Chronos-2 on Synth

Drift-Corrected Price Paths and Live Benchmark Performance of a Fine-Tuned Foundation Model

SYNTH

Bittensor Subnet 50

July 8, 2026

Abstract

The Chronos2 Miner adapts a pretrained time-series foundation model (`amazon/chronos-2` [2, 1]) to probabilistic price-path forecasting for twelve financial assets, using LoRA fine-tuning [3] on a single consumer GPU. From the model's per-step predictive quantiles we generate a Monte-Carlo ensemble of price paths, and we correct a directional-drift tendency of the foundation forecaster at inference so that the central path is not a regime-dependent directional bet, while preserving the forecast's dispersion and tail shape. This note describes the approach at a high level, characterizes the emitted predictive distributions—volatility seasonality, single-step increment shape, cross-asset tail structure, and $\sqrt{t}$ horizon growth—and reports the deployed miner's current standing on the live Synth subnet (Section 4). Implementation-level details—including exact training settings, inference transformations, sampling rules, and serving heuristics—are intentionally omitted: the purpose of this note is to communicate design principles and observed behaviour rather than to provide a replication recipe.

1 System overview

The miner forecasts, for each of twelve assets, a full probability distribution over the asset's price path—not a single number, but 1000 plausible future trajectories that together express its uncertainty. The engine is `amazon/chronos-2` [2], a T5-inspired encoder-only time-series foundation model pretrained on a broad corpus, which we adapt to financial data with a lightweight LoRA adapter [3]. The subnet scores three separate competitions—crypto at a 1-hour horizon, crypto at a 24-hour horizon, and commodities & equities at a 24-hour horizon—which we serve with two models, one per horizon:

- A **24-hour model** (5-minute bars) covering the twelve assets (BTC, ETH, SOL, XRP, HYPE; SPYX, NVDAX, TSLAX, AAPLX, GOOGLX; XAU, WTIOIL), which serves the crypto (24-hour) and commodities & equities (24-hour) competitions.
- A **1-hour model** (1-minute bars) for the liquid crypto assets, which serves the crypto (1-hour) competition.

Each model is trained only on data that precedes its evaluation window, so the live results below are genuinely out-of-sample. All training and inference run on a single NVIDIA RTX 3070 (8 GB).

1.1 The Chronos-2 forecaster

Chronos-2 is a *time-series foundation model* [2, 1]: a single transformer trained once on a large, diverse corpus of series, and capable of being applied to new series *in context* without task-specific retraining; in this system we nevertheless add a lightweight LoRA adaptation to specialize it to financial returns. Given a window of recent observations (and, optionally, covariates), it directly outputs a *probabilistic* forecast: for every future step it emits a set of quantiles of the value at that step, rather than a single point estimate. Mechanically, it slices the input series into fixed-size patches, embeds them as tokens, processes them with an encoder-only transformer using group attention, and maps the resulting representations to multi-step quantile forecasts. Out of the box this yields a predictive band that already captures level, trend and volatility structure. The LoRA adapter is a low-rank weight update—a few megabytes on disk—so adaptation is cheap to iterate on and fits an 8 GB GPU. We forecast the per-bar log-return,

$$r_t = \ln(P_t/P_{t-1}) \,, \tag{1}$$

not the price level: returns are (near-)stationary, so horizon uncertainty grows naturally by cumulating per-step variance, and a fresh anchor price can be reattached at inference.

1.2 Inputs and feature collection

Alongside the return series, the model receives covariates chosen to expose forecast-relevant signal:

- **Recent realized volatility** — measures of how volatile the asset has just been, including one built from each bar's full open/high/low/close range [4] rather than the close-to-close return alone. Recent volatility is the strongest predictor of near-future dispersion, so it informs the width of the predicted band.
- **Volume** — included for assets where a reliable volume feed exists and omitted otherwise.
- **Calendar / session context** — deterministic functions of the timestamp (time-of-day, day-of-week, and trading-session indicators). Because tokenized equities and commodities are far more volatile during their trading session than overnight (Section 3), these features let the model learn strongly session-dependent behavior. Being deterministic and available ahead of time, they are supplied for *both* the observed window and the future horizon.

2 Approach

Drift correction. A foundation forecaster fed recent returns tends to extrapolate the recent trend as a persistent directional drift. On the 24-hour crypto horizon this surfaces as a systematic short tilt in the central path—a directional bet that profits in sustained trends and is penalized on reversals. We remove this tendency at inference, so the central path carries no built-in directional bet, while preserving the forecast's dispersion and tail shape.

Path ensemble. From the per-step predictive quantiles the model emits, we generate a Monte-Carlo ensemble of 1000 sample paths—anchored on the last observed price and accumulating per-step uncertainty over the horizon—which is the object scored by the subnet.

3 Output-distribution characterization

The figures below describe the shape of the distributions the model emits, measured over a recent set of prompts (∼200 per asset, 1000 paths each). They are properties of the emitted per-step marginals and of the sampled ensemble; no value here is scored against a realized outcome.

3.1 Volatility seasonality

Session-bound assets (tokenized stocks and WTIOIL) show a sharp predicted-σ spike at the NYSE open ($\sim$14:30 UTC) and elevated US-session volatility that collapses overnight; crypto is far flatter around the clock (Figure 1). This behaviour is consistent with the information supplied by the calendar/session covariates.

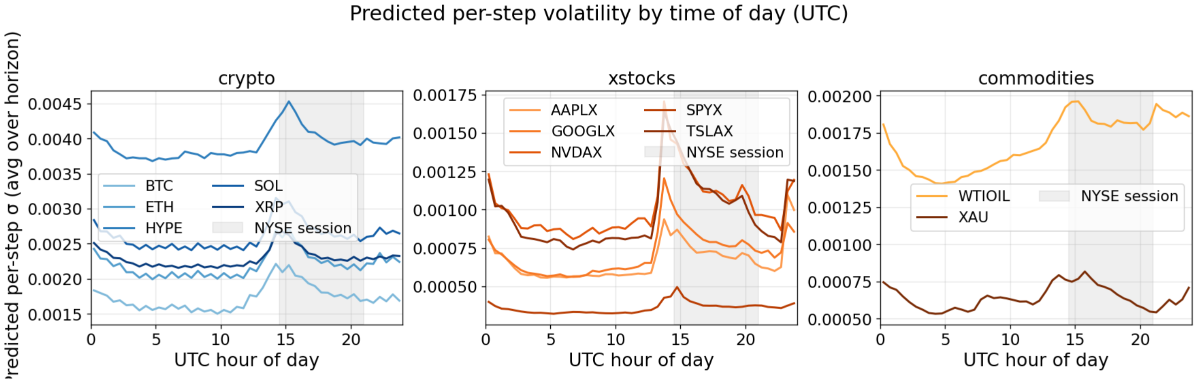

Figure 1: Predicted per-step σ vs. UTC hour, faceted by class; NYSE session shaded (σ averaged over forecast steps and prompts within each 30-min UTC bucket). Strong intraday seasonality for session-bound assets; crypto far flatter.

3.2 Increment shape

Pooling every sampled single-step log-return for an asset (all forecast steps, all paths, all prompts) and standardizing the pooled sample to unit variance, the result is leptokurtic relative to a matched Gaussian (Figure 2): pooled excess kurtosis rises from BTC $\approx$ 0.7 to WTIOIL $\approx$ 2.8 to SPYX $\approx$ 5.0. Two caveats apply. First, the sampled increments are bounded by the model's predictive quantile range, so this is not an unbounded-tail claim. Second, pooling steps of differing conditional variance itself inflates excess kurtosis, so these numbers over-state the per-step *conditional* kurtosis. The ordering across assets—crypto nearest Gaussian, tokenized stocks fattest—is nonetheless a genuine, asset-appropriate difference in emitted shape rather than one shared distribution.

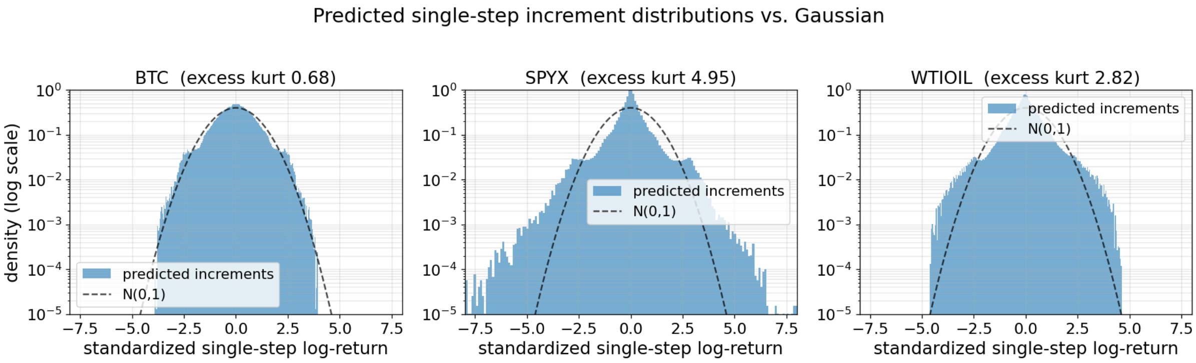

Figure 2: Pooled single-step log-returns per asset (all forecast steps and paths), standardized to unit variance, log-density, with a matched $\mathcal{N}(0,1)$, for BTC, SPYX, WTIOIL. The leptokurtosis shown is partly induced by pooling steps of differing conditional variance (see text).

3.3 Cross-asset tail structure

Table 1 reports per-asset shape statistics. Crypto sits nearest Gaussian (tail ratio ≈ 2.2–2.4 vs. the Gaussian 1.82); the tokenized stocks are fatter-tailed, with SPYX the fattest (3.87, ∼2.1× Gaussian) and carrying the most upside skew (asymmetry 1.05).

Table 1: Distributional shape statistics (24-hour model). Per-step σ in basis points of log-return; tail and asymmetry ratios computed from the sampled ensemble, per step then averaged over steps and prompts. Gaussian reference: tail ratio = 1.82, symmetric asymmetry = 1.0.

Asset	σ_{step1} (bps)	σ_{final} (bps)	$(q_{99}-q_{01})/(q_{90}-q_{10})$	$(q_{90}-q_{50})/(q_{50}-q_{10})$
AAPLX	7.55	7.41	2.81	1.01
BTC	17.20	17.95	2.32	0.99
ETH	21.89	23.19	2.40	1.00
GOOGLX	7.74	7.69	2.87	0.99
HYPE	38.13	40.14	2.19	0.99
NVDAX	11.12	11.01	2.95	1.02
SOL	26.13	27.22	2.32	0.99
SPYX	3.97	3.88	3.87	1.05
TSLAX	10.88	10.24	2.98	1.01
WTIOIL	16.80	18.09	2.60	1.01
XAU	6.66	6.35	2.70	1.00
XRP	22.92	23.92	2.28	1.00

3.4 Horizon growth

Predicted cumulative-return σ grows approximately as $\sqrt{t}$ (Figure 3), with curves near-parallel to a random-walk reference on log–log axes. This reflects how the path ensemble accumulates per-step uncertainty over the horizon, and is a property of the path-construction step rather than of the per-step marginals.

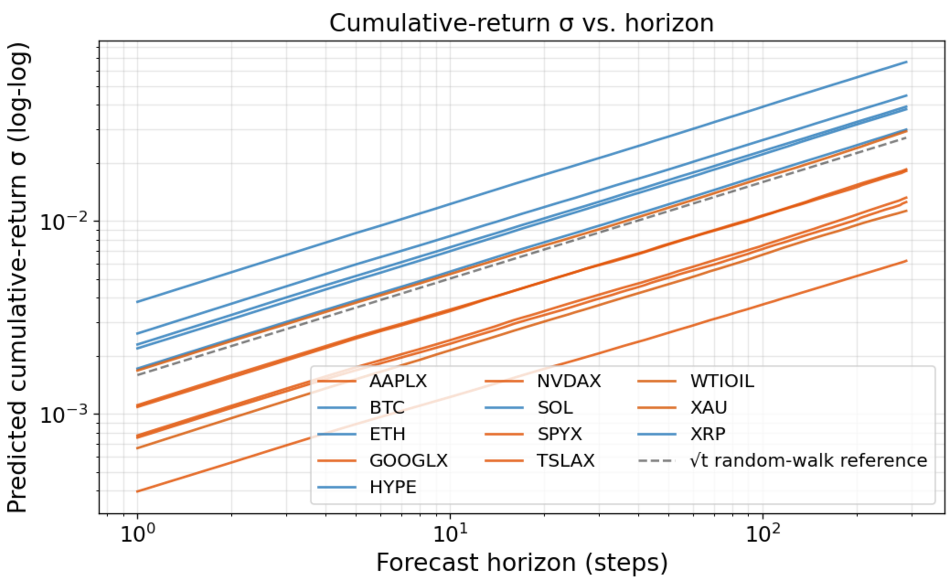

Figure 3: Predicted cumulative-return σ vs. horizon (log–log, 12 assets) with a $\sqrt{t}$ reference.

4 Results: live-subnet standing

The miner runs live on Bittensor Subnet 50 (Synth) [6, 7], competing against a field of roughly 256 miners. All figures below are as of 2026-07-06 and use validator-recorded CRPS [5] (each miner's realized score, with non-responses penalized at the field's 90th percentile); no number is self-reported. Ranks are a time-stamped operational snapshot: they move with the settlement window and the participating field. By on-chain incentive, the subnet's aggregate reward signal, the miner ranks **36th**.

The subnet scores three separate competitions; over each competition's current settlement window the miner places:

- **Crypto, 24-hour horizon — 12th.** Per asset it ranks 1st on XRP, 14th on ETH, and 19th on HYPE.
- **Crypto, 1-hour horizon — 99th** (mid-field).
- **Commodities & equities, 24-hour horizon — 86th**, over the commodity and equity assets it currently covers; per asset it ranks 1st on both AAPLX and WTIOIL, 41st on gold (XAU), and 47th on GOOGLX.

These results are consistent with the design: the drift correction (Section 2) is aimed at the crypto horizons, where an uncorrected directional tilt is most costly, and the calendar/session features (Section 1.2) at the equity and commodity assets, whose volatility is strongly session-bound. The standing is observational and does not isolate the contribution of any single component.

5 Serving

Because the 24-hour-absolute scoring interval evaluates absolute price levels, serving takes care to anchor each forecast on fresh data despite live-feed lag, and to answer every request within the validator's deadline under bursty load.

6 Limitations and future directions

This note describes the emitted forecast distributions and the deployed miner's standing; it is an overview, not a full predictive evaluation. A fuller evaluation would add marginal interval coverage and calibration diagnostics and comparisons against simple baselines. Two directions would deepen the characterization. First, the emitted interval distributions (Section 3) could be compared directly against realized outcomes under stratifications of the input parameters (e.g. by volatility regime, session, or asset class), to test where the predicted shape matches reality and where it does not. Second, the $\sqrt{t}$ horizon-growth curve invites a closer look at its *deviations*: departures from the straight line on log–log axes are, in effect, an estimate of a Hurst-like scaling exponent, and comparing the model's implied scaling to the realized scaling of each asset would show whether the predicted diffusion rate is too fast, too slow, or regime-dependent.

References

[1] A. F. Ansari et al., "Chronos: Learning the Language of Time Series," *Transactions on Machine Learning Research*, 2024. arXiv:2403.07815.

[2] A. F. Ansari et al., "Chronos-2: From Univariate to Universal Forecasting," 2025. arXiv:2510.15821. Model card: `https://huggingface.co/amazon/chronos-2`.

[3] E. J. Hu et al., “LoRA: Low-Rank Adaptation of Large Language Models,” *International Conference on Learning Representations (ICLR)*, 2022. arXiv:2106.09685.

[4] M. B. Garman and M. J. Klass, “On the Estimation of Security Price Volatilities from Historical Data,” *Journal of Business*, 53(1):67–78, 1980.

[5] T. Gneiting and A. E. Raftery, “Strictly Proper Scoring Rules, Prediction, and Estimation,” *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[6] Opentensor Foundation, “Bittensor: A Peer-to-Peer Intelligence Market,” whitepaper, `https://bittensor.com`.

[7] Synth, “Synth Documentation,” Bittensor Subnet 50, accessed July 2026. `https://docs.synthdata.co`.