RF RL Walk-Forward Benchmark Analysis

June 25, 2025

Benchmark testing of proprietary regime-prediction pipeline

Tier-1 & Tier-2 compliant statistics • Significance testing

Pipeline overview
Every model (Random-Forest classification *and* discrete–action CQL) is trained on a sliding window, then produces a one-week, hour-by-hour forecast before the window advances and the most recent week is used for fine-tuning.

380+ engineered features $\longrightarrow$ feature preprocessing $\longrightarrow$ feature selection $\longrightarrow$ model train/predict (walk-forward) $\longrightarrow$ post-processing ensemble $\longrightarrow$ threshold optimisation $\longrightarrow$ **discrete regime output (Long, Short)**.

Contents

1. Per-asset Performance Metrics

-S : strategy | -BH : Buy and Hold

Table 1 — Performance

Symbol	CAGR-S	CAGR-BH	Sharpe-S	Sharpe-BH
Airbnb Inc	0.1013	-0.0594	0.4429	0.0292
Alphabet Inc C	0.0836	0.067	0.4331	0.3759
American Airlines Group	0.3633	-0.1059	1.0081	-0.0974
Apple	0.1276	0.0414	0.6073	0.2892
Bank Of America	0.104	0.0356	0.5288	0.266
Berkshire Hathaway - Class B	0.1328	0.0873	0.9052	0.6207
Boeing Co	0.1394	0.0285	0.5655	0.2489
Broadcom Ltd	0.144	0.3709	0.5527	1.0353
CME Group	0.0405	0.0417	0.3307	0.3392
Caterpillar	0.1475	0.1006	0.6783	0.5068
Coinbase Global Inc	0.8653	0.1233	1.1941	0.5356
Costco	0.0418	0.1345	0.3032	0.7315
Ford Motor	0.2142	-0.0846	0.7514	-0.1055
General Electric	0.3528	0.2465	1.1964	0.9156

Symbol	CAGR-S	CAGR-BH	Sharpe-S	Sharpe-BH
General Motors co	0.0282	0.0481	0.2446	0.3064
Goldman Sachs	0.1112	0.1477	0.5615	0.6973
HP Inc	-0.0032	-0.0874	0.1262	-0.1836
IBM	0.1309	0.1706	0.7066	0.8797
Intel	0.1654	-0.1718	0.5829	-0.2863
J.P. Morgan	0.1432	0.179	0.7481	0.9002
Lululemon	0.0689	-0.0484	0.3645	0.03
Meta Platforms Inc	0.2765	0.2789	0.8267	0.8344
Morgan Stanley	0.0596	0.0949	0.3551	0.4879
NIO Inc	-0.3062	-0.315	-0.3285	-0.3385
Netflix	0.3856	0.3218	1.0312	0.9296
Nvidia	0.4157	0.5015	0.9952	1.1185
Paypal	0.1919	-0.0804	0.6917	-0.0811
PepsiCo	0.0692	-0.0576	0.5263	-0.3242
SONOS Inc	0.2035	-0.1898	0.6518	-0.3015
Tesla Motors	0.3699	0.0136	0.8785	0.2801
Under Armour A	0.0606	-0.1757	0.3481	-0.2096
Wynn Resorts	-0.1019	0.0455	-0.1512	0.3005
Xerox Corp	0.1125	-0.259	0.4614	-0.5211
Zillow Group Inc A	-0.2848	0.0805	-0.5936	0.3916

Table 2 — Risk

Symbol	MaxDD-S	MaxDD-BH	VaR-S	ES-S
Airbnb Inc	-0.4206	-0.515	0.0673	0.0878
Alphabet Inc C	-0.2866	-0.3509	0.0459	0.0671
American Airlines Group	-0.359	-0.5531	0.0644	0.0815
Apple	-0.3704	-0.3261	0.0438	0.0601
Bank Of America	-0.3413	-0.3695	0.0423	0.0569
Berkshire Hathaway - Class B	-0.209	-0.2404	0.0261	0.0332
Boeing Co	-0.4912	-0.4834	0.0539	0.0672
Broadcom Ltd	-0.3889	-0.4038	0.0554	0.0757
CME Group	-0.3409	-0.2948	0.0286	0.0391
Caterpillar	-0.3013	-0.3405	0.0446	0.0568
Coinbase Global Inc	-0.419	-0.785	0.1076	0.1342
Costco	-0.2942	-0.319	0.0406	0.0518
Ford Motor	-0.5042	-0.4678	0.0564	0.067
General Electric	-0.257	-0.338	0.0459	0.0701
General Motors co	-0.6359	-0.3852	0.0541	0.0739
Goldman Sachs	-0.2889	-0.3126	0.0399	0.0489
HP Inc	-0.609	-0.4487	0.0491	0.0642
IBM	-0.3018	-0.1958	0.0331	0.0529
Intel	-0.3681	-0.6338	0.0589	0.0888
J.P. Morgan	-0.3064	-0.2467	0.0393	0.0531
Lululemon	-0.456	-0.5442	0.0578	0.0931

Symbol	MaxDD-S	MaxDD-BH	VaR-S	ES-S
Meta Platforms Inc	-0.4347	-0.592	0.0582	0.0826
Morgan Stanley	-0.2304	-0.3024	0.0451	0.0612
NIO Inc	-0.8762	-0.8728	0.1047	0.1267
Netflix	-0.3823	-0.525	0.0576	0.0789
Nvidia	-0.5534	-0.4847	0.0711	0.0936
Paypal	-0.3824	-0.5124	0.0631	0.079
PepsiCo	-0.1238	-0.3411	0.0237	0.0335
SONOS Inc	-0.5451	-0.7071	0.0593	0.0963
Tesla Motors	-0.4896	-0.6827	0.0789	0.1036
Under Armour A	-0.4853	-0.7081	0.0643	0.0886
Wynn Resorts	-0.6415	-0.4272	0.065	0.0863
Xerox Corp	-0.4877	-0.8155	0.0648	0.0918
Zillow Group Inc A	-0.8493	-0.4223	0.0704	0.1031

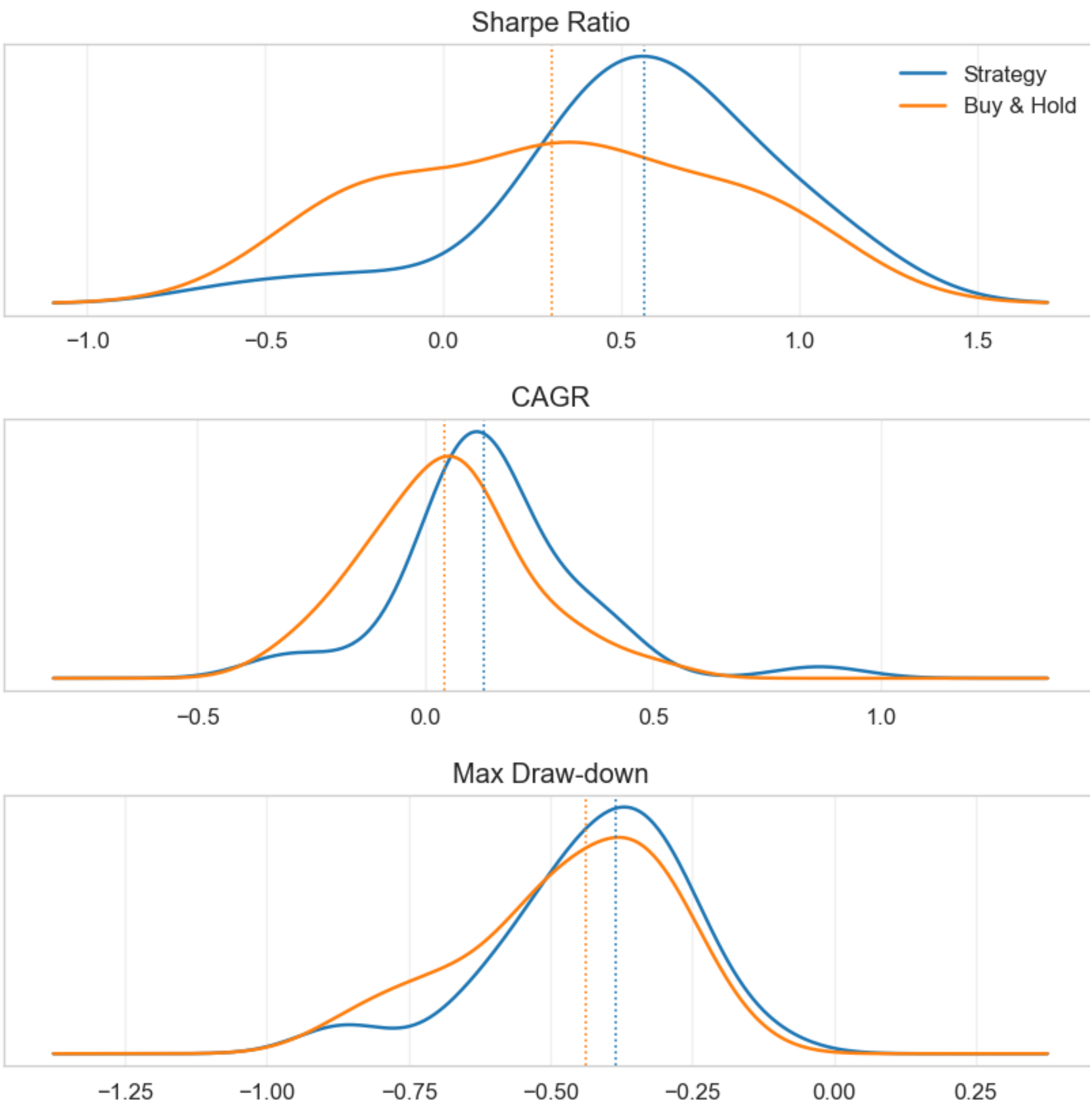
Table 3 — Significance

Symbol	SharpeCI	t-stat	p-value
Airbnb Inc	[-0.49,1.33]	0.6222	0.5339
Alphabet Inc C	[-0.48,1.37]	0.0856	0.9318
American Airlines Group	[0.08,1.90]	1.6463	0.0999
Apple	[-0.33,1.52]	0.503	0.615
Bank Of America	[-0.40,1.43]	0.396	0.6922
Berkshire Hathaway - Class B	[0.00,1.83]	0.3991	0.6899
Boeing Co	[-0.36,1.47]	0.4758	0.6342
Broadcom Ltd	[-0.39,1.40]	-0.7712	0.4407
CME Group	[-0.61,1.25]	-0.0103	0.9918
Caterpillar	[-0.26,1.61]	0.2456	0.806
Coinbase Global Inc	[0.31,2.05]	0.94	0.3473
Costco	[-0.60,1.22]	-0.6369	0.5243
Ford Motor	[-0.16,1.61]	1.2931	0.1961
General Electric	[0.26,2.18]	0.4235	0.672
General Motors co	[-0.65,1.17]	-0.0869	0.9307
Goldman Sachs	[-0.34,1.46]	-0.2072	0.8359
HP Inc	[-0.79,1.02]	0.4663	0.6411
IBM	[-0.20,1.59]	-0.2538	0.7997
Intel	[-0.35,1.47]	1.3086	0.1908
J.P. Morgan	[-0.15,1.67]	-0.2218	0.8245
Lululemon	[-0.56,1.28]	0.5071	0.6122
Meta Platforms Inc	[-0.04,1.66]	-0.0117	0.9907
Morgan Stanley	[-0.56,1.25]	-0.1906	0.8488
NIO Inc	[-1.25,0.62]	0.0232	0.9815
Netflix	[0.15,1.83]	0.1498	0.8809
Nvidia	[0.08,1.89]	-0.2046	0.8379
Paypal	[-0.24,1.61]	1.1595	0.2464
PepsiCo	[-0.39,1.47]	1.2776	0.2015

Symbol	SharpeCI	t-stat	p-value
SONOS Inc	[-0.27,1.54]	1.4367	0.151
Tesla Motors	[-0.03,1.79]	0.8805	0.3787
Under Armour A	[-0.58,1.20]	0.8385	0.4018
Wynn Resorts	[-1.06,0.78]	-0.684	0.494
Xerox Corp	[-0.48,1.37]	1.4766	0.1399
Zillow Group Inc A	[-1.50,0.31]	-1.4751	0.1403

Highlight Stats

- **Return edge** – Strategy beats buy-&-hold on CAGR in **22/34 assets** and on Sharpe ratio in **22/34 assets**.
- **Risk profile** – It reduces max draw-down in **22/34 assets**

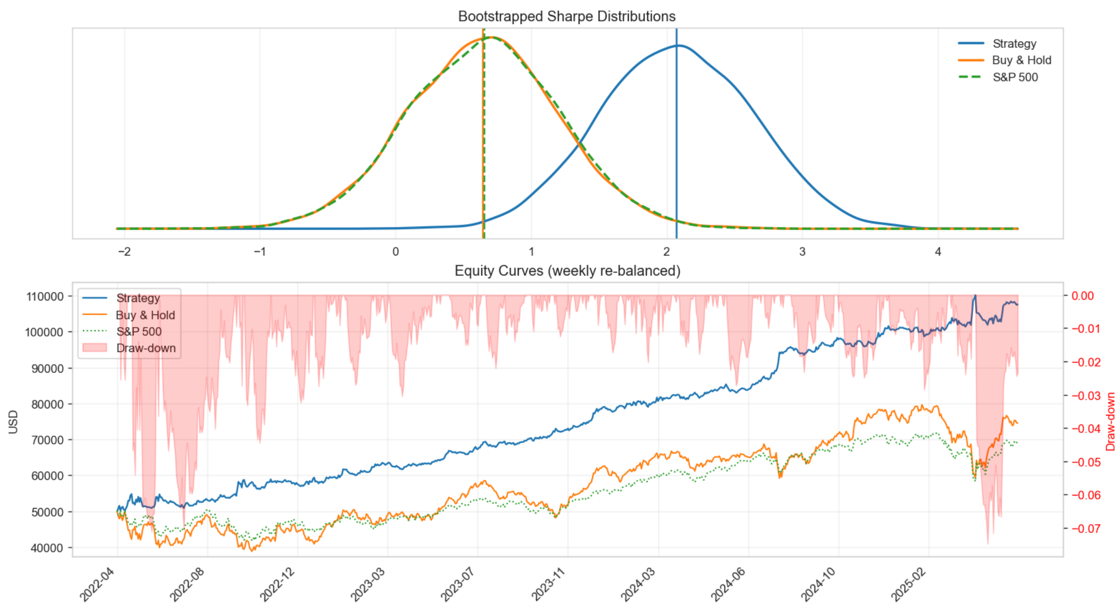


2. Equal Weighted Portfolio.

Equal weighted portfolio with all *strategy* legs, rebalanced weekly.

Table A — Portfolio vs Benchmarks Summary

	CAGR	Sharpe	Sharpe-L	Sharpe-H	MaxDD	VaR99%	ES99%	VaR/ES
Strategy	0.2806	2.0716	1.0007	3.1558	-0.0748	0.0186	0.0275	0.6755
Buy & Hold	0.1376	0.6451	-0.4403	1.7348	-0.2537	0.0423	0.0506	0.8357
S&P 500	0.1096	0.6539	-0.455	1.7546	-0.1889	0.0328	0.0409	0.8037



	Total P/L ($)	Total P/L (%)
Strategy	57542.2	115.08
Buy & Hold	24536.7	49.07
S&P 500	18994.1	37.99

The results clearly demonstrate the advantages of stacking asset edge while rebalancing weekly.

3. Portfolio Statistical Significance

Table B — Statistical Significance Tests

Test	Stat	p-value
t-test (i.i.d.)	3.7044	0.0001
HAC t-test (lag 20 d)	5.1859	<0.0001
Bootstrap > 0 (5-day blocks)	3.7044	<0.0001
ES back-test 1 % (A-S)	7.9132	<0.0001

Statistical Tests — Interpretation

All four robustness checks converge on the same verdict:

- **Persistent positive edge**
  Daily mean return is far above zero whether we assume i.i.d. data
  (classical t-test, $\mathbf{t = 3.70}$, $p = 0.0001$) **or** allow for
  realistic serial-correlation and heteroskedasticity
  (Newey-West, $\mathbf{t = 5.19}$, $p = <0.0001$).

- **Week-level dependence doesn't dampen results**
  A 5-day block bootstrap—which keeps whole trading weeks intact—
  still yields a highly significant one-sided statistic
  (**z** **3.70**, $p = <0.0001$).

- **Tail-risk passes inspection**
  The 1 % Acerbi–Szekely Expected-Shortfall back-test rejects the null of model inadequacy
  ($\mathbf{t = 7.91}$, $p = <0.0001$), indicating that the realised worst-case
  daily losses are **smaller** than a well-calibrated risk model would predict.

  Bottom-line:
  Even after accounting for autocorrelation, clustered volatility,
  week-to-week regime shifts and extreme-tail behaviour,
  the strategy shows a statistically robust, economically meaningful positive expectancy.

4. Final Commentary

In Summary

Over the **3.1-year** walk-forward window
(2022-04-25 → 2025-05-30)

an equal-weight, weekly-rebalanced basket of *all* strategy legs turned
$50,000 into **$107,542** — a CAGR of **28.06%**.

The realized **Sharpe ratio is 2.07**
(bootstrap 95 %-CI 1.00 … 3.16);
maximum draw-down is **-7.48%** and the 99 % Expected-Shortfall only **2.75%**.

For comparison, an equal-weight *buy-and-hold* of the same underlyings delivered
13.76% CAGR with Sharpe **0.65** and a far deeper draw-down of **-25.37%**.

Statistical Significance

Test	Stat	p-value
t-test (i.i.d.)	3.7044	0.0001
HAC t-test (lag 20 d)	5.1859	<0.0001
Bootstrap > 0 (5-day blocks)	3.7044	<0.0001
ES back-test 1 % (A-S)	7.9132	<0.0001

Methods at a Glance

- **Model families** – Random-Forest classifiers & discrete CQL reinforcement-learning agents on **380 +** engineered features.
- **True walk-forward** – Train 52 weeks → predict 1 wk; slide window forward one week and repeat.
- **Post-processing** – Ensemble probability calibration + threshold optimization; signals scaled to unit volatility.
- **Benchmarking** – Equal-weight buy-and-hold of every underlying plus **S&P 500** reference.
- **Compliance metrics** – Full Tier-1 / Tier-2 set: CAGR, Sharpe, Sortino, Max-DD, Calmar, 99 % VaR & ES, HAC tests, block bootstrap, Acerbi–Szekely ES back-test.