

Factor-Based Index Tracking: A Direct Indexing Approach Using PCA Leader Stocks

Ross Cunningham, Emmy Plage, Ethan Davis

Summary of Approach

- Question: Can we replicate the S&P 500 using a small subset of stocks?
- Method: PCA + “leader stock” selection + optimization
- Key Result: 30–40 stocks can closely track the index
- Main Insight: Performance depends on market stability

What is Direct Indexing?

- Invest directly in individual stocks
- Instead of buying an ETF or mutual fund
- Direct indexing managed accounts are expected to grow 12.3% annually and reach \$825B in assets by 2026¹

The logo for S&P Global, featuring a horizontal line above the text "S&P Global" in red.

What are the benefits?

- **Customization** (exclude stocks, tilt sectors)
- **Tax efficiency** (tax-loss harvesting)
- Greater portfolio control

The BlackRock ADVISOR CENTER logo, with "BlackRock" in large bold black font and "ADVISOR CENTER" in smaller bold black font below it.The Vanguard logo, featuring the word "Vanguard" in a red serif font with a registered trademark symbol.

Our Objective

Step 1: Factor Extraction (PCA)

- Identify main drivers using universe of 488 stocks

Step 2: Leader Stock Selection

- Select representative stocks for factors

Step 3: Construct portfolio via 3 optimization methods:

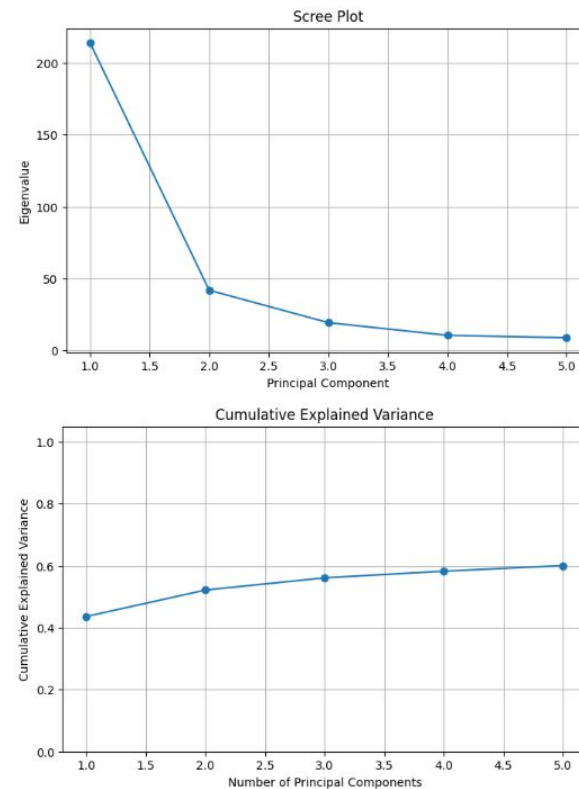
- Unconstrained Optimization
- No Short Selling
- Shrunked Beta

Goal:

- Evaluate which method tracks the index best

PCA & Factor Insight

- PCA decomposes returns into orthogonal factors
- Few factors explain most variation
- Eigenvalue spectrum shows diminishing returns
- Top 5 factors sufficient
- Methodology follows Jiang and Perez (2021)



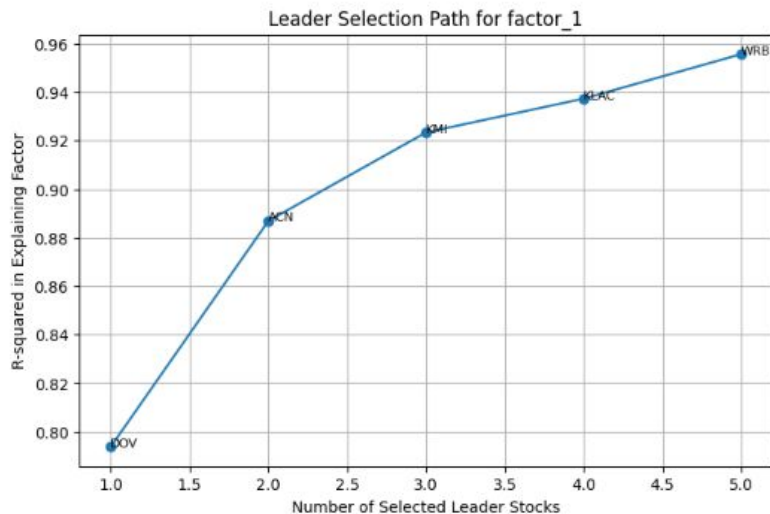
Leader Selection

Rank stocks by correlation with each factor

Iteratively add stocks via regression on residuals

Stop when R^2 threshold reached ($R^2 \geq 95\%$)

Ensures no redundant stocks and maximum explanatory power



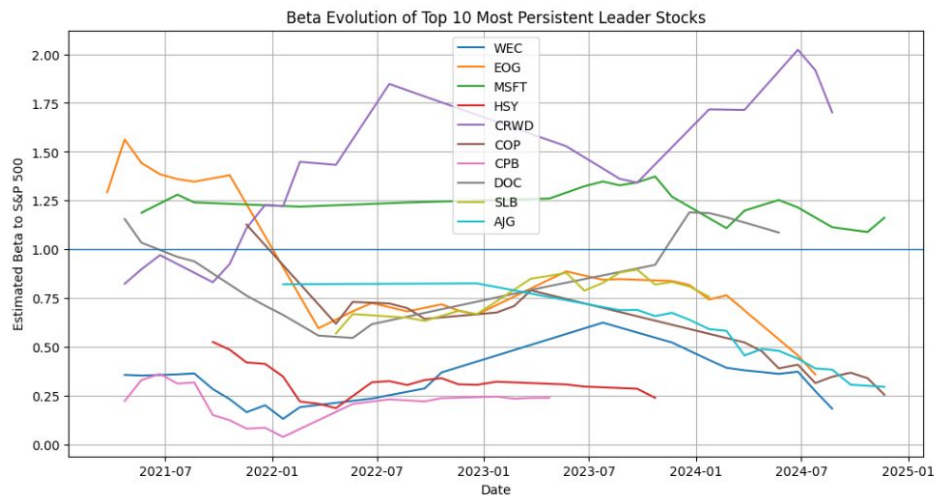
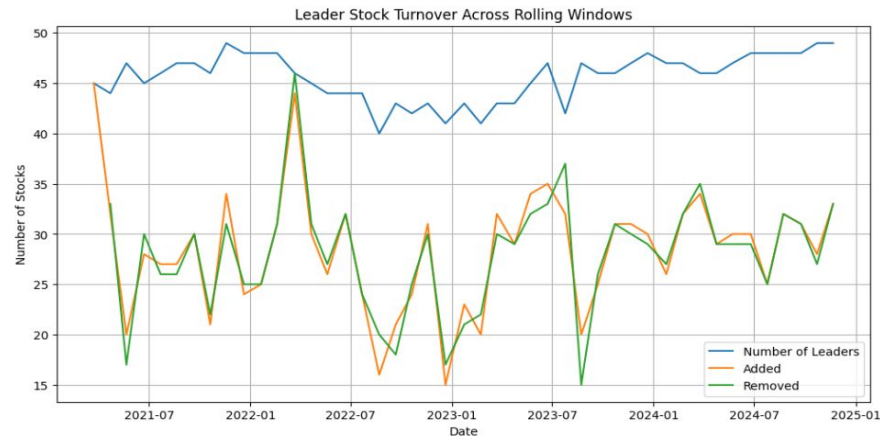
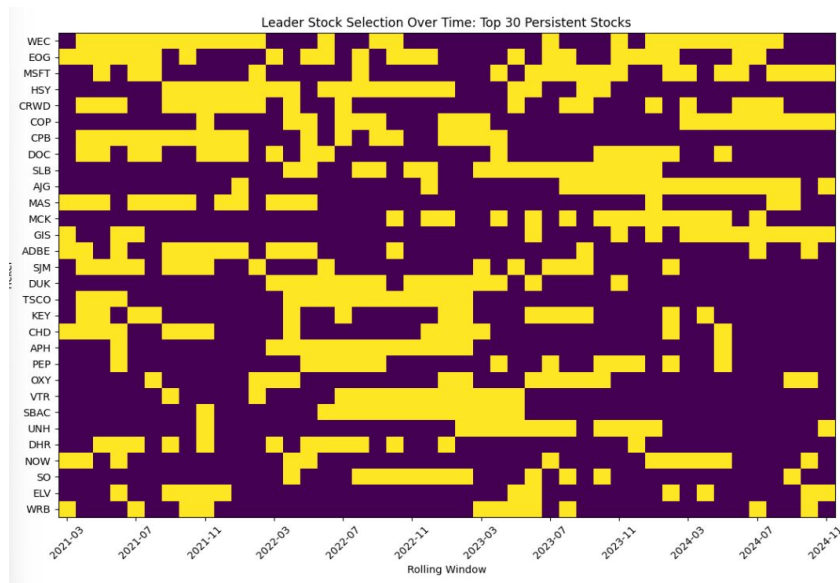
Optimization & Constraints

Objective: maximize Sharpe ratio

Expected returns proxied via beta

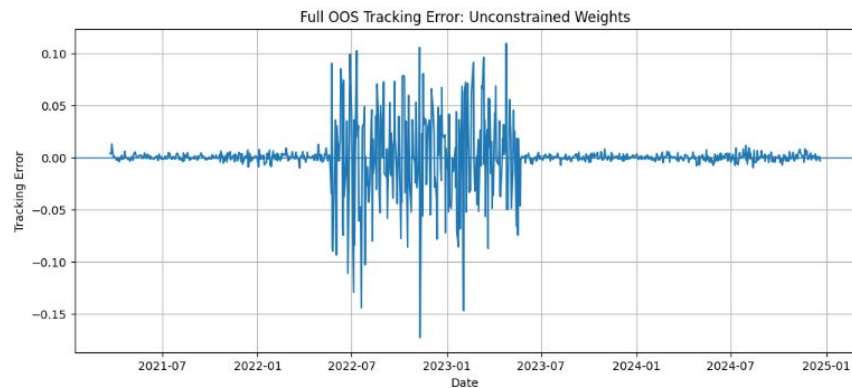
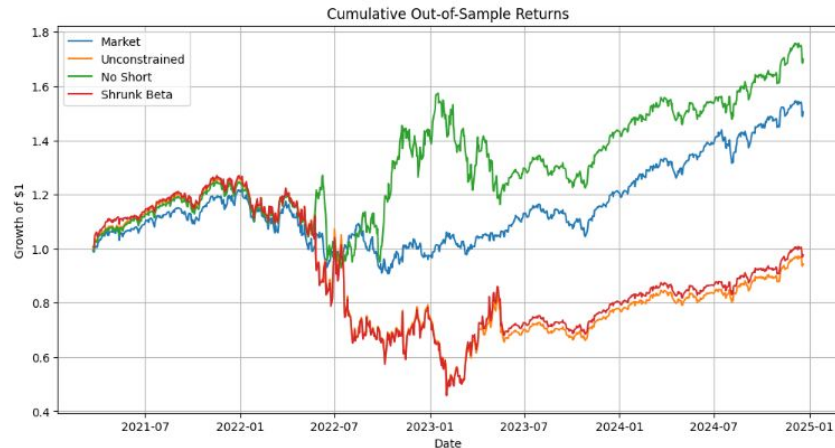
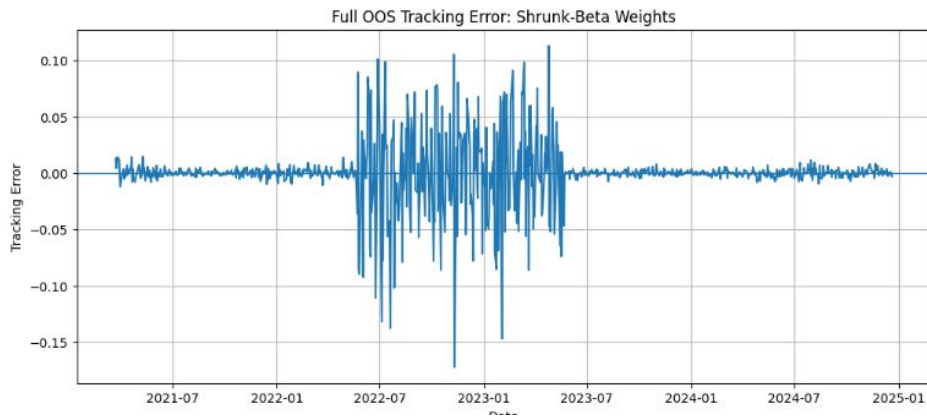
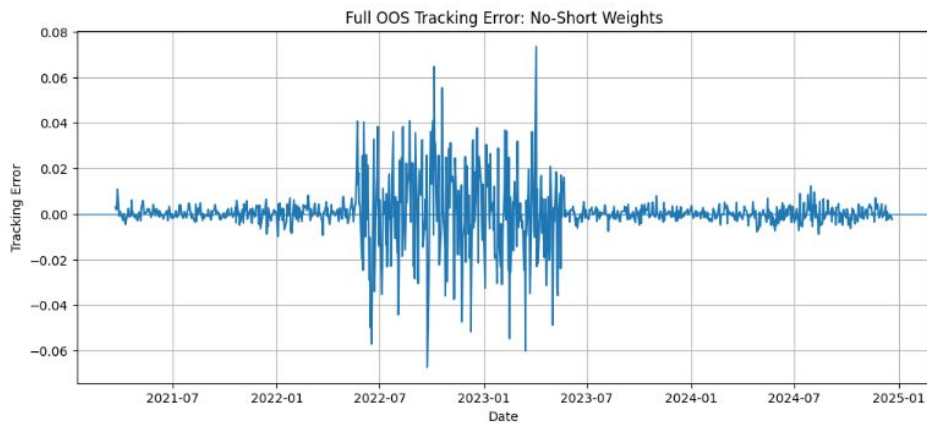
Three approaches: Unconstrained; No short-selling; Shrunk beta (stability)

Weight bounds imposed ($\pm 25\%$)



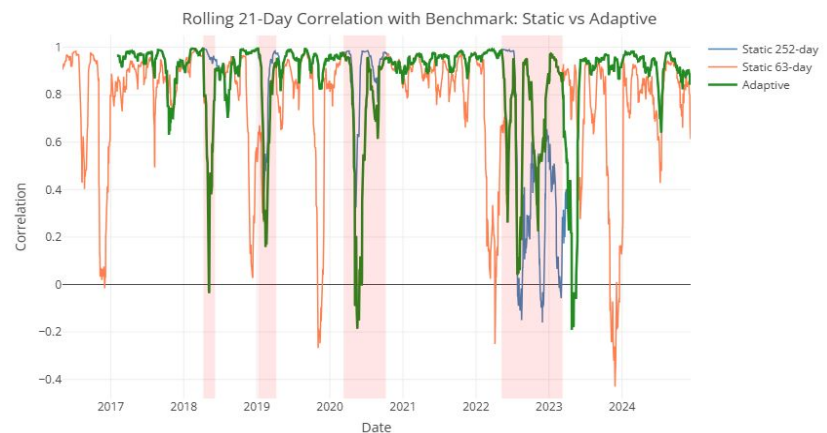
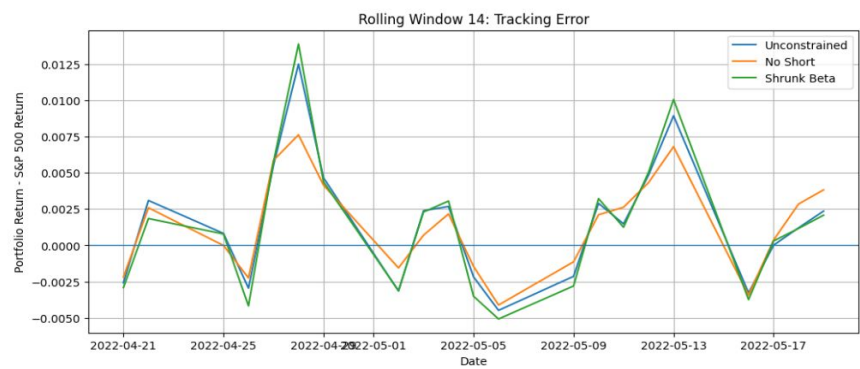
Baseline Performance

- 40-50 stocks replicate index effectively
- Low tracking error in stable periods
- All strategies perform well initially



Structural Breaks

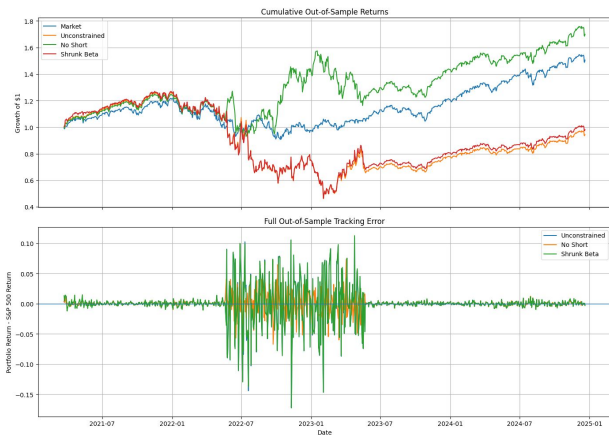
- Factor leadership rotates during regime shifts
- Tracking error increases significantly (252-day correlation drops from 0.945 to 0.413 during 2022-2023)
- Cause: change in correlation structure as it spikes from 0.25 to 0.5 with regime change
- PCA assumes stable covariance
- Model adapts with lag



Strategy	Regime	Start	End	N Days	Cumulative Return	Market Cumulative Return	Annualized Return	Annualized Volatility	Annualized Sharpe	Correlation with Market	Tracking Error Mean	Tracking Error Vol	Mean Absolute Tracking Error	MSE Tracking Error	
0	Unconstrained	Full OOS	2021-03-23	2024-12-20	945	-0.056901	0.505066	0.053159	0.368807	0.144138	-0.042854	-0.000276	0.410594	0.012269	0.000668
1	Unconstrained	Structural Break	2022-06-01	2023-06-30	272	-0.262331	0.077013	-0.082075	0.628886	-0.130508	-0.451513	-0.000680	0.743411	0.035553	0.002185
2	Unconstrained	Outside Structural Break	2021-03-23	2024-12-20	673	0.278485	0.397445	0.107815	0.177591	0.607099	0.751269	-0.000113	0.118020	0.002859	0.000055
3	No Short	Full OOS	2021-03-23	2024-12-20	945	0.699471	0.505066	0.171709	0.245750	0.698715	0.640809	0.000194	0.188827	0.006304	0.000141
4	No Short	Structural Break	2022-06-01	2023-06-30	272	0.095541	0.077013	0.158336	0.384012	0.412321	0.462048	0.000274	0.341573	0.016237	0.000461
5	No Short	Outside Structural Break	2021-03-23	2024-12-20	673	0.551262	0.397445	0.177114	0.159276	1.111996	0.938365	0.000162	0.055099	0.002289	0.000012
6	Shrunk Beta	Full OOS	2021-03-23	2024-12-20	945	-0.022571	0.505066	0.062111	0.367406	0.169053	-0.044207	-0.000241	0.409513	0.012443	0.000665
7	Shrunk Beta	Structural Break	2022-06-01	2023-06-30	272	-0.233431	0.077013	-0.049036	0.625179	-0.078435	-0.453384	-0.000549	0.740138	0.035387	0.002166
8	Shrunk Beta	Outside Structural Break	2021-03-23	2024-12-20	673	0.275071	0.397445	0.107032	0.178824	0.598534	0.740729	-0.000116	0.121039	0.003170	0.000058

Results - Stability vs Adaptability

Annual estimation and 1-month holding period



Monthly estimation and 1-month holding Period

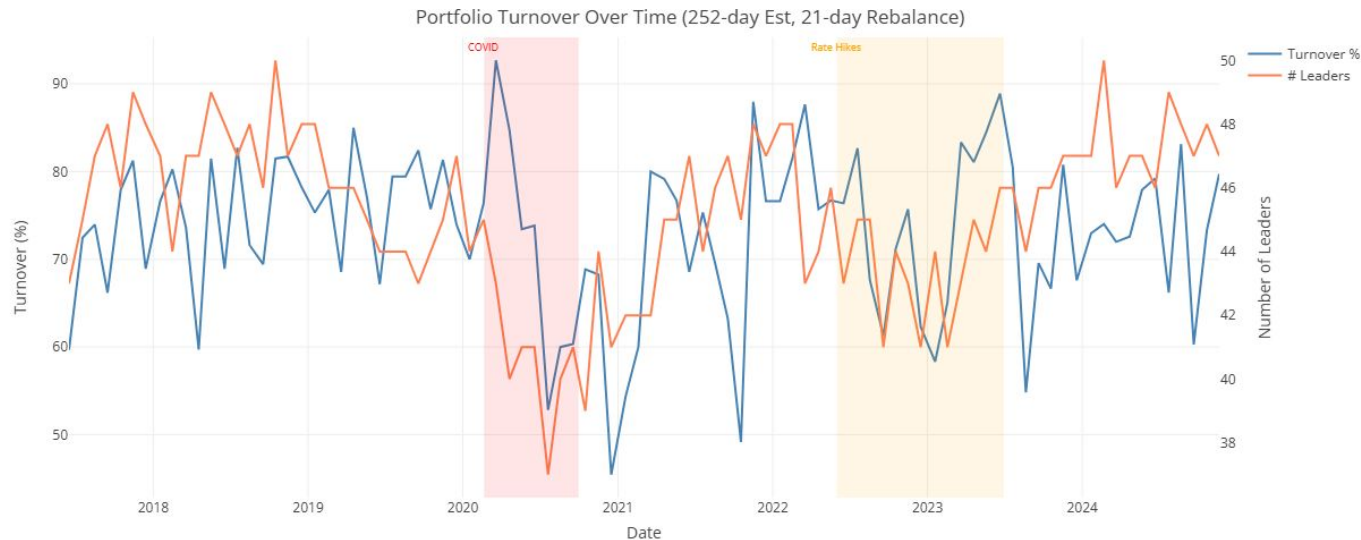


Annual estimation and 1-year holding period



- shorter windows exhibit lower tracking error during structural breaks
- longer windows exhibit lower mean tracking error in stable periods

Turnover



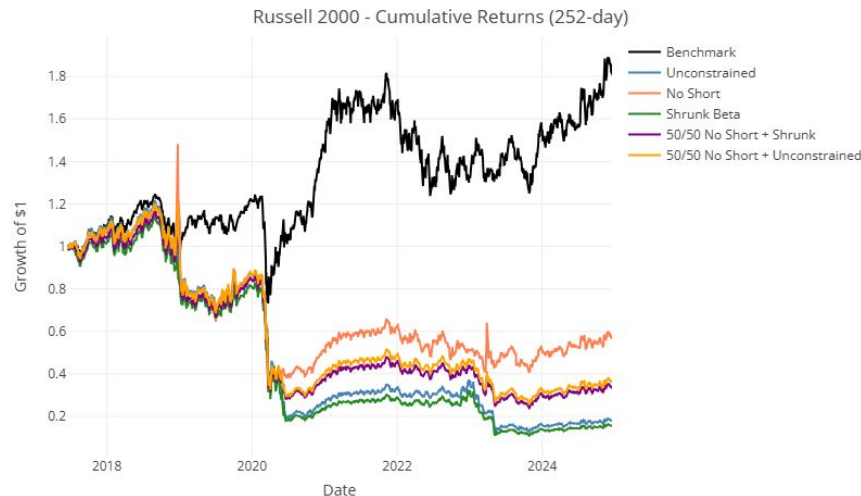
- Turnover gets worse the longer we make the rebalance period
- Top held all energy companies, stocks that move independently to tech
- No stock appears more than 50%
- Regime doesn't affect turnover

Window (21 Day Rebalance)	Stock Turnover	Avg Stocks Held
21-Day	99%	6.7
63-Day	94%	20.9
126-Day	85%	43.8
252-Day	73%	45.2

Expanded Universe - Russell 1000, 2000 indexes

- Better for Indexes with less Stocks. Hard to Properly Show without Index Constituents Historical Data

	S&P 500	Russ 1000	Russ 2000
Avg Pairwise Correlation	0.29	0.27	0.19
PC1 Explains	32%	30%	22%
Best Correlation	0.865	0.836	0.562
Best TE Vol	9.8%	11.0%	22.9%



Key Insights

- A small number of stocks can replicate the index
- PCA effectively captures factor structure
- Stock selection is critical
- Objective function matters
- Minimizing tracking error alone is insufficient

	Ticker	Times_Selected	First_Selected	Last_Selected	Selection Frequency
0	WEC	24	2021-04-22	2024-08-23	0.533333
1	EOG	23	2021-03-23	2024-07-25	0.511111
2	MSFT	20	2021-05-21	2024-11-21	0.444444
3	HSY	20	2021-09-21	2023-10-23	0.444444
4	CRWD	19	2021-04-22	2024-08-23	0.422222
5	COP	18	2021-11-18	2024-11-21	0.400000
6	CPB	18	2021-04-22	2023-04-24	0.400000
7	DOC	17	2021-04-22	2024-05-23	0.377778
8	SLB	17	2022-04-21	2024-01-24	0.377778
9	AJG	17	2022-01-20	2024-11-21	0.377778
10	MAS	15	2021-03-23	2024-09-24	0.333333
11	MCK	15	2022-10-20	2024-07-25	0.333333
12	GIS	15	2021-03-23	2024-11-21	0.333333
13	ADBE	15	2021-03-23	2024-10-23	0.333333
14	SJM	15	2021-04-22	2024-02-23	0.333333
15	DUK	15	2022-03-22	2023-11-21	0.333333
16	TSCO	14	2021-04-22	2023-02-22	0.311111
17	KEY	14	2021-04-22	2024-04-24	0.311111
18	CHD	14	2021-03-23	2024-05-23	0.311111
19	APH	14	2021-06-22	2024-05-23	0.311111

PCA-based replication is effective in stable markets; Small subset of stocks can capture index behaviour; Performance deteriorates under regime shifts

Trade-off: stability vs adaptability

No Short strategy performs optimally against other constraints. Jagannathan and Ma (2003) show that adding nonnegativity constraints (no-short-selling constraint) to portfolio optimization helps reduce estimation error, essentially acting as a regularization tool.

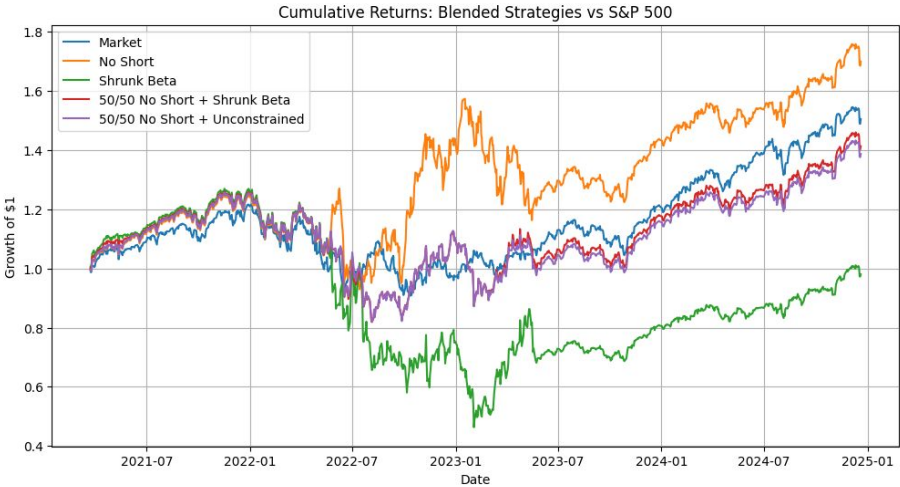
Open questions:

- Can we detect regime changes?
- Can we improve factor stability?
- Can we use a different metric instead of R^2 to determine the stocks that are chosen to represent the factors since R^2 is not a reliable metric?
- Is a blended strategy successful in different periods?

Best approach:

The 1 year estimation window with monthly holding period has the lowest mean tracking error across the entire period. But, this strategy has a high turnover rate. Instead, we recommend using the 1 year estimation period for a 1 year holding period as the most effective trading strategy.

Conclusions



	Setup	Strategy	N Days	Mean Absolute Tracking Error	Final Cumulative Tracking Error	Absolute Final Cumulative Tracking Error	Max Absolute Cumulative Tracking Error	Tracking Error Vol Annualized	Correlation with Market
7	Annual / non-rolling hold	No Short	920	0.004935	-0.023051	0.023051	0.147523	0.141686	0.743201
4	252-day rolling estimate	No Short	945	0.006304	0.183718	0.183718	0.486334	0.188827	0.640809
0	21-day rolling estimate	Unconstrained	1176	0.006827	0.486031	0.486031	0.664417	0.160946	0.672982
2	21-day rolling estimate	Shrunk Beta	1176	0.006834	0.722681	0.722681	0.876436	0.161541	0.695632
1	21-day rolling estimate	No Short	1176	0.009450	0.176971	0.176971	0.353419	0.300931	0.498812
6	Annual / non-rolling hold	Unconstrained	920	0.009649	-0.066586	0.066586	0.436907	0.317318	0.275173
3	252-day rolling estimate	Unconstrained	945	0.012270	-0.261089	0.261089	0.678788	0.410603	-0.042854
5	252-day rolling estimate	Shrunk Beta	945	0.012428	-0.225273	0.225273	0.678971	0.408885	-0.043862
8	Annual / non-rolling hold	Shrunk Beta	920	0.012961	-0.457553	0.457553	0.760443	0.389127	0.065003