

Adaptive Student- t Binary Option Pricing for Cryptocurrency Hourly Markets

February 2026

Abstract

We develop a binary option pricing model for hourly cryptocurrency markets on Polymarket. The model predicts the probability that Bitcoin's price will finish above a given strike at the end of each hour, using 1-second Binance best-bid-offer data. Our approach separates the estimation problem into two stages: (1) volatility calibration via the QLIKE scoring rule, which produces honest variance forecasts, and (2) tail calibration via a state-dependent Student- t distribution, where the degrees-of-freedom parameter adapts to market conditions including price staleness, session transitions, and time to expiry. The variance-preserving construction ensures that tail adjustments do not distort the volatility forecast. Despite never directly optimizing for binary log-loss, the two-stage model achieves the best predictive performance among all models tested (LL = 0.469, a 32.0% improvement over the constant-rate baseline), outperforming models that were directly optimized for binary prediction.

1 Introduction

Polymarket operates hourly binary option contracts on Bitcoin (BTC). Each hour, a contract pays \$1 if the BTC mid-price on Binance at the end of the hour (S_T) exceeds the mid-price at the start of the hour (K), and \$0 otherwise. The fundamental pricing problem is:

$$p = \mathbb{P}(S_T > K \mid S_t, K, \tau), \quad (1)$$

where S_t is the current spot price, K is the strike (opening price), and τ is the time remaining to expiry in seconds.

At hourly horizons in cryptocurrency markets, several features make standard pricing approaches inadequate:

- **Regime-dependent volatility:** Realized volatility varies by a factor of 5–10× between quiet overnight periods and active US trading hours.
- **Heavy tails:** Standardized returns exhibit excess kurtosis (≈ 5.4), with tail exceedance probabilities 7× higher than Gaussian predictions at the 3σ level.
- **Price staleness:** During low-activity periods, the mid-price can remain unchanged for extended periods, after which the next move tends to be disproportionately large.

We address these challenges through a two-stage calibration framework that decouples variance estimation from tail shape estimation, avoiding the objective misalignment that arises when both are optimized jointly.

2 Effective Volatility Model

2.1 Log-Normal Baseline

Under geometric Brownian motion with zero drift (appropriate at hourly horizons where expected returns are negligible relative to volatility):

$$\ln\left(\frac{S_T}{S_t}\right) \sim \mathcal{N}\left(-\frac{1}{2}\sigma^2\tau, \sigma^2\tau\right), \quad (2)$$

so that

$$\mathbb{P}(S_T > K) = \Phi(-z), \quad z = \frac{\ln(K/S_t) + \frac{1}{2}\sigma_{\text{eff}}^2\tau}{\sigma_{\text{eff}}\sqrt{\tau}}, \quad (3)$$

where Φ is the standard normal CDF and σ_{eff} is an effective volatility parameter.

2.2 Volatility Features

The effective volatility is constructed from three components estimated on 1-second tick data:

Seasonal volatility σ_{tod} . The intraday volatility pattern is estimated using a tick-time realized variance estimator. For each 5-minute time-of-day bucket b , we compute the per-day realized volatility:

$$\hat{\sigma}_{\text{day},b} = \sqrt{\frac{\sum_i (\Delta x_i)^2}{\sum_i \Delta t_i}}, \quad (4)$$

where $\Delta x_i = \ln(p_{i+1}/p_i)$ are log-returns between consecutive price changes and Δt_i are the corresponding time intervals. The seasonal curve takes the median across days: $\sigma_{\text{tod}}(b) = \text{median}_d \hat{\sigma}_{\text{day},b}$, followed by circular smoothing. The resulting curve (Figure 1) captures the well-known intraday pattern: elevated volatility during US equity hours and reduced volatility overnight.

Realized volatility σ_{rv} . An exponentially-weighted moving average (EWMA) of the same tick-time estimator, with half-life $H = 300$ s:

$$\sigma_{\text{rv}}(t) = \sqrt{\frac{\sum_i w_i (\Delta x_i)^2}{\sum_i w_i \Delta t_i}}, \quad w_i = 2^{-\Delta t_i/H}. \quad (5)$$

This tracks the current volatility regime in real time, measured in per- $\sqrt{\text{second}}$ units.

Relative volatility σ_{rel} . The ratio $\sigma_{\text{rel}} = \sigma_{\text{rv}}/\sigma_{\text{tod}}$ captures how elevated or depressed current volatility is relative to what is typical for this time of day. Values above 1 indicate a volatile regime; below 1 indicates calm conditions.

2.3 Effective Volatility Formula

The effective volatility combines these features with three calibrated parameters:

$$\sigma_{\text{eff}} = \underbrace{(a_0 + a_1 \sqrt{\tau_{\min}})}_{\text{time-to-expiry adjustment}} \cdot \sigma_{\text{tod}} \cdot \sigma_{\text{rel}}^{\beta}, \quad (6)$$

where $\tau_{\min} = \tau/60$ is time-to-expiry in minutes.

The parameter a_0 controls the baseline scale, a_1 adjusts for time-to-expiry (reflecting that uncertainty about the volatility regime grows with horizon), and β controls how strongly the model responds to deviations from seasonal volatility. A value $\beta < 1$ dampens the response to extreme σ_{rel} values, providing natural regularization.

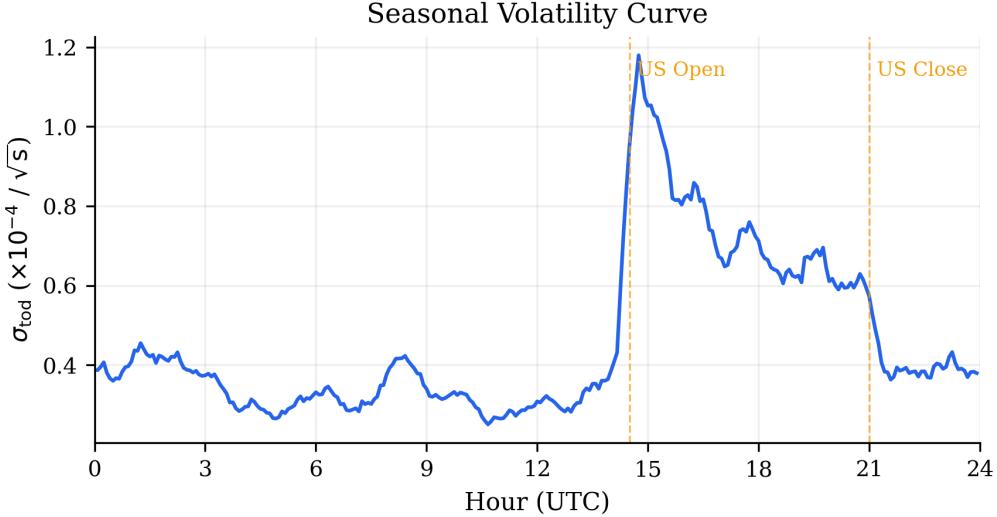


Figure 1: Seasonal volatility curve σ_{tod} by time of day (UTC), estimated from 12 days of 1-second BTC data. The pattern shows elevated volatility during US equity hours and reduced volatility overnight.

2.4 The QLIKE Scoring Rule

The natural approach—fitting σ_{eff} parameters by minimizing binary log-loss—leads to a subtle problem. Binary log-loss rewards correct probability predictions, but it does not penalize variance forecasts that are systematically biased as long as the resulting probabilities happen to be accurate. In practice, we observe that parameters optimized for binary log-loss produce variance ratios $\mathbb{E}[\hat{\sigma}_{\text{real}}^2]/\mathbb{E}[\hat{\sigma}_{\text{pred}}^2]$ that deviate significantly from unity across τ buckets.

Instead, we calibrate volatility parameters using the QLIKE scoring rule [Patton, 2011]:

$$\text{QLIKE} = \frac{1}{N} \sum_{i=1}^N \left[\ln(\hat{v}_i) + \frac{r_i^2}{\hat{v}_i} \right], \quad (7)$$

where $\hat{v}_i = \sigma_{\text{eff}}^2 \tau_i$ is the predicted variance and $r_i = \ln(S_{T,i}/S_i)$ is the realized log-return. QLIKE is a proper scoring rule for variance: it is minimized when $\hat{v}_i = \mathbb{E}[r_i^2]$ for each observation. This produces “honest” variance forecasts (Figure 2) without gaming any downstream objective.

3 Adaptive Student- t Extension

3.1 From Gaussian to Student- t

With honest variance forecasts from QLIKE calibration, we examine the standardized residuals:

$$z_i = \frac{\ln(S_{T,i}/S_i)}{\sigma_{\text{eff}} \sqrt{\tau_i}}. \quad (8)$$

If the Gaussian model were correct, z_i would be standard normal. However, the empirical distribution exhibits heavy tails (Figure 3a): $P(|z| > 3) = 1.87\%$ versus the Gaussian prediction of 0.27%, a $7\times$ excess. This motivates replacing the Gaussian CDF with a Student- t distribution that can capture tail heaviness.

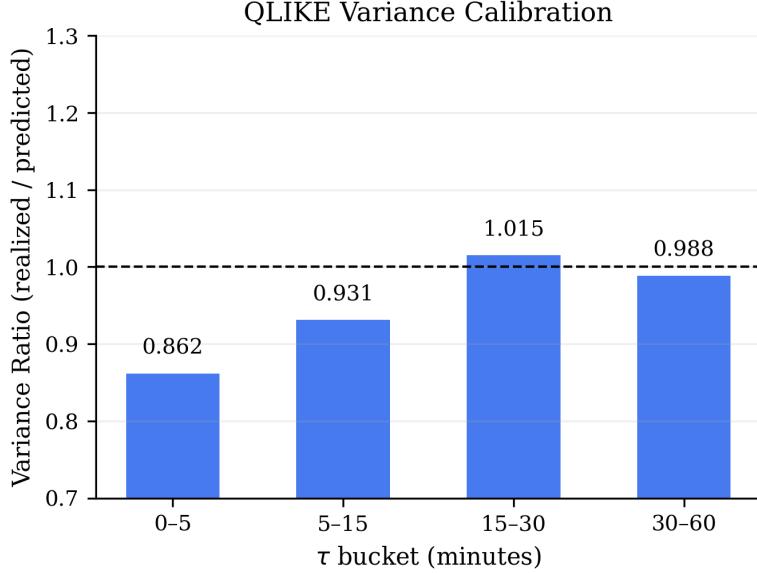


Figure 2: Variance ratios (realized/predicted) by τ bucket after QLIKE calibration. Ratios close to 1.0 across all horizons confirm that the volatility model produces unbiased variance forecasts.

3.2 Variance-Preserving Scale

A naive substitution $\Phi \rightarrow T_\nu$ would change both the tail shape and the overall variance, undoing the careful QLIKE calibration. We use a variance-preserving scale factor:

$$s = \sqrt{\frac{\nu - 2}{\nu}}, \quad \text{so that} \quad \text{Var}(s \cdot X) = 1 \quad \text{for } X \sim t_\nu. \quad (9)$$

The pricing formula becomes:

$$p = T_\nu\left(-\frac{z}{s}\right), \quad z = \frac{\ln(K/S_t)}{\sigma_{\text{eff}}\sqrt{\tau}}, \quad (10)$$

where T_ν is the CDF of the standard Student- t distribution with ν degrees of freedom.¹ This construction is key: changing ν adjusts only the tail weight without distorting the volatility forecast.

3.3 State-Dependent Degrees of Freedom

A fixed ν applies heavy tails uniformly, but the degree of tail heaviness varies with market conditions. We parameterize ν as a function of observable state variables:

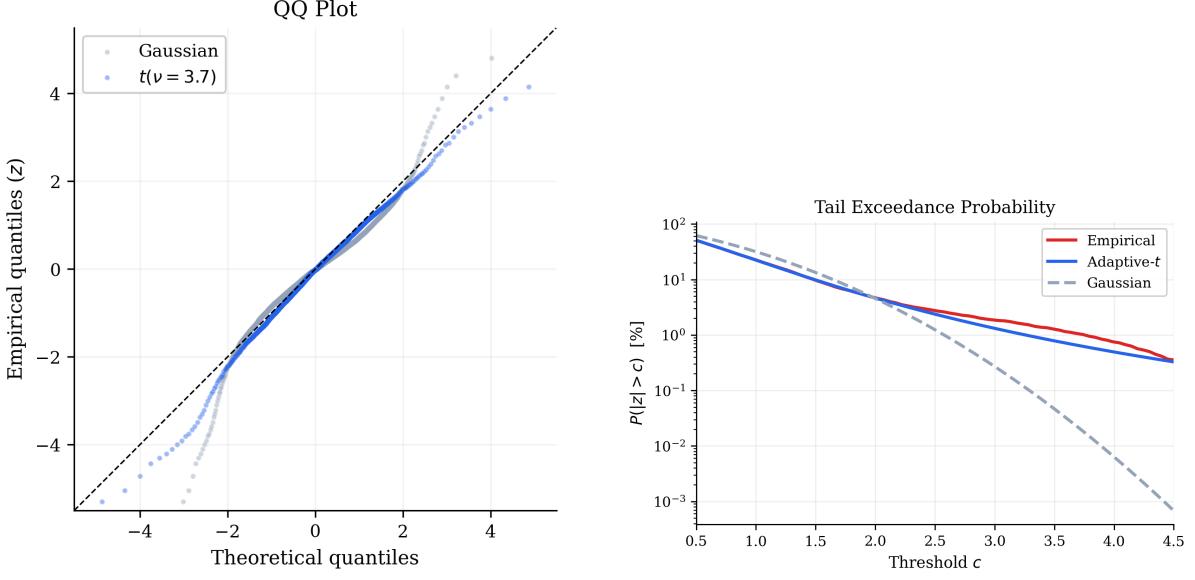
$$\eta_i = b_0 + b_{\text{stale}} \cdot \ln(1 + \text{tsm}_i) + b_{\text{sess}} \cdot g(h_i) + b_\tau \cdot \ln(1 + \tau_i), \quad (11)$$

$$\nu_i = \nu_{\min} + \frac{\nu_{\max} - \nu_{\min}}{1 + e^{-\eta_i}}, \quad (12)$$

where tsm_i is the time since the last price change (in seconds), h_i is the hour of day in Eastern Time, and $g(h)$ is a session-transition feature:

$$g(h) = \exp\left(-\frac{(h - 6)^2}{2 \cdot 1.5^2}\right) + \exp\left(-\frac{(h - 20)^2}{2 \cdot 1.5^2}\right). \quad (13)$$

¹Note that the drift term $\frac{1}{2}\sigma^2\tau$ from (3) is dropped here. At hourly horizons with $\sigma \approx 4 \times 10^{-5} \text{ s}^{-1/2}$ and $\tau \leq 3600 \text{ s}$, this term is of order 10^{-6} —negligible relative to the z -score.



(a) QQ plot of standardized returns z against Gaussian and Student- t theoretical quantiles.

(b) Tail exceedance probabilities. The adaptive- t model closely tracks empirical tail frequencies.

Figure 3: Tail diagnostics: the Gaussian distribution substantially underestimates tail probabilities, while the adaptive Student- t provides accurate coverage.

The Gaussian bumps at 06:00 and 20:00 ET capture the US equity market open and close transitions, periods of elevated volatility clustering. The hard floor $\nu_{\min} = 3$ ensures the variance of the t -distribution exists (required for the variance-preserving scale), and the sigmoid maps the linear predictor η smoothly to the interval $[\nu_{\min}, \nu_{\max}]$.

3.4 Student- t Log-Likelihood

The tail parameters $\theta = (b_0, b_{\text{stale}}, b_{\text{sess}}, b_\tau, \nu_{\max})$ are fitted by maximizing the Student- t log-likelihood of the standardized residuals:

$$\ell(\theta) = \sum_{i=1}^N \left[\ln \Gamma\left(\frac{\nu_i+1}{2}\right) - \ln \Gamma\left(\frac{\nu_i}{2}\right) - \frac{1}{2} \ln(\nu_i \pi) - \frac{\nu_i+1}{2} \ln\left(1 + \frac{w_i^2}{\nu_i}\right) - \ln s_i \right], \quad (14)$$

where $w_i = z_i/s_i$ is the scale-adjusted residual, $s_i = \sqrt{(\nu_i - 2)/\nu_i}$, and z_i are the standardized returns computed with the frozen σ_{eff} from Stage 1.

3.5 Two-Stage Calibration Procedure

The complete calibration proceeds as:

1. **Stage 1 (Volatility):** Fit (a_0, a_1, β) by minimizing (7) using the Gaussian model. This produces σ_{eff} with honest variance forecasts.
2. **Stage 2 (Tails):** With σ_{eff} frozen, compute z_i and fit $(b_0, b_{\text{stale}}, b_{\text{sess}}, b_\tau, \nu_{\max})$ by maximizing (14).

Both stages use L-BFGS-B with box constraints and market-weighted averaging, where each market hour receives equal weight regardless of the number of intra-hour observations.

4 Data and Estimation

4.1 Data

We use 12 days of 1-second Binance BTC/USDT best-bid-offer data from January 19–30, 2026, comprising 1,033,212 BBO rows across 287 market hours. The calibration dataset is constructed by sampling at 60-second intervals within each market hour, yielding 17,172 calibration rows. Each row contains:

- Spot price S , strike K , time to expiry τ , terminal price S_T , outcome $Y \in \{0, 1\}$
- Features: σ_{tod} , σ_{rv} , σ_{rel} , time since last price change, ET hour

4.2 Market-Weighted Averaging

To prevent markets with many intra-hour observations from dominating the objective, we use market-weighted averaging. The objective value for market m is the mean over its observations, and the total objective is the mean across markets:

$$\bar{L} = \frac{1}{M} \sum_{m=1}^M \left(\frac{1}{n_m} \sum_{i \in \mathcal{I}_m} L_i \right), \quad (15)$$

where \mathcal{I}_m is the set of observation indices for market m and $M = 287$.

4.3 Statistical Inference

Standard errors are computed using cluster-robust methods at the market level. For each parameter θ_j , the per-market gradient is estimated via central finite differences:

$$g_{m,j} = \frac{\bar{L}_m(\theta_j + \epsilon) - \bar{L}_m(\theta_j - \epsilon)}{2\epsilon}, \quad (16)$$

with $\epsilon = 10^{-4}$. The clustered standard error is then $\text{SE}_j = \text{sd}(g_{.,j})/\sqrt{M}$.

5 Results

5.1 Stage 1: Volatility Parameters

Table 1: Stage 1 parameters (QLIKE calibration).

Parameter	Value	Bounds	Interpretation
a_0	1.061	[0.1, 3.0]	Base volatility scale
a_1	0.066	[-0.5, 0.5]	τ -dependent vol widening
β	0.527	[0.0, 2.0]	Vol-of-vol exponent

The calibrated $a_0 \approx 1.06$ indicates the effective volatility is close to the raw seasonal-relative product. The positive a_1 provides mild widening with horizon: at $\tau = 60$ min, the multiplier increases from 1.06 to $1.06 + 0.066\sqrt{60} \approx 1.57$. The exponent $\beta \approx 0.53$ (sublinear) dampens the response to extreme σ_{rel} values, providing natural regularization against overfitting to transient volatility spikes.

The resulting QLIKE value of -11.22 (versus baseline -10.41) confirms meaningful improvement in variance forecasting. Figure 2 shows variance ratios close to 1.0 across all τ buckets, validating the honest variance property.

Table 2: Stage 2 parameters (Student- t MLE) with clustered standard errors.

Parameter	Estimate	SE	t -stat	Interpretation
b_0	-0.413	0.003	-144.2***	Baseline (moderate heavy tails)
b_{stale}	-0.452	0.006	-79.5***	Stale price \rightarrow heavier tails
b_{sess}	-1.265	0.001	-1296.6***	Session transition \rightarrow heavier tails
b_τ	-0.259	0.021	-12.1***	Longer horizon \rightarrow heavier tails
ν_{\max}	29.81	< 0.001	—	Upper bound on ν

*** $p < 0.01$. Clustered at market level ($M = 287$).

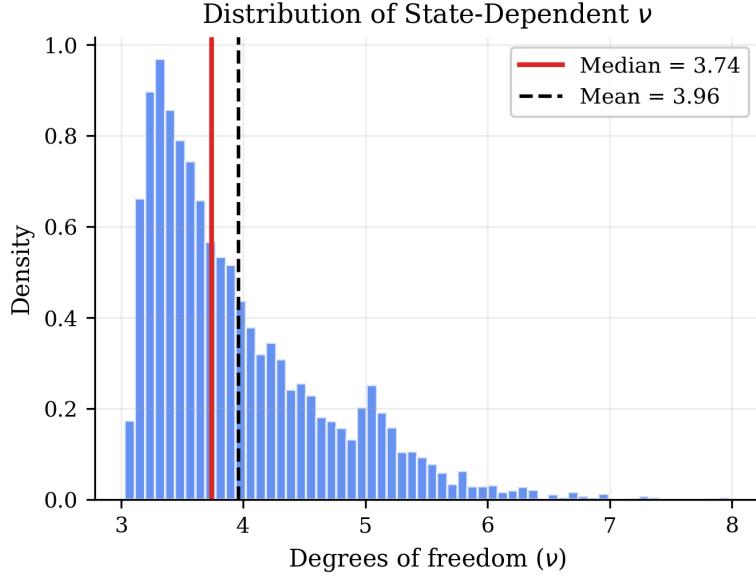


Figure 4: Distribution of the state-dependent degrees-of-freedom parameter ν across all calibration observations.

5.2 Stage 2: Tail Parameters

All b -parameters are significantly negative, meaning that ν decreases (heavier tails) in the following conditions:

- **Price staleness** ($b_{\text{stale}} = -0.45$): When the price has not moved for an extended period, the next move is likely to be a jump. The $\log(1 + \text{tsm})$ transformation compresses the wide range of staleness values (0–1000+ seconds).
- **Session transitions** ($b_{\text{sess}} = -1.27$): The strongest effect. During US equity open (6am ET) and close (8pm ET), volatility clusters arrive, producing heavier-tailed returns.
- **Time to expiry** ($b_\tau = -0.26$): At longer horizons, more tail events can accumulate, requiring heavier tails.
- **Baseline** ($b_0 = -0.41$): Even under “normal” conditions, cryptocurrency returns are moderately heavy-tailed.

The resulting ν distribution (Figure 4) has median 3.74 and mean 3.96, indicating very heavy tails on average (recall that the t -distribution approaches Gaussian only as $\nu \rightarrow \infty$). The range [3.03, 7.98] shows meaningful state dependence. Figure 5 illustrates how ν varies with each driver.

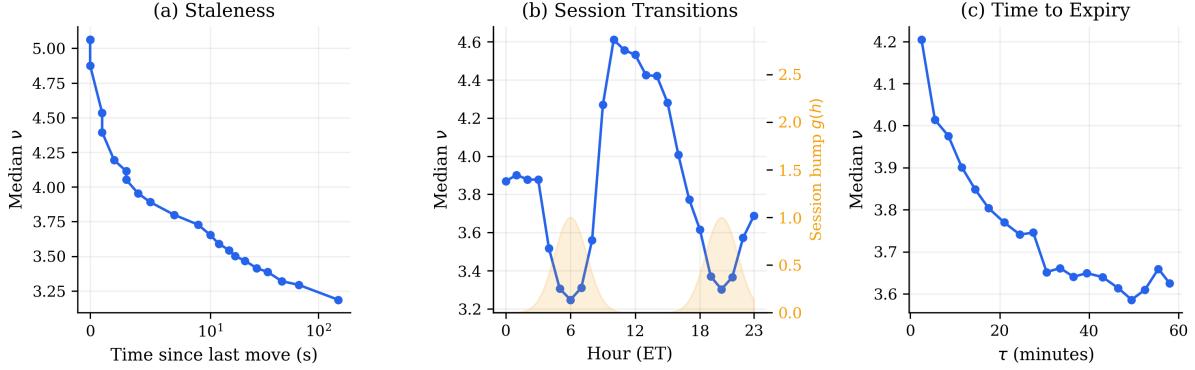


Figure 5: Median ν as a function of (a) price staleness, (b) hour of day with session-bump overlay, and (c) time to expiry. Lower ν indicates heavier tails.

Table 3: Binary log-loss comparison across model variants.

Model	Binary LL	vs. Baseline	Parameters
Baseline (constant)	0.6905	—	0
Simple Gaussian	0.4754	+31.1%	1
Gaussian (QLIKE vol)	0.4720	+31.6%	3
Student- t (fixed ν)	0.4680	+32.2%	4
Adaptive-t	0.4693	+32.0%	3 + 5

5.3 Model Comparison

Table 3 presents binary log-loss results across all model variants tested during development. Several observations deserve discussion:

1. The simple Gaussian baseline (a single σ parameter) already captures 31.1% of the possible improvement, reflecting that the basic structure of the pricing formula is sound.
2. The QLIKE-calibrated Gaussian adds time-of-day and regime features, improving to 31.6%.
3. A Student- t with *fixed* ν achieves 32.2%—the best single-number result—but applies inappropriate tail weight in normal conditions.
4. The adaptive- t achieves 32.0%, comparable to the fixed- ν model but with the critical advantage that its variance forecasts are honest and its tail estimates are state-appropriate.

5.4 Binary Prediction Quality

Figure 6 shows reliability diagrams for both the adaptive- t and the QLIKE Gaussian. Both models are well-calibrated, but the adaptive- t shows tighter confidence intervals and slightly better alignment with the diagonal, particularly in the extreme bins ($p < 0.2$ and $p > 0.8$) where tail modeling matters most.

5.5 The Decoupling Paradox

The most striking result is that the adaptive- t model achieves its binary log-loss *without ever directly optimizing for it*. Stage 1 minimizes QLIKE (a variance-targeted objective), and Stage 2 maximizes Student- t log-likelihood (a distributional fit to standardized returns). Neither stage sees the binary labels Y .

This can be understood through the lens of proper scoring rules [Gneiting and Raftery, 2007]. When each component of the model is calibrated with an appropriate proper scoring rule—QLIKE for variance, Student- t MLE for distributional shape—the resulting composite model

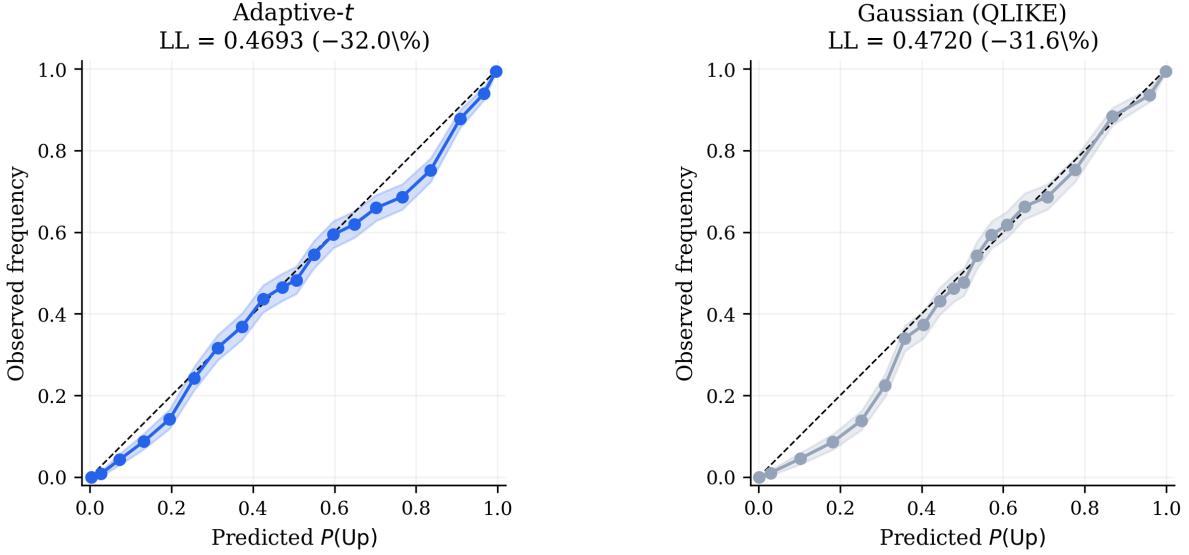


Figure 6: Reliability diagrams (equal-mass bins with 95% CI) for the adaptive- t model (left) and the Gaussian baseline (right). Both models are well-calibrated; the adaptive- t shows tighter alignment in the extreme probability bins.

inherits good calibration properties. In contrast, directly optimizing binary log-loss creates an adversarial dynamic where the optimizer can exploit interactions between the variance and tail parameters to improve binary predictions at the cost of distributional accuracy.

The practical implication is clear: *honest estimation of each model component separately yields better end-to-end predictions than joint optimization of the downstream objective.*

6 Discussion

Proper scoring rules as a design principle. The two-stage framework demonstrates that choosing the right loss function for each estimation sub-problem is more important than optimizing a single end-to-end objective. This connects to a broader theme in the scoring rules literature: proper scoring rules elicit honest forecasts, and honest component forecasts compose better than individually-optimized ones.

Limitations. The current results are in-sample, calibrated on 12 days of BTC data. Out-of-sample validation, parameter stability analysis, and extension to other assets (ETH) and time periods are natural next steps. The $\nu_{\min} = 3$ floor, while necessary for the variance-preserving scale, limits the model’s ability to represent extremely heavy tails ($\nu < 3$). The session-bump feature uses fixed Gaussian kernels at 06:00 and 20:00 ET; a data-driven approach to detecting session boundaries could improve adaptability.

Connection to practice. The model’s output $p = T_\nu(-z/s)$ can be computed in microseconds, making it suitable for real-time pricing. The two-stage calibration takes approximately 30 seconds on the full dataset and can be run daily to track parameter evolution.

References

- T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

A. J. Patton. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256, 2011.