Application of LEAP Model in A-Stock Market

Hengbo LIANG

Tsinghua University

Content

- 1. Quantitative Multi-factor stock selection
- 2. Basic information of LEAP model
- 3. Practical application in A-stock market
    - 3.1 Analysis of rolling window period
    - 3.2 Analysis of the amount of dummy variables
    - 3.3 Adding time series regression
    - 3.4 Adding macro-economic variables

I. Quantitative Multi-Factor Stock Selection

A glance at CAPM

- If M is the Market Portfolio, then the expected return μ_i of any asset i satisfies

$$\mu_i = r_f + \ \beta_i \left[\mu_m - r_f\right]$$

$$\beta_i = \frac{Cov(r_i, r_m)}{Var(r_m)}$$

- r_f: risk-free return;

- The average return is **linear** to the price premium.

From CAPM to APT

- The CAPM does not completely explain the cross-section of expected asset returns.
- The evidence suggests that one or more factors may be required to characterize expected returns.
- Arbitrage Pricing theory allows for multiple risk factors.
- APT assumes the return is

$$R_i = \alpha_i + \beta_i' f + \epsilon_i \, , \qquad with \; E[\epsilon_i | f] = 0 \, , E\left[\epsilon_i^2\right] = \sigma_i^2 \leq \sigma^2 < \infty$$

R_i : return for asset i

α_i : intercept of the factor model

β_i : $(\mathrm{K} \times 1)$: vector of factor sensitivities for asset i

f : a $(\mathrm{K} \times 1)$ vector of common factor realizations

ϵ_i : disturbance term.

Quant Multi-factor Stock Selection

Pros:

- Anecdotal evidence suggests ~80% of stock picking is done ‘by hand’ (individuals making calls on fundamentals)
    - Relies heavily on talent (or luck) of individual analyst
    - Individuals can only process limited information with little academic research on subject
    - Human nature suggests cognitive biases likely
- Certain fundamental, technical, and market emotional factors may contain valuable information in predicting stock returns
- Quantitative selection is accurate, flexible, unbiased

Methodology

<u>Back test</u>

- Consider factor performance and consistency (both long and short candidates) in predicting returns
- Evaluate the performance with Sharp Ratio, Spearman IC, and etc.
- Select most promising factors for inclusion in the model

<u>Weight</u>

- Once individual factors selected must decide on weights for final model by either:
  a) ‘Eye balling’ best factors and assigning weights for a scoring model
  b) Pushing individual factor portfolios into a mean-variance optimizer
  c) Machine learning

Stock Selection Training

Given information **prior to** period *t*, predict performance of stocks of period *t*

- Training set

Predictor 1a	Predictor 1b	Predictor 2a	Predictor 2b	…	Goal
Return of period *t*-1	Factor of period *t*-1	Return of period *t*-2	Factor of period *t*-2	…	Return of period *t*

Learning a ranking function to rank testing data

– Select n highest to buy, n lowest to short-sell

Fama-Macbeth Regression

$$r_{i,t} = \alpha_i + \sum_{k=1}^{K_t} \beta_{k,t} f_{i,k,t-1} + \epsilon_{i,t}$$

$$\beta_{k,t} = \mu_k + \sum_{k=1}^{K_t} \varphi_{k,l} \beta_{k,t-l} + \epsilon_{k,t}$$

- First regress: asset returns --- factors (before period t) → asset's beta (at period t)
- Second regress: betas --- betas (before period t) (Time Series: AR model)

II. A Summary of LEAP Model

A summary to LEAP

Yin Luo, CFA, CPA
Wolfe research

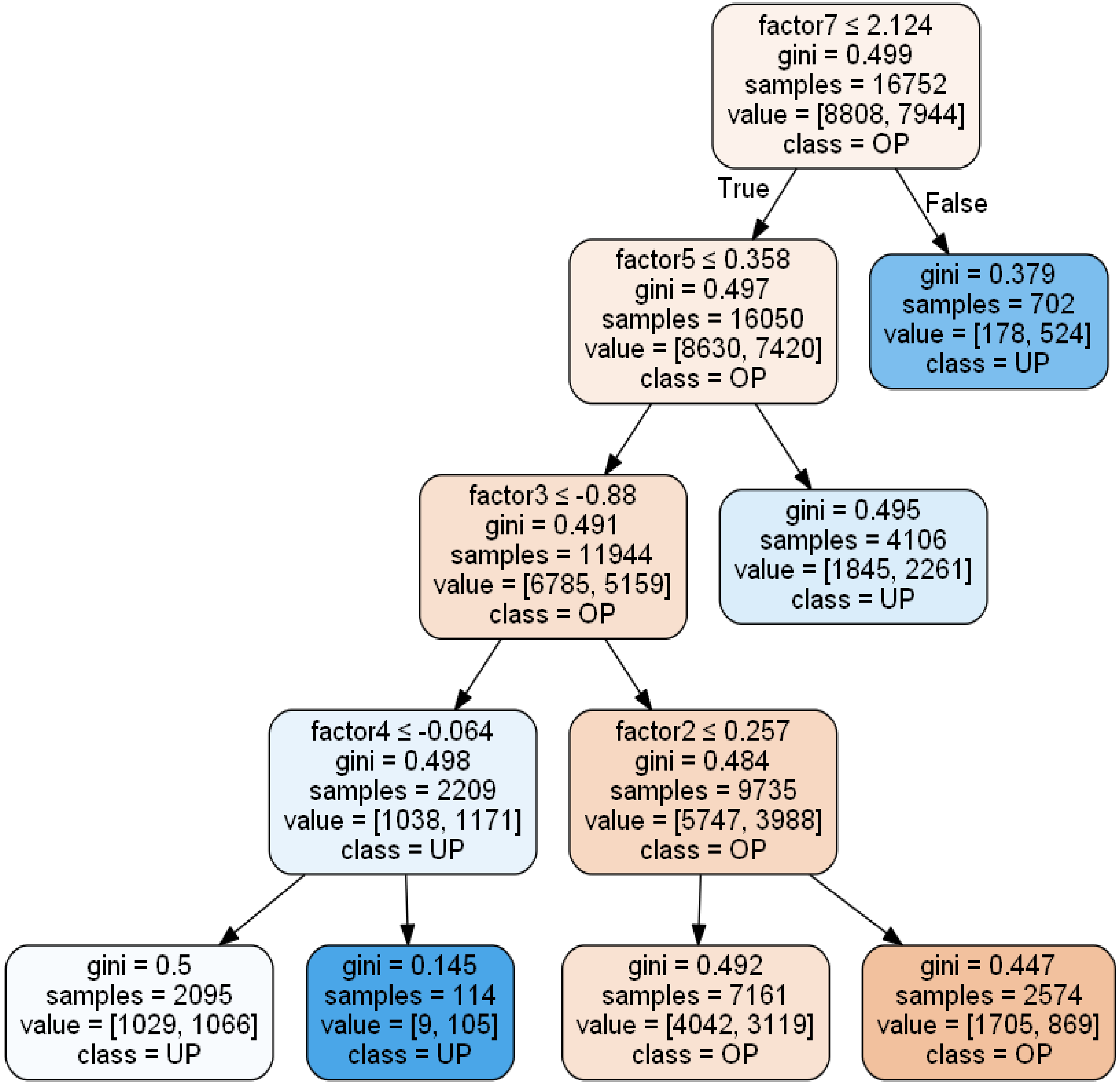
- Linear-Economic Alpha Processing

- Machine Learning
- Fama-Macbeth Regression
- Macro-economic variables

Nonlinear machine learning component

- Random Forest:
  - Automatically conducts feature selection.

- CART Decision Tree:
  - Based on the factor selected through RF.
  - Classification: Top 30% as “Out Performance”, Bottom 30% as “Under Performance”
  - Keep the model simple and parsimonious. Only 4-6 terminal nodes.
  - Create dummy variables for each of the terminal nodes.
  - Argument: stocks falling into each of the terminal node may have different payoff patterns.



$$D_{i,m,t} = \begin{cases} 1, if\ true \\ 0, if\ false \end{cases}$$

Stands for the dummy variable for the *m*th terminal node from the CART model in period *t*, for stock *i*.

Modified Fama-Macbeth Regression

- On each month, we perform the cross-sectional Fama-Macbeth regression, on the previous 120 months of data. (Rolling)

- $$r_{i,t} = \sum_{k=1}^{K_t} \underbrace{\beta_{k,t} f_{i,k,t-1}}_{\text{fixed}} + \sum_{m=1}^{M_t} \underbrace{\gamma_{m,t} D_{i,m,t-1}}_{\text{Change with periods}} + \epsilon_{i,t}$$

Where,

$r_{i,t}$ is the normalized and neutralized return for stock *i* in period *t*.

$\beta_{k,t}$ is the estimated coefficient for factor *k* in period *t*.

$f_{i,k,t-1}$ is the normalized and neutralized score of stock *i*, for factor *k*, in period *t-1*;

$\gamma_{m,t}$ is the estimated coefficient for the *m*th terminal node from our CART model in period t;

$D_{i,m,t-1}$ is the dummy variable for the *m*th terminal node from the CART model in period t-1;

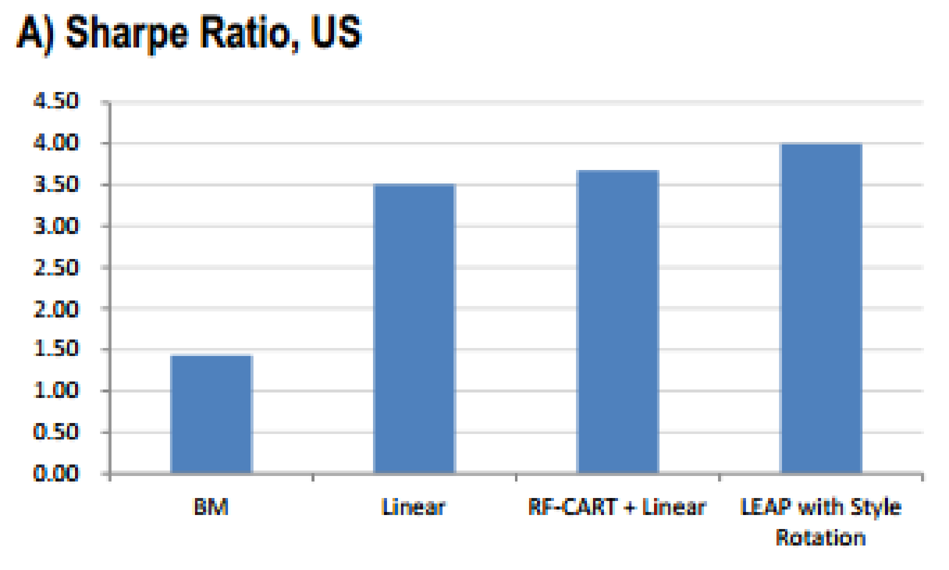
$\epsilon_{i,t}$ is the regression residual, for stock *i* in period *t*.

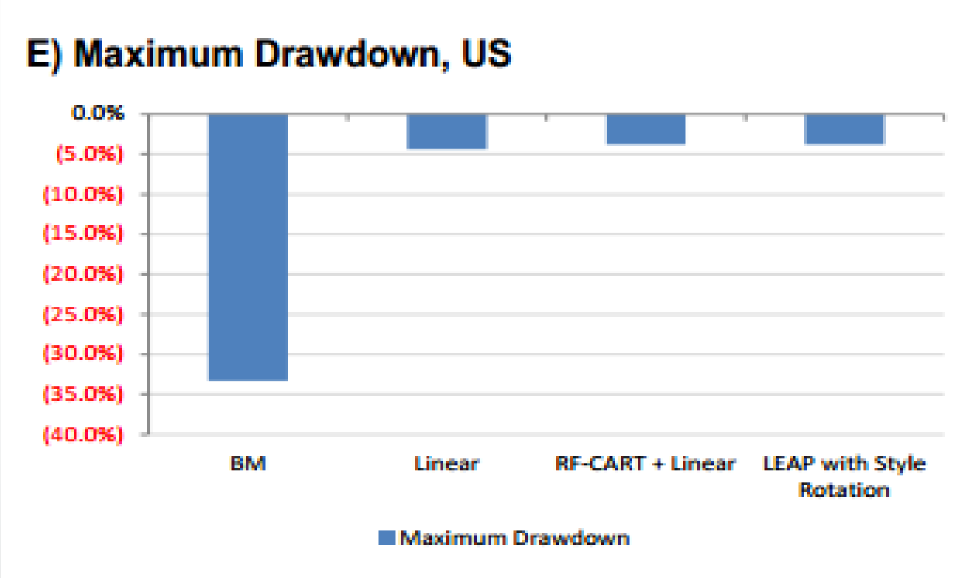
Improvement: A style rotation

- Adding time-series regression and Macro-economic variables:
  - $\beta_{k,t} = \mu_k + \sum_{h=1}^{H} \psi_{k,h} E_{h,t-1} + \sum_{l=1}^{L} \varphi_{k,l} \beta_{k,t-l} + \epsilon_{k,t}$
  - $\gamma_{m,t} = \theta_m + \sum_{h=1}^{H} \upsilon_{m,h} E_{h,t-1} + \sum_{l=1}^{L} \chi_{m,l} \gamma_{m,t-l} + \varepsilon_{m,t}$
- Where,
- $\beta_{k,t}$ is the factor return (estimated from the previous section) for factor k in period t;
- μ_k is the intercept term, i.e., the unconditional factor return, expected to be positive ;
- $E_{h,t-1}$ is the *h*th macro variable at period t-1,i.e., we use previous period's macro variable to predict the subsequent period's factor return;

Region	VRP	Political Uncertainty	Economic Uncertainty	Interest Rate	Equity Market	Credit Spread	Business Cycle	Currency	Oil & Commodities	Seasonality
US	US VRP	US	US	US	US	US Corporate	Philly Fed ADS			√
Canada	US VRP	Canada	Canada	US				CAD/USD		
LATAM	US VRP	US	US		US			Trade Wgted USD		
Europe	US VRP	Europe	US		Europe	Europe High Yield	Exp Europe GDP	EUR/USD		√
UK		UK	UK	UK	Global	US High Yield		GBP/USD		√
Emerging EMEA	US VRP	Europe	US		Global			Trade Wgted USD		
AxJ		US	AxJ	US	AxJ	US High Yield	Exp AxJ GPD		Oil Shock	
Japan			Japan			US High Yield	Japan Nowcasting			√
ANZ	US VRP	Australia	ANZ	Australia	Australia		Exp ANZ GDP	AUD/USD	Baltic	

Result in US market





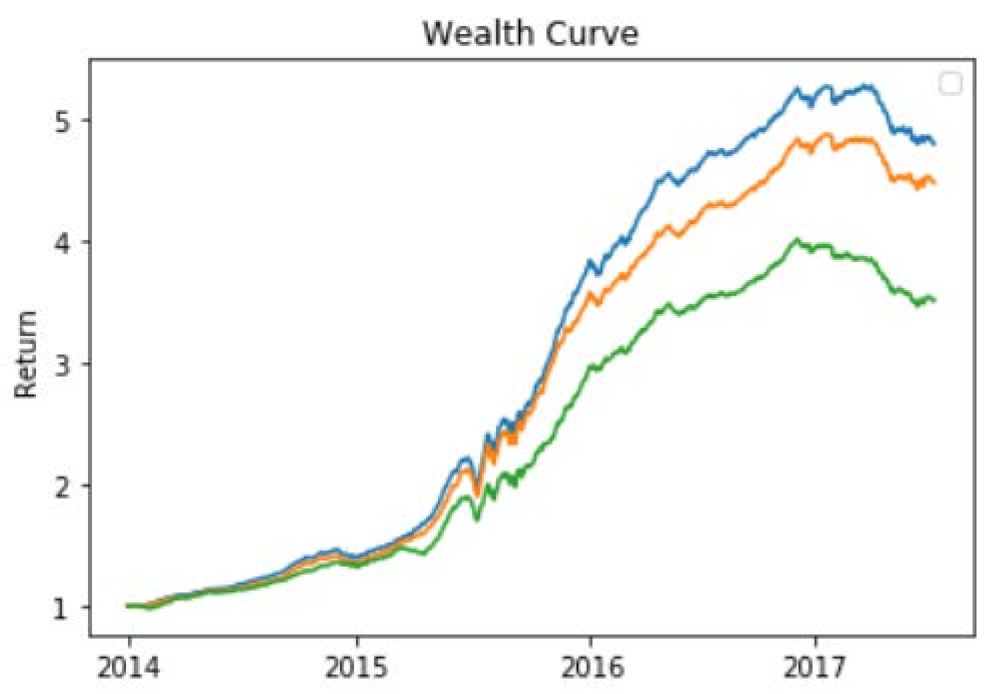
III. A practical Application

编号	37	36\	5\	35	33	11	7	16	34
IR	5.55	5.37	5.11	4.69	4.41	4.3	3.84	3.8	3.75
Rtn/%	37.73	38.97	41.84	33.22	34.61	34.02	26.84	36.93	36.22

3.1 Study of Rolling Window Period

- Set the rolling window period

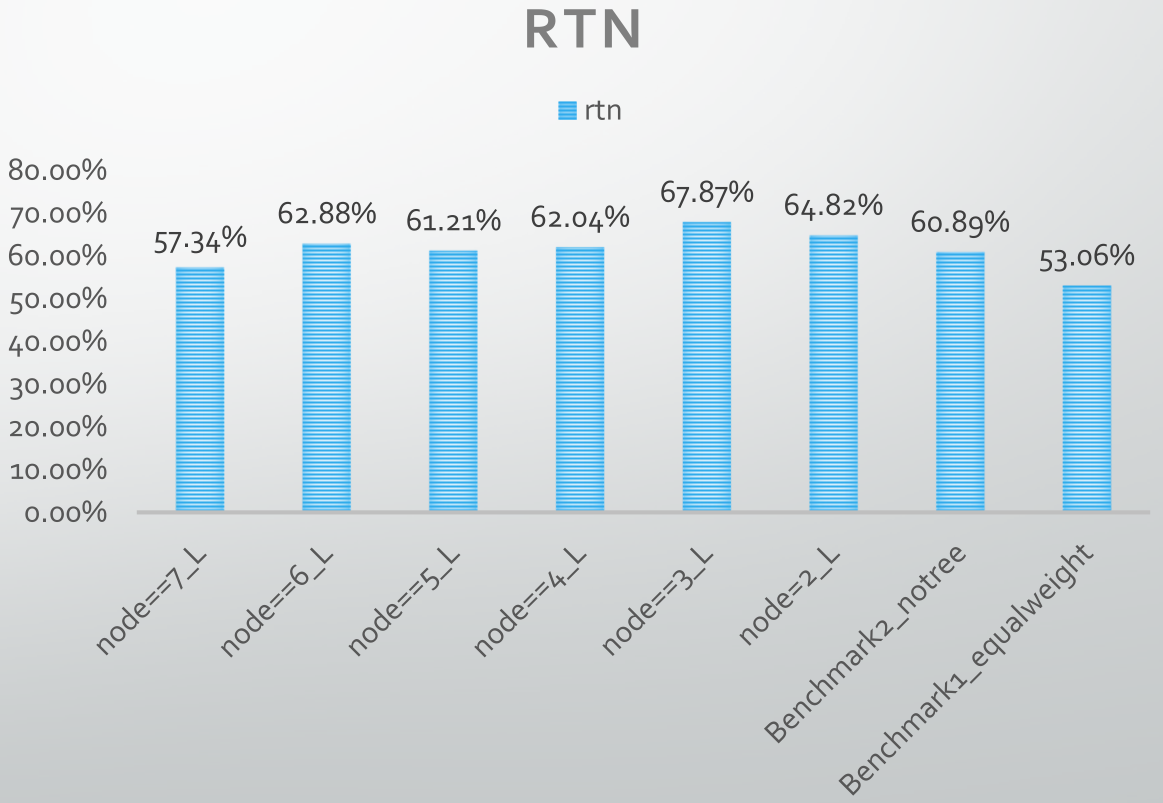
  as 20D, 60D, 120D



Rolling window period	IR	Rtn	Max drawdown	std	IC_avg/std
20D	4.07	42.67 %	-13.8%	8.82%	1.14
60D	4.7	52.91%	-10.58%	9.12%	1.34
120D	4.85	55.84%	-10.94%	9.24%	1.41

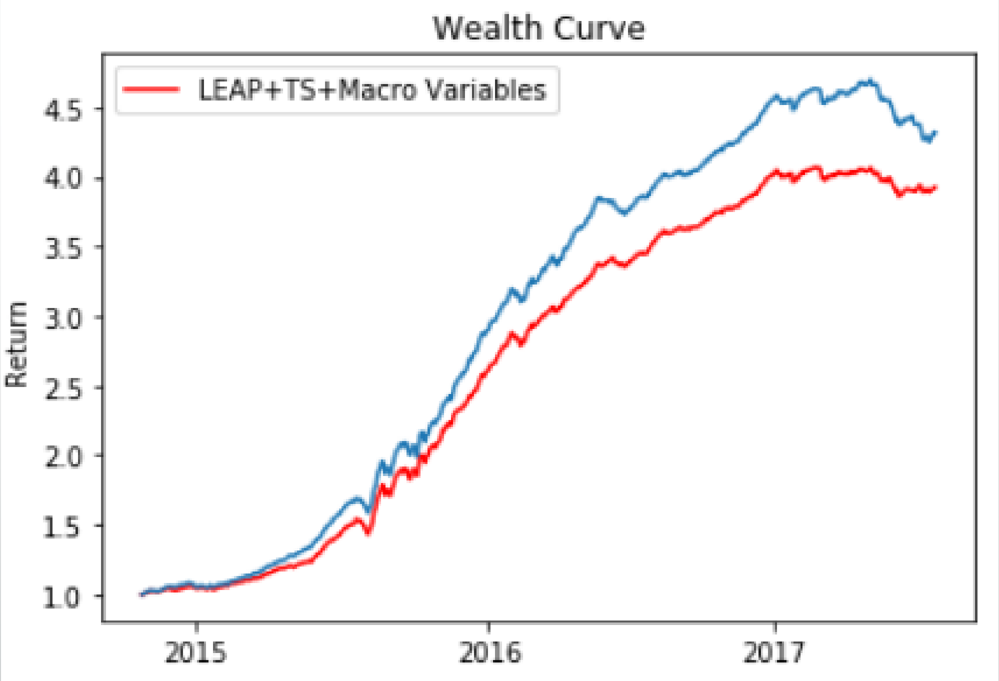
3.2 Study of amount of nodes

- Choosing 13 factors into the regression model.
- Set the amount of nodes as 2-7.
- Backtest period: 2014.9-2017.7



3.3 Adding macro-economic variables

- Choose 3 macro-economic variables as an attempt: GDP, PPI , CPI
- However, the result is not so good.
- Maybe macro-economic variables are not so suitable to A-stock market…
- China :   High Frequency instead of weekly, monthly data.



A conclusion

	IR	Rtn	Max drawdown	std
Node + TS +Macro	5.37	65.24%	-7.07 %	9.43 %
Node=3 with TS	5.53	66.97%	-6.6%	9.35%
node==3_L	5.51	67.87%	-6.1 %	9.49 %
Benchmark 2 Only regression	5.26	60.89%	-7.01 %	9.12 %
Benchmark 1 Equal Weight	4.61	53.06%	-8.87 %	9.32 %

- Thanks for listening！