Unlocking Liquidity on Prediction Markets

William Flanders[*‡] Stephen Flanders[†‡]

Lattica Finance[‡]

March 2026

Abstract

In this paper, we design and calibrate a shortfall pricing risk engine that enables solvent lending against collateralized prediction market positions. Prediction markets have experienced breakout success, yet locked capital and bounded returns leave substantial value on the table. Freeing this capital would strengthen informational quality by letting informed traders redeploy across markets, improve execution and retail profitability through deeper books, and open the door to institutional derivative products. However, because positions are bounded on $[0, 1]$, reprice discontinuously, and exhibit especially heavy tails, solvency cannot be guaranteed from liquidation alone, rendering existing decentralized credit solutions unviable. We take a different approach by building an engine, calibrated on historical Polymarket data, that generates an at-origination-time loss distribution from market state features, enforces fixed-duration epochs, and prices lender shortfall ex ante through a Wang distortion. In a dual-pool Monte Carlo simulation on held-out market data, the risk engine pool earned a 92.5% annualized return versus -27.2% for a flat-rate baseline, limited max drawdown to 3.6% versus 85%, originated roughly twelve times as many loans, and halved the liquidation rate, with state-contingent premiums ranging from 63 bps to 4,000 bps per epoch under the most conservative lending parameters.

[*] william@lattica.fi
[†] stephen@lattica.fi
[‡] https://lattica.fi

1 Introduction

Prediction markets have emerged as powerful belief engines, turning monetary instruments into positions that innately carry predictive power (Hanson, 2003; Arrow et al., 2008). Their impact continues to grow as these platforms expand into increasingly niche subjects and broader cultural domains. For instance, their effect on political science and prevailing polling theory has already begun to reshape the discipline in the wake of the 2024 U.S. presidential election (Clinton and Huang, 2025).

As demand for trading these positions has skyrocketed, so has the cost of capital lockup. A trader who enters a position in a long-horizon market ties up cash in an asset with capped upside, no interim yield, and resolution risk. The longer the horizon, the larger the opportunity cost of holding that position rather than redeploying capital elsewhere. This pushes liquidity toward short-dated events, or simply off the platform, and leaves many longer-dated books thin. A credit layer that lets traders borrow against locked positions would address this directly, and the case for building one extends well beyond simple capital efficiency.

First, letting informed traders unlock and redeploy capital allows them to express and update views more frequently across more markets, strengthening the informational quality that makes prediction markets valuable (Wolfers and Zitzewitz, 2004). Second, thin books worsen execution for all participants; Della Vedova (2026) show that retail traders who correctly forecast outcomes still lose money because they arrive late to illiquid books and pay unfavorable prices, a problem that deeper liquidity directly mitigates. Third, the absence of a functioning credit layer blocks derivative products, hedges, and secondary instruments that traditional asset classes take for granted. Without the ability to borrow against positions, prediction markets remain a spot-only venue, and the full composability required for institutional participation cannot develop.

Existing platforms have attempted to solve the capital lockup problem by paying position holders and market makers "holding" and "liquidity" rewards from treasuries. These programs serve better as band-aids than sustainable long-term solutions. Locked capital still remains inefficient, reward rates are arbitrarily calculated, and their net contribution to market quality is still uncertain [1].

This paper proposes a direct fix: lending against prediction-market positions. A trader

[1] For long-term markets with competitive odds and holding rewards enabled, we observe that while spread improves, L2 orderbook data shows $> 80\%$ of liquidity sits at or above the 80th price level passively earning holding rewards, leaving the majority price levels closer to the midprice sparse. This means any trader with an informational advantage that doesn't at least push their true odds to 80% (for many markets that's a $\Delta \geq 50\%$) will likely face significant execution friction.

posts their position as collateral, borrows cash against it, and gets to keep the upside exposure while regaining liquidity. If such borrowing can be underwritten safely, many positions stop being dead capital, and liquidity can then recycle through the rest of the market.

Standard lending solutions rely on the lending platform's operational ability to liquidate underwater collateral in time. This makes them inherently optimistic. To this end, platforms enforce conservative loan-to-value (LTV) ratios, both at origination and for liquidation. These provide an additional safety margin and make lending over-collateralized. These solutions work because collateral price movements are generally smooth; losses typically occur over long enough time frames, and automated-market-makers (AMMs) and/or deep books help ensure that most price levels on the downward path likely present some opportunity to execute before reaching the solvency line. Thus, protocols shift their executional burden to after the adverse movement has already happened. The objective for these optimistic solutions is ex post: ensure solvency once a drawdown has occurred (Merton, 1976; Gould et al., 2013).

This does not work for prediction markets; the obstacle being that these positions are not ordinary collateral. Their prices are bounded on $[0, 1]$, they often reprice in jumps rather than smooth paths due to how information arrives, and they exhibit excessive regime switching with esoteric, market-specific pricing dynamics. Unlike traditional lending, there is no optimistic solution; downward price movements are nearly instantaneous as informational shocks almost always precipitate a liquidity vacuum, followed by a brief period where the market attempts to find a new equilibrium. Once a drawdown has occurred, it is already too late. Thus, our objective must be ex ante: ensure solvency before the drawdown has occurred.

Our solution is a calibrated risk engine that prices epoch shortfall risk directly. Instead of trying to model the full lifetime value of a loan with a single global process, we break the problem into set-length underwriting horizons, or epochs (Rasmussen, 2025). The risk engine is fed an at-origination-time state-conditional feature vector and estimates the conditional loss distribution of collateral over each epoch. It then prices that loss with a Wang distortion, and produces a premium that scales when tail risk or execution risk worsens.

Section 2 states the lending problem and the pricing setup. Section 3 summarizes the data and calibration pipeline. Section 4 presents the state-contingent premium schedule and the engine's performance in a dual-pool Monte Carlo comparison against a baseline flat-rate model.

2 Model

For a lender, the relevant question is not whether a market is interesting or whether a given price level is directionally correct. The underwriting question is narrower: *if a loan is opened in state x, what is the distribution of lender shortfall over the next epoch?* This type of loss-based framing is common in actuarial finance. It reveals a wonderful simplification: once lender shortfall is defined on realized outcomes, the resulting shortfall distribution can incorporate collateral moves, liquidation timing, and execution quality. This distribution is then central to assessing whether capital remains solvent after adverse repricing. For our purposes, this is advantageous because it lets us model the conditional shortfall distribution directly, rather than derive a fully general closed-form model of higher-order market dynamics and microstructure.

We also work with discrete epochs rather than a single terminal maturity. At origination time t, state x_t contains the information available to the engine such as current price region, epoch length, time to resolution, market category, and liquidity or execution features. A policy θ specifies the loan terms, most importantly collateral intensity and any underwriting limits attached to that state.

Given (x_t, θ), let $L_t^\theta \geq 0$ denote lender loss over the next epoch. The engine's task is to choose policies that balance borrower utility while keeping the tail of L_t^θ inside acceptable limits. This is naturally an insurance problem. Rather than forecasting the full path of "true" event probabilities, we price the conditional law of shortfall directly.

2.1 Admissibility and pricing

Let X denote the origination state as a random variable and let x denote a realized state. At each state x, admissible policies must satisfy three conditions:

$$\Pr(L^\theta > \bar{l}(x) \mid X = x) \leq \epsilon(x), \qquad \mathrm{ES}_{\alpha,\theta}(x) \leq \bar{e}(x), \qquad U(\theta, x) \geq \underline{U}(x), \tag{2.1}$$

where $\bar{l}(x)$ is the maximum allowable loss threshold in state x, $\epsilon(x)$ is the maximum allowable probability of exceeding that threshold, $\mathrm{ES}_{\alpha,\theta}(x)$ is expected shortfall at tail level α under policy θ conditional on state x, $\bar{e}(x)$ is the maximum allowable expected shortfall in state x, $U(\theta, x)$ is borrower utility under policy θ in state x, and $\underline{U}(x)$ is the minimum acceptable borrower utility in state x. The first condition limits breach probability, the second limits tail severity, and the third rules out policies that are safe but not economically worth offering.

This formulation matters because collateral is regime-sensitive. Two loans with the same

tenor can have very different risk if one originates in a deep, central, far-from-resolution market and the other in a shallow, boundary-priced, near-resolution market. The admissible set must therefore be state dependent.

To price that state dependence we estimate the conditional loss distribution and then apply a Wang distortion to the physical law. Let $F_{L|X}(\ell \mid x) = \Pr(L^\theta \le \ell \mid X = x)$ denote the conditional cumulative distribution function of lender loss in state x. The quoted premium is generated from a distorted measure that overweights adverse outcomes:

$$F^*_{L|X}(\ell \mid x) = \Phi\big(\Phi^{-1}\big(F_{L|X}(\ell \mid x)\big) + \eta\big)\,, \tag{2.2}$$

where $\eta > 0$ is the Wang distortion parameter and Φ is the standard normal cumulative distribution function. Intuitively, η is the price of tail capital: larger η implies more weight on bad states and therefore higher quoted premium.

Risk is measured and premiums are priced per epoch rather than as a single flat APR. For each origination state x, the engine returns a state-specific underwriting quote, so the platform can lend cheaply in calm states without subsidizing dangerous ones. There is a caveat, however: because the engine consumes per-epoch (and therefore per-loan) state, it has a blind spot with respect to diagnosing and pricing correlation risk across the book. While the Wang distortion accurately prices single-loan shortfall, the model lacks a correction for how concentrated or correlated collateral exposure among outstanding loans can amplify portfolio losses. Under normal operations this risk is negligibly small, but it is precisely the left-tail outcomes central to this paper where correlated drawdowns become dangerous. The 2007 financial crisis illustrates the consequences when portfolio-level correlation risk is not properly accounted for, as individually well-underwritten positions can produce systemic losses when their collateral deteriorates simultaneously.

In production, a platform has two solutions: i.) price in the risk, or ii.) minimize exposure. For this paper, we have taken option ii.). This includes a filter on markets or positions available to be collateralized given the support of loans already issued. In production, this can be as simple as applying pool caps or stricter LTV ratios as correlation risk grows, for which a separate controller can moderate. For option i.), it is not enough to simply append book-level features to the state input and retrain. Correlation risk is a property of the joint distribution across loans. Recovering that joint structure from per-loan marginals is a nontrivial modeling problem, particularly when the dependence itself is state-dependent and concentrates in the tails. We have privately explored solutions pertinent to option i.), but for now they are beyond the scope of this paper.

2.2 Semi-parametric loss model

A single global parametric law is poorly suited to these markets. Prediction-market returns are boundary-constrained, strongly state dependent, and visibly heavy tailed. The engine therefore uses a semi-parametric conditional law: a flexible bulk fit for the center of the distribution and an EVT tail model for exceedances. This keeps the estimator responsive where most observations live while still treating the underwriting tail as a first-class object.

That choice lines up with the economics of the problem. The center of the distribution tells the engine how often routine repricing occurs. The tail determines whether the lender survives. A good underwriting model must do both, but the tail is the binding constraint. Appendix A provides the full contract-to-loss mapping, proofs, and formal admissibility results; the main point here is simpler: we quote lending terms from the distribution of shortfall, and we price that shortfall with an explicit tail-risk operator.

3 Calibration

We calibrate the engine on Polymarket trade data spanning 2022–2026.

Table 1: *Validation-set calibration summary across quantiles.*

Metric	Temporal test	New-market test
$Q_{0.01}$ empirical coverage	0.0094	0.0091
$Q_{0.05}$ empirical coverage	0.0580	0.0557
$Q_{0.95}$ empirical coverage	0.9477	0.9472
$Q_{0.99}$ empirical coverage	0.9884	0.9888
Crossing after repair	0.0%	0.0%

The raw sample contains roughly 397M trades, which are transformed into an epoch-level panel of about 885M observations and then filtered to a resolved, eligible universe used for fitting, validation, and testing. Detailed implementation information can be found in **??**.

Three empirical features of the data drive the design. First, tails are extremely non-Gaussian. Across epoch lengths from 1 hour to 168 hours, empirical kurtosis remains far above the Gaussian benchmark of 3, with the shortest horizons exhibiting the most violent tail behavior. Second, risk is strongly state dependent: the conditional CDFs in Figure 1 and the premium surface in Figure 2 show that time to resolution, starting price bucket, and market category interact nonlinearly — the distribution shape changes across states, not merely its location or scale. Third, many epochs are effectively zero-return, so any method that assumes smooth homoskedastic noise is badly misspecified.

Those facts motivate the calibration pipeline. We estimate conditional quantiles of logit returns using gradient-boosted quantile regression (Koenker and Bassett, 1978; Ke et al., 2017), then splice generalized Pareto tails onto the extremes. Quantile outputs are recalibrated with conformal shifts and monotonicity repair so the lower tail is usable for underwriting. This gives a fitted conditional law that is flexible in the bulk, disciplined in the tail, and stable enough to map into loss space.

To test generalization, we reserve three disjoint held-out subsets: two for validation and one for testing. The two validation splits are a forward-time holdout with purge gaps and a held-out-market split containing markets absent from training. All held-out sets were filtered to ensure isolation and no leakage from correlated cross-event markets. The first tests robustness to temporal drift; the second tests whether the model generalizes beyond markets seen in fitting. Final simulation results are then reported separately on a third held-out test set.

Table 1 shows that lower-tail coverage is close to nominal on both held-out validation sets, especially at the 1% and 99% quantiles that matter most for underwriting. Similar results across the temporal and new-market splits indicate that the model generalizes out of sample. The remaining error is concentrated away from the underwriting tail and is less important for premium setting.

Once the conditional return law is calibrated, it is mapped into lender loss using the contract and liquidation rules, then pushed through the distorted pricing operator in Equation (2.2). The output is not a scalar "platform rate." It is a state-contingent underwriting surface that can be queried at origination and updated when state changes.

3.1 Model Diagnostics

Before turning to simulation, we examine three intermediate outputs of the fitted pipeline that illustrate how the engine characterizes risk across origination states.

Figure 1 shows representative conditional CDFs of logit-transformed returns for four origination states. We emphasize that these are return distributions, not the loss distributions directly used in pricing; the latter are derived from the former via the contract-to-loss mapping which is shown in more detail in Appendix A.2. Sharp separation across origination states is clearly visible. Safer states sit to the right of risky and near-resolution states, especially in the lower tail. The difference is not a simple widening or narrowing of the same distribution; distribution shape changes fairly dramatically with respect to state. Some states have mild routine variation but much thicker adverse tails, while others remain tighter throughout.

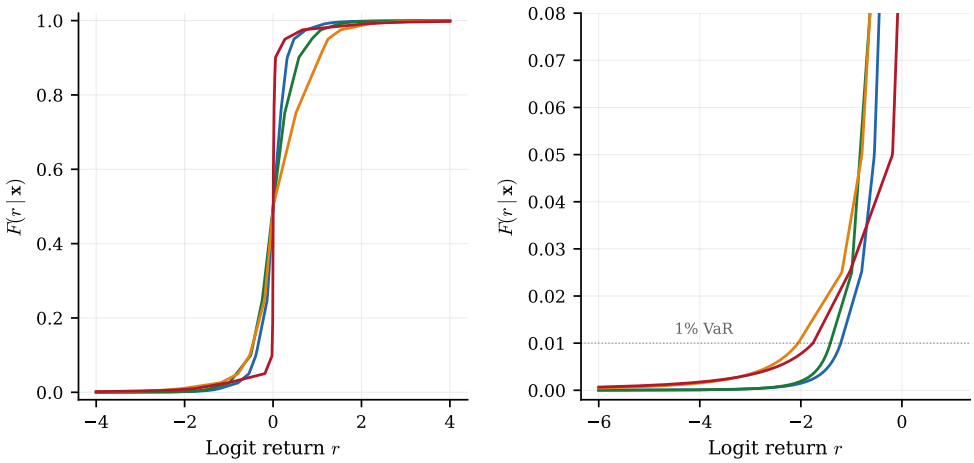

Figure 1: *Conditional return distributions $F(r \mid x)$ for four representative origination states evaluated from the fitted pipeline. Left: full-support CDFs. Right: zoomed-in view of the left-tail below the 1% VaR threshold (dotted line). The figure highlights regime-sensitive tail thickening as origination state becomes more adverse. Blue: safe (politics, $p = 0.50$, 30 d to resolution, 24 h epoch). Green: moderate (crypto, $p = 0.70$, 7 d, 24 h). Orange: risky (other, $p = 0.85$, 12 h, 4 h). Red: near-resolution (crypto, $p = 0.60$, 4 h, 1 h).*

A single volatility parameter, or even set of parameters, applied uniformly across all origination states cannot capture this. Boundary-priced crypto near resolution, central political markets far from resolution, and thin "other" markets do not belong in the same pricing bucket. Collateral also does not deteriorate smoothly: as markets approach resolution or prices move toward the boundaries, downside risk can change quickly, and small state changes can produce large changes in left-tail exposure.

Figure 2 gives a two-dimensional slice of the premium surface across starting price and time to resolution. Even this restricted slice is clearly non-flat. Premium is lowest in far-from-resolution central states, rises as time to resolution shortens, and rises again as prices move toward the boundaries. There is therefore no single sensible platform-wide borrowing rate for this collateral class. A flat quote necessarily underprices some states and overprices others.

The dangerous states are often the ones that look safest under naive intuition: contracts near \$0.95 or \$0.05, close to resolution, with seemingly little room left to move. The surface says otherwise. A small price move in probability space can still translate into severe loss in return and liquidation space. "Almost resolved" is not the same thing as "almost guaranteed."

Interestingly, the surface is also very structured with minimal noise. Premium loads along interpretable dimensions tied to market mechanics, like boundary effects, shrinking time, and state-dependent jump risk, rather than jumping idiosyncratically from cell to cell.

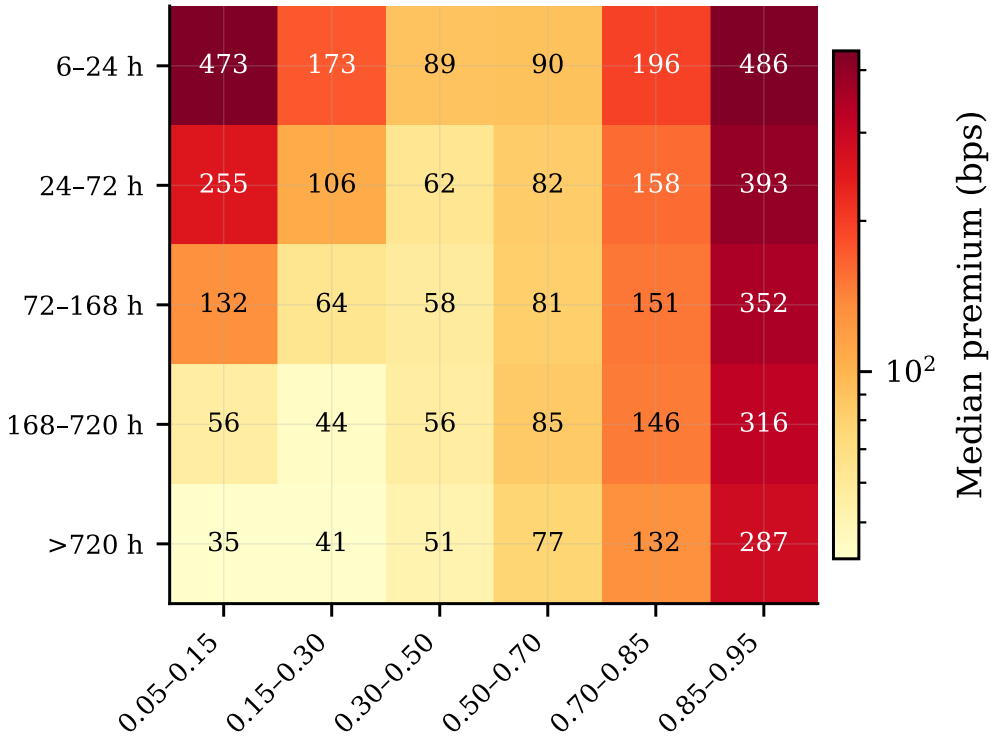

Figure 2: *Two-dimensional slice of the quoted premium surface by starting-price bin and time-to-resolution bin. The remaining state variables are held fixed. Premium is lowest in far-from-resolution central states and rises near the price boundaries and near resolution.*

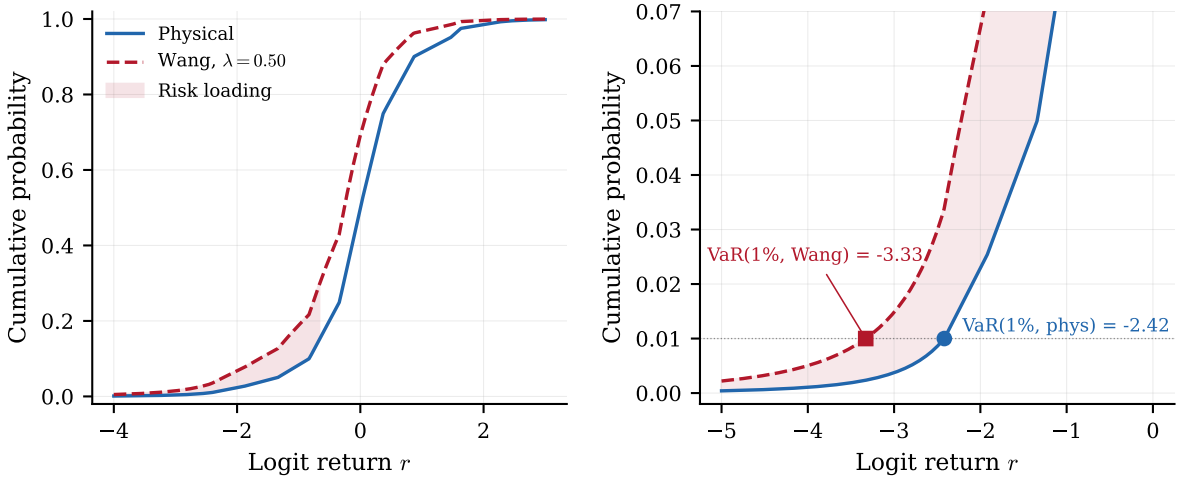

Figure 3: *Physical and Wang-distorted conditional laws for a representative origination state. The distortion increases weight on adverse outcomes while preserving state ordering.*

The engine produces a surface with economically comprehensible shape.

Figure 3 shows the pricing operator applied to a representative state. The Wang transformation shifts probability weight toward adverse outcomes while preserving state ordering,

converting the actuarial loss distribution into a quoted premium by assigning an explicit cost to tail-capital consumption. Expected loss alone is too weak an underwriting object: two states can have similar average loss but very different tail exposure. The distortion separates those cases by charging more where adverse outcomes are more dangerous to lender capital, pricing the balance-sheet cost of tail risk rather than pretending those losses will diversify away.

4 Simulation

We compare two otherwise identical lending pools that differ only in pricing policy. The naïve pool charges a flat 2.5% APR with no explicit risk premium. The risk engine pool charges state-contingent epoch premiums implied by the calibrated engine. Both pools start with $1M capital and replay identical historical price paths over 32 stochastic demand draws, with a held-out test market sample serving as the price-generating process.

Table 2: *Charged premium per epoch in basis points of loan notional ($\eta = 1.0$, 20,000 holdout grid points). Values are mean [P25, P75].*

Epoch	LTV = 40%	LTV = 50%	LTV = 60%	LTV = 70%	LTV = 75%
1h	151.7 [63, 96]	194.9 [63, 139]	252.1 [63, 209]	328.5 [63, 330]	376.9 [63, 409]
4h	308.7 [63, 199]	390.6 [63, 312]	492.4 [63, 477]	620.8 [63, 693]	698.3 [69, 828]
12h	611.1 [63, 539]	743.9 [63, 714]	899.1 [80, 932]	1082.9 [115, 1253]	1188.9 [147, 1455]
24h	886.2 [63, 763]	1058.2 [87, 1070]	1253.1 [125, 1436]	1474.2 [199, 1875]	1597.7 [261, 2117]
72h	1416.9 [160, 1848]	1667.7 [247, 2349]	1932.2 [356, 2861]	2212.5 [521, 3363]	2360.4 [623, 3617]
168h	1837.8 [253, 3061]	2139.0 [398, 3588]	2443.0 [596, 4111]	2752.2 [776, 4563]	2910.2 [866, 4777]

Table 2 shows the state-contingent premium schedule that makes this possible. Premiums range from around 63 bps at the safest 1-hour, low-LTV states up to over 4,000 bps for the riskiest 168-hour, high-LTV states. The interquartile range within each cell reflects the remaining state heterogeneity not captured by epoch and LTV alone. A flat 2.5% APR — roughly 48 bps per week — sits well below the mean premium for any epoch beyond 4 hours at LTV $\geq 60\%$, which is precisely where the naive pool's losses concentrate.

Table 3 shows a large difference in lender outcomes. The flat-rate pool loses money, suffers severe drawdowns, and opens far fewer loans. The risk engine pool remains profitable, keeps drawdowns contained, and supports much higher origination volume.

Importantly, the risk engine pool is not improving outcomes by avoiding risk, it actually lends more, far more. It opens roughly twelve times as many loans while cutting the liquidation rate in half.

Flat pricing, and other optimistic-liquidation solutions fail because they collect too little

Table 3: *Dual-pool Monte Carlo simulation comparing a naïve flat-rate pool against the risk engine Wang-premium pool. Both pools start with $1M capital and replay identical historical price paths over 32 stochastic demand draws. The naïve pool charges a fixed 2.5% APR with no additional risk premium; the risk engine pool charges state-contingent epoch premiums at* $\eta = 1.0$. *Metrics are means across runs; P5–P95 bands are tight (relative SE* $< 1\%$*) and omitted for clarity.*

Metric	Naïve	Risk Eng.	Δ
Net P&L ($K)	−850	+2,891	+3,741
Ann. Return	−27.2%	+92.5%	+119.7 pp
Max Drawdown	−85.0%	−3.6%	+81.4 pp
Loans Opened	92,404	1,081,066	—
Liquidations	5,853	33,280	—
Liquidation Rate	6.3%	3.1%	−3.2 pp
Cum. Losses ($K)	850	4,956	—
Worst Daily Loss ($K)	37	73	—
Worst Weekly P&L ($K)	−64	−24	—

compensation in the states where losses are concentrated. Adverse states consume capital faster than the pool can rebuild it through ordinary interest income. Once capital falls, origination capacity falls with it, and the book enters a slow solvency spiral. It is cumulative underpricing of the left tail. The risk engine pool behaves differently because it charges where the left tail is actually located. That premium buffer absorbs losses in the same states that would otherwise deplete lender capital. Capital remains available, origination continues, and the book can scale without becoming a disguised short-volatility position.

References

Arrow, Kenneth J., Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O. Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D. Nelson, George R. Neumann, Marco Ottaviani, Thomas C. Schelling, Robert J. Shiller, Vernon L. Smith, Erik Snowberg, Cass R. Sunstein, Paul C. Tetlock, Philip E. Tetlock, Hal R. Varian, Justin Wolfers, and Eric Zitzewitz, "The Promise of Prediction Markets," *Science*, May 2008, *320* (5878), 877–878.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath, "Coherent Measures of Risk," *Mathematical Finance*, 1999, *9* (3), 203–228.

Balkema, A. A. and L. de Haan, "Residual Life Time at Great Age," *The Annals of Probability*, October 1974, *2* (5), 792–804.

Clinton, Joshua D and TzuFeng Huang, "Prediction Markets? The Accuracy and Efficiency of $2.4 Billion in the 2024 Presidential Election," December 2025.

Czado, Claudia, Tilmann Gneiting, and Leonhard Held, "Predictive Model Assessment for Count Data," *Biometrics*, December 2009, *65* (4), 1254–1261.

Dalen, Shaw, "Toward Black Scholes for Prediction Markets: A Unified Kernel and Market Maker's Handbook," October 2025.

Gould, Martin D., Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison, "Limit Order Books," *Quantitative Finance*, November 2013, *13* (11), 1709–1742.

Hanson, Robin, "Combinatorial Information Market Design," *Information Systems Frontiers*, January 2003, *5* (1), 107–119.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," 2017.

Koenker, Roger and Gilbert Bassett, "Regression Quantiles," *Econometrica*, January 1978, *46* (1), 33.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6), 1315–1335.

McNeil, Alexander, Rüdiger Frey, and P Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools*, Vol. 101 October 2005.

Meinshausen, Nicolai, "Quantile Regression Forests," 2006.

Merton, Robert C., "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *The Journal of Finance*, 1974, *29* (2), 449–470.

_ , "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Finan-*

cial Economics, January 1976, *3* (1), 125–144.

Pickands, James, "Statistical Inference Using Extreme Order Statistics," *The Annals of Statistics*, January 1975, *3* (1), 119–131.

Rasmussen, Kaleb, "Enabling Leverage on Prediction Markets," November 2025.

Roll, Richard, "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market," 1984.

Romano, Yaniv, Evan Patterson, and Emmanuel J. Candès, "Conformalized Quantile Regression," May 2019.

Vedova, Joshua Della, "Who Profits from Prediction Markets? Execution, Not Information," February 2026.

Wang, Shaun S., "A Class of Distortion Operators for Pricing Financial and Insurance Risks," *The Journal of Risk and Insurance*, 2000, *67* (1), 15–36.

Wolfers, Justin and Eric Zitzewitz, "Prediction Markets," *Journal of Economic Perspectives*, June 2004, *18* (2), 107–126.

A Loss Distributions & Tail-Risk Pricing

A.1 Conceptual Framework

Lending, at its simplest, is the purchase of money now in exchange for discounted money later at a premium, typically paid through coupon-like interest payments over a fixed term. Classical valuation approaches price a loan by modeling the loan quantitatively, and then approximate its expected payoff over its full (or remaining) term. Expectations are nice for a number of reasons. Most pertinent to this section, and the broader binary asset discussion, is that they inherently incorporate adverse outcomes such as default risk. However, as previously mentioned, pricing this way requires a general model that can then be used to somewhat-reliably forecast. In practice, accurate valuation depends on credibly capturing the asset's value-generating process, which is often latent and/or incredibly complex, by analyzing and theorizing about market observables and their interrelationships. Depending on the choice of model, one can perform many simulations via Monte Carlo to capture the whole range of outcomes and paths, or validate an analytical solution against an experimental distribution via a Kolmogorov-Smirnov (KS) test, or perform an out-of-sample, real-time backtest, and so on.

For most traditional assets, this value-generating process is anchored by fundamentals. Corporate bonds and equities are tied to firm cash flows as well as the information about these flows (earnings, balance-sheet evolution, etc.), while derivatives inherit structures from their underlyings. Prediction markets are different. There is no conventional cash-flow anchor or fundamental underlying structure; the value-generating process is the *information process*. To add complexity, prices are bounded on $[0, 1]$, and repricing usually is discontinuous. It should be noted that these price dynamics are likely a function of current platform architectures and not necessarily an innate property of prediction markets themselves. Limit orderbooks (LOB) are the preferred architectures that the largest platforms by volume currently use to facilitate trading. The evolution of orders on LOBs follows a càdlàg process; discontinuity is baked into the very design of current platforms (Gould et al., 2013).

Architectural and platform design choices are beyond the scope of this paper. While meaningful work is being done to improve the design of prediction markets[2] [3], a true liquidity solution must be both composable and tangible with existing markets, not just hypothetical.

Regardless, jumping price movements pose a serious problem for collateralized financing (Merton, 1976): given a large enough price movement, single-sided liquidity can suddenly dry up, and prices can rocket past liquidation barriers and solvency lines as the market

[2] `https://substack.com/home/post/p-184769183`

[3] `https://www.bedlamresear.ch/posts/hip-4-event-futures/`

searches for a new equilibrium.

Due to these idiosyncrasies, fitting a single global dynamic model for all prediction markets typically yields a miscalibrated result. It is more productive and computationally efficient to treat individual markets (or clusters of markets) as distinct regimes with shared structural features, rather than as one homogeneous asset class that a general form can approximate. Accordingly, we shift from a terminal expected-payoff lens to an insurance/credit lens. We model the loss distribution and charge a premium for bearing shortfall risk. Once written in lender-solvency terms, the relevant object is not the instantaneous value of collateral, but the distribution of shortfall over a finite horizon, conditional on origination state and execution conditions. The advantage of the insurance formulation is that we need not reliably model information arrival, nor become entirely dependent on order execution and liquidation recoveries. Instead, we can capture the worst behavior of the tails and cover lenders while still providing borrowers the opportunity to unlock capital without extremely punitive rates (McNeil et al., 2005).

A.2 Contract-to-Loss Mapping

Borrowing ideas from Rasmussen (2025), we split term risk into discrete epochs and price tail risk at each epoch.

Definition A.1 (Model primitives)**.** Fix an epoch horizon $\Delta > 0$ and a filtered probability space

$$(\Omega, \mathcal{F}, (\mathcal{F}_u)_{u \geq 0}, \mathbb{P}).$$

Let $\mathcal{X} \subset \mathbb{R}^d$ be the state space with realized state $x \in \mathcal{X}$, and let

$$\Theta \subset \mathbb{R}^m$$

be the policy space with $\theta \in \Theta$. For each (x, θ), epoch loss is a nonnegative random variable

$$L_x^\theta : \Omega \to \mathbb{R}_+,$$

with conditional CDF

$$F_{L|X,\theta}(l \mid x) := \mathbb{P}(L_x^\theta \leq l \mid X = x).$$

Let

$$U : \Theta \times \mathcal{X} \to \mathbb{R}_+$$

denote utilization (economic usefulness), and fix state-dependent limits

$$\epsilon : \mathcal{X} \to (0, 1), \qquad \bar{l} : \mathcal{X} \to \mathbb{R}_+, \qquad \bar{e} : \mathcal{X} \to \mathbb{R}_+, \qquad \underline{U} : \mathcal{X} \to \mathbb{R}_+.$$

At fixed x, L_x^θ is the policy-induced epoch-loss random variable, $F_{L|X,\theta}(\cdot \mid x)$ is its conditional distribution, and $U(\theta, x)$ is the objective functional. The state-dependent quantities

$\epsilon(x), \bar{l}(x), \bar{e}(x), \underline{U}(x)$ are constraint primitives: respectively, a breach-probability tolerance at threshold $\bar{l}(x)$, an ES cap, and a minimum-utility requirement.

Definition A.2 (State-conditional tail risk functionals)**.** For $\alpha \in (0,1)$, $x \in \mathcal{X}$, and $\theta \in \Theta$, define

$$\mathrm{VaR}_{\alpha,\theta}(x) := \inf\{l \in \mathbb{R}_+ : F_{L|X,\theta}(l \mid x) \geq \alpha\}, \tag{A.1}$$

$$\mathrm{ES}_{\alpha,\theta}(x) := \mathbb{E}\big[L_x^\theta \,\big|\, L_x^\theta > \mathrm{VaR}_{\alpha,\theta}(x)\big]. \tag{A.2}$$

Equation (A.1) fixes the α-quantile under the conditional loss law, while Equation (A.2) evaluates conditional tail severity beyond that quantile (Artzner et al., 1999). The pair separates quantile location from tail severity beyond that quantile.

Definition A.3 (Admissible policies and selection)**.** For each $x \in \mathcal{X}$, define

$$\mathcal{A}(x) := \big\{\theta \in \Theta : \mathbb{P}(L_x^\theta > \bar{l}(x) \mid X = x) \leq \epsilon(x),\ \mathrm{ES}_{\alpha,\theta}(x) \leq \bar{e}(x),\ U(\theta, x) \geq \underline{U}(x)\big\}. \tag{A.3}$$

A policy is *loss-admissible at* x iff $\theta \in \mathcal{A}(x)$. The selection problem is

$$\theta^*(x) \in \arg\max_{\theta \in \Theta} U(\theta, x) \quad \text{s.t.} \quad \theta \in \mathcal{A}(x). \tag{A.4}$$

Equation (A.3) is the intersection of a breach-probability cap, an ES cap, and a utility floor. Equation (A.4) is therefore a constrained utility maximization over that intersection.

Proposition A.4 (Well-posedness and existence)**.** *Fix $x \in \mathcal{X}$. Suppose:*

1. *Θ is nonempty and compact;*
2. *$U(\cdot, x)$ is upper semicontinuous on Θ;*
3. *$\theta \mapsto \mathbb{P}(L_x^\theta > \bar{l}(x) \mid X = x)$ is lower semicontinuous;*
4. *$\theta \mapsto \mathrm{ES}_{\alpha,\theta}(x)$ is lower semicontinuous;*
5. *$\mathcal{A}(x) \neq \varnothing$.*

Then $\mathcal{A}(x)$ is compact and the argmax set in Equation (A.4) *is nonempty.*

Proof. See Appendix D.1. ■

Under Proposition A.4, the constrained problem admits at least one maximizer at each fixed state.

Corollary A.5 (Uniqueness under convexity)**.** *Fix $x \in \mathcal{X}$. In addition to the assumptions of Proposition A.4, suppose:*

1. *$\Theta \subset \mathbb{R}^m$ is convex;*
2. *$U(\cdot, x)$ is strictly concave on Θ;*
3. *the constraint maps*

   $$g_1(\theta) := \mathbb{P}(L_x^\theta > \bar{l}(x) \mid X = x) - \epsilon(x), \quad g_2(\theta) := \mathrm{ES}_{\alpha,\theta}(x) - \bar{e}(x), \quad g_3(\theta) := \underline{U}(x) - U(\theta, x)$$

   are convex in θ;
4. *(Slater condition) there exists $\tilde{\theta} \in \Theta$ such that*

   $$g_i(\tilde{\theta}) < 0, \qquad i = 1, 2, 3.$$

Then:

1. *the feasible set $\mathcal{A}(x)$ is convex and nonempty;*
2. *problem Equation (A.4) has a unique optimizer $\theta^*(x)$;*
3. *Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient for optimality; i.e., there exist multipliers $\mu_1, \mu_2, \mu_3 \geq 0$ such that*

   $$\nabla_\theta U(\theta^*, x) - \sum_{i=1}^{3} \mu_i \nabla_\theta g_i(\theta^*) = 0, \qquad \mu_i g_i(\theta^*) = 0, \ i = 1, 2, 3,$$

   with $g_i(\theta^*) \leq 0$, $i = 1, 2, 3$.

Proof. See Appendix D.2. ■

Corollary A.5 implies that under convex constraints and strict concavity of $U(\cdot, x)$, Equation (A.4) admits a unique optimizer $\theta^*(x)$. Slater regularity rules out boundary-only degeneracy, so KKT conditions are both necessary and sufficient. The multipliers (μ_1, μ_2, μ_3) admit a shadow-value interpretation for, respectively, breach-probability tolerance, tail-severity tolerance, and minimum-utility requirement.

For contract instantiation, fix origination time t, epoch horizon $\Delta > 0$, and identify $x \equiv X_t$, $L_x^\theta \equiv L_{t,\Delta}^\theta$.

Definition A.6 (Epoch contract cash-flow map)**.** Fix t, Δ, θ. Principal is B_t^θ, epoch recovery is $R_{t,\Delta}^\theta$, and epoch loss is

$$L_{t,\Delta}^\theta := (B_t^\theta - R_{t,\Delta}^\theta)^+. \tag{A.5}$$

At state $x \equiv X_t$, this realizes the primitive L_x^θ from Definition A.1 via the identification $L_x^\theta \equiv L_{t,\Delta}^\theta$.

The pure-premium floor is

$$\Pi_{\text{pure},\theta}(x) \geq \mathbb{E}[L_{t,\Delta}^\theta \mid X_t = x].$$

For implementation, epochs are evaluated on the discrete menu

$$\Delta \in \{1, 4, 12, 24, 72, 168\} \text{ hours}. \tag{A.6}$$

This menu is a numerical discretization, not a structural restriction. The contract-pricing object is the maturity-indexed surface

$$\Delta \mapsto \Pi_\theta(x, \Delta), \qquad \Delta > 0,$$

which is defined on continuous maturity in principle; Equation (A.6) is a computational grid used to approximate that continuous object.

Let T_{res} denote resolution time and

$$\text{TTR}_t := T_{\text{res}} - t.$$

Empirically, as $\text{TTR}_t \to 0$, realized volatility typically rises, so TTR_t is a protocol-level control variable in addition to a state coordinate. We impose a resolution buffer $\tau_{\text{buf}} > 0$ and define the origination gate as follows.

Definition A.7 (Origination gate and admissible origination times)**.** Given buffer $\tau_{\text{buf}} > 0$, define

$$\mathcal{G}_t := \mathbf{1}\{\text{TTR}_t \geq \tau_{\text{buf}}\}.$$

The admissible origination set is

$$\mathcal{T}_{\text{orig}} := \{t : \mathcal{G}_t = 1\} = \{t : T_{\text{res}} - t \geq \tau_{\text{buf}}\}.$$

Epochs may be originated only on $\mathcal{T}_{\text{orig}}$; if $\mathcal{G}_t = 0$, origination is disabled and only previously originated epochs remain active.

Let $p_t \in (0, 1)$ be YES-equivalent collateral price, $q > 0$ collateral units, $V_t = q\, p_t$. Policy sets

$$B_t^\theta = \lambda_\theta V_t(1 - h_\theta) = \lambda_\theta(1 - h_\theta) q\, p_t,$$

with $\lambda_\theta \in (0, 1)$, $h_\theta \in [0, 1)$. Initial coverage is

$$C_t^\theta = \frac{V_t}{B_t^\theta} = \frac{1}{\lambda_\theta(1 - h_\theta)}.$$

Given maintenance requirement $C_{\min,\theta} > 1$, the liquidation trigger follows a first-passage structure analogous to structural credit models (Merton, 1974). Trigger at $u \in [t, t + \Delta]$ is

$$\frac{V_u}{B_t^\theta} \leq C_{\min,\theta} \iff p_u \leq \ell_t^\theta, \qquad \ell_t^\theta := \frac{C_{\min,\theta} B_t^\theta}{q}. \tag{A.7}$$

Substituting B_t^θ into Equation (A.7),

$$\ell_t^\theta = C_{\min,\theta}\lambda_\theta(1 - h_\theta)p_t. \tag{A.8}$$

Define barrier ratio

$$b_\theta := \frac{\ell_t^\theta}{p_t} = C_{\min,\theta}\lambda_\theta(1 - h_\theta). \tag{A.9}$$

Definition A.8 (Maintenance barrier representation)**.** For policy θ, define level barrier

$$\ell_t^\theta := \frac{C_{\min,\theta} B_t^\theta}{q},$$

ratio barrier

$$b_\theta := \frac{\ell_t^\theta}{p_t} = C_{\min,\theta}\lambda_\theta(1-h_\theta),$$

and first-passage time

$$\tau_t^\theta := \inf\{u \in [t, t+\Delta] : p_u \le \ell_t^\theta\}.$$

Maintenance breach over $[t, t+\Delta]$ is equivalent to $\tau_t^\theta \le t+\Delta$, or equivalently $\inf_{u\in[t,t+\Delta]}(p_u/p_t) \le b_\theta$.

Hence trigger over the epoch can be written as

$$\inf_{u\in[t,t+\Delta]} \frac{p_u}{p_t} \le b_\theta.$$

Because $p \in (0,1)$, set

$$y_t := \log\frac{p_t}{1-p_t}, \qquad r_{t,u} := y_u - y_t.$$

Then $p_u = \sigma(y_t + r_{t,u})$, $\sigma(z) = 1/(1+e^{-z})$, and

$$p_u \le \ell_t^\theta \iff \log\frac{p_u}{1-p_u} \le \log\frac{\ell_t^\theta}{1-\ell_t^\theta} \iff r_{t,u} \le a_t^\theta,$$

where

$$a_t^\theta := \log\frac{\ell_t^\theta}{1-\ell_t^\theta} - \log\frac{p_t}{1-p_t}. \tag{A.10}$$

Definition A.9 (Log-odds first-passage form)**.** Let $y_t = \log\big(p_t/(1-p_t)\big)$, $r_{t,u} = y_u - y_t$, and

$$a_t^\theta := \log \frac{\ell_t^\theta}{1-\ell_t^\theta} - \log \frac{p_t}{1-p_t}.$$

Then the trigger event admits the first-passage representation

$$\tau_t^\theta \le t + \Delta \iff \inf_{u\in[t,t+\Delta]} r_{t,u} \le a_t^\theta.$$

If trigger occurs, define execution gap

$$\delta_{\tau_t^\theta} := p_{\tau_t^\theta} - p^{\mathrm{ex}}_{\tau_t^\theta} \ge 0,$$

and let $\kappa_\theta \in [0,1)$ summarize fees/costs. Trigger-path recovery is

$$R^\theta_{\tau_t^\theta} = q\, p^{\mathrm{ex}}_{\tau_t^\theta}(1-\kappa_\theta).$$

Using $p^{\mathrm{ex}}_{\tau_t^\theta} = p_{\tau_t^\theta} - \delta_{\tau_t^\theta}$,

$$L^\theta_{\tau_t^\theta} = q\Big[\lambda_\theta(1-h_\theta)p_t - (p_{\tau_t^\theta} - \delta_{\tau_t^\theta})(1-\kappa_\theta)\Big]^+. \tag{A.11}$$

At barrier touch $p_{\tau_t^\theta} \approx \ell_t^\theta$, substitute Equation (A.8) into Equation (A.11):

$$L^\theta_{\tau_t^\theta} \approx q\Big[\lambda_\theta(1-h_\theta)p_t\big(1 - C_{\min,\theta}(1-\kappa_\theta)\big) + (1-\kappa_\theta)\delta_{\tau_t^\theta}\Big]^+. \tag{A.12}$$

If no trigger occurs by $t+\Delta$, recovery is rule-dependent:

$$R^\theta_{t,\Delta} = \mathcal{R}_\theta\big(\{p_u\}_{u\in[t,t+\Delta]}, X_t\big)\,.$$

Then

$$L^\theta_{t,\Delta} = \ell(r_{t,\Delta}, X_t; \theta), \tag{A.13}$$

Definition A.10 (State-conditioned loss map)**.** For fixed (t,Δ,θ), the measurable map

$$\ell(\cdot,\cdot;\theta) : (r_{t,\Delta}, X_t) \mapsto L^\theta_{t,\Delta}$$

is the contract-to-loss map. Conditional on $X_t = x$, it induces the loss law $F_{L|X,\theta}(\cdot \mid x)$ of Definition A.1.

with

$$X_t = \big(\mathrm{TTR}_t,\ \mathcal{G}_t,\ p_t,\ \Delta,\ \mathrm{category}_t,\ \mathrm{activity/liquidity}_t,\ \mathrm{size}_t,\ \mathrm{market\ age}_t\big). \tag{A.14}$$

Let $A_\theta := \{\tau_t^\theta \le t + \Delta\}$. Then

$$F_{L|X,\theta}(l \mid x) = \pi_\theta(x) F_\theta^{\mathrm{tr}}(l \mid x) + \big(1 - \pi_\theta(x)\big) F_\theta^{\mathrm{nt}}(l \mid x), \quad \pi_\theta(x) := \mathbb{P}(A_\theta \mid X = x), \tag{A.15}$$

and, by conditional expectation under the same partition,

$$\mathrm{EL}_\theta(x) = \pi_\theta(x)\, \mathrm{EL}_\theta^{\mathrm{tr}}(x) + \big(1 - \pi_\theta(x)\big)\, \mathrm{EL}_\theta^{\mathrm{nt}}(x). \tag{A.16}$$

Definition A.11 (Trigger mixture decomposition)**.** Let $A_\theta := \{\tau_t^\theta \le t+\Delta\}$, $\pi_\theta(x) := \mathbb{P}(A_\theta \mid X = x)$, and $F_\theta^{\mathrm{tr}}(\cdot \mid x)$, $F_\theta^{\mathrm{nt}}(\cdot \mid x)$ denote trigger/no-trigger conditional loss laws. Then

$$F_{L|X,\theta}(\cdot \mid x) = \pi_\theta(x) F_\theta^{\mathrm{tr}}(\cdot \mid x) + \big(1 - \pi_\theta(x)\big) F_\theta^{\mathrm{nt}}(\cdot \mid x),$$

with the corresponding decomposition of conditional mean loss.

Comparative statics through b_θ:

$$\frac{\partial b_\theta}{\partial \lambda_\theta} = C_{\min,\theta}(1 - h_\theta) > 0, \qquad \frac{\partial b_\theta}{\partial h_\theta} = -C_{\min,\theta}\lambda_\theta < 0, \qquad \frac{\partial b_\theta}{\partial C_{\min,\theta}} = \lambda_\theta(1 - h_\theta) > 0.$$

For numerical stability, prices are clipped to $[\varepsilon, 1 - \varepsilon]$, $\varepsilon > 0$, before log-odds transforms.

A.3 Tail Model and Premium Construction

Given Equation (A.15), and writing $\mathrm{EL}_\theta(x) := \mathbb{E}[L_x^\theta \mid X = x]$ for the conditional expected loss, the tail functionals $\mathrm{VaR}_{\alpha,\theta}(x)$ and $\mathrm{ES}_{\alpha,\theta}(x)$ are as in Definition A.2.

Because tails are sparse and state-dependent, we use a bulk-tail splice. Choose threshold $u_\theta(x)$, and define exceedance

$$Y_\theta := L_x^\theta - u_\theta(x) \;\big|\; L_x^\theta > u_\theta(x), X = x.$$

Under the Pickands–Balkema–de Haan theorem (Balkema and de Haan, 1974; Pickands, 1975), exceedances above a sufficiently high threshold converge in distribution to a GPD:

$$\mathbb{P}(Y_\theta \le y \mid L_x^\theta > u_\theta(x), X = x) \approx 1 - \left(1 + \xi_\theta(x)\frac{y}{\beta_\theta(x)}\right)^{-1/\xi_\theta(x)}, \quad y \ge 0, \tag{A.17}$$

with $\beta_\theta(x) > 0$, $\xi_\theta(x) \in \mathbb{R}$, and support condition $1 + \xi_\theta(x) y/\beta_\theta(x) > 0$. Let

$$p_{u,\theta}(x) := \mathbb{P}(L_x^\theta \le u_\theta(x) \mid X = x).$$

Then

$$F_{L|X,\theta}(l \mid x) = \begin{cases} F_\theta^{\text{bulk}}(l \mid x), & l \le u_\theta(x), \\ p_{u,\theta}(x) + (1 - p_{u,\theta}(x))\, G_{\xi_\theta(x),\beta_\theta(x)}\big(l - u_\theta(x)\big), & l > u_\theta(x). \end{cases} \tag{A.18}$$

Definition A.12 (Bulk-tail splice specification)**.** At state x, choose threshold $u_\theta(x)$ and bulk law $F_\theta^{\text{bulk}}(\cdot \mid x)$ on $[0, u_\theta(x)]$. For exceedances above $u_\theta(x)$, impose a GPD tail with parameters $(\xi_\theta(x), \beta_\theta(x))$. The full conditional loss law is the splice in Equation (A.18), with continuity at $u_\theta(x)$ through $p_{u,\theta}(x)$.

For tail region $z_{\alpha,\theta}(x) := \text{VaR}_{\alpha,\theta}(x) \ge u_\theta(x)$, invert the tail branch of Equation (A.18):

$$z_{\alpha,\theta}(x) = u_\theta(x) + \frac{\beta_\theta(x)}{\xi_\theta(x)} \left[\left(\frac{1-\alpha}{1 - p_{u,\theta}(x)} \right)^{-\xi_\theta(x)} - 1 \right], \qquad \xi_\theta(x) \neq 0, \tag{A.19}$$

with log-limit

$$z_{\alpha,\theta}(x) = u_\theta(x) - \beta_\theta(x) \log\left(\frac{1-\alpha}{1 - p_{u,\theta}(x)} \right) \quad \text{as } \xi_\theta(x) \to 0.$$

If $\xi_\theta(x) < 1$, the corresponding conditional tail mean is

$$\text{ES}_{\alpha,\theta}(x) = z_{\alpha,\theta}(x) + \frac{\beta_\theta(x) + \xi_\theta(x)\big(z_{\alpha,\theta}(x) - u_\theta(x)\big)}{1 - \xi_\theta(x)}. \tag{A.20}$$

Proposition A.13 (ES-VaR cap equivalence in the EVT tail region)**.** *Fix (x, θ). Suppose $z_{\alpha,\theta}(x) := \text{VaR}_{\alpha,\theta}(x) \ge u_\theta(x)$ and $\xi_\theta(x) < 1$. Then the ES admissibility condition*

$$\text{ES}_{\alpha,\theta}(x) \le \bar{e}(x)$$

is equivalent to the VaR cap

$$z_{\alpha,\theta}(x) \le (1 - \xi_\theta(x))\, \bar{e}(x) - \beta_\theta(x) + \xi_\theta(x)\, u_\theta(x). \tag{A.21}$$

Proof. See Appendix D.3. ■

Proposition A.13 shows that, in the EVT tail regime, the ES cap is exactly a VaR cap. The practical consequence is that admissibility in Equation (A.3) can be enforced through a quantile threshold (Equation (A.21)) without changing the underlying risk standard, which simplifies calibration and stress-time constraint checks. Physical expected loss gives the pure premium:

$$\text{EL}_\theta(x) = \mathbb{E}[L_x^\theta \mid X = x]$$

Risk compensation is introduced via Wang distortion (Wang, 2000):

$$F^*_{L|X,\theta}(l \mid x) = \Phi\big(\Phi^{-1}\big(F_{L|X,\theta}(l \mid x)\big) + \eta\big), \qquad \eta > 0. \tag{A.22}$$

Quoted premium is

$$\Pi_\theta(x) := \mathbb{E}^*[L^\theta_x \mid X = x]. \tag{A.23}$$

Definition A.14 (Distorted premium operator)**.** Given physical law $F_{L|X,\theta}(\cdot \mid x)$, define the Wang-distorted law

$$F^*_{L|X,\theta}(l \mid x) = \Phi\big(\Phi^{-1}\big(F_{L|X,\theta}(l \mid x)\big) + \eta\big), \qquad \eta > 0,$$

and the quoted premium operator

$$\Pi_\theta(x) := \mathbb{E}^*[L^\theta_x \mid X = x].$$

Remark A.15 (Wang distortion for prediction market collateral)**.** *The quoted premium $\Pi_\theta(x) = \mathbb{E}^*[L^\theta_x \mid X = x]$ is finite under any Wang parameter $\eta > 0$ without imposing moment conditions on the GPD tail index. Because $p_t \in (0,1)$, collateral value satisfies $V_t = qp_t < q$, and principal is $B^\theta_t = \lambda_\theta(1-h_\theta)qp_t < q$. Since $L^\theta_{t,\Delta} = (B^\theta_t - R^\theta_{t,\Delta})^+ \le B^\theta_t < q$, epoch loss is bounded above by a deterministic constant. A bounded random variable has all moments finite under any probability measure, so $\mathbb{E}^*[L^\theta_x \mid X = x] < \infty$ holds unconditionally — the Wang distortion F^* does not alter this. This stands in contrast to equity or crypto collateral, where unbounded upside makes tail integrability a genuine restriction.*

Remark A.16 (Bounded support implies $\xi_\theta(x) < 0$ for the loss GPD)**.** *The same boundedness argument structurally constrains the GPD shape parameter. Prices are clipped to $[\varepsilon, 1-\varepsilon]$, so logit-space returns satisfy*

$$r_{t,\Delta} = y_{t+\Delta} - y_t \in \left[\log\tfrac{\varepsilon}{1-\varepsilon} - \log\tfrac{1-\varepsilon}{\varepsilon},\ \log\tfrac{1-\varepsilon}{\varepsilon} - \log\tfrac{\varepsilon}{1-\varepsilon}\right],$$

a compact interval. Since $L^\theta_{t,\Delta}$ is a measurable function of $r_{t,\Delta}$ and X_t (Definition A.10), the loss random variable inherits this bounded support. The GPD admits a finite upper endpoint if and only if $\xi_\theta(x) < 0$, in which case the endpoint is $u_\theta(x) + \beta_\theta(x)/|\xi_\theta(x)| < \infty$. For $\xi_\theta(x) \ge 0$ the GPD support is unbounded, which is incompatible with bounded loss. Therefore any correctly specified GPD fit to loss exceedances must satisfy $\xi_\theta(x) < 0$ at every state x. This simultaneously validates the $\xi_\theta(x) < 1$ condition required by Proposition A.13 and Equation (A.20)*, and confirms that $\mathbb{E}[L^\theta_x \mid X = x] < \infty$ under the physical measure. Fitted values of $\xi_\theta(x) \ge 0$ in any epoch stratum should be treated as a model misspecification signal rather than a valid tail estimate.*

Using $\mathbb{E}[L] = \int_0^\infty (1 - F_L(l))\, dl$ for $L \ge 0$, apply the identity under $F_{L|X,\theta}$ and $F^*_{L|X,\theta}$,

then subtract:

$$\Pi_\theta(x) - \mathrm{EL}_\theta(x) = \int_0^\infty \left(F_{L|X,\theta}(l \mid x) - F^*_{L|X,\theta}(l \mid x)\right)\, dl. \tag{A.24}$$

The operational decomposition is

$$\Pi_\theta(x) = \mathrm{EL}_\theta(x) + \mathcal{K}_\theta(x) + c_{\mathrm{liq}}(x, S; \theta), \tag{A.25}$$

where $\mathcal{K}_\theta(x)$ captures tail-capital compensation and $c_{\mathrm{liq}}(x, S; \theta)$ captures execution impact at size S. In implementation, $\mathcal{K}_\theta(x)$ is identified from the distorted tail of Equation (A.18), while c_{liq} is identified from size-conditioned execution gaps and depth.

Definition A.17 (Premium decomposition components)**.** The decomposition

$$\Pi_\theta(x) = \mathrm{EL}_\theta(x) + \mathcal{K}_\theta(x) + c_{\mathrm{liq}}(x, S; \theta)$$

separates expected-loss compensation, tail-capital compensation, and size-conditioned execution-friction compensation.

The execution-friction channel $c_{\mathrm{liq}}(x, S; \theta)$ in Equation (A.25) is identified from two microstructure estimators evaluated at origination.

The first is the Roll autocovariance-based spread estimator (Roll, 1984):

$$\hat{s}_{\mathrm{Roll}} = 2\sqrt{-\operatorname{Cov}(\Delta y_t,\, \Delta y_{t-1})}\, \mathbf{1}\{\operatorname{Cov}(\Delta y_t,\, \Delta y_{t-1}) < 0\}, \tag{A.26}$$

where Δy_t is the first difference of the log-odds price process over a trailing window. When the serial covariance is non-negative, we set $\hat{s}_{\mathrm{Roll}} = 0$ and fall back on a minimum premium floor.

The second is a linear price-impact coefficient in the spirit of Kyle (1985):

$$\hat{\lambda}_{\mathrm{Kyle}} = \frac{\operatorname{Cov}(\Delta y_t,\, v_t)}{\operatorname{Var}(v_t)}, \tag{A.27}$$

where v_t is signed trade flow in USD over the same window. The coefficient measures logit-space price displacement per dollar of net order flow.

The two estimators combine into a size-conditioned liquidation cost:

$$c_{\mathrm{liq}}(x, S) = \tfrac{1}{2}\hat{s}_{\mathrm{Roll}}(x) + \hat{\lambda}_{\mathrm{Kyle}}(x) \cdot S, \tag{A.28}$$

where S is the notional position to be unwound. Both quantities are state-dependent through the trailing window and enter the premium surface at origination through Equation (A.25).

Because $\hat{\lambda}_{\text{Kyle}}$ scales linearly in S, the liquidation channel dominates the risk premium for large positions, which motivates explicit position-sizing limits at the protocol level.

B Data Methodology

B.1 Data and Market Context

Polymarket is a decentralized prediction market platform on Polygon where users trade binary claims that settle to \$1 if an event occurs and \$0 otherwise. Trading is organized via a central limit order book (CLOB), with matching handled off-chain and settlement on-chain. Over the sample window, markets span politics, sports, crypto, macro/finance, and miscellaneous topics.

Our empirical panel combines two public data sources: (i) market metadata from the Gamma API (question text, creation/resolution timestamps, status, cumulative volume, and tags), and (ii) trade-level executions from the subgraph index of `OrderFilled` events (timestamp, market identifier, trade price, token quantity, and USD notional). The raw panel covers 2022–2026 and is transformed into an epoch-level risk panel.

Because markets trade complementary YES/NO claims, all executions are converted to a YES-equivalent price convention:

$$p_t^{\text{YES-eq}} = \begin{cases} p_t^Y, & \text{YES-side trade,} \\ 1 - p_t^N, & \text{NO-side trade.} \end{cases}$$

This creates a single $(0, 1)$-supported process per market and avoids side-label artifacts in downstream estimation.

B.2 Epoch Generation, Filtering, & Train/Test Sets

The empirical target is the state-conditional epoch-loss law $F_{L|X,\theta}(\cdot \mid x)$ from Appendix A, implemented through the contract-to-loss map in Equation (A.13) and the trigger/no-trigger mixture structure in Equation (A.15). Operationally, originations are generated on rolling, overlapping windows (not disjoint blocks), with horizon-specific stride so short-tenor loans are sampled at higher temporal resolution than long-tenor loans. Overlap induces serial dependence, so validation uses forward-time blocking with purge gaps rather than IID-style random folds.

Prices are transformed in logit coordinates (Dalen, 2025) with boundary clipping for numerical stability:

$$\tilde{p}_t = \min\{0.995, \max(0.005, p_t)\}, \qquad y_t = \log\left(\frac{\tilde{p}_t}{1 - \tilde{p}_t}\right), \qquad r_{t,\Delta} = y_{t+\Delta} - y_t,$$

Table B.1: Dataset and Modeling Universe Summary

Raw trades (2022–2026)	397M
Constructed epochs	885M
Markets in epoch panel	158K
Eligible resolved markets (post filters)	170K
Training sample (stratified)	9.9M
Temporal test sample	3.0M
Held-out market test sample	2.0M

with horizons fixed to Equation (A.6). New originations are excluded inside the pre-resolution buffer using the origination gate in Definition A.7, reflecting the late-stage rise in volatility/execution fragility encoded through TTR_t in the state vector (Equation (A.14)).

Eligibility filtering is applied before fitting: minimum activity, minimum market duration, minimum cumulative volume, and resolved-status requirements. This removes economically irrelevant sparse markets while retaining broad cross-category heterogeneity. The resulting panel is then stratified to preserve the joint state geometry used by the conditional loss estimator.

Evaluation is done along two independent out-of-sample axes: (i) forward-time holdout with purge gap and (ii) held-out-market split for cross-market generalization.

B.3 Processing Pipeline

Implementation is anchored to the constrained selection program in Equation (A.4). Proposition A.4 ensures that the admissible optimization is attained under the maintained regularity conditions, while Corollary A.5 gives uniqueness and KKT characterization under convexity/Slater conditions. In practice, calibration is therefore a constrained search for $\theta^*(x)$ over $\mathcal{A}(x)$, with multiplier diagnostics used to identify which underwriting constraints are active by state.

Three empirical features drive the estimator design. First, tails are strongly non-Gaussian at all horizons, so the model cannot rely on a single light-tailed parametric family. Table B.3 provides an illustrative view of tail thickness by epoch. Note that for a Gaussian, kurtosis ~ 3. Second, risk is state-dependent through time-to-resolution, category, and price region interactions. Third, activity is zero-inflated: many epochs have no effective repricing, creating point mass around zero returns and destabilizing residual-standardization methods.

These features motivate the bulk-tail splice in Equation (A.18) and the separate tail-capital vs. liquidity channels in Equation (A.25).

The bulk conditional law is estimated with gradient-boosted quantile regression (Koenker

Table B.2: Quantile Calibration Across Training and Held-Out Evaluation Sets. *Each cell reports the empirical coverage fraction at the nominal quantile level, with signed error* $(\hat{\alpha}-\alpha)$ *in parentheses. Negative error indicates conservative coverage (fewer breaches than nominal); positive indicates anti-conservative. The temporal test set is a forward-time holdout with purge gap; the new-market test set contains markets absent from training entirely.*

Nominal α	Train	Temporal test	New-market test
0.01	0.0088 (−0.0012)	0.0094 (−0.0006)	0.0091 (−0.0009)
0.025	0.0246 (−0.0004)	0.0327 (+0.0077)	0.0300 (+0.0050)
0.05	0.0495 (−0.0005)	0.0580 (+0.0080)	0.0557 (+0.0057)
0.10	0.0989 (−0.0011)	0.1010 (+0.0010)	0.0984 (−0.0016)
0.25	0.2370 (−0.0130)	0.2217 (−0.0283)	0.2191 (−0.0309)
0.50	0.4322 (−0.0678)	0.4128 (−0.0872)	0.4069 (−0.0931)
0.75	0.7620 (+0.0120)	0.7802 (+0.0302)	0.7779 (+0.0279)
0.90	0.9022 (+0.0022)	0.9068 (+0.0068)	0.9055 (+0.0055)
0.95	0.9506 (+0.0006)	0.9477 (−0.0023)	0.9472 (−0.0028)
0.975	0.9756 (+0.0006)	0.9737 (−0.0013)	0.9740 (−0.0010)
0.99	0.9902 (+0.0002)	0.9884 (−0.0016)	0.9888 (−0.0012)

and Bassett, 1978) (LightGBM; Ke et al., 2017) at

$$\tau \in \{0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.975, 0.99\}.$$

Covariates are constructed from market state, horizon structure, temporal proximity to resolution, and execution environment, using only information available at origination. In addition, a proprietary microstructure/liquidity block is included to capture execution frictions and state-dependent impact; these signals enter both the conditional loss model and the liquidity channel in Equation (A.25).

Tail exceedances are then fit with horizon-stratified GPD components (MLE), yielding the EVT splice in Equation (A.18). For admissibility screening in the EVT region, we apply Proposition A.13 to rewrite the ES limit as an equivalent VaR bound (Equation (A.21)), which makes tail-constraint checks numerically simpler during calibration and stress runs.

Raw quantile outputs are then recalibrated. In the initial filtered run, the lower tail was anti-conservative ($Q_{0.01}$ breach rate $\approx 1.53\times$ nominal) and exhibited non-trivial crossing (about 6.97%, e.g., $Q_{0.01} > Q_{0.05}$). We correct this in two steps: conformal quantile shifts δ_τ on a calibration fold (Romano et al., 2019), followed by monotonicity enforcement across τ. Post-correction, $Q_{0.01}$ coverage is aligned to nominal (approximately $1.00\times$) and crossing is eliminated.

Table B.2 reports quantile coverage on the training set and both independent evaluation axes. Three features are worth noting.

First, calibration at the critical left tail, which governs shortfall risk and, by extension,

Table B.3: Tail Thickness by Epoch Length.

Epoch length	Empirical kurtosis
1h	1,768
4h	594
12h	263
24h	166
72h	82
168h	50

the premium charges in Equation (A.23), is tighter on the held-out sets than on the training data. At $\alpha = 0.01$, absolute error falls from 12 bp (train) to 6 bp (temporal) and 9 bp (new-market). This is a consequence of the two-stage correction: conformal shifts δ_τ are tuned on a calibration fold drawn from the test window, so the corrected quantiles are, by construction, adapted to the out-of-sample marginal. More importantly, the improvement transfers to the new-market split indicating that the conditional features generalise the tail structure rather than memorise market-specific idiosyncrasies.

Second, the error pattern is stable across the two test axes. Temporal and new-market profiles are nearly parallel: both exhibit a mild conservative bias ($\hat{\alpha} < \alpha$) below the median and a slight anti-conservative lean above the 75th percentile, with maximum absolute discrepancy concentrated in the interquartile region ($\alpha \in \{0.25, 0.50\}$). The interquartile bias reflects the well-documented difficulty of gradient-boosted quantile estimators in regions of high conditional density (Meinshausen, 2006) and is economically benign for the pricing application: the premium operator in Equation (A.22) loads almost entirely on tail quantiles, where calibration is sharpest.

Third, the near-symmetry of the temporal and new-market error profiles suggests that the dominant source of out-of-sample variation is distributional shift in time, not unobserved market heterogeneity. This is consistent with the epoch-stratified GPD design, which pools tail shape across markets within each horizon stratum and therefore does not rely on market identity for extrapolation. In effect, the model learns a shared tail grammar—indexed by horizon, price region, and category—that ports to unseen markets because the structural drivers of tail risk (boundary effects, resolution dynamics, liquidity withdrawal) are common across the Polymarket universe.

Using Equations (A.13) and (A.15), fitted return risk is mapped into $L \mid X = x$, then priced through Equations (A.22) and (A.23). The main specification calibrates λ_w, and quoted charges inherit both tail-capital loading and microstructure-sensitive liquidity adjustments through $c_{\text{liq}}(x, S)$ in Equation (A.25).

The pipeline output is a state-contingent risk surface rather than a scalar platform rate:

$$x \mapsto (\Pi(x),\ \text{collateral intensity}(x),\ \text{eligibility}(x),\ \text{limits}(x))\,,$$

with rollover treated as fresh underwriting under updated state.

Calibration is completed end-to-end by tuning distortion and business-logic thresholds on calibration windows, then validating under forward-time and held-out-market tests with sensitivity and stress perturbations. Policy guardrails (eligibility, limits, and origination gating) are therefore empirical control variables rather than fixed constants, and are re-estimated as state geometry shifts.

C Additional Tables & Figures

C.1 Probability Integral Transform

Figure C.1 reports the probability integral transform (PIT) histogram. Under correct calibration, PIT values are approximately uniform on $[0, 1]$. In aggregate, the histogram is close to uniform, with low-tail frequencies near target: $P(\text{PIT} < 1\%) = 0.0100$ and $P(\text{PIT} < 5\%) = 0.0545$. This indicates accurate extreme-tail calibration with mild conservatism in the broader lower tail, a directionally desirable property for lending applications. The spike at ~ 0.5 is an artifact of a zero-inflated return structure: many epochs see no trades, so the realized logit return sits exactly at the conditional median. This is can be corrected by resampling for each point mass from a uniform distribution over the CDF's spike interval. (Czado et al., 2009). This is unnecessary as it is purely a cosmetic fix, and the PIT tails show adequate fit.

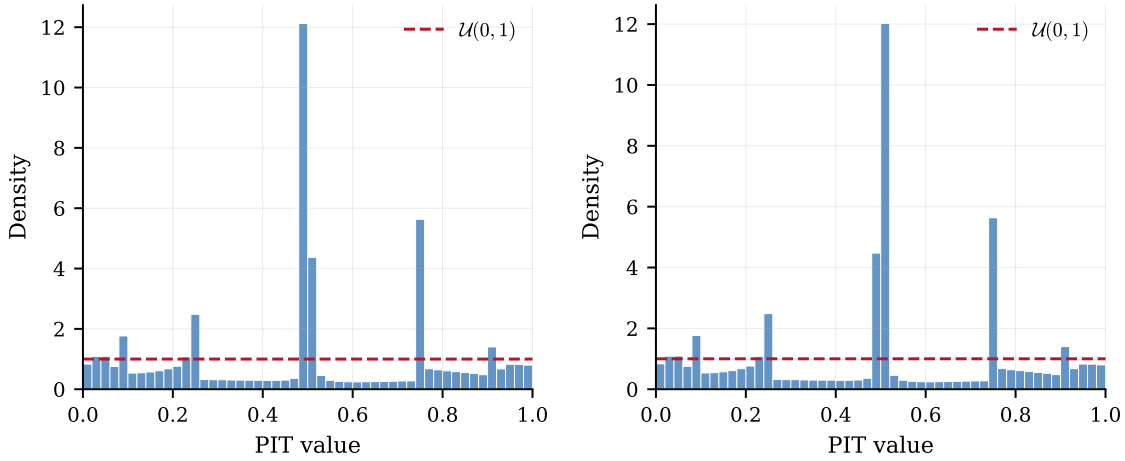

Figure C.1: *PIT histogram versus $U(0, 1)$ benchmark for the fitted conditional law. Bars near the 1% and 5% regions provide a direct visual check of lower-tail coverage accuracy relevant for risk thresholds.*

C.2 Wang Distortion at $\lambda_w = 0.50$: Nominal-to-Physical Tail Mapping

Nominal tail level α	Implied physical level	Severity multiplier
0.001	0.000165	6.1×
0.005	0.001050	4.8×
0.010	0.002354	4.2×
0.025	0.006948	3.6×
0.050	0.015982	3.1×

C.3 Conformal Quantile Shifts δ_τ on Calibration Fold

τ	δ_τ	interpretation
0.010	-0.0042	model too high $\rightarrow$ pull down
0.025	-0.0001	well-calibrated
0.050	-0.0004	well-calibrated
0.100	$+0.0000$	well-calibrated
0.250	$+0.0000$	well-calibrated
0.500	$+0.0000$	well-calibrated (point mass near zero)
0.750	-0.0000	well-calibrated
0.900	-0.0000	well-calibrated
0.950	$+0.0000$	well-calibrated
0.975	$+0.0003$	well-calibrated
0.990	$+0.0031$	model too low $\rightarrow$ push up

C.4 Sensitivity of Backtest Metrics to the Wang Distortion Parameter Λ (LTV $= 50\%$)

λ	Premium (bps)	Comb. Ratio	Shortfall	Sharpe	Avg EL (bps)
0.05	16.9	47.0%	2.0%	0.089	-9.0
0.10	17.9	44.6%	1.9%	0.098	-9.9
0.25	20.7	38.5%	1.9%	0.125	-12.7
0.50	26.2	30.3%	1.8%	0.173	-18.3
0.70	32.0	24.9%	1.7%	0.213	-24.0
1.00	44.0	18.1%	1.5%	0.272	-36.0

D Proofs

D.1 Proof of Proposition A.4

Proposition (Well-posedness and existence)**.** *Fix $x \in \mathcal{X}$. Suppose:*

1. *Θ is nonempty and compact;*
2. *$U(\cdot, x)$ is upper semicontinuous on Θ;*
3. *$\theta \mapsto \mathbb{P}(L_x^\theta > \bar{l}(x) \mid X = x)$ is lower semicontinuous;*
4. *$\theta \mapsto \mathrm{ES}_{\alpha,\theta}(x)$ is lower semicontinuous;*
5. *$\mathcal{A}(x) \neq \varnothing$.*

Then $\mathcal{A}(x)$ is compact and the argmax set in (A.4) *is nonempty.*

Proof. Fix $x \in \mathcal{X}$, and define the maps on Θ:

$$g_1(\theta) := \mathbb{P}(L_x^\theta > \bar{l}(x) \mid X = x), \qquad g_2(\theta) := \mathrm{ES}_{\alpha,\theta}(x), \qquad g_3(\theta) := \underline{U}(x) - U(\theta, x).$$

Then

$$\mathcal{A}(x) = \{\theta \in \Theta : \ g_1(\theta) \leq \epsilon(x), \ g_2(\theta) \leq \bar{e}(x), \ g_3(\theta) \leq 0\}.$$

By assumption, g_1 and g_2 are lower semicontinuous (l.s.c.). Also, $U(\cdot, x)$ is upper semicontinuous (u.s.c.), hence $-U(\cdot, x)$ is l.s.c.; adding the constant $\underline{U}(x)$ preserves l.s.c., so g_3 is l.s.c.

For any l.s.c. function g, the sublevel set $\{\theta : g(\theta) \leq c\}$ is closed. Therefore, each set

$$A_1 := \{\theta \in \Theta : g_1(\theta) \leq \epsilon(x)\}, \quad A_2 := \{\theta \in \Theta : g_2(\theta) \leq \bar{e}(x)\}, \quad A_3 := \{\theta \in \Theta : g_3(\theta) \leq 0\}$$

is closed in Θ. Hence

$$\mathcal{A}(x) = A_1 \cap A_2 \cap A_3$$

is closed in Θ. Since Θ is compact, every closed subset of Θ is compact; thus $\mathcal{A}(x)$ is compact. Assumption (5) gives $\mathcal{A}(x) \neq \varnothing$.

Now consider maximizing $U(\cdot, x)$ over $\mathcal{A}(x)$. Because $U(\cdot, x)$ is u.s.c. and $\mathcal{A}(x)$ is nonempty compact, the Weierstrass extreme value theorem implies that $U(\cdot, x)$ attains its maximum on $\mathcal{A}(x)$. Therefore,

$$\arg\max_{\theta \in \mathcal{A}(x)} U(\theta, x) \neq \varnothing.$$

Since this is equivalent to the constrained problem, the argmax set in (A.4) is nonempty.

Therefore, $\mathcal{A}(x)$ is compact and the policy selection problem is well posed with at least one maximizer. ■

D.2 Proof of Corollary A.5

Corollary (Uniqueness under convexity)**.** *Fix $x \in \mathcal{X}$. In addition to the assumptions of Proposition A.4, suppose:*

1. *$\Theta \subset \mathbb{R}^m$ is convex;*
2. *$U(\cdot, x)$ is strictly concave on Θ;*
3. *the constraint maps*

   $$g_1(\theta) := \mathbb{P}(L_x^\theta > \bar{l}(x) \mid X = x) - \epsilon(x), \quad g_2(\theta) := \mathrm{ES}_{\alpha,\theta}(x) - \bar{e}(x), \quad g_3(\theta) := \underline{U}(x) - U(\theta, x)$$

   are convex in θ;
4. *(Slater condition) there exists $\tilde{\theta} \in \Theta$ such that $g_i(\tilde{\theta}) < 0$ for $i = 1, 2, 3$.*

Then the feasible set $\mathcal{A}(x)$ is convex and nonempty, (A.4) *has a unique optimizer $\theta^*(x)$, and KKT conditions are necessary and sufficient for optimality.*

Proof. By convexity of Θ and convexity of each g_i, the sublevel sets

$$\{\theta \in \Theta : g_i(\theta) \leq 0\}, \quad i = 1, 2, 3,$$

are convex; hence

$$\mathcal{A}(x) = \bigcap_{i=1}^{3} \{\theta \in \Theta : g_i(\theta) \leq 0\}$$

is convex. Nonemptiness follows from Slater (strict feasibility implies feasibility), and compactness follows from Proposition A.4.

Existence of an optimizer follows from Proposition A.4. To show uniqueness, let $\theta_1 \neq \theta_2 \in \mathcal{A}(x)$ and $\lambda \in (0, 1)$. Since $\mathcal{A}(x)$ is convex,

$$\theta_\lambda := \lambda\theta_1 + (1 - \lambda)\theta_2 \in \mathcal{A}(x).$$

Strict concavity of $U(\cdot, x)$ yields

$$U(\theta_\lambda, x) > \lambda U(\theta_1, x) + (1 - \lambda)U(\theta_2, x).$$

Hence two distinct maximizers cannot exist; therefore the maximizer is unique.

Finally, maximizing a concave objective over convex constraints is equivalent to minimizing a convex function $-U$ over a convex feasible set. Under Slater, strong duality holds and KKT conditions are necessary and sufficient. Therefore there exist multipliers $\mu_i \geq 0$ satisfying stationarity, primal feasibility, dual feasibility, and complementary slackness, and any point satisfying KKT is the unique optimizer. ■

D.3 Proof of Proposition A.13

Proposition (ES-VaR cap equivalence in the EVT tail region)**.** *Fix* (x, θ)*. Suppose* $z_{\alpha,\theta}(x) := \mathrm{VaR}_{\alpha,\theta}(x) \geq u_\theta(x)$ *and* $\xi_\theta(x) < 1$*. Then the ES admissibility condition* $\mathrm{ES}_{\alpha,\theta}(x) \leq \bar{e}(x)$ *is equivalent to the VaR cap*

$$z_{\alpha,\theta}(x) \leq (1 - \xi_\theta(x))\,\bar{e}(x) - \beta_\theta(x) + \xi_\theta(x)\,u_\theta(x).$$

Proof. From Equation (A.20),

$$\mathrm{ES}_{\alpha,\theta}(x) = z_{\alpha,\theta}(x) + \frac{\beta_\theta(x) + \xi_\theta(x)\big(z_{\alpha,\theta}(x) - u_\theta(x)\big)}{1 - \xi_\theta(x)}.$$

Collect terms in $z_{\alpha,\theta}(x)$:

$$\mathrm{ES}_{\alpha,\theta}(x) = \frac{z_{\alpha,\theta}(x) + \beta_\theta(x) - \xi_\theta(x)u_\theta(x)}{1 - \xi_\theta(x)}.$$

Because $\xi_\theta(x) < 1$, we have $1 - \xi_\theta(x) > 0$, so

$$\mathrm{ES}_{\alpha,\theta}(x) \leq \bar{e}(x) \iff z_{\alpha,\theta}(x) + \beta_\theta(x) - \xi_\theta(x)u_\theta(x) \leq (1 - \xi_\theta(x))\bar{e}(x).$$

Rearranging gives

$$z_{\alpha,\theta}(x) \leq (1 - \xi_\theta(x))\,\bar{e}(x) - \beta_\theta(x) + \xi_\theta(x)\,u_\theta(x),$$

which is Equation (A.21). ■