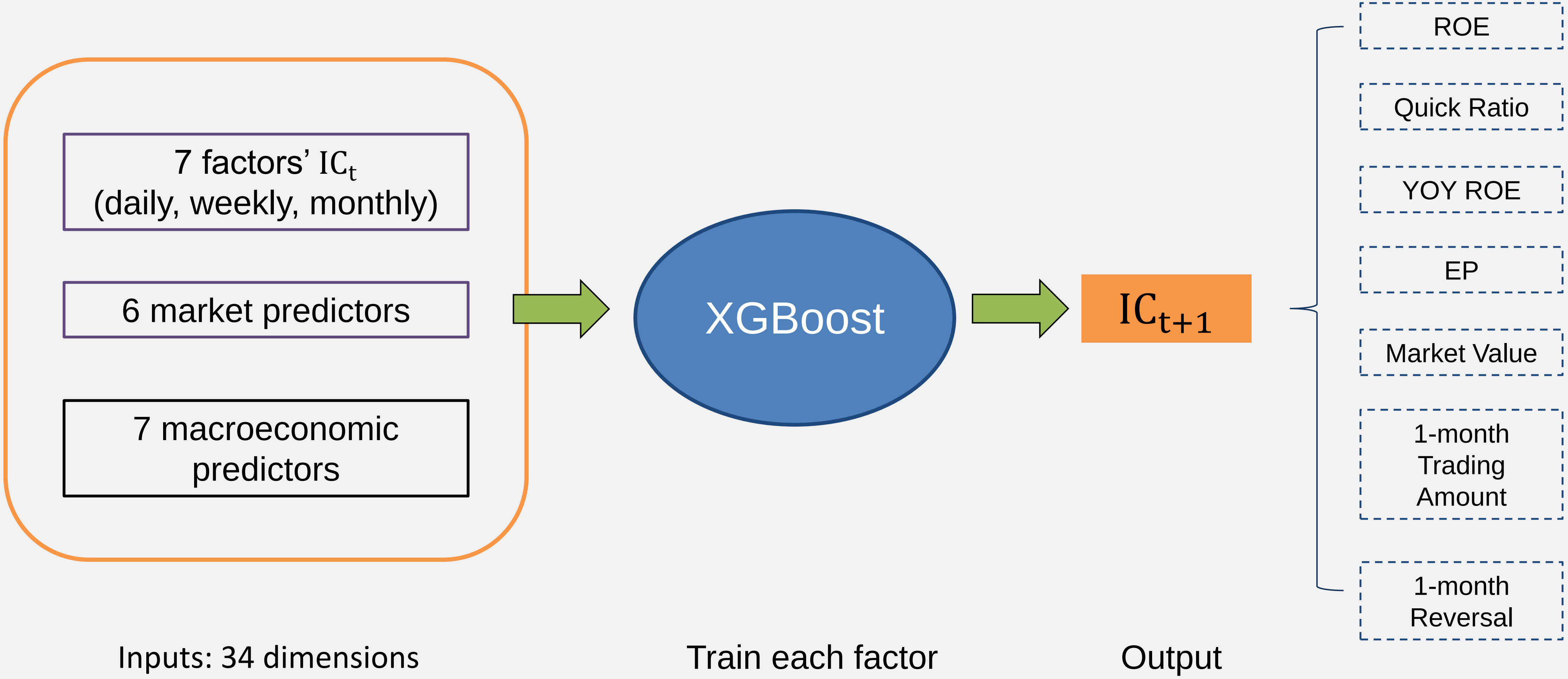

Factor Exposure
Stock Return
Data Processing
Remove Extreme value
Remove Nan value
Normalization
Macro Predictors
Market Predictors
Historical IC
XGBoost
Future IC
Factor Weighting + Stock Weighting
Portfolio
Risk Hedging: Index Futures

Factor Selection

Select factors from 7 categories:

Category	Factor	Description
Profit	ROE	$\frac{\text{Net Income}}{\text{Average Shareholders}'\text{Equity}}$
Quality	Quick Ratio	Ratio in Last Financial Statement
Growth	YOY ROE	Year-on-year Growth of ROE
Valuation	EP	Earning/Price
Scale	Market Value	Market Value
Liquidity	1-month Trading Amount	20-day Average Trading Amount
Technical	Price Reversal	20-day Quote Change

XGBoost Prediction Model



XGBoost Model

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

- Initial boosting tree: $\hat{y}_i^0 = f_0(x_i) = 0$

  At step t, we need to calculate: $\hat{y}_i^t = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i)$

  Goal: minimize objective function: $Obj^t = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i{}^t\right) + \sum_{i=1}^{t} \Omega(f_i)$

  $$= \sum_{i=1}^{n} l(\, y_i, \hat{y}_i{}^{t-1} + f_t(x_i)) + \Omega(f_i) + \text{constant}$$

- For the regression problem, minimizing $l\,(y_i, \hat{y}_i{}^{t-1} + f_t(x_i))$ equivalent to the t tree is fitting the residual between y_i and $\hat{y}_i{}^{t-1}$. In the following iterations, larger residual samples will be fitted which is equivalent to increasing the weights of underperformed samples.
- Aggregate a series of weak learners to achieve better performance.

XGBoost Model Performance

$$\text{MAE}=\frac{1}{N}\sum_{t=1}^{N}\left|\widehat{y_t}-y_t\right|$$

Mean Absolute Errors between Predicted IC and True IC Value

Factor	ROE	Quick Ratio	YOY ROE	EP	Market Value	1-month Trading Amount	Price Reversal
MAE	0.1046	0.1023	0.0631	0.1205	0.0854	0.1256	0.1291

03 Timing Strategy and Back-test

Multi-factor Dynamic Adjustment Strategy
-- Dynamically Weighting Factors

This report selected 7 factors: ROE, quick ratio, YOY ROE and EP are positive direction factors; market value, 1-month trading amount, and 1-month reversal are negative direction factors. The factor weighting method is as following:

1) At time t, predict factor i's next period $IC_{i,t}$ by XGBoost Model;
2) For positive direction factor, when $IC_{i,t} > 0$, $w_{i,t} = IC_{i,t}$, otherwise, $w_{i,t} = 0$;
3) For negative direction factor, when $IC_{i,t} < 0$, $w_{i,t} = -IC_{i,t}$, otherwise, $w_{i,t} = 0$.
4) Normalized the weight to make $\sum w_{i,t} = 1$; If all factors are invalid in the next period, then equal weighting will be used, $w_{i,t}$=1/7.

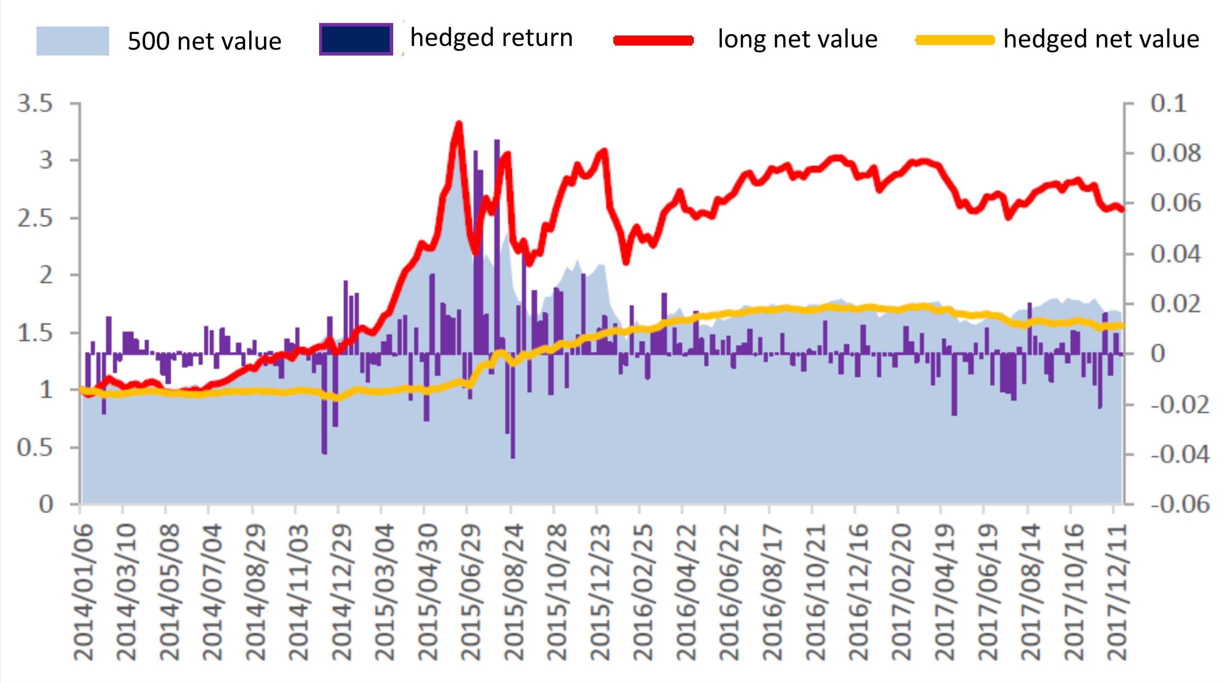
Back-test Parameter Setting

- Adjustment period: 5 trading days
- Stock pool: CSI 500 Index, excluding ST and suspended stocks
- Stock weighting: equal weights
- Factor weighting: take predicted ICs as weights
- In-sample training period: 2008.01-2013.12
- Out-of-sample back-testing: 2014.01-2017.12
- Risk hedging: CSI 500 Index Futures
- Transaction cost: bilateral 3‰

Equally-weighted Multi-factor Model (benchmark)

Since 2017, Equally-weighted Factor Model performs not good with the big drawdown

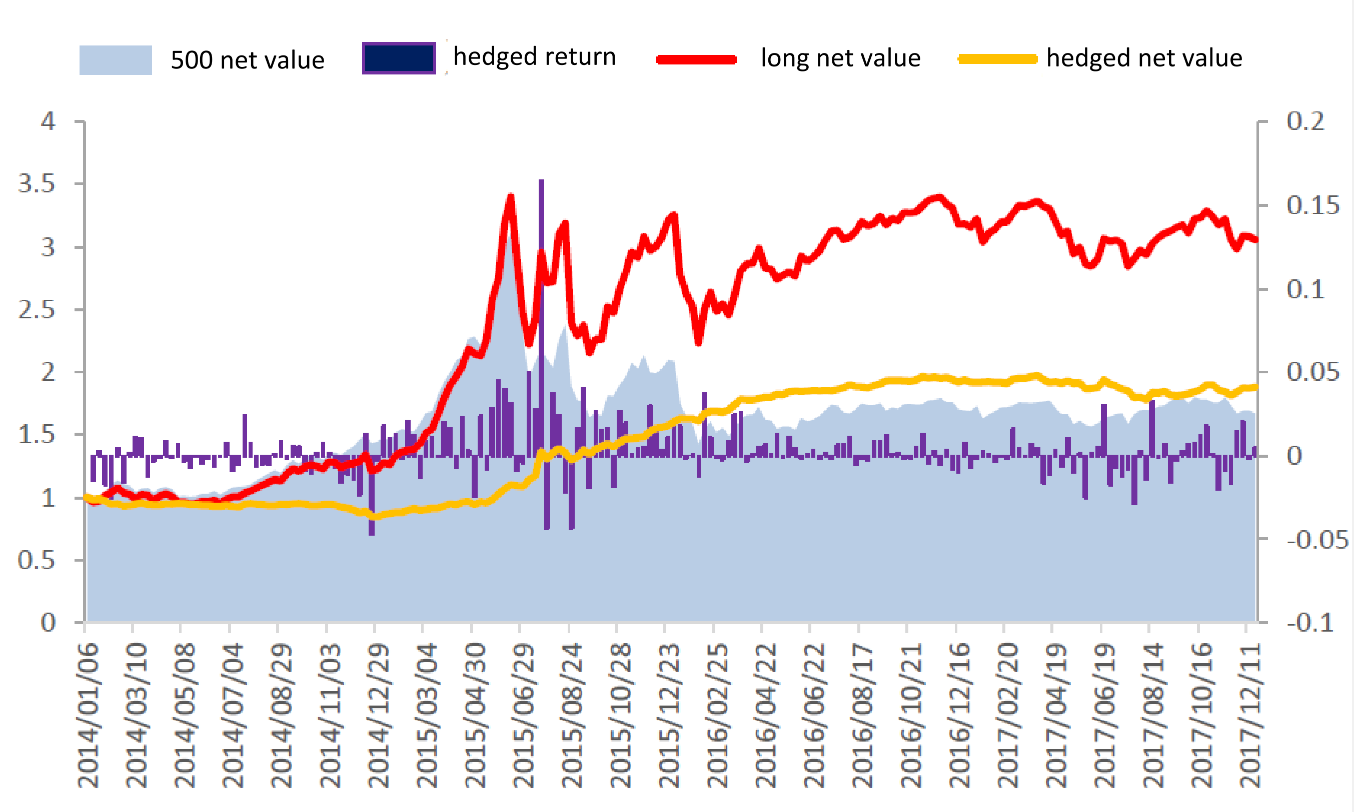
Equally-weighted Factor Strategy Performance



Back-testing results	
Annual Return	11.73%
Annua Volatility	10.75%
Information Ratio	1.09
Drawdown	-11.26%
Cumulative Return	55.83%
Weekly Win Rate	55.44%

Multi-factor Dynamic Timing Strategy (Machine Learning)

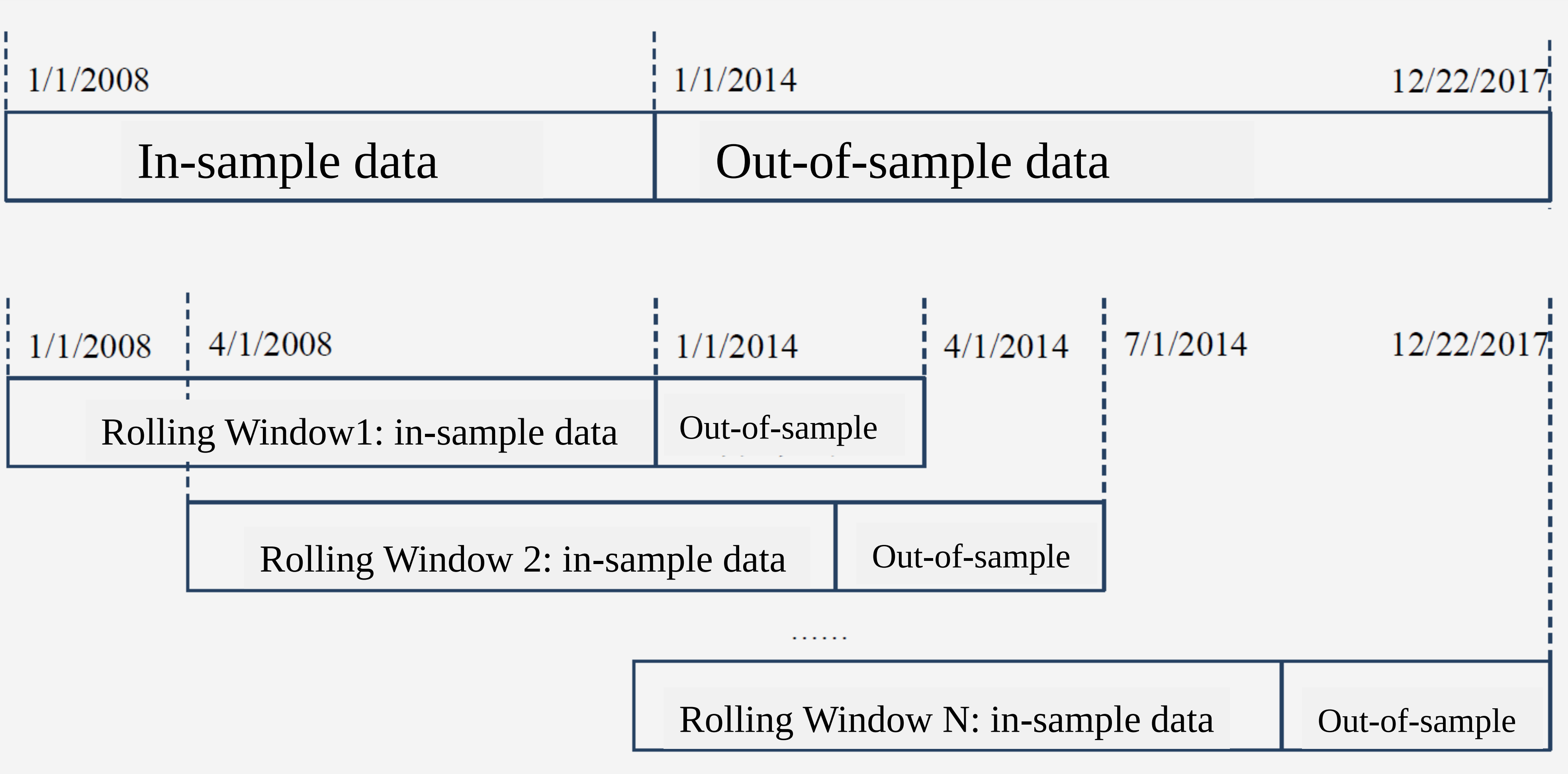
Machine Learning Strategy Performance Since 2014



Back-testing Results Based on XGBoost Model	
Annual Return	17.12%
Annua Volatility	13.22%
Information Ratio	1.30
Drawdown	-15.36%
Cumulative Return	88.18%
Weekly Win Rate	55.96%

Compared with equally-weighted multi-factor model, the multi-factor strategy based on XGBoost model can have significant higher annual return and information ratio, but also higher drawdown.

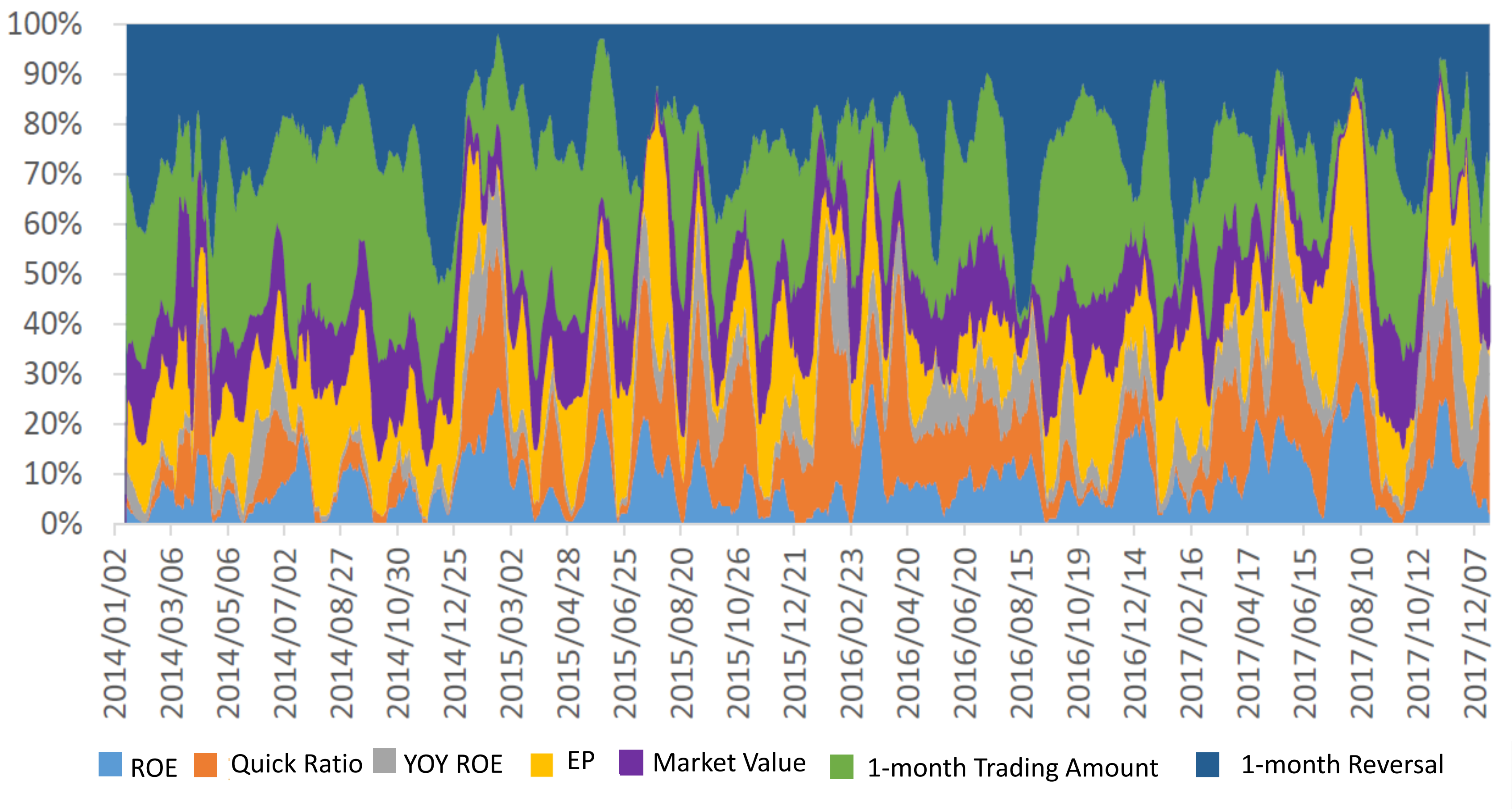
Machine Learning Model Improvement: Rolling Window



Factor weights of the Rolling Window Training Model

Since April 2014, the sum of weights of the market value, 1-month trading amount and 1-month reversal decreased, but the weights of EP and ROE increased significantly.

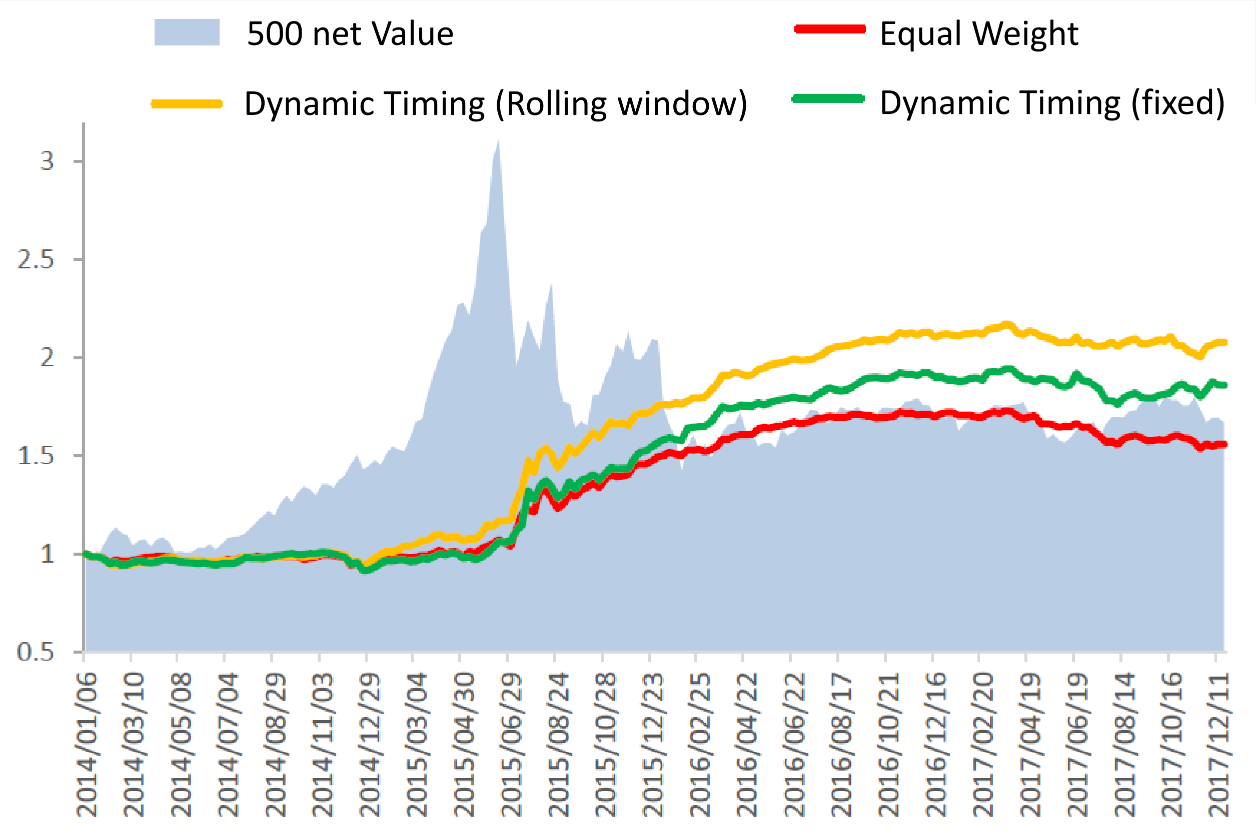
→ Rolling window model can capture the market style shift better



Factor Weights Based on XGBoost Model (Rolling Window)

Back-test Performance Comparison

Rolling window multi-factor strategy's information ratio is 1.74, which is increased by 59.6% and 27.9% than equally-weighted model and fixed XGBoost model, and it has a significant improvement on win rate, annual return and drawback.



Strategy	Equally-weighted	Fixed Window	Rolling Window
Annual Return	11.73%	16.78%	20.08%
Annual Volatility	10.75%	12.30%	11.51%
Information Ratio	1.09	1.36	1.74
Drawback	-11.26%	-9.41%	-7.68%
Cumulative Return	55.83%	85.96%	107.89%
Weekly Win Rate	55.44%	54.92%	63.21%

Back-test Performance Comparison: Year Based

Rolling window multi-factor strategy performs best among three strategies. In the stable market, this strategy can get higher abnormal returns; in the dynamic market, this strategy can get a smaller drawback.

Strategy/Year		2014	2015	2016	2017.12.18
Equally-weighted	Abnormal Return	-6.75%	60.28%	15.14%	-9.45%
	Weekly Win Rate	50.00%	65.31%	63.27%	42.55%%
	Drawback	-7.13%	-7.10%	-1.55%	-11.26%
Dynamic Timing (Fixed Window)	Abnormal Return	-8.28%	70.80%	20.62%	-1.59%
	Weekly Win Rate	43.75%	71.43%	63.27%	40.43%
	Drawback	-9.41%	-6.48%	-1.81%	-9.23%
Dynamic Timing (Rolling Window)	Abnormal Return	-4.56%	83.75%	21.04%	-2.06%
	Weekly Win Rate	52.08%	75.51%	73.47%	51.06%
	Drawback	-6.54%	-6.49%%	-1.21%	-7.68%

04 Conclusion

Conclusion

- Utilizing machine learning model, XGBoost Model to predict factors' future IC value, and weighting the factors with the predicted IC values can have significantly better performance than the equally-weighted factor model;

- Select 1/10 of the stocks from stock pool, CSI 500 Index to construct portfolio, meanwhile, utilize CSI 500 Index Futures to hedge risks. Rolling-window dynamic timing factor strategy has 107.89% cumulative return, 63.21% weekly win rate, 20.08% annual return, and 7.68% drawback;

- In the stable market, rolling-window dynamic timing factor strategy can get higher abnormal returns; in the dynamic market, this strategy can get a smaller drawback.

Thanks !

Cindy (Zixin) Zhou
Washington University in St. Louis