

FORECASTING RESEARCH · BTC · ETH · XRP · LINK

How well each forecasting model predicted whether BTC, ETH, XRP and LINK would end above or below price targets 30 to 90 days out, scored against what actually happened and compared with the simple base rate.

Bottom line

0
models promoted on the sealed holdout. The base rate stays the reference.

4 of 24
asset x horizon tests had enough holdout outcomes to judge (spot prices, 30 days). The rest were too short.

+16.2%
best holdout skill: Lognormal, 90-day vol, BTC 30 days. Its 95% interval (−7.9% to +44.3%) includes zero, so it is not reliable.

5 of 232
development tests where a model was reliably better than the base rate. Most differences were within noise.

Today's 30-day call for each coin (UP, DOWN or FLAT), from the model with the best record here, is on the [Crypto Price Odds](#) page.

Base rate
How often a move of that size happened in the past. The simplest possible forecast, and the bar every model must clear.

Skill
Reduction in forecast error (Brier score) versus the base rate. 0% = no better, +10% = 10% less error, negative = worse.

Sealed holdout
The most recent stretch of history, hidden while models were built and then scored once. The only fully fair test.

P vs Q
P = real-world forecasts from history and ML. Q = implied by option prices, which include a risk premium, so shown for reference only.

Final holdout verdict

What the one-time test decided
The sealed holdout (settlement outcomes from 2025-08-30, spot outcomes from 2025-04-02) was opened once and every model was judged by the rule fixed in advance.
No model was promoted in any of the 4 testable asset x horizon cells. Past event frequency stays the reference: no demonstrated improvement.

0 promotions	4 testable cells	20 untestable cells	42 model x cell tests					
Outcomes	Asset	Horizon	Strongest dev. baseline	Promoted	Best holdout model	Its skill	95% interval	≈ indep. outcomes
spot	BTC	30d	Past frequency	none: past frequency kept	Lognormal, 90-day vol	+16.2%	−7.9% to +44.3%	18
spot	ETH	30d	Lognormal, 90-day vol	none: past frequency kept	Lognormal, 90-day vol	−1.8%	−11.6% to +6.9%	18
spot	LINK	30d	Lognormal, EWMA vol	none: past frequency kept	GARCH(1,1)-t	+3.1%	−5.6% to +15.7%	18
spot	XRP	30d	Lognormal, 90-day vol	none: past frequency kept	Lognormal, 90-day vol	−4.5%	−13.6% to +5.2%	18

Every asset x horizon cell				
A cell is testable only if the holdout has at least 10 independent event and non-event windows. 20 of 24 cells fall short: the holdout covers only the most recent stretch (settlement outcomes from 2025-08-30, spot outcomes from 2025-04-02), too short for enough non-overlapping windows at most horizons. No model can be promoted or rejected there. Select a tested cell to see it in the leaderboard.				
Asset	30d	60d	90d	
Settlement prices (Deribit 08:00 UTC)				
BTC	· too few outcomes	· too few outcomes	· too few outcomes	
ETH	· too few outcomes	· too few outcomes	· too few outcomes	
XRP	· too few outcomes	· too few outcomes	· too few outcomes	
LINK	· too few outcomes	· too few outcomes	· too few outcomes	
Spot closes (exchanges)				
BTC	tested	· too few outcomes	· too few outcomes	
ETH	tested	· too few outcomes	· too few outcomes	
XRP	tested	· too few outcomes	· too few outcomes	
LINK	tested	· too few outcomes	· too few outcomes	

In every tested cell, no model passed all four conditions, so past event frequency is kept.

Outcomes	Asset	h	Model	Skill	95% interval	(1)	(2)	(3)	(4)	Result	Note
spot	BTC	30d	Chronos (price)	+6.2%	−13.0% to +31.1%	✓	✗	✓	✓	not promoted	must also pass in Stage 5 in-band: no
spot	BTC	30d	Chronos (volatility)	+13.7%	−10.3% to +44.0%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	Filtered historical simulation	+9.9%	−5.2% to +25.7%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	GARCH(1,1)-t	+12.3%	−4.4% to +32.1%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	Logistic regression	+10.5%	−17.3% to +43.4%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	Lognormal, 90-day vol	+16.2%	−7.9% to +44.3%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	Lognormal, EWMA vol	+12.8%	−14.0% to +43.9%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	Lognormal, EWMA vol + drift	+5.7%	−22.6% to +32.6%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	Random forest	+3.1%	−25.8% to +36.7%	✓	✗	✓	✓	not promoted	
spot	BTC	30d	XGBoost	−5.9%	−35.9% to +29.0%	✗	✗	✓	✓	not promoted	
spot	BTC	30d	XGBoost + DVOL features	−0.9%	−23.9% to +22.8%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	Chronos (price)	−13.0%	−23.2% to −3.1%	✗	✗	✓	✓	not promoted	must also pass in Stage 5 in-band: no
spot	ETH	30d	Chronos (volatility)	−3.3%	−11.6% to +4.9%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	Filtered historical simulation	−2.5%	−6.4% to +1.4%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	GARCH(1,1)-t	−8.5%	−22.1% to +4.0%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	Logistic regression	−8.6%	−20.8% to +3.6%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	Lognormal, 90-day vol	−1.8%	−11.6% to +6.9%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	Lognormal, EWMA vol	−6.0%	−15.6% to +2.3%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	Lognormal, EWMA vol + drift	−16.4%	−28.4% to −6.4%	✗	✗	✗	✓	not promoted	
spot	ETH	30d	Random forest	−8.8%	−22.0% to +3.9%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	XGBoost	−13.9%	−29.0% to +0.2%	✗	✗	✓	✓	not promoted	
spot	ETH	30d	XGBoost + DVOL features	−19.1%	−31.4% to −6.1%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	Chronos (price)	−11.1%	−18.3% to −4.6%	✗	✗	✓	✓	not promoted	must also pass in Stage 5 in-band: no
spot	LINK	30d	Chronos (volatility)	−0.2%	−13.5% to +15.9%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	Filtered historical simulation	−1.5%	−10.4% to +8.8%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	GARCH(1,1)-t	+3.1%	−5.6% to +15.7%	✓	✗	✓	✓	not promoted	
spot	LINK	30d	Logistic regression	−7.0%	−33.1% to +27.8%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	Lognormal, 90-day vol	−0.2%	−12.5% to +17.5%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	Lognormal, EWMA vol	+0.7%	−12.8% to +20.0%	✓	✗	✓	✓	not promoted	
spot	LINK	30d	Lognormal, EWMA vol + drift	−7.6%	−27.2% to +14.0%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	Random forest	−9.8%	−36.9% to +27.8%	✗	✗	✓	✓	not promoted	
spot	LINK	30d	XGBoost	−8.6%	−36.1% to +28.8%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	Chronos (price)	−9.3%	−15.0% to −2.6%	✗	✗	✓	✓	not promoted	must also pass in Stage 5 in-band: no
spot	XRP	30d	Chronos (volatility)	−5.2%	−12.8% to +3.9%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	Filtered historical simulation	−5.3%	−20.9% to +12.7%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	GARCH(1,1)-t	−16.2%	−40.9% to +0.4%	✗	✗	✗	✓	not promoted	
spot	XRP	30d	Logistic regression	−13.4%	−38.8% to +22.9%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	Lognormal, 90-day vol	−4.5%	−13.6% to +5.2%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	Lognormal, EWMA vol	−5.3%	−17.2% to +7.0%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	Lognormal, EWMA vol + drift	−15.7%	−33.0% to +2.6%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	Random forest	−10.0%	−32.0% to +21.5%	✗	✗	✓	✓	not promoted	
spot	XRP	30d	XGBoost	−13.5%	−36.4% to +18.7%	✗	✗	✓	✓	not promoted	

A challenger is promoted for an (asset, cohort, horizon) only if, on the holdout, (1) its mean Brier is lower than both climatology and the strongest development-period baseline for that cell, (2) the block-bootstrap 95% CI of its Brier difference vs climatology lies entirely below 0 and the verdict is seed-stable, (3) its reliability (mean [observed − forecast] over bins, weighted by count) is no worse than climatology's by more than 0.02, and (4) its coverage is at least 90% of climatology's. Otherwise the simpler baseline is retained and the report says no demonstrated improvement.

Clarifications recorded before opening: Made 2026-10-14 BEFORE the holdout was opened, after an independent pre-flight review that saw no holdout result: coverage (4) is measured on the cell's testable holdout targets; Chronos's price representation is scored on all rows and must also pass on its in-band subset; the strongest development baseline itself is judged against climatology only in (1); cells with no testable holdout targets are reported as untestable; the retained model is the strongest development baseline if it passes against climatology, else climatology.

Leaderboard: every model on one asset and horizon

Brier skill compares each model's probabilities with the outcomes, relative to the past event frequency on the same forecasts. 0% means no better than the base rate; positive is better; negative is worse. Bars show the 95% interval.

Selection printed: Holdout period · Spot closes (exchanges) · BTC · 30 days · History, Market-implied (Q-proxy), Machine learning, Foundation model · calibration: GARCH(1,1)-t. Change the filters on screen to print another cell.

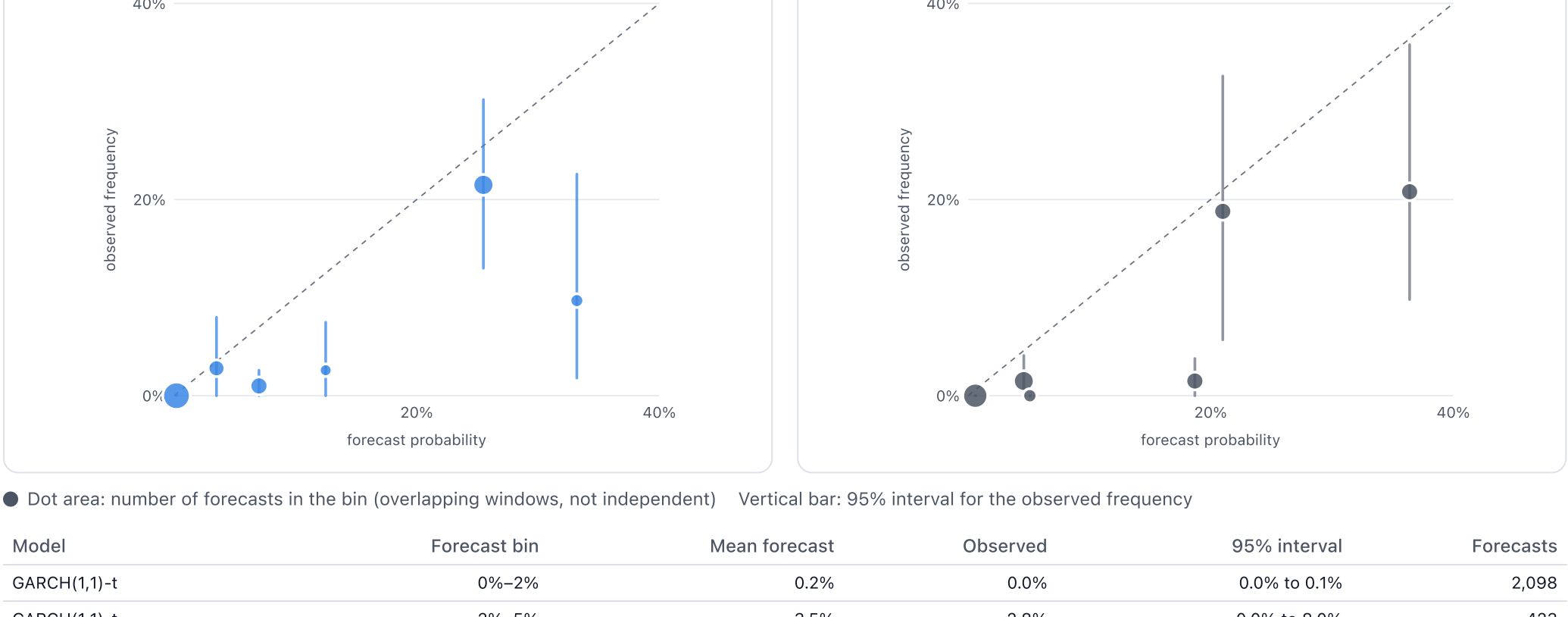


All models in this cell (table view of the chart)						
Model	Separate outcomes	Brier	Base-rate Brier	Skill	95% interval	Verdict
Lognormal, 90-day vol P · scored on P models	18	0.1592	0.1899	+16.2%	−7.9% to +44.3%	within noise
Lognormal, EWMA vol P · scored on P models	18	0.1656	0.1899	+12.8%	−14.0% to +43.9%	within noise
Lognormal, EWMA vol + drift P · scored on P models	18	0.1791	0.1899	+5.7%	−22.6% to +32.6%	within noise
Filtered historical simulation P · scored on P models	18	0.1711	0.1899	+9.9%	−5.2% to +25.7%	within noise
GARCH(1,1)-t P · scored on P models	18	0.1666	0.1899	+12.3%	−4.4% to +32.1%	within noise
DVOL implied vol Q-proxy · scored on P + Q-proxy	18	0.1639	0.1899	+13.7%	−9.3% to +38.8%	within noise
Logistic regression P · scored on Stage 4	18	0.1700	0.1899	+10.5%	−17.3% to +43.4%	within noise
Random forest P · scored on Stage 4	18	0.1840	0.1899	+3.1%	−25.8% to +36.7%	within noise
XGBoost P · scored on Stage 4	18	0.2012	0.1899	−5.9%	−35.9% to +29.0%	within noise
XGBoost + DVOL features P · scored on Stage 4 + DVOL	18	0.1915	0.1899	−0.9%	−23.9% to +22.8%	within noise
Chronos (price) P · scored on Stage 5	18	0.1782	0.1899	+6.2%	−13.0% to +31.1%	within noise
Chronos (volatility) P · scored on Stage 5	18	0.1639	0.1899	+13.7%	−10.3% to +44.0%	within noise

Scored on. All 5 comparison sets shown (P models, P + Q-proxy, Stage 4, Stage 4 + DVOL, Stage 5) cover the same 528 forecast dates from 2025-04-02 to 2026-09-11, 1 price targets, ≈18 independent outcomes. Within a comparison every model is scored on the same forecasts. ≈ independent outcomes counts non-overlapping 30-day windows: daily forecasts overlap, so this, not the number of forecast dates, sets how much the result can be trusted.

Calibration: do the probabilities mean what they say?

When a model says 20%, does the event happen about 20% of the time? Forecasts are grouped into bins; points on the dashed diagonal are perfectly calibrated. Uses the period, outcomes, asset and horizon selected above.



How this was tested

- Walk-forward, no future data.** Each forecast was made using only prices (and, where used, option-implied volatility) available on its date. Models that learn from outcomes were refitted only on outcomes that had already happened by the refit date.
- Same forecasts for every model in a comparison.** Skill is measured against the past event frequency on exactly the same dates and price targets, so differences are not caused by one model covering easier periods.
- Honest uncertainty.** Daily forecasts over 30– to 90-day windows overlap heavily, so intervals come from a block bootstrap (1000 resamples) and the count that matters is the number of non-overlapping windows, usually a few dozen. A verdict counts only if it holds across 5 bootstrap seeds.
- A sealed final test, opened once.** The most recent stretch of history (settlement outcomes from 2025-08-30, spot outcomes from 2025-04-02) was set aside before any model was built. It has been opened once, for all models together, and is logged.
- Promotion rule fixed in advance.** A model replaces the simpler baseline only if, on the holdout, it beats both the base rate and the strongest development baseline, the 95% interval for its improvement excludes zero, its calibration is no worse, and it forecasts nearly as many cases as the base rate. Otherwise the baseline stays.
- Independent reviews at every stage.** Each stage was reviewed by a separate reviewer before moving on. The reviews caught real bugs, which were fixed and re-run before any holdout result was seen.
- Market-implied is not a forecast.** The DVOL model uses option-implied volatility. Option prices include a risk premium, so it is shown as a Q-proxy reference, not as a real-world (P) forecast, and it is not eligible for promotion.