Project for Investment

June 25, 2024

1.1 Fama-French Five-Factor Model

PHBS 北京大学汇丰商学院

From CAPM to Fama-French Five-Factor Model

CAPM

Markowitz (1952)

- Modern Portfolio Theory (MPT)
- Systematic risk explains return
- $E(r_i) - r_f = \beta_i \cdot E(r_M - r_f) + \epsilon_i$
- Restrictions:
  - ✓ Imperfect stock market
  - ✓ Anomaly exists ($\alpha \neq 0$)

Fama-French Three-Factor Model

Fama and French (1993)

- Additional two factors: size & BP
- Model incorporating size & BP factors performed better on the NASDAQ market than other models
- $R_{it} - R_{Ft} = \alpha_i + \beta_i \cdot (R_{Mt} - R_{Ft}) + s_i \cdot SMB_t + h_i \cdot HML_t + \epsilon_i$

Fama-French Five-Factor Model

Fama and French (2015)

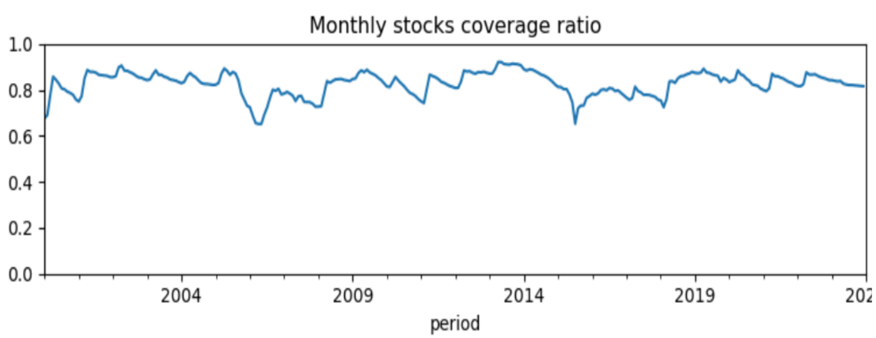
- Adding profitability & investment
- $M_t = \sum_{\tau=1}^{\infty} \frac{E(Y_{t+\tau} - dB_{t+\tau})}{(1+r)^{\tau}}$
- Miller and Modigliani (1961)
- $R_{it} - R_{Ft} = \alpha_i + \beta_i \cdot (R_{Mt} - R_{Ft}) + s_i \cdot SMB_t + h_i \cdot HML_t + r_i \cdot RMW_t + c_i \cdot CMA_t + \epsilon_i$

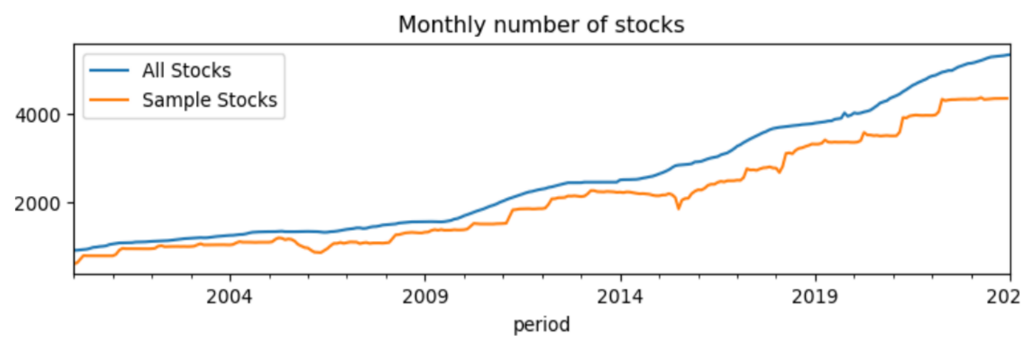
	2 × 3 Factors				2 × 2 Factors				2 × 2 × 2 × 2 Factors			
	GRS	$A\lvert a_i\rvert$	$\frac{A\lvert a_i\rvert}{A\lvert\bar{r}_i\rvert}$	$\frac{A(\hat{\alpha}_i^2)}{A(\hat{\mu}_i^2)}$	GRS	$A\lvert a_i\rvert$	$\frac{A\lvert a_i\rvert}{A\lvert\bar{r}_i\rvert}$	$\frac{A(\hat{\alpha}_i^2)}{A(\hat{\mu}_i^2)}$	GRS	$A\lvert a_i\rvert$	$\frac{A\lvert a_i\rvert}{A\lvert\bar{r}_i\rvert}$	$\frac{A(\hat{\alpha}_i^2)}{A(\hat{\mu}_i^2)}$
Panel A: 25 Size-B/M portfolios												
HML	3.62	0.102	0.54	0.38	3.54	0.101	0.53	0.36	3.40	0.096	0.51	0.36
HML RMW	3.13	0.095	0.50	0.24	3.11	0.096	0.51	0.26	3.29	0.089	0.47	0.24
HML CMA	3.52	0.101	0.53	0.39	3.46	0.100	0.53	0.37	3.18	0.096	0.51	0.35
RMW CMA	2.84	0.100	0.53	0.22	2.78	0.093	0.49	0.19	2.78	0.087	0.46	0.13
HML RMW CMA	2.84	0.094	0.50	0.23	2.80	0.093	0.49	0.23	2.82	0.088	0.46	0.18
Panel B: 25 Size-OP portfolios												
HML	2.31	0.108	0.68	0.51	2.31	0.109	0.68	0.51	1.91	0.089	0.56	0.37
RMW	1.71	0.067	0.42	0.12	1.82	0.078	0.49	0.16	1.73	0.059	0.37	0.05
HML RMW	1.64	0.062	0.39	0.16	1.74	0.058	0.36	0.03	1.62	0.064	0.40	0.06
HML CMA	3.02	0.137	0.86	0.90	2.85	0.135	0.85	0.86	2.06	0.102	0.64	0.49
RMW CMA	1.87	0.075	0.47	0.12	1.67	0.066	0.42	0.05	1.61	0.068	0.43	0.05
HML RMW CMA	1.87	0.073	0.46	0.12	1.73	0.066	0.42	0.06	1.60	0.069	0.43	0.07
Panel C: 25 Size-Inv portfolios												
HML	4.56	0.112	0.64	0.57	4.40	0.107	0.61	0.53	4.32	0.100	0.57	0.56
CMA	4.03	0.105	0.60	0.47	4.05	0.106	0.61	0.47	4.23	0.123	0.70	0.62
HML RMW	4.40	0.106	0.61	0.57	4.26	0.103	0.59	0.52	4.45	0.116	0.66	0.66
HML CMA	4.00	0.099	0.57	0.43	3.97	0.098	0.56	0.41	3.70	0.084	0.48	0.35
RMW CMA	3.33	0.085	0.49	0.29	3.28	0.082	0.47	0.26	3.50	0.082	0.47	0.27
HML RMW CMA	3.32	0.085	0.49	0.29	3.27	0.082	0.47	0.27	3.59	0.082	0.47	0.28

1.1 Fama-French Five-Factor Model

A-share Data Processing

Stock Pool	The entire A-share market(excluding Beijing Stock Exchange), excluding ST and PT stocks at any time t, as well as stocks that are suspended or have been listed for less than one month, and stocks with a negative BP value.
Time Period	From January 2000 to December 2023
Market Factor	Wind All A-share Index
Size Factor	Total market capitalization = Stock price per share × Total number of shares issued
Valuation Factor	Book-to-price ratio (BP), which is the inverse of the price-to-book ratio.
Profitability Factor	ROE excluding non-recurring gains and losses, diluted
Investment Factor	The rate of change in total assets for the current period relative to the previous period
Risk-free Rate	Set to zero, individual stock returns are calculated using Wind's post-adjusted closing prices
Frequency	Calculate factor values on a monthly basis and recalculate the corresponding factor returns through re-layering (details in the next section)





1.1 Fama-French Five-Factor Model

Five-Factors Construction

Grouping method	Group calculation method	Factor construction
Size: divided into two groups according to the median, recorded as large (B) or small (S); **BP, profitability, investment:** divided into three groups according to the 30%, 70% quartile, recorded as BP - High (H), Medium (N), Low (L) profitability - Strong (R), Medium (N), Weak (W) Investment - Conservative (C), Medium (N), Strong (A)	Size & BP double-sorting	$SMB_{BP} = (SH + SN + SL)/3 - (BH + BN + BL)/3$
	Size & profitability double-sorting	$SMP_{op} = (SR + SN + SW)/3 - (BR + BN + BW)/3$
	Size & investment double-sorting	$SMB_{inv} = (SC + SN + SA)/3 - (BC + BN + BA)/3$
	Average of the three groups	$SMB = (SMB_{BP} + SMB_{op} + SMB_{inv})/3$
	Size & BP double-sorting	$HML = (BH + SH)/2 - (BW + SW)/2$
	Size & profitability double-sorting	$RMW = (BR + SR)/2 - (BL + SL)/2$
	Size & investment double-sorting	$CMA = (BC + SC)/2 - (BA + SA)/2$

Large cap stock

- BP/profit/investment top 30% from smallest to largest BL/BW/BC
- BP/profit/investment from small to large 30% to 70% BN
- BP/profit/investment from small to large after 30% BH/BR/BA

Small cap stock

- BP/profit/investment top 30% from smallest to largest SL/SW/SC
- BP/profit/investment from small to large 30% to 70% SN
- BP/profit/investment from small to large after 30% SH/SR/SA

- Market factor construction:
  - ✓ 1) the size-weighted average of all stocks in our sample;
  - ✓ 2) Wind All A-shares index
- The return of the portfolio is equally-weighted calculated.
- Also, we use the five factors calculated by CSMAR for comparison.

1.1 Fama-French Five-Factor Model

Descriptive Statistics

	MARKET	SMB	HML	RMW	CMA
MARKET	1.00	0.15	0.00	-0.27	0.06
SMB	0.15	1.00	-0.14	-0.42	0.25
HML	0.00	-0.14	1.00	-0.35	0.56
RMW	-0.27	-0.42	-0.35	1.00	-0.81
CMA	0.06	0.25	0.56	-0.81	1.00

	MARKET	SMB	HML	RMW	CMA
Count	288	300	300	300	300
Mean	0.006504	0.007024	0.002777	-0.000029	0.000038
Std	0.07803	0.032333	0.031581	0.02924	0.021225
Min	-0.276813	-0.127652	-0.136664	-0.151348	-0.062349
25%	-0.039319	-0.012187	-0.012035	-0.017581	-0.013831
50%	0.006333	0.007776	0.000959	0.000299	0.000148
75%	0.046209	0.025071	0.021286	0.015629	0.012887
Max	0.289248	0.175586	0.133087	0.122123	0.079322

Factor	Ret_mean(%)	T_value	P_value	P_star
MARKET	0.650424	1.132198	0.257551	
SMB	0.68911	3.457133	0.000546	***
HML	0.3119	1.803435	0.07132	*
RMW	-0.017511	-0.109176	0.913063	
CMA	0.026649	0.226446	0.820855	

%	Small Size	2	3	4	Big Size
Valuation					
Low	1.19	0.85	0.49	0.50	0.53
2	1.56	1.30	0.84	0.81	0.56
3	1.93	1.34	0.90	0.74	0.56
4	2.13	1.53	0.96	0.76	0.61
High	2.71	1.67	0.92	0.76	0.64
Profitability					
Low	2.66	1.31	0.51	0.27	0.21
2	2.37	1.40	0.79	0.56	0.32
3	1.70	1.43	0.93	0.73	0.62
4	1.05	1.55	1.16	1.07	0.61
High	0.79	0.85	0.92	0.84	0.73
Investment					
Low	2.47	1.32	0.58	0.37	0.46
2	2.34	1.51	0.79	0.67	0.47
3	1.83	1.50	0.97	0.78	0.64
4	1.58	1.32	0.92	0.94	0.54
5	0.97	1.01	1.01	0.77	0.74

1.2 Model Comparison

We follow Fama and French (2015) and test the traditional three-factor model first, and then test four-factor and five-factor models:

$$R_{it} = \alpha_i + b_i R_{Mt} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \epsilon_{it}.$$

- We do a GRS test on the Alpha from the regression, and give the p-value, notated as PGRS.
- The GRS test is an F-statistic test that tests whether the N asset regression Alpha is jointly zero (Gibbons, Ross and Shanken, 1989).

The five-factor regression model performs better than the other regression models.

Factor	GRS	PGRS
Subgroup A	**Size-value**	
RM SMB HML	3.6211	0.00%
RM SMB HML RMW	1.8847	0.80%
RM SMB HML RMW CMA	1.7775	1.48%
Subgroup B	**Size-earnings**	
RM SMB HML	6.2666	0.00%
RM SMB HML RMW	3.9714	0.00%
RM SMB HML RMW CMA	3.7326	0.00%
Subgroup C	**Size-investment**	
RM SMB HML	3.615	0.00%
RM SMB HML RMW	2.1044	0.21%
RM SMB HML RMW CMA	1.9958	1.48%

1.3 Factorial multicollinearity test and treatment

There may be a problem of multicollinearity, so we did Inter-factor regression.

Variables	SMB	HML	RMW	CMA	Intercept	MARKET	R^2
R_M							
beta	-0.412106	0.076357	-2.07169	-1.903513	0.009499		0.170119
t-value	-2.522001	0.413208	-7.289948	-4.717031	2.18814		
SMB							
beta		-0.426656	-0.899758	-0.350639	0.008518	-0.053339	0.402592
t-value		-6.940329	-9.204679	-2.348068	5.711342	-2.522001	
HML							
beta	-0.340905		0.038092	1.031757	0.005147	0.007897	0.501945
t-value	-6.940329		0.382545	8.596977	3.745087	0.413208	
RMW							
beta	-0.256075	0.013568		-1.038336	0.002339	-0.076313	0.793183
t-value	-9.204679	0.382545		-20.13044	2.822969	-7.289948	
CMA							
beta	-0.0545	0.200705	-0.567064		0.000175	-0.038294	0.788386
t-value	-2.348068	8.596977	-20.13044		0.28245	-4.717031	

- CMA has strong covariance with RMW, and we remove the portion of the investment factor that is covariant with the other factors

$$CMAO = CMA - \widehat{CMA}$$
$$\widehat{CMA} = \alpha + bR_M + sSMB + hHML + rRMW$$

$$R_{it} = \alpha_i + b_i R_{Mt} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMAO_t + \epsilon_{it}$$

1.4 Size-Valuation (BP) Grouping

		Small	2	3	4	Big
				Alpha		
Low		-0.000842	-0.002815	-0.003861	-0.001128	0.003874
	2	0.001225	0.001301	-0.001203	-0.000297	0.003446
	3	0.002223	0.000795	-0.00122	-0.00053	0.002592
	4	0.001063	0.000528	-0.002078	8.13E-05	-0.000341
High		0.002448	-0.000198	-0.002296	-0.001681	0.001059
				beta_RM		
Low		1.050782	1.03783	1.043481	1.036438	0.939801
	2	1.027772	1.006791	1.024301	1.040264	1.019913
	3	1.033385	1.027302	1.038891	1.03392	1.036884
	4	1.032977	1.050815	1.008727	1.062059	1.028341
High		1.013446	1.027976	1.023904	1.043888	0.904082
				beta_SMB		
Low		1.435603	0.904379	0.605424	0.261243	-0.454722
	2	1.491563	0.942342	0.681858	0.31003	-0.431001
	3	1.411425	1.014653	0.750006	0.348678	-0.386459
	4	1.426219	1.035514	0.697503	0.322751	-0.283274
High		1.400036	1.044356	0.56371	0.324872	-0.248153
				beta_HML		
Low		-0.419666	-0.553523	-0.732338	-0.806593	-0.675589
	2	-0.291267	-0.405649	-0.519866	-0.464902	-0.397735
	3	-0.103631	-0.107489	-0.16757	-0.270022	0.092634
	4	0.114854	0.072477	-0.011049	0.038319	0.554477
High		0.35036	0.267944	0.357661	0.498357	0.841182
				beta_RMW		
Low		-0.202273	-0.371338	-0.393219	-0.258195	0.04595
	2	-0.099745	-0.405995	-0.337218	-0.186796	0.051455
	3	-0.110854	-0.266367	-0.240934	-0.188808	0.108925
	4	-0.082938	-0.23825	-0.315643	-0.236074	0.169757
High		-0.263633	-0.341467	-0.409266	-0.247358	0.175776
				beta_CMAO		
Low		0.045015	0.445698	0.220263	-0.170568	0.123887
	2	-0.022385	0.308124	-0.026151	0.066579	0.213934
	3	0.1186	0.090311	0.117524	0.051247	0.036373
	4	-0.051946	0.195231	0.215579	0.015138	0.118492
High		0.111903	0.232274	0.306946	0.104655	0.020687

		Small	2	3	4	Big
				t(Alpha)		
Low		-0.634793	-1.834009	-2.243768	-0.755773	2.567249
	2	0.94555	0.91134	-0.741465	-0.174801	2.541842
	3	1.728026	0.485523	-0.779813	-0.334141	1.795996
	4	0.831825	0.370475	-1.350816	0.051132	-0.241594
High		1.478972	-0.146253	-1.646539	-1.234719	0.728097
				t(R_M)		
Low		62.01171	52.9152	47.4499	54.3378	48.73933
	2	62.0931	55.20254	49.39404	47.97458	58.87343
	3	62.85355	49.08018	51.98758	51.01482	56.21803
	4	63.24725	57.66954	51.31523	52.29838	57.02872
High		47.91107	59.46623	57.45087	59.99659	48.63344
				t(SMB)		
Low		29.02525	15.79738	9.431712	4.692277	-8.079233
	2	30.87225	17.70143	11.26474	4.898369	-8.523443
	3	29.41073	16.60756	12.85804	5.894062	-7.178419
	4	29.91696	19.46958	12.15624	5.444865	-5.382005
High		22.67539	20.69743	10.8361	6.396829	-4.573265
				t(HML)		
Low		-8.348403	-9.513252	-11.22538	-14.2545	-11.81041
	2	-5.93166	-7.497361	-8.450403	-7.227162	-7.739077
	3	-2.124687	-1.731058	-2.826605	-4.491046	1.69299
	4	2.370482	1.340791	-0.189467	0.636053	10.36523
High		5.583277	5.224813	6.764678	9.654999	15.25303
				t(RMW)		
Low		-3.353499	-5.318918	-5.02326	-3.802823	0.669473
	2	-1.692921	-6.253742	-4.56833	-2.420106	0.834413
	3	-1.894172	-3.575104	-3.387096	-2.617159	1.6591
	4	-1.426612	-3.673268	-4.510964	-3.265788	2.644746
High		-3.501348	-5.549269	-6.451241	-3.993909	2.656359
				t(CMAO)		
Low		0.352597	3.016154	1.329382	-1.1869	0.852761
	2	-0.179499	2.242348	-0.167373	0.407532	1.639054
	3	0.957439	0.572672	0.780572	0.335612	0.261745
	4	-0.422142	1.422089	1.455586	0.098935	0.872173
High		0.702158	1.783387	2.285903	0.798347	0.147699

- The modified five-factor model explains basically all of the excess returns.
- Portfolio returns follow market factors closely.
- The regression coefficients of SMB are monotonically decreasing from left to right.
- The regression coefficients of the valuation factor increase monotonically from top to bottom.

1.5 Size-Earnings Grouping

	Small	2	3	4	Big
Alpha					
Low	0.005129	0.001098	-0.005311	-0.004169	-0.001966
2	0.003806	-0.000394	-0.001362	-0.000874	-0.001546
3	0.000393	-0.00077	-0.001239	-0.001728	0.002076
4	-0.001753	0.002342	-0.000371	0.001074	0.001854
High	-0.005968	-0.002753	-0.002168	0.000899	0.003875
beta_RM					
Low	1.027161	1.036508	1.035143	1.050095	1.051537
2	1.018112	1.038165	1.034275	1.042718	1.066957
3	1.03405	1.023722	1.006629	1.05675	0.994599
4	1.033809	1.016159	1.026441	1.041063	0.939437
High	1.05232	1.050536	1.035414	1.034752	0.919906
beta_SMB					
Low	1.400244	0.89774	0.679584	0.299295	-0.296942
2	1.301662	1.015093	0.601098	0.294984	-0.35824
3	1.456211	1.165839	0.750726	0.377235	-0.425355
4	1.524036	0.949314	0.680551	0.408347	-0.389279
High	1.624324	0.890824	0.639839	0.243258	-0.491062
beta_HML					
Low	-0.085287	-0.163545	-0.144503	-0.128628	0.192358
2	-0.185641	-0.12581	-0.17173	-0.224639	0.062844
3	-0.098689	-0.105657	-0.229958	-0.153698	0.283878
4	-0.162493	-0.185088	-0.207361	-0.128265	0.303678
High	-0.120687	-0.084589	-0.095455	-0.151749	0.233061
beta_RMW					
Low	-0.645656	-0.807817	-0.790011	-0.771367	-0.829407
2	-0.339924	-0.427977	-0.528488	-0.501222	-0.467697
3	0.026626	-0.097908	-0.389201	-0.18967	-0.011214
4	0.216481	-0.064239	0.011295	0.032041	0.098667
High	0.626645	0.231691	0.184165	0.197431	0.408548
beta_CMAO					
Low	0.408148	0.457054	0.159135	-0.279027	-0.22336
2	0.194526	0.131796	0.211392	0.040197	0.307171
3	-0.061532	0.117426	0.238455	0.085647	0.214825
4	-0.106889	0.09506	0.004254	0.07178	0.283626
High	-0.360044	0.207645	0.126786	0.169881	0.054583

	Small	2	3	4	Big
t(Alpha)					
Low	4.070238	0.873114	-3.928719	-2.851459	-0.916084
2	3.127389	-0.280279	-0.831857	-0.547074	-0.968299
3	0.286269	-0.498856	-0.780908	-1.153351	1.209182
4	-1.141148	1.472406	-0.24608	0.697957	1.469131
High	-3.718468	-1.544479	-1.388212	0.583988	3.030556
t(R_M)					
Low	63.78908	64.48353	59.92734	56.20355	38.35179
2	65.46192	57.77159	49.43912	51.06368	52.29003
3	58.94292	51.87547	49.65928	55.20002	45.34453
4	52.65066	49.99911	53.27912	52.9307	58.27157
High	51.30961	46.11736	51.87314	52.59236	56.30021
t(SMB)					
Low	29.79144	19.13407	13.47869	5.488013	-3.710328
2	28.67288	19.35236	9.843742	4.949082	-6.014878
3	28.4377	20.23945	12.68799	6.750862	-6.643679
4	26.59127	16.0026	12.10219	7.112798	-8.272376
High	27.13338	13.39756	10.98196	4.23578	-10.29634
t(HML)					
Low	-1.785368	-3.429663	-2.819945	-2.320648	2.364884
2	-4.023519	-2.359948	-2.767062	-3.70825	1.038191
3	-1.896263	-1.804752	-3.824006	-2.706291	4.362619
4	-2.789571	-3.069853	-3.628174	-2.198251	6.349519
High	-1.983588	-1.251723	-1.61201	-2.599869	4.808117
t(RMW)					
Low	-11.26442	-14.11853	-12.84865	-11.59834	-8.498229
2	-6.140085	-6.690644	-7.096927	-6.89565	-6.439268
3	0.426377	-1.393799	-5.393924	-2.78333	-0.143634
4	3.097294	-0.887965	0.164711	0.457647	1.719344
High	8.583654	2.857346	2.591996	2.81904	7.024404
t(CMAO)					
Low	3.364208	3.774	1.22278	-1.982165	-1.081245
2	1.660081	0.973437	1.34116	0.261276	1.998069
3	-0.46553	0.789775	1.561331	0.593796	1.299932
4	-0.722525	0.620804	0.029308	0.484384	2.335034
High	-2.330048	1.209855	0.843058	1.146015	0.443383

- The modified five-factor model explains essentially all of the excess returns.
- Portfolio returns follow market factors closely.
- The regression coefficients of SMB are monotonically decreasing from left to right.
- The regression coefficient of the ROE factor increases monotonically from top to bottom.

1.6 Size-investment grouping

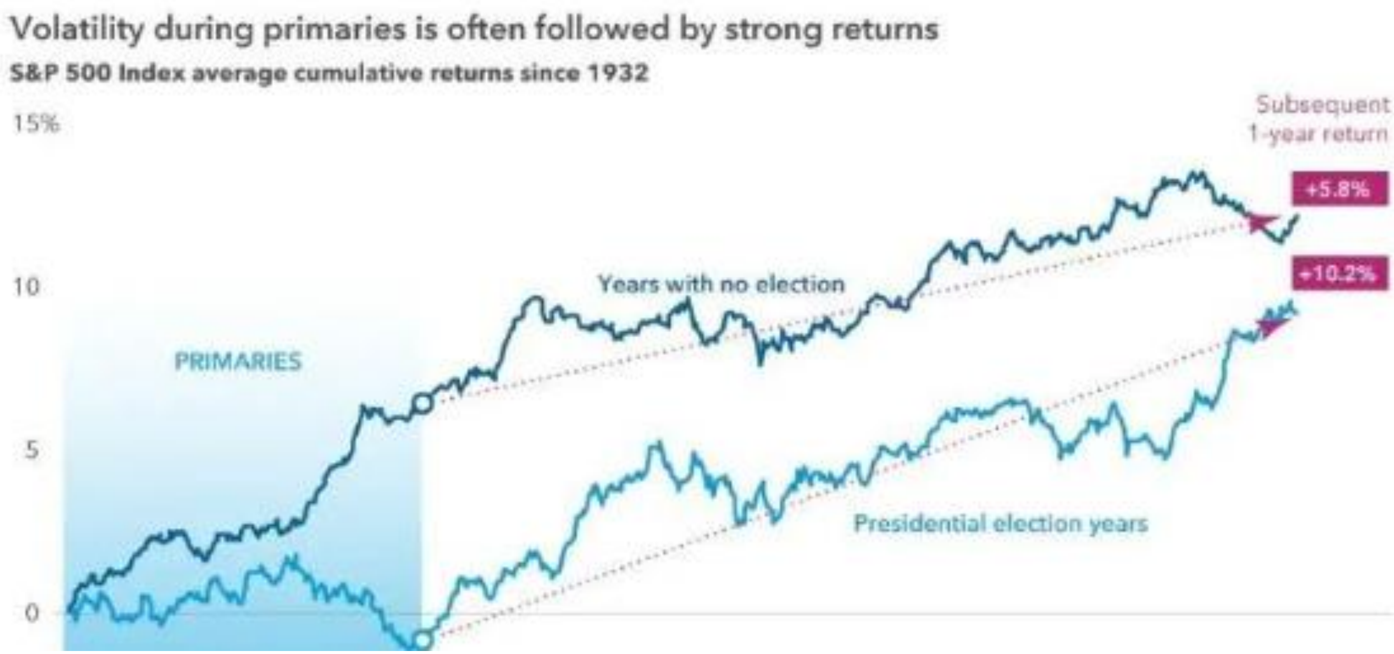
	Small	2	3	4	Big
			Alpha		
Low	0.003201	-0.000187	-0.00417	-0.003255	0.000775
2	0.002801	0.000504	-0.001091	-0.001279	0.000776
3	0.001569	0.000965	-0.001132	-0.000295	0.000626
4	-0.000498	0.000221	-0.002615	0.000555	0.003005
High	-0.00152	-0.001675	-0.001202	-6.28E-05	0.004176
			beta_RM		
Low	1.022334	1.037047	1.008447	1.029118	1.008831
2	1.012633	1.014818	1.013925	1.027349	0.931463
3	1.014568	1.049122	1.03326	1.010543	0.893157
4	1.05392	1.028804	1.042934	1.06333	0.964664
High	1.060974	1.030689	1.045438	1.101945	1.059498
			beta_SMB		
Low	1.376259	0.972958	0.661002	0.372672	-0.370043
2	1.40484	0.985552	0.584701	0.329692	-0.41307
3	1.376433	1.033651	0.653736	0.276669	-0.353695
4	1.497978	0.993344	0.701374	0.312037	-0.371364
High	1.522008	1.00328	0.740639	0.345649	-0.609344
			beta_HML		
Low	-0.103395	-0.099	-0.062195	0.011665	0.260689
2	-0.105146	-0.061818	-0.175971	-0.046396	0.442679
3	-0.131351	-0.099319	-0.187734	-0.160119	0.436838
4	-0.119652	-0.176812	-0.174832	-0.20556	0.19651
High	-0.188327	-0.291385	-0.333741	-0.383396	-0.234714
			beta_RMW		
Low	-0.510887	-0.582366	-0.563598	-0.497408	-0.147036
2	-0.2877	-0.446757	-0.575259	-0.309207	-0.096591
3	-0.207805	-0.277462	-0.302579	-0.157078	0.191643
4	0.055157	-0.221223	-0.221204	-0.112643	0.326416
High	0.258059	0.024156	0.020496	0.00946	0.163161
			beta_CMAO		
Low	0.639193	0.771107	0.546314	0.197422	0.606114
2	0.227324	0.37089	0.421495	0.253034	0.601657
3	0.242955	0.187341	0.218673	0.250715	0.362425
4	-0.173141	-0.009246	-0.11773	-0.061308	0.028783
High	-0.751768	-0.538164	-0.426573	-0.685673	-0.60003

	Small	2	3	4	Big
			t(Alpha)		
Low	2.832403	-0.144677	-2.970493	-2.183056	0.504292
2	2.264537	0.345733	-0.69961	-0.843327	0.508151
3	1.1489	0.665633	-0.7212	-0.203947	0.490899
4	-0.368845	0.155016	-1.785646	0.367514	2.422082
High	-1.039799	-1.089107	-0.713028	-0.038162	3.127571
			t(R_M)		
Low	70.79015	62.76491	56.21277	54.01326	51.38957
2	64.06259	54.45368	50.89477	53.01872	47.71024
3	58.13344	56.61327	51.51568	54.7452	54.82002
4	61.08333	56.38519	55.72761	55.10748	60.84006
High	56.80085	52.45272	48.52665	52.38671	62.09657
			t(SMB)		
Low	32.64829	20.17403	12.62307	6.701041	-6.457859
2	30.44806	18.11754	10.05499	5.829082	-7.248536
3	27.01966	19.10938	11.16638	5.134903	-7.437399
4	29.74408	18.65144	12.83935	5.540241	-8.024059
High	27.91561	17.49213	11.77793	5.629585	-12.23517
			t(HML)		
Low	-2.413337	-2.019725	-1.168622	0.206375	4.476284
2	-2.242248	-1.118128	-2.977465	-0.807111	7.64317
3	-2.536974	-1.806613	-3.155095	-2.923976	9.037968
4	-2.337619	-3.266503	-3.149011	-3.591039	4.177689
High	-3.398605	-4.998574	-5.22192	-6.143947	-4.637083
			t(RMW)		
Low	-9.938137	-9.901805	-8.825742	-7.334115	-2.104169
2	-5.113196	-6.734584	-8.112042	-4.482919	-1.389903
3	-3.345035	-4.206258	-4.238074	-2.390593	3.304489
4	0.898077	-3.406147	-3.320515	-1.64001	5.78342
High	3.881231	0.345362	0.267274	0.126342	2.686478
			t(CMAO)		
Low	5.874485	6.194288	4.041867	1.375271	4.097969
2	1.908783	2.641448	2.80813	1.7332	4.090279
3	1.84769	1.341787	1.447049	1.802728	2.952484
4	-1.331905	-0.067258	-0.834944	-0.421713	0.240942
High	-5.341851	-3.635072	-2.62805	-4.3265	-4.667652

- The regression coefficient of the CMAO factor decreases monotonically from top to bottom.
- The regression coefficients of the RMW factor from top to bottom are also quite regular - monotonically increasing.

2.1 Introduction

In order to examine the January Effect and the Monday Effect in the Chinese Shanghai and Shenzhen A-share market by analyzing stock price data from January 1, 2000, to January 1, 2024. These effects refer to the phenomena where stock market returns are abnormally high in January and abnormally low on Mondays, respectively.



2.2 Data and Methodology

- **Data Source**

  Stock prices of listed companies in Shanghai and Shenzhen A-share market.

- **Time Range**

  January 1, 2000, to January 1, 2024.

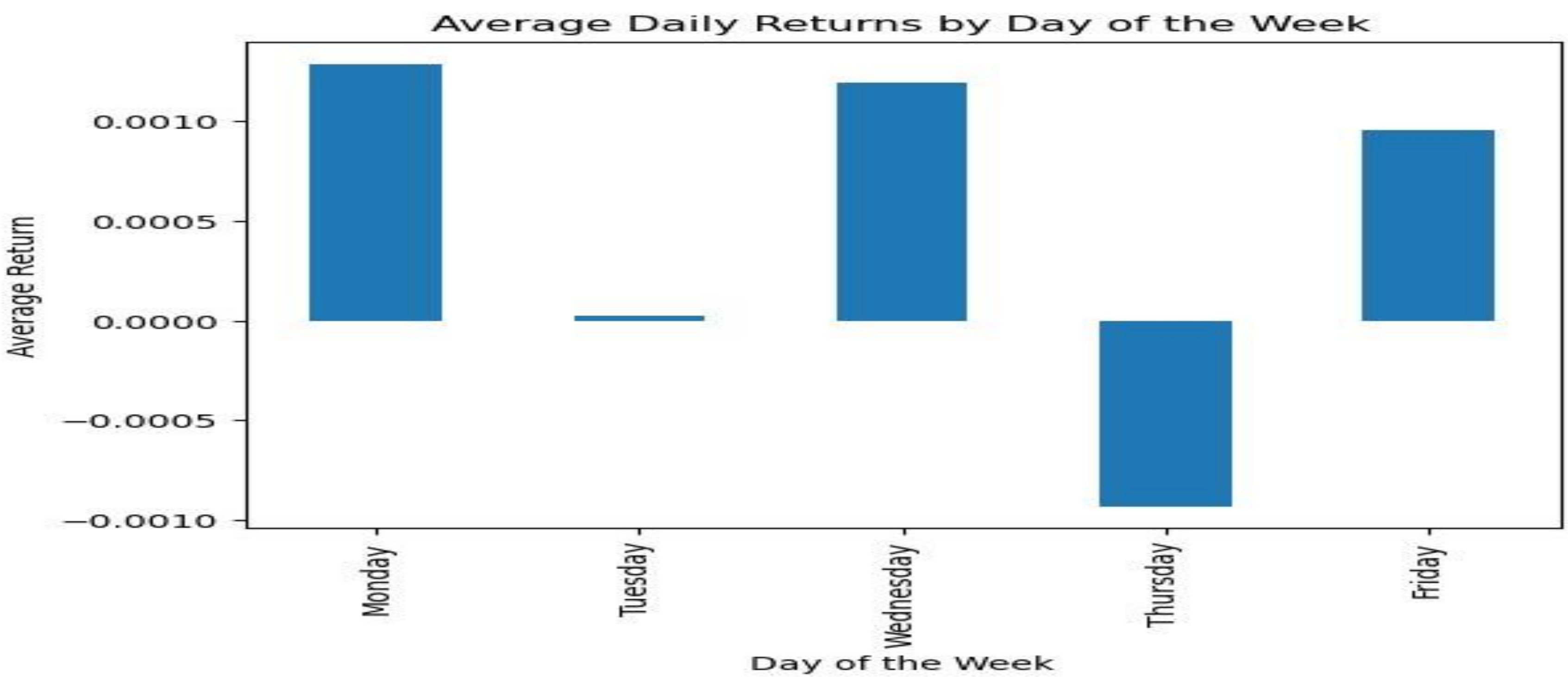
- **Analysis Method**

  Python was used to perform statistical analysis, including T- tests and ANOVA (Analysis of Variance) to test for the January Effect and the Monday Effect.

2.3 Statistical Graphical Analysis

- **Average Daily Returns by Day of the Week**

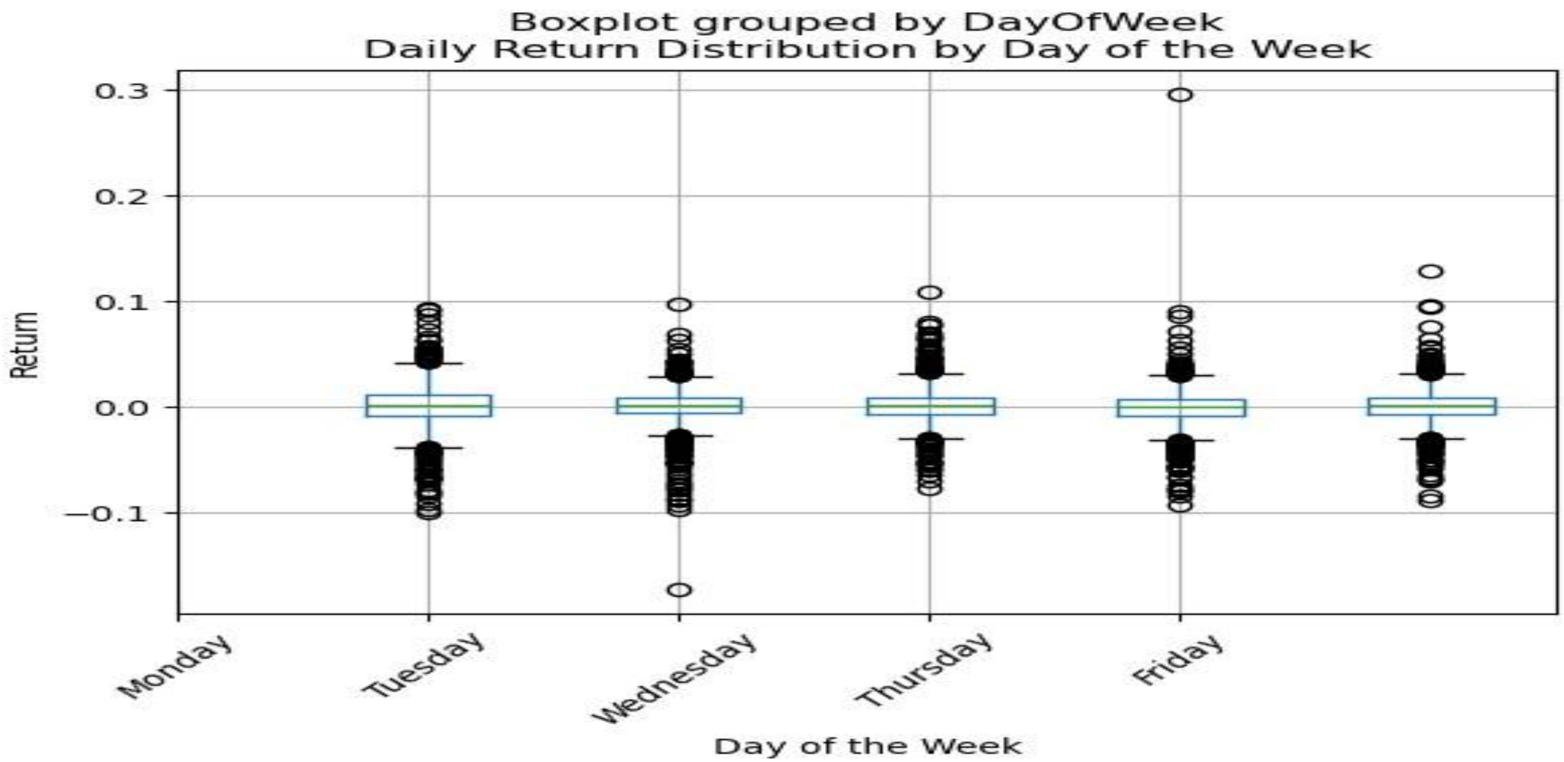
The graph shows that Thursday has the highest return rate at 0.0010. Friday has the second highest return rate at 0.0005. Monday has the lowest return rate at -0.0010.



2.3 Statistical Graphical Analysis

- **Boxplot Grouped by Day of the Week**

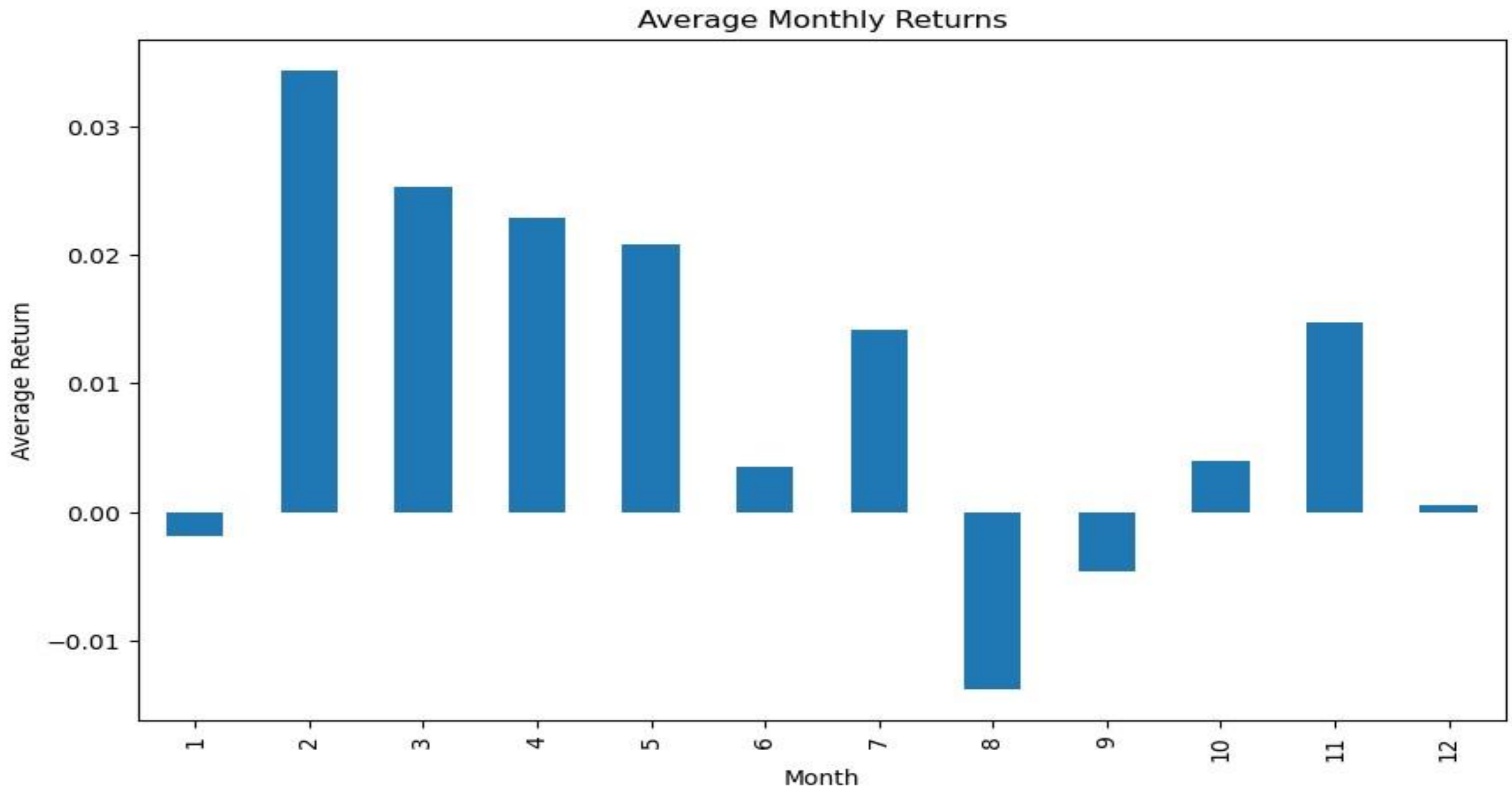
The boxplot for Monday indicates a lower median and larger dispersion, suggesting greater volatility in returns on Mondays.



2.3 Statistical Graphical Analysis

- **2.3.3 Average Monthly Returns**

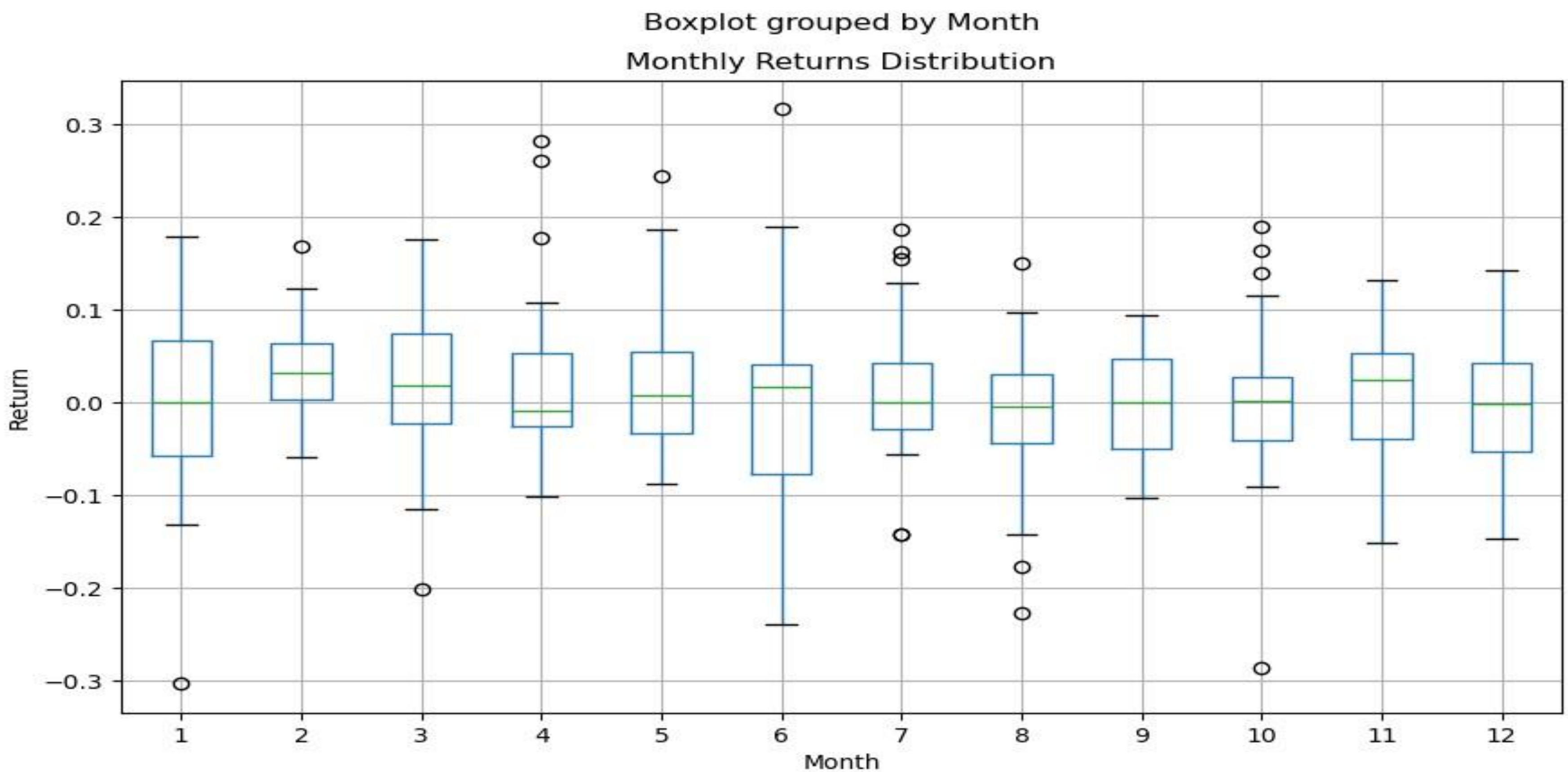
January has the highest return rate at 0.03.



2.3 Statistical Graphical Analysis

- **Boxplot Grouped by Month**

The boxplot for January has a higher median and smaller lower quartile.



2.4 Statistical Test Results

- **Monday Effect Test**

✓ The T-test statistic is 1.6632 with a p-value of 0.0964, which does not pass the significance level of 0.05.

✓ The ANOVA F-statistic is 4.0759 with a p-value of 0.0027, which passes the significance level of 0.05.

- **January Effect Test**

✓ The T-test statistic is -0.8371 with a p-value of 0.4031, which does not pass the significance level of 0.05.

✓ The ANOVA F-statistic is 0.8887 with a p-value of 0.5517, which does not pass the significance level of 0.05.

2.5 Conclusion and Discussion

Based on the visual analysis of the statistical graphics and the results of the statistical tests, the following conclusions can be drawn:

- **Monday Effect**

✓ Although the graphical analysis suggests that Monday's returns are lower, the p-value from the T-test is greater than 0.05, indicating insufficient statistical evidence to support the existence of the Monday Effect.

✓ The p-value from the ANOVA is less than 0.05, which may indicate some difference in returns among different days of the week. However, since the T-test did not pass the significance test, we cannot conclude that the Monday Effect is statistically significant.

- **January Effect**

✓ The graphical analysis shows that January has the highest return rate, but both the T-test and ANOVA p-values are greater than 0.05, indicating insufficient statistical evidence to support the existence of the January Effect.

2.6 Recommendations

- ✓ The calendar effects in the stock market may be influenced by a variety of factors, including market sentiment and economic policies. Therefore, investors should not rely solely on calendar effects when analyzing the stock market but should consider a comprehensive range of information.
- ✓ Further research can be conducted by expanding the sample range, considering other potential market factors, or using more complex statistical models.

3.1 Co-skewness Factor of Stock Return

From CAPM to Co-skewness Factor

CAPM

- Large error in pricing prediction
- Due to strong assumptions
- $E(r_i) - r_f = \beta_i \cdot E(r_M - r_f) + \epsilon_i$

Higher-order Moment CAPM Model

- Harvey, Campbell & Siddique, Akhtar. (2000).
- Kraus, Alan & Litzenberger, Robert. (1976).
- $m_{t+1} = a_t + b_t R_{M,t+1} + c_t R_{M,t+1}^2$

Conditional Co-skewness Factor

- Harvey, Campbell & Siddique, Akhtar. (2000).
- $E_t[r_{i,t+1}] = \gamma_{1,t} Cov_t[r_{i,t+1}, r_{M,t+1}] + \gamma_{2,t} Cov_t[r_{i,t+1}, r_{M,t+1}^2]$

Co-skewness factor $CSK_{XYY} = \frac{E[(X - E[X])(Y - E(Y)^2]}{\sigma_X \sigma_Y^2}$

X: Individual stock return sequence

Y: Market return sequence

Factor	Factor calculation
CSK_XYY_20D	Using the past 20 trading days as the resurgence window, calculate the sensitivity of individual stock returns to extreme market fluctuations.
CSK_XYY_60D	Using the past 60 trading days as the resurgence window, calculate the sensitivity of individual stock returns to extreme market fluctuations.
CSK_XYY_120D	Using the past 120 trading days as the resurgence window, calculate the sensitivity of individual stock returns to extreme market fluctuations.
CSK_XYY_240D	Using the past 240 trading days as the resurgence window, calculate the sensitivity of individual stock returns to extreme market fluctuations.

Kraus, Alan & Litzenberger, Robert. (1976). Skewness Preference and the Valuation of Risk Assets. Journal of Finance. 31. 1085-1100.
Harvey, Campbell & Siddique, Akhtar. (2000). Conditonal Skewness in Asset Pricing Tests. The Journal of Finance. 55.

3.1 Co-skewness Factor of Stock Return

In extremely falling markets and extremely rising markets, investors' risk preferences and tolerances are different, and the co-skewness factor may behave differently.

The **upward co-skewness factor** is defined as

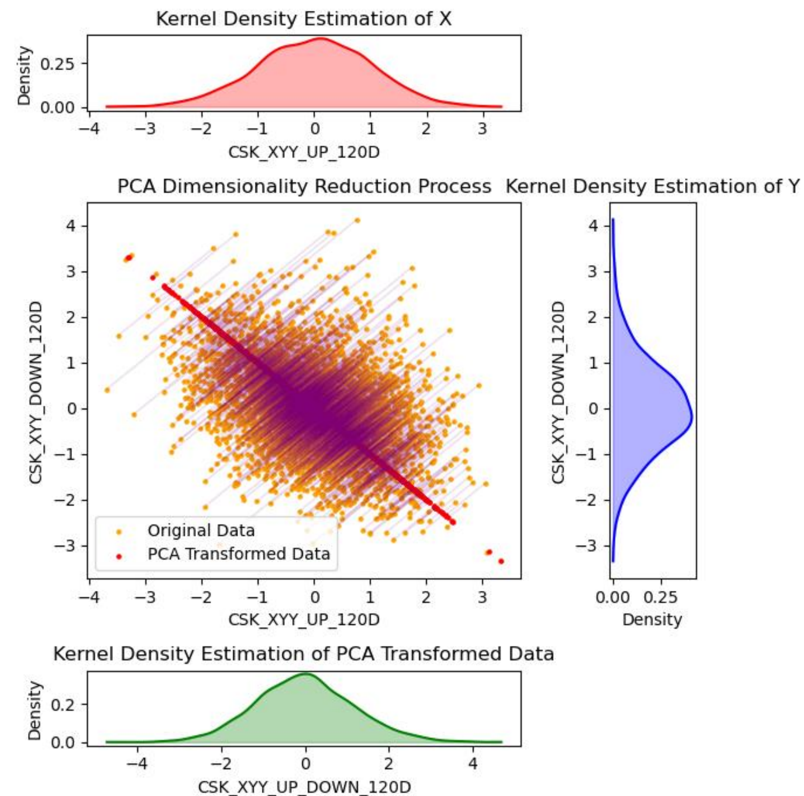
$$CSK\_UP_{XYY} = \frac{E[(X - E[X])(Y - E[Y])^2)|Y > E[Y]]}{\sqrt{E[(X - E[Y])^2|Y > E[Y]]E[(Y - E[Y])^2|Y > E[Y]]}}$$

The **downside co-skewness factor** is defined as

$$CSK\_DOWN_{XYY} = \frac{E[(X - E[X])(Y - E[Y])^2)|Y < E[Y]]}{\sqrt{E[(X - E[Y])^2|Y < E[Y]]E[(Y - E[Y])^2|Y < E[Y]]}}$$

- We calculated the synthesis factor CSK_UP_DOWN_XYY_{X}D of the up-down factor through **PCA dimensionality** reduction synthesis.

The following figure is synthesized based on 120 trading days as an example:



3.1 Co-skewness Factor of Stock Return

Definition of stock return Upside and Downside co-skewness factors

Factor	Factor calculation
CSK_XYY_UP_20D	Taking the past 20 trading days as the rolling window, calculate the upward co-skewness of individual stock returns.
CSK_XYY_UP_60D	Taking the past 60 trading days as the rolling window, calculate the upward co-skewness of individual stock returns.
CSK_XYY_UP-120D	Taking the past 120 trading days as the rolling window, calculate the upward co-skewness of individual stock returns.
CSK_XYY_UP-240D	Taking the past 240 trading days as the rolling window, calculate the upward co-skewness of individual stock returns.
CSK_XYY_DOWN_20D	Using the past 20 trading days as the rolling window, calculate the downward co-skewness of individual stock returns.
CSK_XYY_DOWN_60D	Using the past 60 trading days as the rolling window, calculate the downward co-skewness of individual stock returns.
CSK_XYY_DOWN_120D	Using the past 120 trading days as the rolling window, calculate the downward co-skewness of individual stock returns.
CSK_XYY_DOWN_240D	Using the past 240 trading days as the rolling window, calculate the downward co-skewness of individual stock returns.
CSK_XYY_UP_DOWN_20D	Obtained by PCA dimensionality reduction of CSK_XYY_UP-20D and CSK_XYY_DOWN_20D.
CSK_XYY_UP_DOWN_60D	Obtained by PCA dimensionality reduction of CSK_XYY_UP-60D and CSK_XYY_DOWN_60D.
CSK_XYY_UP_DOWN_120D	Obtained by PCA dimensionality reduction of CSK_XYY_UP-120D and CSK_XYY_DOWN_120D.
CSK_XYY_UP_DOWN_240D	Obtained by PCA dimensionality reduction of CSK_XYY_UP-2 40D and CSK_XYY_DOWN_240D.

3.2 Factor Data Analysis

1. Factor distribution plot

✓ **MAD method** to remove the extreme values of the factors.

The specific steps are as follows and we set n=5.

$$\tilde{x}_l = \begin{cases} x_M + n * D_{MAD}, & if \quad x_i > x_M + n * D_{MAD} \\ x_M - n * D_{MAD}, & if \quad x_i < x_M - n * D_{MAD} \\ x_i, & else \end{cases}$$

x_M: The median of the sequence x_i

D_{MAD}: The median of the sequence $|x_i - x_M|$

$\tilde{x}_l$: The value of x_i after de-extreme correction

The sequence after data de-extremed is standardized.

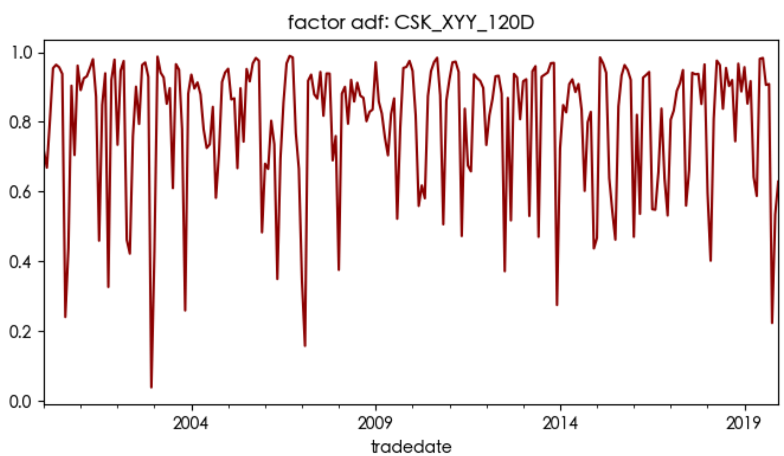
$$\tilde{x}_l = \frac{x_i - u}{\sigma}$$

u: The mean of the sequence x_i

σ: The standard deviation of the sequence x_i

$\tilde{x}_l$: The value of the sequence x_i after standardization

2. Factor Autocorrelation

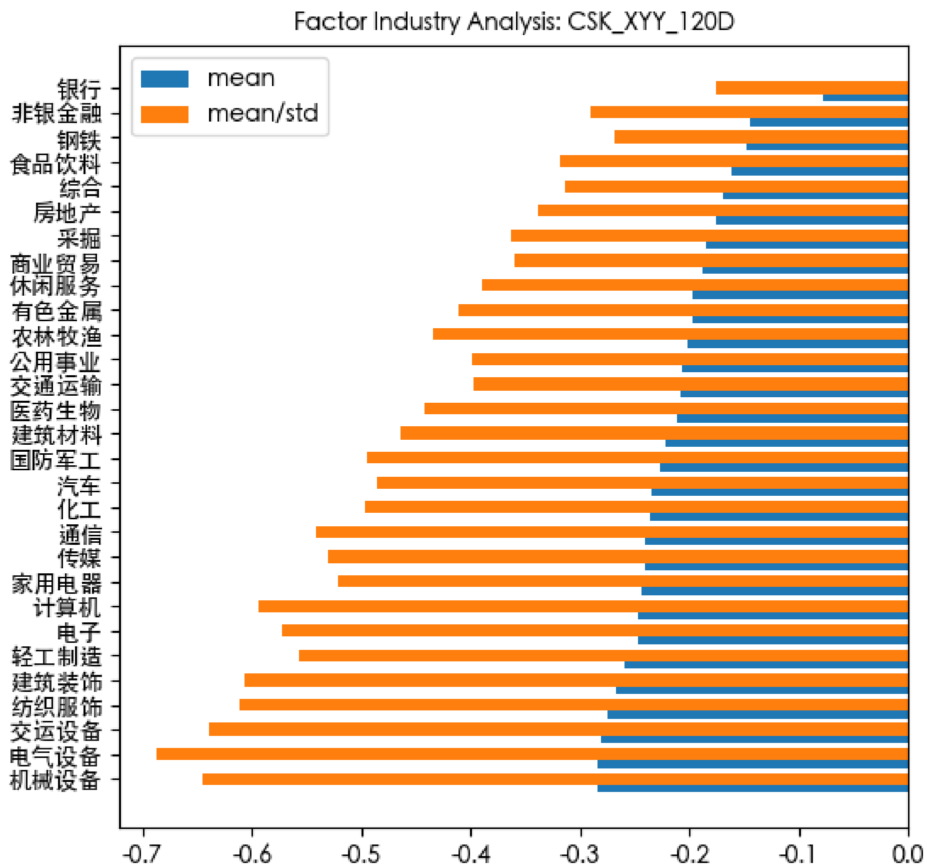


- Due to the calculation method of factors, the autocorrelation of high-day factors is naturally very large.

3.2 Factor Data Analysis

3. Industry factor mean analysis

- There are indeed significant differences in valuation factors across industries. Therefore, we will neutralize the factors later.



3.3 Factor cross-sectional analysis

1. Factor Industry/Market Cap Neutralization

- We neutralize factors with the purpose of **eliminating or weakening the correlation** between factors, market capitalization, and industries to ensure that the source of factor returns is mainly related to the factor being studied rather than being interfered by other factors.

✓ The factor-neutral model is as follows:

$$\boldsymbol{F_i = \alpha_i + \beta_{industry} \cdot I_{industry} + \beta_{size} \cdot S_{size}}$$

F_i: Factor value of stock i

α_i: Intercept term, represents the specific factor value of stock i

$I_{industry}$: A dummy variable indicating the industry to which stock i belongs

S_{size}: Market value factor of stock I (we use log Market value)

$\beta_{industry}$: Coefficient of industry factor

β_{size}: Coefficient of market value factor

✓ The formula for calculating the industry and market value neutral factor value is

$$\boldsymbol{F_{neutralized} = F_i - \beta_{industry} \cdot I_{industry} - \beta_{size} \cdot S_{size}}$$

3.3 Factor cross-sectional analysis

2. Single Factor Test - Regression Method

Regression model $$\boldsymbol{r_i^{T+1} = \sum_j X_j^T f_{ji}^T + X_d^T d_i^T + \mu_i^T}$$

r_i^{T+1}: The rate of return of stock i in period T+1
d_i^T: Stock i's exposure to factor d in period T
f_{ji}^T: Exposure of stock i to the jth industry factor in period T (1 if it belongs to the industry, 0 otherwise)
X_j^T (or X_d^T): The factor return rate of the jth industry factor (or factor d) in the Tth period, requires four regression fitting
μ_i^T: The residual return of stock i in period T

Hypothetical test $\boldsymbol{H_0: X_d^T = 0}$
Alternative hypothesis $\boldsymbol{H_1: X_d^T \neq 0}$

Stock Pool	Excluding ST stocks, all A-shares in the Shanghai and Shenzhen stock markets.
Time Period	From December 31, 1999 to December 31, 2019
Cross-sectional period	Factor exposure is calculated on the last trading day of each calendar month and regressed with the returns of individual stocks for the next full calendar month.
Regression weights	We refer to the **Barra** manual and adopt weighted least squares regression (**WLS**), using the square root of the individual stock's liquid market capitalization as the weight.

3.3 Factor cross-sectional analysis

Table: Summary of Valuation Factor Regression Test Results (1999.12.31-2019.12.31)

Factor	|t| mean	Proportion of |t|>2	|t| mean/|t| std	Return mean	Return t-value
CSK_XYY_20D	2.59	51.25%	-0.045	-0.049%	-0.864
CSK_XYY_60D	2.984	60%	-0.062	-0.084%	-1.311
CSK_XYY_120D	3.395	60.42%	-0.037	-0.065%	-0.928
CSK_XYY_240D	3.551	67.5%	-0.048	-0.095%	-1.33
CSK_XYY_UP_20D	3.0	55.42%	0.425	0.383%	6.637
CSK_XYY_UP_60D	3.146	57.08%	0.314	0.302%	5.537
CSK_XYY_UP_120D	3.45	62.92%	0.216	0.231%	3.767
CSK_XYY_UP_240D	3.435	63.75%	0.174	0.187%	2.918
CSK_XYY_DOWN_20D	2.984	55.42%	-0.314	-0.288%	-4.254
CSK_XYY_DOWN_60D	3.181	60.42%	-0.251	-0.244%	-4.017
CSK_XYY_DOWN_120D	3.366	62.08%	-0.161	-0.179%	-2.829
CSK_XYY_DOWN_240D	3.372	63.33%	-0.156	-0.184%	-3.023
CSK_XYY_UP_DOWN_20D	3.244	56.67%	0.16	0.166%	2.329
CSK_XYY_UP_DOWN_60D	3.323	61.25%	0.095	0.145%	2.32
CSK_XYY_UP_DOWN_120D	3.547	65.42%	0.207	0.237%	3.972
CSK_XYY_UP_DOWN_240D	3.527	67.5%	0.103	0.094%	1.36

- **CSK_XYY_UP_DOWN_120D** factor has the best effect in statistical terms.

✓ The average value of $|t|$ obtained by regression method is 3.547, which is the second largest among all factors.

✓ If we ignore the goodness of fit and only look at the size of the factor yield, the absolute value of the average monthly yield of CSK_XYY_UP_DOWN_120D exceeds 0.2%,and the factor can pass the hypothesis test.

3.3 Factor cross-sectional analysis

3. Factor Test - IC Test

We conducts IC test on the co-skewness factor of stock returns to test the stock selection ability of this factor

IC test results (1999.12.31-2019.12.31)

Factor	Rank IC	ICIR	Half IC value
CSK_XYY_20D	-0.0067	-0.3376	2
CSK_XYY_60D	-0.0116	-0.4902	1
CSK_XYY_120D	-0.0084	-0.3406	4
CSK_XYY_240D	-0.0055	-0.2303	11
CSK_XYY_UP_20D	0.056	2.5548	2
CSK_XYY_UP_60D	0.0479	1.9999	5
CSK_XYY_UP_120D	0.0415	1.6521	12
CSK_XYY_UP_240D	0.0361	1.4189	12
CSK_XYY_DOWN_20D	-0.049	-2.0879	1
CSK_XYY_DOWN_60D	-0.0431	-1.7634	6
CSK_XYY_DOWN_120D	-0.0354	-1.4066	12
CSK_XYY_DOWN_240D	-0.0329	-1.3605	12
CSK_XYY_UP_DOWN_20D	0.0312	1.0539	3
CSK_XYY_UP_DOWN_60D	0.0166	0.5626	12
CSK_XYY_UP_DOWN_120D	0.0251	0.8839	3
CSK_XYY_UP_DOWN_240D	0.0133	0.4845	1

$$IC_d^T = Corr\left(R^{T+1}, d^T\right) = \frac{Cov\left(R^{T+1}, d^T\right)}{\sqrt{Var\left(R^{T+1}\right) \cdot Var\left(d^T\right)}}$$

IC_d^T: IC value of factor d in the T period
R^{T+1}: Return vector of individual stock in T+1 period
d^T: Exposure vector in the T-period on factor d

- The co-skewness factor does have a **negative premium effect** in the A-share market.
- The co-skewness factor **CSK_XYY_120D** has a good stock selection effect.
- The stock selection effects of the upward co-skewness factor CSK_XYY_UP_120D and the downward co-skewness factor CSK_XYY_DOWN_120D are improved compared with CSK_XYY_120D.

3.3 Factor cross-sectional analysis

Out of sample (2020.1.1-now)

- **T-test**

Factor	|t| mean	Proportion of |t|>2	|t| mean/|t| std	Return mean	Return t-value
CSK_XYY_120D	5.682	87.04%	-0.243	-0.315%	-2.497
CSK_XYY_UP_120D	4.2	61.11%	0.022	0.015%	0.157
CSK_XYY_DOWN_120D	3.838	66.67%	-0.323	-0.313%	-2.756
CSK_XYY_UP_DOWN_120D	3.609	61.11%	0.101	0.133%	0.969

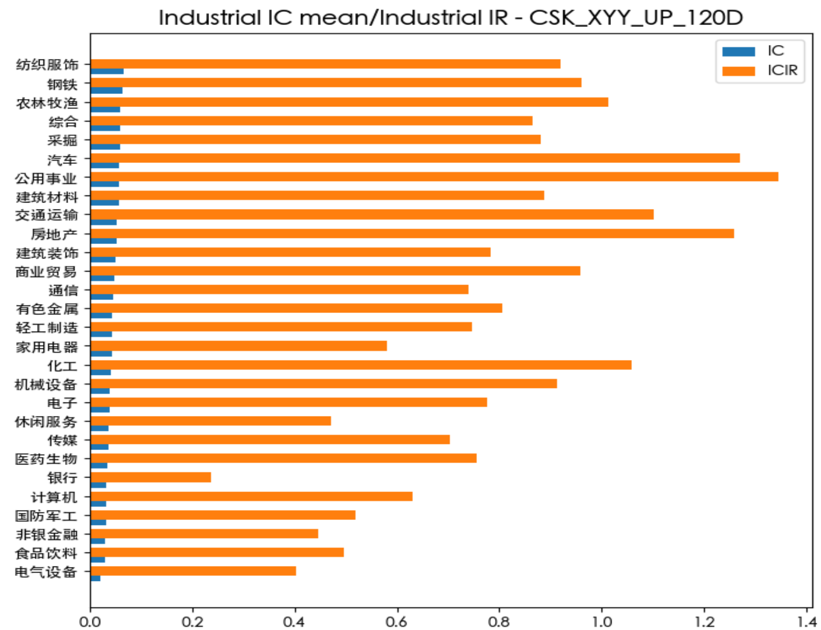
- **IC test**

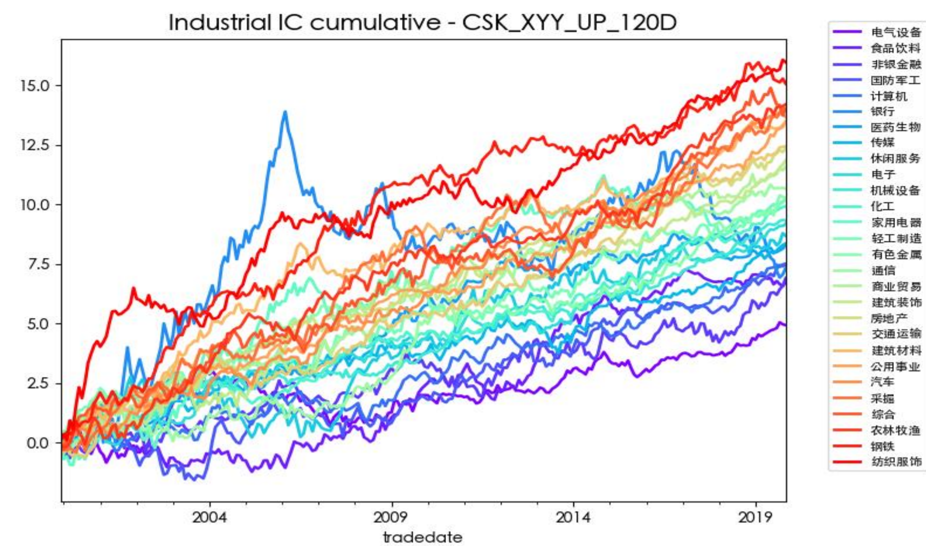
Factor	Rank IC	ICIR	Half IC value
CSK_XYY_120D	-0.0251	-1.0301	12
CSK_XYY_UP_120D	0.0405	1.7747	3
CSK_XYY_DOWN_120D	-0.0516	-3.3521	3
CSK_XYY_UP_DOWN_120D	0.0265	1.1502	3

- **CSK_XYY_120D** factor has the best effect in a statistical sense. The mean value of $|t|$ obtained by the regression method is 5.682, which is the largest among these factors. In addition, the proportion of $|t| > 2$ in the t-value exceeds 80%, and the absolute value of the average monthly return exceeds 0.2%.
- **CSK_XYY_DOWN_120D** factor has a better prediction effect than the other three factors.

3.4 IC Industry Analysis

- This paper also conducts IC analysis based on industries.
- As is shown in the following graph, most of the industries exhibit similar IC level in a reasonable range, except for banking industry.
- This is because most stocks in banking industry show stable overall performance, with small fluctuations in returns. As a result, its correlation with the market performance could be rather small.
- Especially when the market goes up (CSK_UP), most of the investors would not change the proportion of banking stocks holding since it generates relative stable performance.

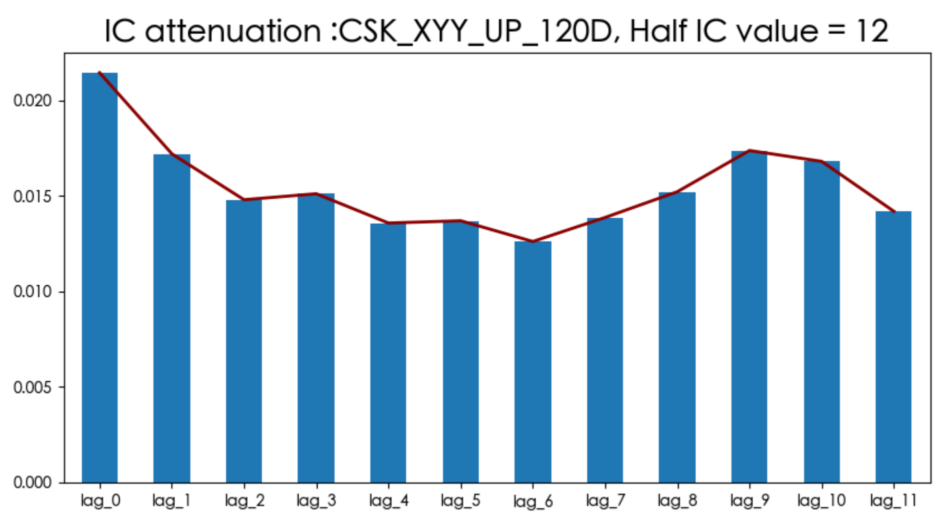


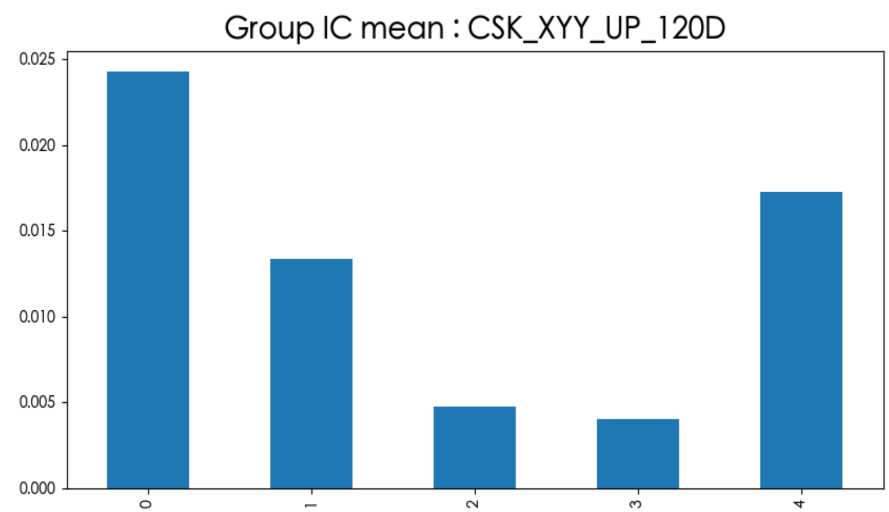


3.5 IC Attenuation and IC Grouping

- The time decay of factor IC is the most commonly mentioned concept of factor decay, **used to measure how long a factor's predictive ability for the future can last.**
- This part extends the calculation formula of IC by changing the time window between factor and stock return from 1 period to 1-11periods.
- Demonstrated by the graph, it is shown that IC decreases when the time window between factor and stock return goes broadening, meaning that **there exists ideal timing requirement for co-skewness.**

- The following graph shows the mean of CSK_XYY_UP_120D in **5 IC groups divided by IC value size.**





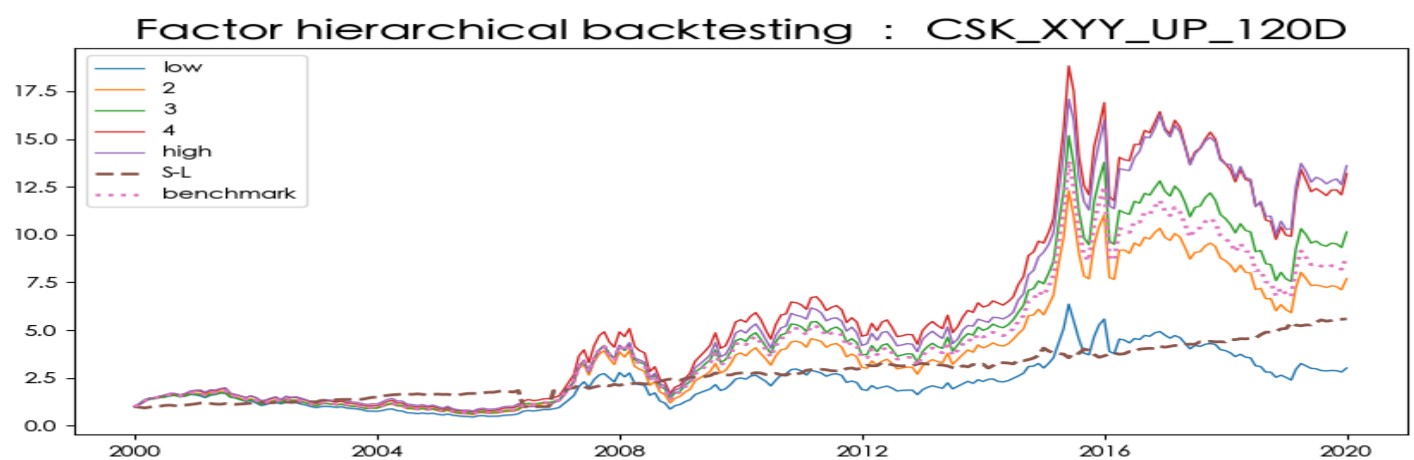
3.6 Constructing Group Factor Portfolio

- In general, the method of **constructing group factor portfolio is to sort factors from low to high to construct an investment portfolio**. The basic information of portfolio used in this section is shown as below.
- -Sample space: Excluding ST stocks, all A-shares in the Shanghai and Shenzhen stock markets
- -Neutralization or not: Neutralization of industry market value (Shenwan level 1 industry)
- -**Benchmark: Equal weighted return of all stocks in sample**
- -IC category: Spearman related sparse
- -Adjustment frequency: Adjust positions at the end of each month based on the closing price
- -**Number of layers: 5**
- -Transaction fees: None
- -Weighting method: equal weight

- **Using Equal weighted return of all stocks in sample as benchmark portfolio, the five groups of return rate based on co-skewness sorting are shown as below.**

3.6 Constructing Group Factor Portfolio

- Sample stocks are divided into 5 groups based on CSK_XYY_UP_120D, from highest to lowest denoted as 5 (high), 4, 3, 2, 1 (low).
- Then **long-short combination portfolio** is constructed by shorting the relative low return group and at the same time longing the relative high return group between 5 (high) and 1 (low), **denoted as Small-Minus-Large (S-L) or Large-Minus-Small (L-S).**
- The purpose of constructing this long-short combination portfolio is to **discover the potential trend of alpha return rate generated by CSK_XYY_UP_120D.** The return rate generated by long-short combination portfolio experiences a growth from 2000 to 2020, **which means the alpha return caused by co-skewness is becoming more significant over time.**
- However, the alpha return rate is relatively low. **Return rate related to market, or to say beta return, cannot be ignored.**



3.6 Constructing Group Factor Portfolio

- The primary characteristics of this constructed portfolio is shown as below.
- In addition to annual return rate, beta value, alpha value, Sharpe ratio and so on, this paper adds the **turnover rate of long portfolio on both sides** in constructed S-L or L-S.
- Since **turnover rate can be used to measure transaction cost**, adding it into consideration can include the impact of transaction cost in this constructed portfolio and be closer to reality.

Characteristics using different market index: 1999.12-2019.12

Factor	Market index	Strategy annual return rate	Strategy beta	Strategy alpha	Strategy volatility	Strategy Sharpe ratio	Strategy maximum drawdown	Strategy excess annual return rate	Strategy excess annual volatility	Strategy excess maximum drawdown	Strategy excess Sharpe Ratio	Long turnover rate
CSK_XYY_120D	CS1000	8.38%	0.02	0.06	7.06%	0.90	-4.52%	0.09%	1.04%	-7.19%	-1.83	8.43
	HS300	-0.05%	0.04	-0.02	8.34%	-0.25	-17.24%	-8.02%	0.72%	-70.68%	-13.99	8.22
	CS500	3.59%	0.05	0.01	7.98%	0.20	-12.60%	-5.49%	0.61%	-51.75%	-12.35	8.53
	GZ2000	10.52%	-0.01	0.09	8.24%	1.03	-7.72%	-4.12%	0.64%	-21.51%	-9.51	8.37
CSK_XYY_UP_120D	CS1000	7.16%	-0.15	0.05	9.21%	0.56	-18.95%	-1.98%	0.95%	-13.07%	-4.19	8.06
	HS300	4.20%	0.09	0.01	9.84%	0.22	-15.99%	-6.79%	0.65%	-64.56%	-13.59	7.85
	CS500	5.83%	-0.01	0.04	9.84%	0.39	-12.32%	-5.08%	0.62%	-49.04%	-11.51	7.61
	GZ2000	7.39%	-0.14	0.07	9.46%	0.57	-19.24%	-5.10%	0.64%	-26.00%	-11.05	7.84
CSK_XYY_DOWN_120D	CS1000	10.41%	-0.07	0.08	7.36%	1.14	-4.51%	-0.04%	1.13%	-8.31%	-1.80	7.13
	HS300	3.58%	0.06	0.01	10.50%	0.15	-12.09%	-6.71%	0.68%	-64.16%	-12.75	6.51
	CS500	6.59%	0.02	0.04	8.44%	0.54	-13.67%	-4.20%	0.62%	-42.55%	-10.07	7.04
	GZ2000	12.35%	-0.09	0.11	8.65%	1.20	-9.87%	-3.87%	0.71%	-20.34%	-8.28	6.96
CSK_XYY_UP_DOWN_120D	CS1000	3.59%	-0.14	0.02	10.42%	0.15	-13.39%	-2.22%	0.67%	-11.68%	-6.26	13.39
	HS300	5.57%	0.03	0.03	8.69%	0.41	-19.39%	-6.39%	0.61%	-62.07%	-13.77	14.14
	CS500	8.60%	-0.03	0.07	9.56%	0.69	-11.05%	-2.96%	0.67%	-32.20%	-7.45	14.23
	GZ2000	6.82%	-0.14	0.07	10.64%	0.45	-15.21%	-5.20%	0.57%	-26.43%	-12.63	13.8

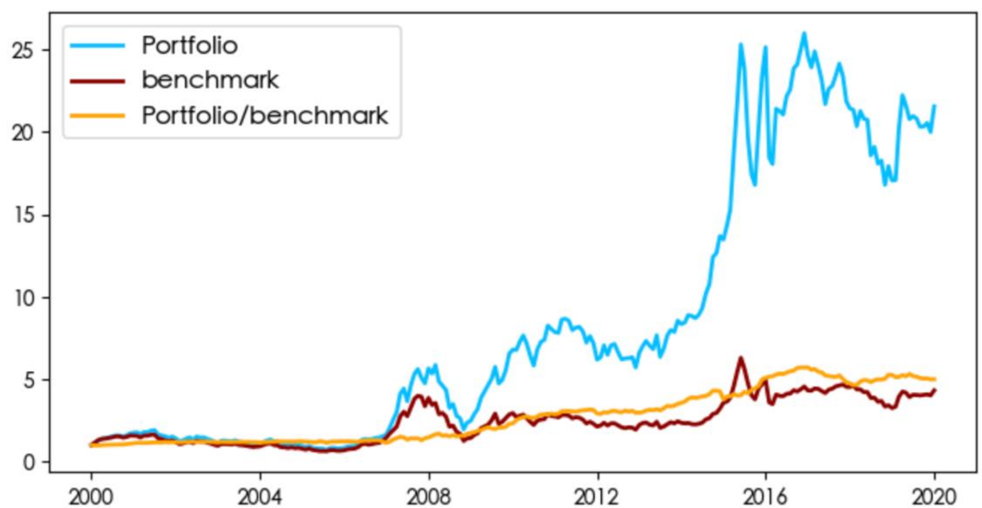
Co-skewness factor stratification test results: 1999.12-2019.12

	Strategy annual return rate	Strategy beta	Strategy alpha	Strategy volatility	Strategy Sharpe ratio	Strategy maximum drawdown	Strategy excess annual return rate	Strategy excess annual volatility	Strategy excess maximum drawdown
CSK_XYY_20D	2.41%	0.05	0.00	7.18%	0.06	-14.83%	-3.83%	1.06%	-54.43%
CSK_XYY_60D	3.07%	0.05	0.01	9.02%	0.12	-25.21%	-4.12%	0.90%	-56.87%
CSK_XYY_120D	3.21%	0.07	0.01	9.75%	0.12	-29.81%	-4.59%	0.87%	-60.95%
CSK_XYY_240D	1.59%	0.07	-0.01	12.93%	-0.03	-30.72%	-5.38%	0.97%	-66.94%
CSK_XYY_UP_20D	14.40%	0.03	0.12	18.95%	0.65	-40.13%	9.87%	2.88%	-6.48%
CSK_XYY_UP_60D	10.53%	0.03	0.08	18.97%	0.45	-41.77%	3.14%	1.51%	-6.13%
CSK_XYY_UP_120D	8.99%	0.02	0.07	22.93%	0.30	-45.72%	2.74%	1.56%	-5.36%
CSK_XYY_UP_240D	7.35%	0.01	0.05	18.18%	0.29	-38.22%	0.33%	1.38%	-16.04%
CSK_XYY_DOWN_20D	13.39%	0.06	0.11	18.25%	0.62	-38.84%	8.04%	2.41%	-7.32%
CSK_XYY_DOWN_60D	10.86%	0.04	0.09	14.47%	0.61	-30.92%	3.42%	1.48%	-8.22%
CSK_XYY_DOWN_120D	8.62%	0.04	0.06	14.10%	0.47	-28.74%	0.56%	1.11%	-8.78%
CSK_XYY_DOWN_240D	8.72%	0.04	0.07	17.49%	0.38	-36.45%	1.44%	1.22%	-6.13%
CSK_XYY_UP_DOWN_20D	6.71%	0.01	0.05	10.53%	0.45	-23.89%	-1.47%	1.16%	-32.91%
CSK_XYY_UP_DOWN_60D	3.59%	0.00	0.02	12.25%	0.13	-27.71%	-1.76%	1.40%	-41.56%
CSK_XYY_UP_DOWN_120D	6.02%	-0.03	0.04	10.81%	0.37	-23.11%	-1.36%	1.19%	-33.52%
CSK_XYY_UP_DOWN_240D	3.14%	0.00	0.01	10.22%	0.11	-22.93%	-3.12%	1.27%	-51.34%

3.7 Constructing Top N Portfolio

- Due to the alpha return rate generated by co-skewness exists but is relatively small during some time, the impact of beta cannot be ignored.
- **This section selects the top N largest (or smallest) stocks with factor values** for each period to construct an investment portfolio, based on four market index including CS1000, HS300, CS500 and GZ2000. Also based on the analysis above, four representative factors including CSK_XYY_120D, CSK_XYY_UP_120D, CSK_XYY_DOWN_120D, CSK_XYY_UP_DOWN_120D are chosen for constructing Top N portfolio.

Top N Portfolio: CSK_XYY_DOWN_120D

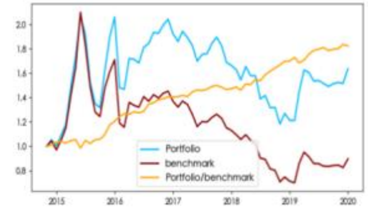


3.7 Constructing Top N Portfolio

- As is shown in the graph, the performance of Top N Portfolio exceeds that of benchmark portfolio, **which implies that co-skewness could be a high-quality factor for return rate.**

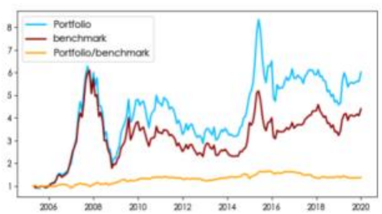
CSK_XYY_120D: CS1000

CSK_XYY_UP_120D: CS1000

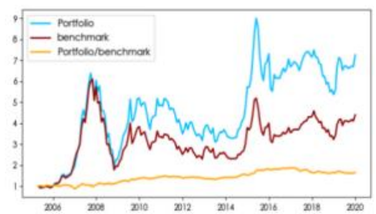


CSK_XYY_DOWN_120D: CS1000

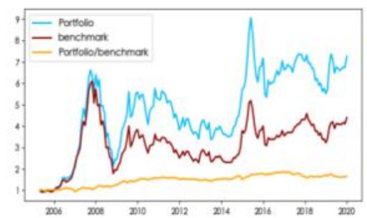
CSK_XYY_UP_DOWN_120D: CS1000



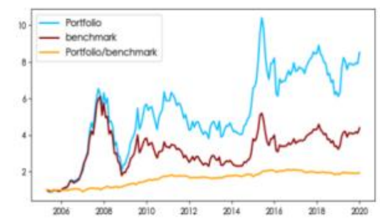
CSK_XYY_120D: HS300



CSK_XYY_UP_120D: HS300



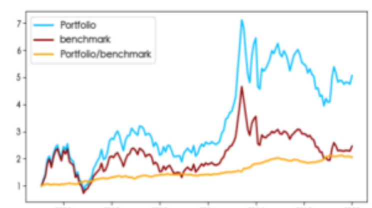
CSK_XYY_DOWN_120D: HS300



CSK_XYY_UP_DOWN_120D: HS300

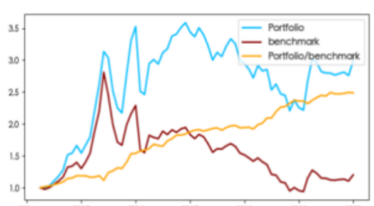
CSK_XYY_120D: CS500

CSK_XYY_UP_120D: CS500

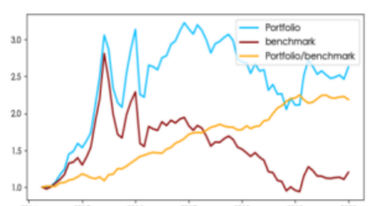


CSK_XYY_DOWN_120D: CS500

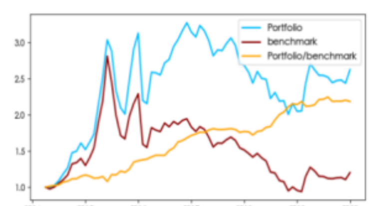
CSK_XYY_UP_DOWN_120D: CS500



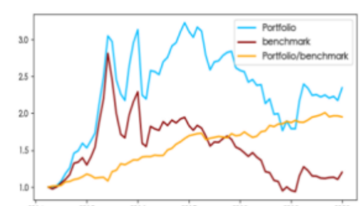
CSK_XYY_120D: GZ2000



CSK_XYY_UP_120D: GZ2000



CSK_XYY_DOWN_120D: GZ2000



CSK_XYY_UP_DOWN_120D: GZ2000

3.8 Robust Test -- Fama-Macbeth regression

Time Period	From December 31, 1999, to April 30, 2024
Method	Fama-Macbeth Regression(1973)
Factors	Fama-French five factors come from CSMAR

	CSK_XYY_120D	*CSK_XYY_UP_120D*	*CSK_XYY_DOWN_120D*	*CSK_XYY_UP_DOWN_120D*
Anomaly Factor	-0.007^{*} (-1.696)	0.007^{**} (2.208)	-0.007^{***} (-3.349)	0.002^{***} (2.773)
RiskPremium	0.058 (1.184)	0.036 (0.635)	0.076 (1.533)	0.050 (1.055)
SMB	-0.064 (-1.628)	-0.101^{**} (-2.204)	-0.065 (-1.610)	-0.034 (-0.938)
HML	-0.103^{***} (-3.023)	-0.094^{***} (-2.653)	-0.075^{**} (-2.233)	-0.081^{**} (-2.551)
RMW	-0.016 (-0.525)	-0.022 (-0.693)	-0.010 (-0.375)	-0.020 (-0.754)
CMA	-0.006 (0.195)	0.010 (0.308)	0.002 (0.049)	0.006 (0.204)
$Avg.R^2$	0.0123	0.0131	0.0107	0.0122

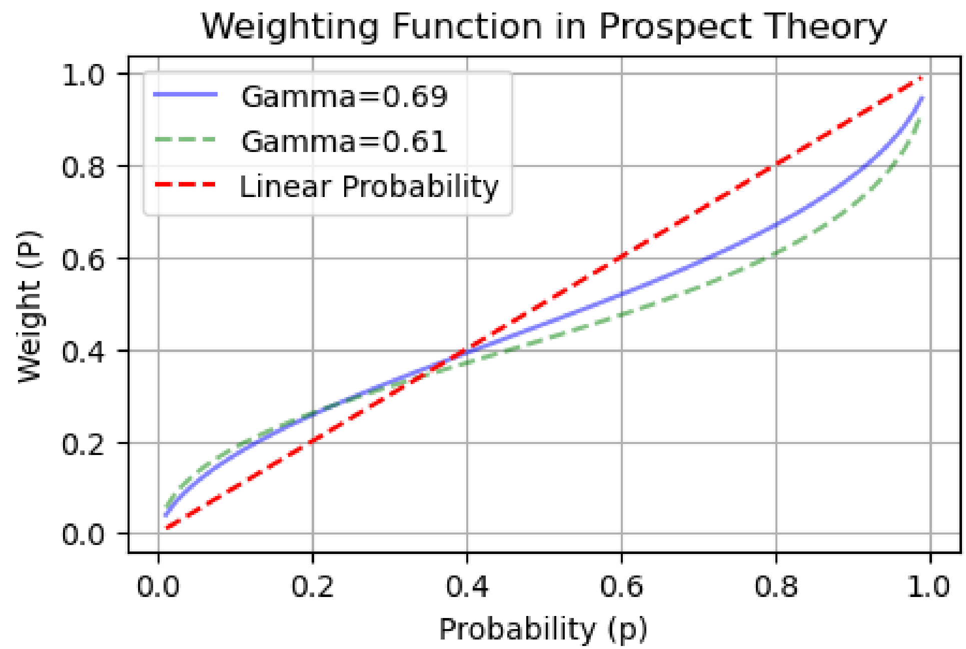
- All anomaly factors remain steady.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: empirical tests. Journal of Political Economy, Vol. 81(3), 607 – 636.

3.9 How to explain anomaly factor? – Prospect Theory

***Assumption* 1:** under the framework of Prospect Theory. The weighted function of investors can be represented as:

$$weight(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}}$$



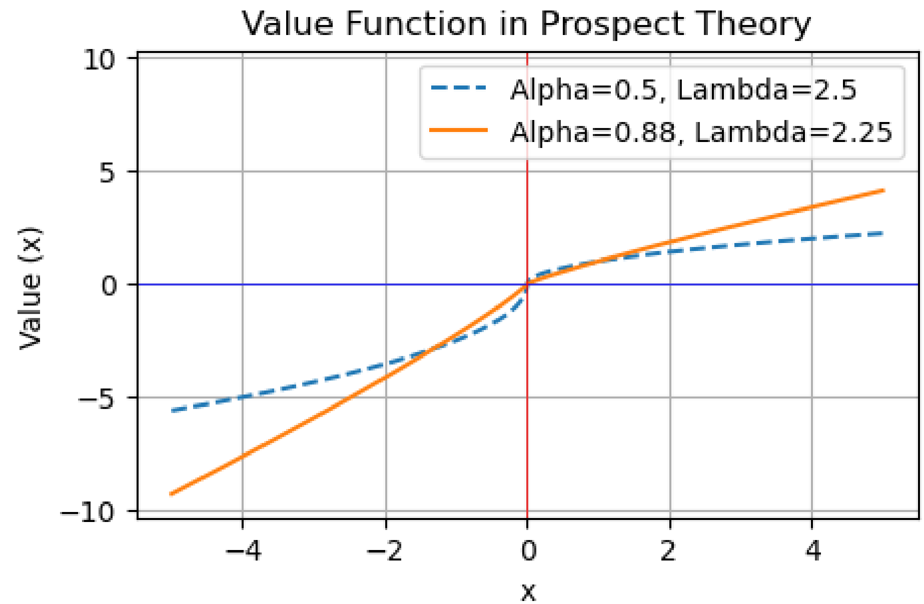
- When calculating the weight of each outcome. Kahneman and Tversky (1979) stated that *weight(p)* should satisfy *weight(0)=0* and *weight(1)=1*, but **when *0<p<1*, *weight(p)* is a nonlinear function of *p*.**
- When p is very small, *weight(p)>p*, indicating that **people often overestimate the likelihood of outcomes with very low probabilities.**

Kahneman, D. and A. Tversky (1979). Prospect Theory: an analysis of decision under risk. *Econometrica*, Vol. 47(2), 263 – 292.

3.9 How to explain anomaly factor? – Prospect Theory

***Assumption* 2:** under the framework of Prospect Theory. The value function of asset can be represented as:

$$\text{Value}(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -\lambda(-x)^{\alpha} & \text{if } x < 0 \end{cases}$$



- The outcome *x*, which refers to gains and losses, is relative to a given reference point, ***value(x)* is a nonlinear function of *x*.**
- The value function reflects people's loss aversion. **The value function *value(x)* is not symmetric on the left and right sides**; the negative growth of the loss part is faster than the positive growth of the gain part: *value(x)<-value(-x)*. Empirical research has shown that the pain caused by losses is about twice the happiness brought by gains.
- **Whether it is a gain or a loss, the value function exhibits diminishing sensitivity**. This means that when the outcome is a gain, the value function is concave; when the outcome is a loss, the value function is convex.

3.9 How to explain anomaly factor? – Prospect Theory

***Assumption* 3:** the weight function in Cumulative Prospect Theory (Tversky and Kahneman (1992)):

$$\pi(x_i) = \begin{cases} weight^{+}(P_i + P_{i+1} \dots + P_{max}) - weight^{+}(P_{i+1} + \cdots + P_{max}) & if\ 0 \le X_i \le X_{max} \\ weight^{-}(P_{min} + \cdots + P_i) - weight^{-}(P_{min} + \cdots + P_i + P_{i+1}) & if\ X_{min} \le X_i < 0 \end{cases}$$

$$weight^{+}(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}}, \gamma = 0.61\ ;\ weight^{-}(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{\frac{1}{\gamma}}}, \gamma = 0.69$$

- Contrast to Prospect Theory, in Cumulative Prospect Theory, **the value of *π(x)* is determined by the cumulative weighting functions *weight+(.)* or *weight-(.)*.**
- Due to Cumulative Prospect Theory allowing for multiple outcomes at both the gain and loss ends, this implies that **people tend to overestimate the probabilities of outcomes occurring at the tails of the distribution.**
- **The sum of the weights of all outcomes does not necessarily equal 1**.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, Vol. 5(4), 297 – 323.

3.9 How to explain anomaly factor? – Prospect Theory

***Experiment*:** We refer to the construction method of Barberis, Mukherjee, and Wang (2016). The monthly returns of a stock over the past 36 months are sorted from smallest to largest, denoted as r_{min} , r_{min+1} , …… , r_{max-1} , r_{max}, and it is assumed that the probability of each outcome occurring is $\frac{1}{36}$.

Step 1: The distribution of outcomes for each stock can be expressed as:

$$\left(r_{min}, \frac{1}{36}; r_{min+1}, \frac{1}{36}; \ \dots; r_{max-1}, \frac{1}{36}; r_{\max}, \frac{1}{36}\right)$$

Step 2: Consider the value function and cumulative weighting function of Prospect Theory, we can express the distribution of outcomes as:

$$(Value(r_{min}), \pi(r_{min}); Value(r_{min+1}), \pi(r_{min+1}); \ \dots; Value(r_{max-1}), \pi(r_{max-1}); Value(r_{\max}), \pi(r_{max}))$$

Step 3: Calculating it, we obtain the TK value for each stock:

$$TK = \sum_{i=min}^{max} Value(r_i) \times \pi(r_i)$$

Barberis, N., A. Mukherjee, and B. Wang (2016). Prospect theory and stock returns: An empirical test. *Review of Financial Studies*, Vol. 29(11), 3068 – 3107.

3.9 How to explain anomaly factor? – Prospect Theory

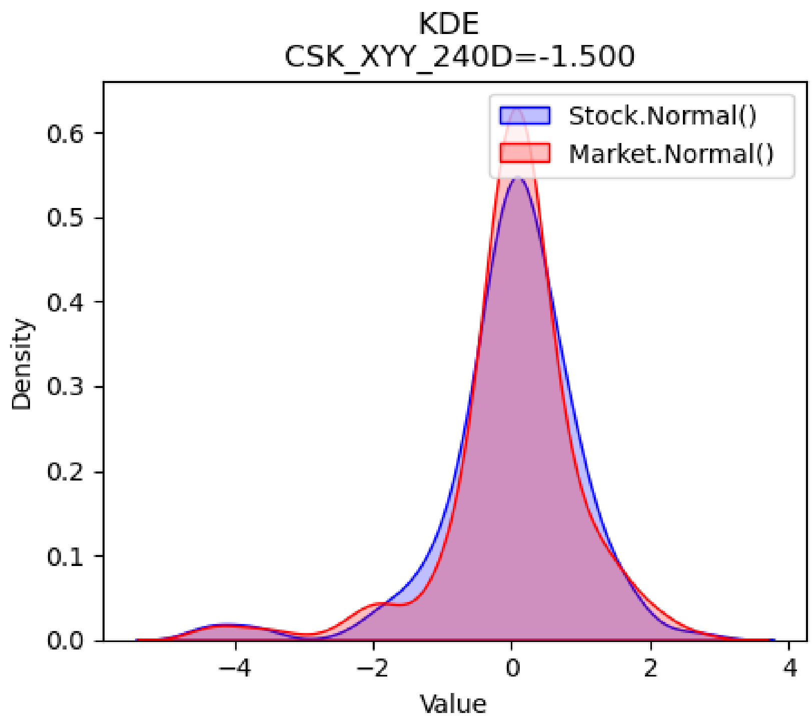
***Test*:** We use $CSK\_XYY\_DOWN\_120D$ factor for back testing over the period from *December 31, 1999*, to *May 30, 2024*, and divide it into 10 layers. (With all other conditions being the same as previously stated).

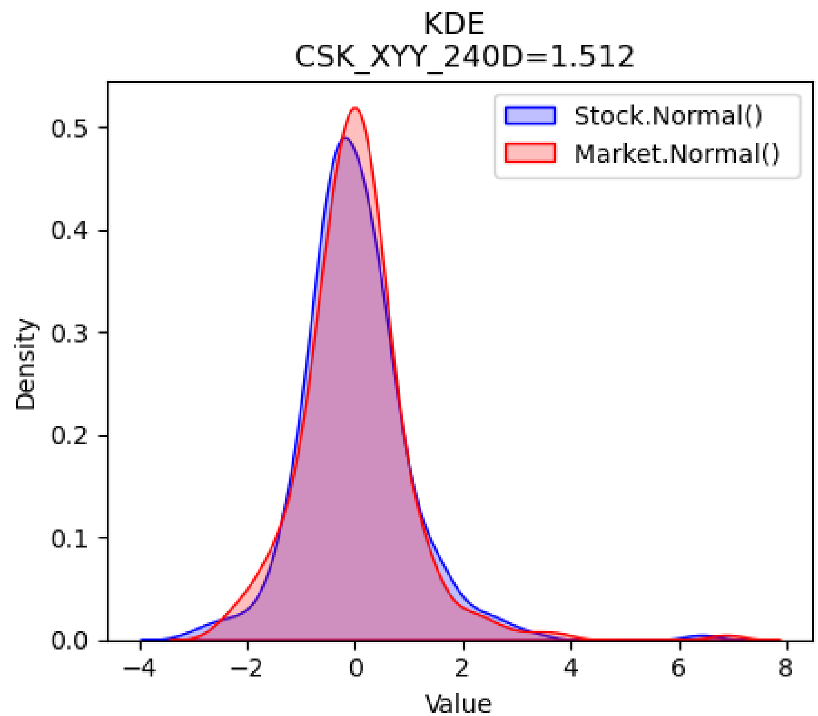
We report the back testing results along with the average TK and $CSK\_XYY\_DOWN\_120D$ values for each group of stocks:

Group	***Annualized Return***	***Factor(mean)***	***TK(mean)***
Low	*15.74%*	*-1.4316*	*-0.06568*
2	*13.70%*	*-0.9626*	*-0.06458*
3	*11.86%*	*-0.6883*	*-0.06355*
4	*12.06%*	*-0.4502*	*-0.06176*
5	*10.97%*	*-0.2185*	*-0.06049*
6	*10.79%*	*0.0256*	*-0.05886*
7	*9.03%*	*0.3044*	*-0.05734*
8	*7.69%*	*0.6596*	*-0.05521*
9	*4.98%*	*1.1390*	*-0.05265*
High	*-0.69%*	*2.1885*	*-0.05013*

3.9 How to explain anomaly factor? – Prospect Theory

***Intuition* 1:** When exhibiting positive co-skewness, the right tail of the stock distribution is thicker compared to the market distribution, and vice versa.





***Intuition* 2:** Stocks with high TK values are more attractive. Stocks with high TK values tend to be overestimated, and vice versa.

3.9 How to explain anomaly factor? – Prospect Theory

PHBS 北京大学汇丰商学院

***Result*:**

Group	*Annualized Return*	*Factor(mean)*	*TK(mean)*
Low	*15.74%*	*-1.4316*	*-0.06568*
2	*13.70%*	*-0.9626*	*-0.06458*
3	*11.86%*	*-0.6883*	*-0.06355*
4	*12.06%*	*-0.4502*	*-0.06176*
5	*10.97%*	*-0.2185*	*-0.06049*
6	*10.79%*	*0.0256*	*-0.05886*
7	*9.03%*	*0.3044*	*-0.05734*
8	*7.69%*	*0.6596*	*-0.05521*
9	*4.98%*	*1.1390*	*-0.05265*
High	*-0.69%*	*2.1885*	*-0.05013*

- **The *TK* values almost perfectly increase with the number of groups.**
- We hypothesis: Within the framework of Prospect Theory, individuals frequently overestimate the tail return, thereby demonstrating a **preference for lottery-like stocks characterized by positive co-skewness**, which consequently leads to their overvaluation, and vice verse.
- This dynamic contributes to the persistent co-skewness anomaly.