Engine vs. the Literature

A Best-Practices Review of the trading-app-saas Options Engine (v3.95.0) Against Published Quantitative Research

Prepared: August 16, 2026

Scope: Cross-reference of ENGINE_SPEC.md (v3.95.0, 2026-08-15) against academic and industry research on options signal generation, volatility modeling, Greeks computation, execution quality, VIX and regime conditioning, probability-of-profit estimation, price-range forecasting, position sizing, and ML validation. Approximately 130 primary sources were reviewed; the consolidated bibliography appears in Section 10.

1. Executive Summary

The engine's core economic thesis — systematically selling defined-risk, 30–45 DTE, 0.15–0.30-delta credit structures on liquid underlyings — is the single best-documented edge in the options literature: the volatility risk premium (Bakshi & Kapadia [1]; Carr & Wu [3]; Coval & Shumway [2]; CBOE PUT index studies [4][5]). Its risk architecture (deterministic gates, defined-risk-only, kill switches, de-risking after vol spikes, inverse-variance-style sizing, calibration monitoring) aligns with the literature's strongest prescriptions, in several places more closely than typical institutional practice. The López de Prado ML validation stack (triple-barrier labels, purged CV with embargo, deflated-Sharpe backtesting, drift monitoring) is textbook state-of-the-art in design.

Against that strong baseline, the review found **4 divergences** (places where the engine is on the wrong side of published evidence), **9 material gaps** (best practices absent), and **12 refinements** (supported features that the literature would tune differently). The highest-priority items:

1. **The technical-composite entry signal is the weakest link.** Five of its six votes (RSI, MACD, Bollinger, MA-cross state, RSI-extreme) come from a signal family whose profitability does not survive data-snooping correction or transaction costs in modern data (Sullivan, Timmermann & White [24]; Park & Irwin [25]). The literature-supported core of the composite is the trend/momentum component — everything else is mostly re-counting one correlated signal.
2. **The pre-earnings short-vol calendar builder sells the run-up window that the best evidence says is *underpriced*.** Gao, Xing & Zhang [32] show ATM straddles bought ~3 days before announcements earn +3.34% on average into the event. Short-front-month structures entered 1–7 days pre-earnings are on the wrong side of this unless conditioned on implied-move ≥ historical realized move (Dubinsky et al. [31]).
3. **The PEAD scanner targets an anomaly that is documented as dead in its own universe.** Martineau [30] shows post-earnings-announcement drift disappeared in large/mid caps around 2006; it survives only in microcaps the engine's liquidity gates would exclude.
4. **Greeks inputs, not Greeks math, are the pricing problem**: European BSM on American-style equity options with q = 0 and a hardcoded r = 5.3% produces class-level errors — 1–3 vol points on dividend payers around ex-dates [45][40], put early-exercise premia of 5–15% of option value at 5% rates (Barraclough & Whaley: \$1.9 B in foregone value from unexercised puts [43]), and √T-scaled IV distortion on LEAPS from rate staleness [46].

5. **The engine's own IVR ≥ 35 entry gate has no published support, while the well-supported gate (its NVRP metric) is computed but not used as the gate.** Goyal & Saretto [13] and Bollerslev, Tauchen & Zhou [11] support conditioning on IV *minus realized vol* — the engine should promote NVRP from display flag to primary vol-entry condition.

The single highest-value cheap addition, supported across three research areas: **close the calibration loop** — log forecast POP on every trade and continuously test it against realized outcomes with reliability diagrams and Brier decomposition (Gneiting & Raftery [111]; Riley et al. [110]). The engine's calibration_report module already implements most of this; it should be promoted from diagnostic to gating input.

Alignment scorecard

Domain	Verdict	Headline
VRP harvesting core	**Aligned**	Best-documented edge in the literature; expect compressed premia post-2006
Technical composite	**Divergent**	Oscillator votes don't survive data-snooping tests; trend component is the supported core
Skew/option-implied signals	**Aligned (partial)**	Matches Cremers-Weinbaum / Xing-Zhang-Zhao families; expect post-publication decay
PEAD scanner	**Divergent**	Anomaly documented dead in liquid names since ~2006
Pre-earnings calendars	**Divergent**	Sells an underpriced window; needs implied-vs-historical-move condition
RV estimation	**Deviation**	Yang-Zhang should be primary, not confined to NVRP
Vol forecasting	**Gap**	No HAR/GARCH; IV-anchored forecast is the cheap upgrade
IV surface / skew metrics	**Partial**	Delta-interpolated 25Δ is fine; normalize skew by ATM (Mixon), fit slices in total variance
IV rank	**Partial**	True-IV-rank store is the right fix; prefer percentile over min-max rank
Greeks engine math	**Aligned**	Merton formulas, vectorization, conditioning guard all sound
Greeks inputs (r, q, exercise style, T)	**Gap**	Stale rate, q=0, no American overlay, calendar-day T at short DTE
GEX	**Aligned**	Index-only sign trust matches the evidence precisely
Execution tactics	**Aligned**	Mid-anchored package limits, open avoidance, spread gates all supported
Slippage measurement	**Partial**	Add decision-time benchmark + opportunity cost of unfilled orders
Paper-fill realism	**Partial**	Concede in half-spread units; cap at natural price; calibrate vs live
VIX regime bands	**Partial**	Fixed bands weak; term-slope sizing is the engine's best-supported VIX rule
Macro/OPEX blackouts	**Partial**	Justified as tail control, not EV; OPEX ban weakest
POP methodology	**Partial**	N(d2)-at-breakeven correct but risk-neutral and single-sigma; label it, then calibrate
Exit rules (50%/2×/21 DTE)	**Aligned (practitioner-grade)**	Variance control, not alpha; verify on own fills
Position sizing	**Aligned (partial)**	Half-Kelly + 2% floor is textbook; add portfolio-level correlation constraint

Domain	Verdict	Headline
ML validation stack	**Aligned (partial)**	Design is best-in-class; small-sample settings (isotonic, embargo, PSI thresholds) need tightening

---

2. Methodology

The reference point is ENGINE_SPEC.md generated 2026-08-15 from static analysis of the full v3.95.0 source tree (328 Python files). Literature was gathered by six parallel research sweeps (signal generation; volatility modeling; pricing/Greeks; execution quality; VIX/regimes/events; POP/sizing/ML validation), prioritizing peer-reviewed journals (JF, RFS, JFE, JFQA, Management Science, Journal of Business), then NBER/SSRN working papers, then named practitioner research where academic coverage is absent (noted explicitly wherever the evidence is practitioner-grade). Magnitudes are quoted gross of costs unless stated. Verdict vocabulary: **Aligned** (engine matches best practice), **Partial** (supported direction, tuning differs), **Deviation** (engine choice measurably inferior to a documented alternative), **Divergent** (engine is on the wrong side of the published evidence), **Gap** (documented best practice absent).

---

3. Signal Generation

3.1 The core edge: volatility risk premium harvesting — ALIGNED

Engine: Default strategy set = bull put / bear call spreads and iron condors, 30–45 DTE, short delta 0.15–0.30, on curated liquid universes, gated on vol richness.

Literature: Delta-hedged option positions lose money for buyers systematically — the negative volatility risk premium (Bakshi & Kapadia [1]). Zero-beta ATM S&P straddles lose ≈3% per week (Coval & Shumway [2]); index variance-swap sellers earned persistently high Sharpe ratios (Carr & Wu [3]). The CBOE PUT index earned equity-like returns at two-thirds the volatility over 1986–2018 (Sharpe 0.65 vs 0.49), with an average VIX-minus-realized gap of 4.2 vol points [4]. Israelov & Nielsen [7] show the rewarded component of option-selling strategies is precisely the short-vol exposure, not the residual delta. Israelov & Tummala [8] find the per-unit-risk premium concentrated in short-dated, moderately-OTM options, with puts richer than calls.

Cross-reference: The engine's design is directly in the supported family. Three caveats the spec should carry: (a) the premium has compressed materially post-2006 (PUT Sharpe 0.50 vs SPX 0.51 over 2006–2018 [4]) — expectations should be set off the modern subsample; (b) the evidence base is index/large-cap — the engine's liquid-universe restriction is therefore load-bearing, not cosmetic; (c) there is **no peer-reviewed validation of the specific 45-DTE / 50%-profit conventions** — they are defensible gamma/transaction-cost compromises and variance-control rules (see §8.2), not documented optima.

3.2 The IV-rank entry gate — DEVIATION

Engine: Credit entries require IV rank ≥ 35 (min-max rank over 252 days; true-IV-rank store replacing an HV-proxy after 60 observations). NVRP = (IV−RV)/RV with Yang-Zhang RV is computed and displayed (CHEAP_VOL < 0, FAIR < 0.10, PREMIUM above) but is not the entry gate.

Literature: No peer-reviewed evidence supports IV rank as an entry filter — it is a practitioner construct with documented pathologies: a single vol spike compresses rank for a full year, and the spike day's exit from the window causes a discontinuity [63][64]. What *is* documented: conditioning on the IV−RV spread. Goyal & Saretto [13] show sorting options on (HV − IV) produces economically large straddle returns (a net-of-costs replication reports ≈3.9%/month on the decile spread [14]); Bollerslev, Tauchen & Zhou [11] show the variance risk premium predicts returns. Cheng [77] adds the crucial regime caveat: the ex-ante premium *falls or turns negative right after vol spikes* — exactly when min-max IV rank reads highest.

Recommendation: Promote NVRP (already horizon-matched and Yang-Zhang-based — better-built than most practitioner VRP proxies) to the primary vol-entry condition, with IV percentile (not min-max rank) as the secondary display metric. The failure mode this fixes is concrete: post-spike entries where IVR ≥ 35 passes while IV−RV is thin or negative.

3.3 The technical composite — DIVERGENT

Engine: Six equal votes (RSI signal, MACD signal, EMA-50/200 cross, Bollinger touch, RSI extremes, price vs SMA200) with a ×1.25/×0.75 trend-regime multiplier; BUY/SELL at ±2.0.

Literature: The classic pro-technical result (Brock, Lakonishok & LeBaron [23]) does not survive modern scrutiny: Sullivan, Timmermann & White's reality-check bootstrap over 7,846 rules finds the best rules provide no out-of-sample edge post-1986 [24], and Park & Irwin's survey concludes most positive studies are contaminated by data snooping and vanish after costs [25]. **No robust peer-reviewed evidence exists for RSI or MACD specifically at daily frequency after costs.** What survives: time-series momentum (Moskowitz, Ooi & Pedersen [27] — Sharpe > 1 diversified, robust to costs) and trend/momentum indicators as *market-level* forecasters complementary to fundamentals (Neely et al. [28]). On combining: the forecast-combination literature (Timmermann [35]) supports equal weighting — but only over *diverse, individually validated* signals; five correlated momentum-family oscillators have an effective N of roughly two.

Cross-reference: The engine's trend multiplier and MA-stack/price-vs-200MA components are the supported core. The RSI/MACD/Bollinger votes add correlated noise with no documented standalone edge. Mitigating context: in this engine the composite is a *candidate ranker* feeding a VRP-based strategy, not a standalone alpha — the harm is misallocated scan attention rather than direct losses. Still, the vote design overstates its own information content.

Recommendation: Collapse the composite toward its trend/momentum core (MA stack, price vs 200MA, ADX regime, plus a 12-month TSMOM term per [27]); validate any retained oscillator out-of-sample on the engine's own selection-ledger history before it keeps a vote.

3.4 Option-implied signals (skew, call-put spread) — ALIGNED (PARTIAL)

Engine: Call-skew inversion detection (rich OTM calls → sell into bull call spreads), 25Δ risk reversal, put-skew steepness for bull-put selection; skew_25d as an agent feature.

Literature: This is a documented signal family: stocks with relatively expensive calls outperform (call−put IV spread ≈ +51 bps/week, Cremers & Weinbaum [18]); steep put smirks predict underperformance of ≈10.9%/yr (Xing, Zhang & Zhao [17]); ΔIV innovations predict returns (An, Ang, Bali & Cakici [20]); O/S volume ratio predicts (Johnson & So [21]); single-name IV term-structure slope predicts straddle returns (Vasquez [22]).

Cross-reference: Selling *rich* wings detected by skew shape is consistent with this literature, and the engine's usage (structure/strike selection tilt rather than standalone entry) matches the recommended application. Caveats: all these are cross-sectional signals with documented post-publication decay, strongest in smaller/info-asymmetric names where the engine's liquidity gates bind; and the skew_analyzer's spot-normalized slope is good practice, but the surface-level put/call skews should be normalized by ATM IV before any cross-name comparison (Mixon [56]).

3.5 Chart patterns and the Launcher scanner — PARTIAL

Literature: Lo, Mamaysky & Wang [26] — the strongest pro-pattern study — finds patterns carry statistically significant *conditional-distribution* information but explicitly does not establish profitability. Breakout/range rules fail modern data-snooping tests [24]. The Launcher scanner's ingredients (proximity to 52-week high, prior run, relative strength) sit adjacent to documented momentum effects [27], but the 7-criteria bundle as such is unvalidated practitioner methodology.

Cross-reference: The engine's use of patterns is defensible *because of where they sit*: the tape gate uses them to veto short strikes inside measured-move zones (risk filtering), not to generate entries. That is the literature-consistent use of pattern information. The whipsaw demotion and implausible-target penalties are sensible guards with no academic analog to contradict them. Flag: the range-regime module's **Hurst exponent** input is unreliable at 60–252-bar windows — random walks routinely produce $\hat{H}$ far from 0.5 at these sample sizes (Couillard & Davison [61]) — so it should carry the least weight of the three regime conditions (ADX, BB-width percentile, Hurst).

3.6 Earnings-adjacent scanners — DIVERGENT (two of three)

PEAD scanner: Bernard & Thomas's drift [29] is documented as *dead in large/mid caps since ~2006* (Martineau [30]); it survives only in microcaps that fail the engine's own optionability/liquidity gates. As built (EPS surprise ≥ 10%, liquid optionable names), the scanner targets an extinct anomaly. Recommendation: retire it, or re-scope explicitly to the LEAPS-entry timing role with the NVRP gate doing the real work.

Pre-earnings short-vol calendars: The builder sells front-month IV 1–7 days before announcements. Gao, Xing & Zhang [32] show the pre-announcement run-up window is on average *underpriced* — long straddles bought 3 days out earn +3.34% into the event — while the crush is harvestable only *through* the event and only when the implied move exceeds a good forecast of the realized move (Dubinsky et al. [31] provide the extraction method; Wright [86] shows scheduled-event variance is overpriced on average, roughly half the FOMC-day implied bump being premium). Recommendation: add an implied-move ≥ historical-realized-move condition as a hard gate on the calendar builder, and note that its short leg dying *before* the announcement (the builder's current design for AMC prints) partially mitigates but does not eliminate the wrong-side exposure.

Earnings blackout (agent): Supported for unattended operation — single-name gap risk is unhedgeable overnight, and the blackout is correctly framed as tail control. The whitelist mechanism is the right hook for the documented middle path (defined-risk premium through events only when implied ≥ historical move).

---

4. Volatility Modeling

4.1 Realized-vol estimation — DEVIATION

Engine: 20-day close-to-close log-return std × $\sqrt{252}$ is the primary RV everywhere (regime bands, expiry scoring, cones); Yang-Zhang is used only inside NVRP; the spec explicitly freezes close-to-close because downstream bands are tuned to it.

Literature: Efficiency per observation vs close-to-close: Parkinson ≈5× [50], Garman-Klass ≈7.4× [51], Rogers-Satchell ≈6× (drift-robust) [52], Yang-Zhang up to ≈14× theoretical, ≈7–8× practical [16][53]. Molnár [53] adds that single-period √variance estimates are biased low ≈25% for close-to-close vs 3–4% for GK/Parkinson. A 20-day YZ estimate carries roughly the precision of a 140-day close-to-close window without the staleness.

Cross-reference: The engine has the better estimator implemented and quarantined. The "bands are tuned to CC" freeze is a legitimate migration constraint, not a justification — the literature-consistent path is a one-time recalibration of the 1.20/0.85 regime cutoffs against YZ-based ratios, then promotion of YZ to primary. Until then, every iv_hv_ratio consumed by the selector, conviction engine, and expiry optimizer carries ≈2–3× more estimation noise than necessary.

4.2 Volatility forecasting — GAP

Engine: No GARCH/HAR forecasting; HV20 is a backward-looking estimate; EWMA (20-day half-life) only for sizing.

Literature: Implied vol is the best single forecaster of future realized vol in the majority of horse races (Poon & Granger's 93-study survey [55]; Christensen & Prabhala [12]; Blair, Poon & Taylor [17b]); GARCH(1,1) remains the time-series benchmark (Hansen & Lunde [54]); HAR-RV (Corsi [15]) — three OLS coefficients on daily/weekly/monthly RV — is the modern default from daily data and typically beats GARCH out of sample.

Recommendation (cheap, ordered): (1) where a *forward* vol is needed (cones, POP, expected move), use ATM IV de-biased by the ticker's trailing IV/RV ratio rather than raw HV20; (2) add a 3-coefficient HAR on daily YZ estimates; (3) skip heavier GARCH stacks — the marginal value at this scale is negligible per [54]. The EWMA inverse-variance sizing is separately well-supported (Moreira & Muir [79]; Fleming, Kirby & Ostdiek [57] — realized-measure inputs roughly double the economic value of vol timing).

4.3 IV surface mechanics — PARTIAL

Engine: Per-expiry ATM from nearest strike; 25Δ via linear interpolation of IV on chain deltas (moneyness proxy fallback); term slope = back − front ATM IV; no arbitrage checks.

Literature: Industry standard is per-slice SVI/SSVI in *total implied variance vs log-moneyness* with explicit no-butterfly/no-calendar constraints (Gatheral [58]; Gatheral & Jacquier [59]); naive interpolation on sparse noisy retail quotes readily produces butterfly violations [60]. Delta-space linear interpolation of two smile points is essentially a crude Vanna-Volga input set [62] — acceptable for *monitoring*, not pricing. Skew comparability requires normalizing by ATM IV (Mixon [56]).

Cross-reference: For a monitoring/selection engine, current practice is defensible. Three cheap upgrades: fit a parabola or SSVI-lite per expiry across all liquid quotes (noise suppression against wide retail spreads); interpolate in total variance; verify w(k,T) monotonicity across expiries before trusting the term-slope value. The agent's skew_25d feature (already ATM-normalized) is the right convention — extend it to the surface module's raw put_skew/call_skew fields.

4.4 Regime detection — PARTIAL

The engine's roadmap note ("HMM regime seam reserved") matches the literature's direction: 2–3 state Markov-switching vol models are robustly identified and improve short-horizon forecasts (Hamilton [65]; Ang & Timmermann [66]); statistical jump models reduce regime flickering [67]. ADX/BB-width heuristics have no forecasting literature and the Hurst input is unreliable at these windows [61] — the current range-regime module should be treated as a tradability *heuristic* (which is how the tape gate uses it), and the planned HMM should become the statistical regime layer.

---

5. Greeks Computation & Pricing

5.1 Engine math — ALIGNED

The Merton formula set (with q plumbed), the vectorized bisection IV engine with no-arbitrage masks, the vega-conditioning guard, and closed-form second-order Greeks are all sound. The May-era put-charm defect is confirmed fixed. The conditioning guard (reject IV where a $0.01 tick moves IV > 2 points) is a legitimate quality filter for exactly the deep-OTM regime the literature warns about; the refinement is to keep it as a *data-quality flag* rather than a solver failure, since a robust solver still produces an exact answer to a noisy input.

5.2 Exercise style — GAP

Engine: European BSM on American-style equity options; ex-dividend *dates* fetched for an assignment guard only.

Literature: The early-exercise premium is near zero for calls on non-dividend names but material for (a) **ITM puts at today's ~5% rates** — put EEP grows with rates and moneyness, commonly 5–15% of option value for deep-ITM medium-dated puts [41][42]; Barraclough & Whaley [43] document $1.9 B in value destroyed by *failure* to exercise puts optimally, i.e., the premium is economically first-order; and (b) **calls spanning ex-div dates** — Pool, Stoll & Whaley [44] show over half of should-be-exercised call OI goes unexercised, with market makers harvesting the difference. Cboe itself computes American IVs on CRR trees with discrete dividends [40]. Second-order effect: solving *European* IV from *American* market prices biases IV high for ITM puts, corrupting every downstream Greek and the put side of skew metrics.

Recommendation: Add a closed-form American overlay (Barone-Adesi-Whaley [38] or Bjerksund-Stensland [39], both vectorizable) or a ~100-step CRR tree for ITM contracts; flag puts where $K\cdot(1-e^{-rT})$ exceeds remaining time value as early-exercise candidates. The existing ex-div guard covers the call side's *event* but not the pricing bias.

5.3 Dividends — GAP

q is plumbed but always 0. Ignoring dividends biases call IV low / put IV high by commonly 1–3 vol points near ATM for quarterly payers with an ex-date inside 30–60 DTE [45][40] — the same order as the engine's own conditioning threshold. Minimum fix: populate q from trailing yield (closed-form, first-order correction); better: fetch amounts (dates are already fetched) and use S − PV(dividends) as the forward input, reserving tree handling for ITM/long-dated. Vendor-IV research (Wallmeier [47][48]) shows dividend mishandling is a leading cause of cross-vendor IV disagreement — relevant because the engine mixes Schwab, ORATS-derived (Tradier), and self-computed (Alpaca) IVs.

5.4 Risk-free rate — DEVIATION

A hardcoded r = 0.053 will drift 100+ bp across a rate cycle. The IV error scales with $\sqrt{T}$ (≈0.2–0.3 vol pt at 60 DTE per 50 bp, 0.5–1.0 pt at LEAPS tenors) and every EEP calculation depends on r. The standard is OIS/SOFR discounting (Hull & White [46b]); for option-embedded rates specifically, box-spread rates run ≈35–40 bp above Treasuries (van Binsbergen, Diamond & Grotteria [46]). **Cheapest material fix in this section:** pull SOFR (or CMT) daily and interpolate per expiry — Cboe's methodology uses spline-interpolated CMT [40]. Affects LEAPS modules most.

5.5 Time convention — PARTIAL

T = calendar/365 is standard and fine ≥ 60 DTE; below ~30 DTE it produces the documented Friday IV-collapse artifact (weekend variance is priced out before the weekend — French & Roll [49]; practitioner treatments [49b][49c] quantify calendar-vs-business-time divergence at up to ~5 vol points inside the final 3 months). The engine trades 30–45 DTE mainstream (mild exposure) but its 7–14 DTE short-DTE scanner sits squarely in the distortion zone. Recommendation: minutes-based T (the VIX convention) for the short-DTE path, and treat Friday-afternoon short-dated IVs with vol-time skepticism.

5.6 IV solver — REFINEMENT

Brent/bisection is functional but strictly dominated by Jäckel's "Let's Be Rational" [37] — machine precision in two iterations for all inputs, open-source via py_vollib, faster for vectorized chains, robust exactly where bisection loses digits (price near intrinsic). Drop-in swap; low priority relative to §5.2–5.4 because solver error is dwarfed by input error.

5.7 GEX — ALIGNED

The engine trusts net-GEX *sign* only for index tickers and uses walls as placement nudges. This matches the evidence precisely: the customers-long-puts/short-calls sign convention is defensible at the index level (SqueezeMetrics [68]; Barbon & Buraschi's gamma-fragility mechanism [69]; Baltussen et al.'s hedging-demand intraday momentum across 60+ markets [70]) and known-unreliable for single names, where flow-based inventory estimates frequently contradict it (formalized by [71]). The v3.84 index-only restriction is exactly the published-best-practice boundary. Residual refinement: document the sign convention on the GEX surface (the flip *level* is convention-invariant; its *interpretation* is not [71]).

---

6. Execution Quality

6.1 What the engine gets right — ALIGNED

The literature's strongest execution findings map directly onto existing engine behavior:

- **Quoted spreads overstate achievable cost.** Effective half-spreads run ≈76% of quoted; execution-timing traders pay ≈42% of the naive effective spread (Muravyev & Pearson [90]). A patient mid-anchored limit strategy is precisely the documented cost-capture behavior.
- **Package execution for multi-leg.** Complex order books exist to eliminate leg risk; the engine never legs. Correct — with the caveat that package effective spreads are empirically *wider* than sum-of-mids suggests (Chaput & Ederington [95]; Flint et al. [96]), so condors should be expected to cost roughly legs × per-leg concession, and 2-leg verticals are structurally cheaper to trade than 4-leg condors.
- **Avoiding the open.** Options spreads are widest at the open and decline through the morning with no closing U-shape (Chan, Chung & Johnson [94]) — the 15-minute entry quiet window (and 20-minute sandbox exit window) is directly supported.
- **The natural price as worst-case bound** (short bid − long ask) is the correct pessimistic benchmark; retail auction fills average roughly mid + 25–50% of half-spread (Hendershott et al. [93]; Bryzgalova et al. [92]; Ernst & Spatt [92b]).
- **Spread-percentage as the binding liquidity gate with OI corroboration** matches the determinants literature (spread is the direct cost measure; OI/volume are indirect, stale corroborators — Wei & Zheng [98]; Christoffersen et al. [99]). The engine's $0.10 tick-tolerance floor already fixes the known percentage-gate pathology on cheap wings.

6.2 Refinements the literature supports

1. **Reprice cadence and units.** Fill hazard decays quickly with time-in-book (Lo, MacKinlay & Zhang [100]); a 60-minute stale threshold is slow. Reprice on *quote movement* or a minutes-scale cadence, and express concessions in fractions of the half-spread rather than percent-of-mid (percent-of-mid conflates premium level with spread width — this applies to both the live escalation ladder and the paper fill ladder's 5%-of-mid steps).

2. **Implementation-shortfall accounting.** Measuring slippage as fill-vs-mid on *completed* orders is the classic survivorship distortion Perold's framework exists to fix [101]: unfilled/expired orders must be charged opportunity cost, and the benchmark should be the decision-time net mid, not the fill-time mid. The engine's execution-stats module has the data (intents carry staged-time premium and terminal state); adding an expired-order opportunity-cost line closes the gap.
3. **Delayed-feed relaxation is backwards.** The 25% spread gate for yfinance-sourced chains relaxes the gate exactly where the *mid itself* is least reliable. The literature-consistent posture is stricter treatment (or refusal) on delayed feeds — which the auto-agent already implements via its Schwab-only chain filter; the relaxation should not leak into any traded path.
4. **Paper-fill realism.** Mid-fill-always is the documented optimism bias (Lo et al. [100]; practitioner convention brackets fills between mid and natural [103][104]). The engine's ladder is structurally right; bound simulated fills at the natural price, apply a sub-100% mid-fill probability, and calibrate against live fills using the per-plane slippage telemetry it already records. The sim-vs-sandbox gap metric in journal insights is exactly the right instrument — promote it to a calibration input for the paper plane.

---

7. VIX, Regimes & Event Risk

7.1 Term-structure sizing — ALIGNED (the engine's best VIX rule)

Backwardation-conditioned de-risking is the best-supported element of the VIX stack: the basis strongly predicts VIX-futures/variance returns (Simon & Campasano [80]); the slope summarizes essentially all term-structure information and prices variance risk (Johnson [81]); backwardation onset is the crisis marker (Alexander & Korovilas [82]). Refinement: use percent slope ((3M−30D)/30D) rather than a fixed −2-point threshold (the spread scales with level), and consider a full halt (not 30% size) in deep backwardation, where expected short-vol returns are negative-to-ambiguous [77][80].

7.2 Level bands — PARTIAL

Fixed VIX bands are a reasonable UX taxonomy but weak as signal: the same level means opposite things depending on prevailing realized vol [76], and the VIX premium *falls or inverts* right after spikes (Cheng [77]) — the "sell premium because VIX is high" intuition fails precisely post-spike. The PANIC halt is consistent with this (and with Moreira-Muir inverse-variance logic [79] and the Volmageddon record [83]). Recommendation: augment bands with VIX percentile and the VIX−RV spread; the engine's VIX-guard already caps credit-side scores in FEAR/PANIC + backwardation, which is the right composite condition.

7.3 Event blackouts — PARTIAL (correctly justified, wrongly framed)

Wright [86] shows scheduled-event variance is *overpriced* on average — macro blackouts cost positive EV for premium sellers and are justified as tail-variance control for an unattended engine, which is a legitimate and sufficient justification, but the spec should state it as such. The pre-FOMC drift [84] has attenuated post-2011 [85] and should not be traded directly. **OPEX/triple-witching blackout is the weakest rule**: expiration effects are microstructure noise plus strike pinning (Stoll & Whaley [87]; Ni, Pearson & Poteshman [88]) — the literature supports avoiding expiring-week near-the-money strikes, not banning entries. The static macro calendar (hardcoded through 2027) is operationally fragile; dates should refresh from a live source.

7.4 Tail risk and hedging — ALIGNED

The engine's preference hierarchy — de-risk via halts and sizing rather than continuously-held protection, with the tail-hedge ladder as an opt-in recommend-only program — matches Israelov's "Pathetic Protection" result (continuously-held puts reduce returns more than drawdowns; sizing down dominates hedge-buying [89]) and the practitioner consensus that convexity pays only with disciplined monetization and small fixed bleed budgets [89b]. The stress-test scenarios (−10%/+15 vol, −20%/+30 vol) are the right instrument class for short-premium books, whose returns are severely negatively skewed (Volmageddon: VIX futures +116% intraday, XIV −96% [83]). VVIX is correctly positioned as a "hedges are expensive" input [78].

---

8. POP, Price Ranges, Exits & Sizing

8.1 POP methodology — PARTIAL

Engine: POP = lognormal N(d2) at the break-even (per-leg smile IV where available), replacing delta proxies; prob-of-touch via GBM first-passage; cones = $S \cdot \exp((r-\sigma^2/2)t \pm k\sigma\sqrt{t})$.

Literature: Three layered corrections apply. (1) **Risk-neutral ≠ physical**: risk-neutral densities fail forecast tests of realized outcomes; utility-tilted densities pass (Bliss & Panigirtzoglou [105]); Ross recovery does not rescue this (misspecified per Borovička, Hansen & Scheinkman [106], fails empirically [106b]). For index short premium the two biases (risk-free drift instead of equity drift; VRP-inflated σ) both make risk-neutral POP *conservative* — which is why 30-delta structures win more often than nominal POP suggests. (2) **Single-sigma lognormal vs the smile**: the market's own density (Breeden-Litzenberger [107]) is left-skewed and fat-tailed; single-σ tail probabilities at 10–25-delta strikes are off by several percentage points — the same order as the whole edge. The engine's per-leg smile IV is partial credit; full credit is integrating the smile-implied CDF. (3) **Prob-of-touch inherits both errors plus the leverage effect** — treat downside-barrier touch probabilities as lower bounds. The straddle≈0.85σ convention point [108] argues for documenting the k↔coverage mapping on cones, and cone *coverage* should be validated empirically (price finishing inside ±1σ materially more than 68% of the time = cones too wide, the expected signature of VRP-inflated σ).

Recommendation: Label POP explicitly as "risk-neutral, smile-referenced"; optionally publish a physical POP using de-biased σ; and — the highest-value item — **close the calibration loop**: the engine already logs expected_pop per trade and has a calibration_report with POP-vs-realized quintiles and the exit-policy caveat. Promote this from diagnostic to a monitored gauge with reliability diagrams, Wilson CIs, and Brier decomposition (Gneiting & Raftery [111]; Murphy [112]); a persistent realized-above-predicted gap at the 70% bin is the VRP measured in the engine's own fills.

8.2 Exit rules — ALIGNED (practitioner-grade)

There is no peer-reviewed study of 50%-profit / 21-DTE management; the evidence is practitioner backtests (tastytrade lineage; independent replications [113][114][115]) consistently showing higher win rate, higher P&L-per-day, and materially reduced variance/tail at slightly lower per-trade EV, with ATM gamma several times higher inside 21 DTE. The correct framing — which the engine's calibration report already carries — is variance control, not alpha. The 2×-credit stop is likewise a risk rule; the practitioner record suggests naked-short stop-losses can reduce long-run EV while capping tails [115b]. The engine's obligation, given the ML layer, is to verify these numbers on its own fills — the triple-barrier labels are structurally identical to the exit rules, which is exactly the label-coherence property López de Prado prescribes.

8.3 Position sizing — ALIGNED (PARTIAL)

Half-Kelly capped at 2× base with a 2%-of-capital hard cap is squarely within best practice: fractional Kelly is the documented response to edge-estimation error (Thorp [117]; MacLean, Thorp & Ziemba [118] — half-Kelly earns ~75% of full-Kelly growth with drawdown-probability falling from x to x^3 for halving); the 2% rule itself has no rigorous evidence but functions as a ruin-avoidance floor [120]. The documented deficiency of per-trade rules: **they ignore correlation across concurrent positions** — ten 2%-risk short puts on correlated names ≈ one 20% bet in a crash. The engine's per-symbol caps, 20% open-risk cap, beta-weighted balancer envelope, and stress tests partially address this; the literature-complete version adds a portfolio-level drawdown/CVaR constraint (convex formulation in Busseti, Ryu & Boyd [119]) and feeds the *calibrated* (not raw) model probability into the Kelly fraction. Inverse-variance scaling of base size is separately supported (Moreira & Muir [79]; Harvey et al. [121]).

8.4 ML validation — ALIGNED (PARTIAL)

The stack (triple-barrier, meta-labeling, purged k-fold with embargo, uniqueness weights, isotonic calibration, PSI drift, PSR/DSR backtests) matches López de Prado [122][123] and the backtest-overfitting corrections [125][126]. Four small-sample settings need tightening against the literature:

1. **Isotonic calibration below ~1,000 samples overfits** (Niculescu-Mizil & Caruana [127]); with MIN_ROWS_TO_TRAIN = 150, the engine should use Platt/logistic (or beta) calibration until the labeled pool clears ~1,000.
2. **Embargo vs label horizon**: a 1% embargo on a small sample is shorter than the 21–45-day label horizon; embargo should be ≥ the label horizon, an admittedly expensive constraint at retail sample sizes.
3. **Class minima**: the 30-per-class minority floor is below the ~100 events + 100 non-events the validation literature requires just to estimate a calibration curve acceptably (Riley et al. [110]); tier thresholds should reflect that grades are unreliable below it.
4. **PSI thresholds** (0.10/0.25) are credit-scoring folklore; they mis-fire at small n and should be sized from PSI's sampling distribution (Yurdakul & Naranjo [128]).

Additionally, DSR's trial count must include every configuration ever tried, and conviction letter grades should ultimately map to calibrated probability bands with reported confidence intervals rather than raw scores. The engine's decision to have the ML gate *only subtract autonomy* is a governance pattern the literature implicitly endorses: in low-signal finance, hundreds of trades suffice to validate calibration but thousands are needed to certify a small edge [110][126].

9. Consolidated Findings & Priority Matrix

Divergences (wrong side of published evidence):

#	Finding	Evidence	Action
D1	Oscillator-heavy technical composite; no standalone validity for RSI/MACD/BB votes	[24][25][27]	Collapse to trend/momentum core; validate survivors on own ledger
D2	Pre-earnings calendar sells an on-average underpriced window	[31][32][86]	Gate on implied ≥ historical announcement move
D3	PEAD scanner targets an anomaly dead in its own universe since ~2006	[30]	Retire or re-scope
D4	Delayed-feed spread-gate relaxation loosens exactly where mids are least reliable	[90]	Tighten or refuse; keep traded paths realtime-only

Gaps (documented best practice absent):

#	Finding	Evidence	Action
G1	No American-exercise overlay; ITM put EEP 5–15% of value at 5% rates	[38][41][43]	BAW/BS2002 overlay or small CRR for ITM
G2	q = 0 on dividend payers: 1–3 vol-pt IV class error	[45][40]	Populate q / S−PV(div); amounts alongside dates
G3	Hardcoded r = 0.053: √T-scaled IV drift, worst on LEAPS	[46][46b]	Daily SOFR/CMT per-expiry
G4	No vol forecast (HAR/IV-anchored); HV20 backward-looking	[15][55][12]	De-biased IV first; 3-coefficient HAR second
G5	NVRP computed but not the entry gate; IVR ≥ 35 is	[13][11][77]	Promote NVRP to primary vol gate
G6	No implementation-shortfall accounting (unfilled-order opportunity cost)	[101]	Add decision-time benchmark + expired-order P&L
G7	No portfolio-level correlated-tail constraint beyond caps/stress	[119][118]	Drawdown/CVaR constraint over concurrent premium
G8	Calibration loop diagnostic-only	[111][110]	Reliability diagrams + Brier as monitored gauges
G9	Static macro calendar	[86]	Live refresh of FOMC/CPI/NFP dates

Refinements (supported features, tune per literature): YZ→primary RV with band recalibration (R1, [16][53]); IV percentile over min-max rank (R2, [63][64]); ATM-normalized surface skews + total-variance interpolation (R3, [56][59]); minutes-based T for the 7–14 DTE path (R4, [49][49c]); Let's-Be-Rational solver swap (R5, [37]); reprice cadence/units in half-spread terms (R6, [100]); paper fills capped at natural with calibrated fill probability (R7, [100][103]); percent term-slope + deep-backwardation halt (R8, [80][81]); OPEX strike hygiene instead of entry ban (R9, [87][88]); Platt-before-isotonic under 1k samples (R10, [127]); embargo ≥ label horizon (R11, [122]); n-aware PSI thresholds (R12, [128]).

Where the engine is ahead of common practice: index-only GEX sign trust (§5.7); fail-closed earnings-unknown gating; the natural-price worst-case bound and tick-tolerance floor; per-plane slippage telemetry separating sim optimism from live fill quality; label-coherent triple barriers matching actual exit rules; ML gate restricted to subtracting autonomy; Wilson intervals on every displayed rate in fleet metrics.

---

10. References

Organized by section of first citation; numbering is continuous. URLs verified August 2026.

Signal generation & VRP

[1] Bakshi, G. & Kapadia, N. (2003). "Delta-Hedged Gains and the Negative Market Volatility Risk Premium." *Review of Financial Studies* 16(2), 527–566.

[2] Coval, J. & Shumway, T. (2001). "Expected Option Returns." *Journal of Finance* 56(3), 983–1009.

[3] Carr, P. & Wu, L. (2009). "Variance Risk Premiums." *Review of Financial Studies* 22(3), 1311–1341.

[4] Bondarenko, O. (2019). "Historical Performance of Put-Writing Strategies." CBOE white paper.

[5] Ungar, J. & Moran, M. (2009). "The Cash-Secured PutWrite Strategy and Performance of Related Benchmark Indexes." *Journal of Alternative Investments*.

[6] Whaley, R. (2002). "Return and Risk of CBOE Buy Write Monthly Index." *Journal of Derivatives*.

[7] Israelov, R. & Nielsen, L. (2015). "Covered Calls Uncovered." *Financial Analysts Journal* 71(6).

[8] Israelov, R. & Tummala, H. (2017). "Which Index Options Should You Sell?" SSRN 2990542.

[11] Bollerslev, T., Tauchen, G. & Zhou, H. (2009). "Expected Stock Returns and Variance Risk Premia." *Review of Financial Studies* 22(11), 4463–4492.

[12] Christensen, B.J. & Prabhala, N.R. (1998). "The Relation Between Implied and Realized Volatility." *Journal of Financial Economics* 50(2), 125–150.

[13] Goyal, A. & Saretto, A. (2009). "Cross-Section of Option Returns and Volatility." *Journal of Financial Economics* 94(2), 310–326.

[14] BSIC (2023). Replication note on Goyal-Saretto net-of-costs decile-spread returns.

[17] Xing, Y., Zhang, X. & Zhao, R. (2010). "What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns?" *JFQA* 45(3), 641–662.

[18] Cremers, M. & Weinbaum, D. (2010). "Deviations from Put-Call Parity and Stock Return Predictability." *JFQA* 45(2), 335–367.

[19] Bali, T. & Hovakimian, A. (2009). "Volatility Spreads and Expected Stock Returns." *Management Science* 55(11), 1797–1812.

[20] An, B.-J., Ang, A., Bali, T. & Cakici, N. (2014). "The Joint Cross Section of Stocks and Options." *Journal of Finance* 69(5).

[21] Johnson, T. & So, E. (2012). "The Option to Stock Volume Ratio and Future Returns." *Journal of Financial Economics* 106(2), 262–286.

[22] Vasquez, A. (2017). "Equity Volatility Term Structures and the Cross Section of Option Returns." *JFQA* 52(6), 2727–2754.

[23] Brock, W., Lakonishok, J. & LeBaron, B. (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." *Journal of Finance* 47(5), 1731–1764.

[24] Sullivan, R., Timmermann, A. & White, H. (1999). "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap." *Journal of Finance* 54(5), 1647–1691.

[25] Park, C.-H. & Irwin, S. (2007). "What Do We Know About the Profitability of Technical Analysis?" *Journal of Economic Surveys* 21(4), 786–826.

[26] Lo, A., Mamaysky, H. & Wang, J. (2000). "Foundations of Technical Analysis." *Journal of Finance* 55(4), 1705–1765.

[27] Moskowitz, T., Ooi, Y.H. & Pedersen, L. (2012). "Time Series Momentum." *Journal of Financial Economics* 104(2), 228–250.

[28] Neely, C., Rapach, D., Tu, J. & Zhou, G. (2014). "Forecasting the Equity Risk Premium: The Role of Technical Indicators." *Management Science* 60(7), 1772–1791.

[29] Bernard, V. & Thomas, J. (1989). "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27 (Suppl.), 1–36.

[30] Martineau, C. (2021). "Rest in Peace Post-Earnings Announcement Drift." *Critical Finance Review*.

[31] Dubinsky, A., Johannes, M., Kaeck, A. & Seeger, N. (2019). "Option Pricing of Earnings Announcement Risks." *Management Science / RFS* 32(2), 646–687.

[32] Gao, C., Xing, Y. & Zhang, X. (2018). "Anticipating Uncertainty: Straddles Around Earnings Announcements." *JFQA* 53(6), 2587–2617.

[35] Timmermann, A. (2006). "Forecast Combinations." *Handbook of Economic Forecasting*, Ch. 4.

[36] Gu, S., Kelly, B. & Xiu, D. (2020). "Empirical Asset Pricing via Machine Learning." *Review of Financial Studies* 33(5), 2223–2273.

Pricing & Greeks

[37] Jäckel, P. (2015). "Let's Be Rational." *Wilmott* 2015(75), 40–53.

[38] Barone-Adesi, G. & Whaley, R.E. (1987). "Efficient Analytic Approximation of American Option Values." *Journal of Finance* 42(2), 301–320.

[39] Bjerksund, P. & Stensland, G. (2002). "Closed Form Valuation of American Options." NHH working paper.

[40] Cboe. "American-Style Options Implied Volatility Calculations Methodology" (CRR trees, discrete dividends, spline-interpolated CMT rates).

[41] McMurray, L. & Yadav, P.K. "The Early Exercise Premium in American Option Prices: Direct Empirical Evidence." SSRN 7110.

[42] Li et al. (2024). "An Empirical Study on the Early Exercise Premium of American Options: Evidence from OEX and XEO Options." *Journal of Futures Markets* 44(7), 1117–1153.

[43] Barraclough, K. & Whaley, R.E. (2012). "Early Exercise of Put Options on Stocks." *Journal of Finance* 67(4).

[44] Pool, V.K., Stoll, H.R. & Whaley, R.E. (2008). "Failure to Exercise Call Options." *Journal of Financial Markets* 11(1), 1–35.

[45] Haug, E., Haug, J. & Lewis, A. (2003). "Back to Basics: A New Approach to the Discrete Dividend Problem." *Wilmott Magazine*.

[46] van Binsbergen, J.H., Diamond, W.F. & Grotteria, M. (2022). "Risk-Free Interest Rates." *Journal of Financial Economics* 143(1), 1–29.

[46b] Hull, J. & White, A. (2013). "LIBOR vs. OIS: The Derivatives Discounting Dilemma." *Journal of Investment Management*.

[47] Wallmeier, M. (2024). "Quality Issues of Implied Volatilities of Index and Stock Options in the OptionMetrics IvyDB Database." *Journal of Futures Markets* 44(5), 854–875.

[48] Wallmeier, M. (2025). "Distorted Implied Volatilities and Borrowing Rates in Version 6.0 of the OptionMetrics IvyDB US Database." SSRN 6394358.

[49] French, K.R. & Roll, R. (1986). "Stock Return Variances: The Arrival of Information and the Reaction of Traders." *Journal of Financial Economics* 17, 5–26.

[49b] Harwood, V. "The Myth of Option Weekend Decay." *Six Figure Investing* (practitioner).

[49c] Abdelmessih, K. "252 vs 365-Day Annualization and Vol Time." *Moontower* (practitioner).

Volatility modeling

[50] Parkinson, M. (1980). "The Extreme Value Method for Estimating the Variance of the Rate of Return." *Journal of Business* 53(1), 61–65.

[51] Garman, M. & Klass, M. (1980). "On the Estimation of Security Price Volatilities from Historical Data." *Journal of Business* 53(1), 67–78.

[52] Rogers, L.C.G. & Satchell, S. (1991). "Estimating Variance from High, Low and Closing Prices." *Annals of Applied Probability* 1(4), 504–512.

[16] Yang, D. & Zhang, Q. (2000). "Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices." *Journal of Business* 73(3), 477–492.

[53] Molnár, P. (2012). "Properties of Range-Based Volatility Estimators." *International Review of Financial Analysis* 23, 20–29.

[54] Hansen, P.R. & Lunde, A. (2005). "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?" *Journal of Applied Econometrics* 20(7), 873–889.

[15] Corsi, F. (2009). "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7(2), 174–196.

[55] Poon, S.-H. & Granger, C. (2003). "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature* 41(2), 478–539.

[17b] Blair, B., Poon, S.-H. & Taylor, S. (2001). "Forecasting S&P 100 Volatility." *Journal of Econometrics* 105(1), 5–26.

[56] Mixon, S. (2011). "What Does Implied Volatility Skew Measure?" *Journal of Derivatives* 18(4), 9–25.

[57] Fleming, J., Kirby, C. & Ostdiek, B. (2003). "The Economic Value of Volatility Timing Using 'Realized' Volatility." *Journal of Financial Economics* 67(3), 473–509.

[58] Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide.* Wiley.

[59] Gatheral, J. & Jacquier, A. (2014). "Arbitrage-Free SVI Volatility Surfaces." *Quantitative Finance* 14(1), 59–71.

[60] Le Floc'h, F. (2020). "An Arbitrage-Free Interpolation of Class C^2 for Option Prices." arXiv:2004.08650.

[61] Couillard, M. & Davison, M. (2005). "A Comment on Measuring the Hurst Exponent of Financial Time Series." *Physica A* 348, 404–418.

[62] Castagna, A. & Mercurio, F. (2007). "The Vanna-Volga Method for Implied Volatilities." *Risk*, January, 106–111.

[63] projectfinance. "IV Rank vs. IV Percentile" (practitioner; documented rank pathologies).

[64] Barchart / Option Samurai. IV rank vs percentile education pages (practitioner).

[65] Hamilton, J.D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57(2), 357–384.

[66] Ang, A. & Timmermann, A. (2012). "Regime Changes and Financial Markets." *Annual Review of Financial Economics* 4.

[67] Shu, Y., Yu, C. & Mulvey, J. (2024). "Downside Risk Reduction Using Regime-Switching Signals: A Statistical Jump Model Approach." *Journal of Asset Management* 25, 493–507.

GEX & dealer positioning

[68] SqueezeMetrics. "Gamma Exposure (GEX)" white paper (practitioner).

[69] Barbon, A. & Buraschi, A. (2021). "Gamma Fragility." SSRN 3725454.

[70] Baltussen, G., Da, Z., Lammers, S. & Martens, M. (2021). "Hedging Demand and Market Intraday Momentum." *Journal of Financial Economics* 142(1), 377–403.

[71] Chilingarian, A. (2026). "The Sign of Dealer Gamma: A Reproducible, Auditable Framework for Computing S&P 500 Gamma Exposure." SSRN 7131778.

VIX, regimes & events

[72] Britten-Jones, M. & Neuberger, A. (2000). "Option Prices, Implied Price Processes, and Stochastic Volatility." *Journal of Finance* 55(2), 839–866.

[73] Cboe. *VIX White Paper: Cboe Volatility Index Methodology*.

[74] Dew-Becker, I., Giglio, S., Le, A. & Rodriguez, M. (2017). "The Price of Variance Risk." *Journal of Financial Economics* 123(2), 225–250.

[76] Edwards, T. & Preston, H. (2017). "Reading VIX: Does VIX Predict Future Volatility?" S&P Dow Jones Indices research.

[77] Cheng, I.-H. (2019). "The VIX Premium." *Review of Financial Studies* 32(1), 180–227.

[78] Park, Y.-H. (2015). "Volatility-of-Volatility and Tail Risk Hedging Returns." *Journal of Empirical Finance* 30, 38–58.

[79] Moreira, A. & Muir, T. (2017). "Volatility-Managed Portfolios." *Journal of Finance* 72(4), 1611–1644.

[80] Simon, D.P. & Campasano, J. (2014). "The VIX Futures Basis: Evidence and Trading Strategies." *Journal of Derivatives* 21(3), 54–69.

[81] Johnson, T.L. (2017). "Risk Premia and the VIX Term Structure." *JFQA* 52(6), 2461–2490.

[82] Alexander, C. & Korovilas, D. (2012). "The Hazards of Volatility Diversification." ICMA Centre DP 2012-07.

[83] Augustin, P., Cheng, I.-H. & Van den Bergen, L. (2021). "Volmageddon and the Failure of Short Volatility Products." *Financial Analysts Journal* 77(3).

[84] Lucca, D.O. & Moench, E. (2015). "The Pre-FOMC Announcement Drift." *Journal of Finance* 70(1), 329–371.

[85] Kurov, A. et al. (2021). "The Disappearing Pre-FOMC Announcement Drift." *Finance Research Letters*.

[86] Wright, J.H. (2021). "Event-Day Options." NBER Working Paper 28306.

[87] Stoll, H.R. & Whaley, R.E. (1990). "Program Trading and Individual Stock Returns: Ingredients of the Triple-Witching Brew." *Journal of Business* 63(1), S165–S192.

[88] Ni, S.X., Pearson, N.D. & Poteshman, A.M. (2005). "Stock Price Clustering on Option Expiration Dates." *Journal of Financial Economics* 78(1), 49–87.

[89] Israelov, R. (2019). "Pathetic Protection: The Elusive Benefits of Protective Puts." *Journal of Alternative Investments* 21(3).

[89b] AQR (2020). "Tail Risk Hedging: Contrasting Put and Trend Strategies." White paper.

Execution quality

[90] Muravyev, D. & Pearson, N.D. (2020). "Options Trading Costs Are Lower Than You Think." *Review of Financial Studies* 33(11), 4973–5014.

[92] Bryzgalova, S., Pavlova, A. & Sikorskaya, T. (2023). "Retail Trading in Options and the Rise of the Big Three Wholesalers." *Journal of Finance* 78(6).

[92b] Ernst, T. & Spatt, C.S. (2022). "Payment for Order Flow and Asset Choice." NBER Working Paper 29883.

[93] Hendershott, T., Khan, S.A. & Riordan, R. (2023). "Option Auctions." Working paper.

[94] Chan, K.C., Chung, Y.P. & Johnson, H. (1995). "The Intraday Behavior of Bid-Ask Spreads for NYSE Stocks and CBOE Options." *JFQA* 30(3), 329–346.

[95] Chaput, J.S. & Ederington, L.H. (2003). "Option Spread and Combination Trading." *Journal of Derivatives* 10(4), 70–88.

[96] Flint, A., Lepone, A. & Yang, J.Y. (2014). "Do Option Strategy Traders Have a Disadvantage?" *Journal of Futures Markets* 34(9), 838–852.

[97] de Fontnouvelle, P., Fishe, R.P.H. & Harris, J.H. (2003). "The Behavior of Bid-Ask Spreads and Volume in Options Markets During the Competition for Listings in 1999." *Journal of Finance* 58(6), 2437–2463.

[98] Wei, J. & Zheng, J. (2010). "Trading Activity and Bid-Ask Spreads of Individual Equity Options." *Journal of Banking & Finance* 34(12), 2897–2916.

[99] Christoffersen, P., Goyenko, R., Jacobs, K. & Karoui, M. (2018). "Illiquidity Premia in the Equity Options Market." *Review of Financial Studies* 31(3), 811–851.

[100] Lo, A.W., MacKinlay, A.C. & Zhang, J. (2002). "Econometric Models of Limit-Order Executions." *Journal of Financial Economics* 65(1), 31–71.

[101] Perold, A.F. (1988). "The Implementation Shortfall: Paper Versus Reality." *Journal of Portfolio Management* 14(3), 4–9.

[102] Almgren, R. & Chriss, N. (2001). "Optimal Execution of Portfolio Transactions." *Journal of Risk* 3(2), 5–39.

[103] spintwig. "Methodology / Backtesting Mechanics" (practitioner fill-parameterization standard).

[104] QuantConnect. "Reality Modeling — Trade Fills" (practitioner).

POP, sizing & ML validation

[105] Bliss, R.R. & Panigirtzoglou, N. (2004). "Option-Implied Risk Aversion Estimates." *Journal of Finance* 59(1), 407–446.

[106] Borovička, J., Hansen, L.P. & Scheinkman, J.A. (2016). "Misspecified Recovery." *Journal of Finance* 71(6), 2493–2544.

[106b] Jackwerth, J. et al. (2021). "Does the Ross Recovery Theorem Work Empirically?" *Journal of Financial Economics*.

[107] Breeden, D. & Litzenberger, R. (1978). "Prices of State-Contingent Claims Implicit in Option Prices." *Journal of Business* 51(4), 621–651.

[108] Brenner, M. & Subrahmanyam, M. (1988). "A Simple Formula to Compute the Implied Standard Deviation." *Financial Analysts Journal* 44(5) ($0.8 \cdot S \cdot \sigma\sqrt{t}$ straddle approximation lineage).

[109] Malz, A. (2014). "A Simple and Reliable Way to Compute Option-Based Risk-Neutral Distributions." FRBNY Staff Report 677.

[110] Riley, R.D. et al. (2021). "Minimum Sample Size for External Validation of a Clinical Prediction Model with a Binary Outcome." *Statistics in Medicine* 40, 4230–4251.

[111] Gneiting, T. & Raftery, A.E. (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation." *JASA* 102(477), 359–378.

[112] Murphy, A.H. (1973). "A New Vector Partition of the Probability Score." *Journal of Applied Meteorology* 12, 595–600.

[113] tastytrade/tastylive Market Measures research on managing winners at 50% and 21-DTE management (practitioner).

[114] spintwig SPY/SPX short-put and strangle 45-DTE management-variant backtests (practitioner).

[115] Sweet Volatility. Independent tastytrade-method replication (practitioner).

[115b] Luckbox Magazine / tastylive. "Managing Losers" (practitioner).

[117] Thorp, E.O. (2006). "The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market." *Handbook of Asset and Liability Management*, Vol. 1.

[118] MacLean, L., Thorp, E. & Ziemba, W. (2010). "Good and Bad Properties of the Kelly Criterion." In *The Kelly Capital Growth Investment Criterion*, World Scientific.

[119] Busseti, E., Ryu, E.K. & Boyd, S. (2016). "Risk-Constrained Kelly Gambling." *Journal of Investing* 25(3), 118–134.

[120] Van Tharp position-sizing framework; fixed-fractional lineage (practitioner).

[121] Harvey, C.R., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M. & Van Hemert, O. (2018). "The Impact of Volatility Targeting." *Journal of Portfolio Management*.

[122] López de Prado, M. (2018). *Advances in Financial Machine Learning.* Wiley.

[123] López de Prado, M. (2018). "The 10 Reasons Most Machine Learning Funds Fail." *Journal of Portfolio Management*.

[125] Bailey, D.H., Borwein, J., López de Prado, M. & Zhu, Q.J. (2017). "The Probability of Backtest Overfitting." *Journal of Computational Finance* 20(4).

[126] Bailey, D.H. & López de Prado, M. (2014). "The Deflated Sharpe Ratio." *Journal of Portfolio Management* 40(5), 94–107.

[127] Niculescu-Mizil, A. & Caruana, R. (2005). "Predicting Good Probabilities with Supervised Learning." *ICML 2005*.

[128] Yurdakul, B. & Naranjo, J. (2020). "Statistical Properties of the Population Stability Index." *Journal of Risk Model Validation* 14(4).

Prepared from ENGINE_SPEC.md v3.95.0 and ~130 primary sources. Magnitudes quoted from the cited studies; practitioner-grade evidence is labeled as such throughout. This document is research commentary, not investment advice.