

Skill or Variance?

What Trading Evaluations Actually Select

Francisco Matilla Serrano

Independent Researcher, Varna, Bulgaria

Corresponding author: Francisco Matilla Serrano (franchimatilla@gmail.com)

ORCID: <https://orcid.org/0009-0004-9491-6374>

16 September 2026

Abstract

Passing a trading evaluation is an outcome, not a direct measurement of trading skill. This paper makes that distinction operational with a stochastic profit-and-loss process in which trader types differ in expected daily profit, volatility, and risk scale. Each type faces the same fictitious “50K” two-stage contract: a \$3,000 target, a \$2,000 drawdown, a 20-day horizon per stage, a \$90 evaluation fee, a \$100 activation fee, a 90% payout share, and a \$3,000 gross first payout. Finite-horizon barrier probabilities are estimated using 40,000 antithetic Monte Carlo paths on a discrete intraday grid and are checked against the corresponding infinite-horizon Brownian first-passage formula. Monte Carlo uncertainty is reported from antithetic pair means.

Three findings matter. First, higher variance can sharply increase the chance that a zero- or negative-drift trader completes the contract. At the baseline risk scale, the negative-drift, high-variance type passes 33.84% of evaluations, compared with 11.71% for the negative-drift, lower-variance type. The high-variance negative-drift type also has positive participant contract expectancy in the deliberately permissive model despite losing \$1,500 in expected market P&L over 20 days. Second, the evaluation is an imperfect classifier. In a population containing 5% positive-drift traders and 95% zero- or negative-drift traders, 91.39% of passers and 87.30% of first-payout events come from the nonpositive group under the stated type mixture. Third, the common claim that “the losing 95% finance the winning 5%” is incomplete. With the prices and payouts used here, the provider loses an expected \$137.41 per attempt in that 5/95 scenario before operating costs. The evaluation fee would need to rise from \$90 to \$227.41, or the gross first payout fall from \$3,000 to \$1,389.64, for the closed-transfer benchmark to break even. Changing the positive-drift share alone does not restore break-even under equal within-group weights: even a 1/99 population loses \$121.68 per attempt. A separate population frontier shows that the variance composition of the nonpositive group is decisive.

An external, vendor-reported industry funnel is used only as a calibration check: 14% reportedly obtained a funded account and 45% of that funded group received a payout, yielding roughly 6.3% before rounding. The 45% is a conditional funded-to-payout rate; the often-quoted 7% is not q . The model remains illustrative rather than an estimate of any provider’s customer base. The specific contribution is to map heterogeneous drift, volatility, and risk scale into pass composition, false selection, and provider transfer arithmetic, rather than to introduce first-passage valuation itself. Within the program, this step is necessary because the probabilities treated as inputs in the opening benchmark become outcomes of an explicit trader-type model.

Keywords: trading evaluation; skill selection; variance; first-passage probability; false discovery; limited liability; provider economics

JEL Classification: C15; D81; D86; G11; G23

1. Introduction

Trading evaluations are commonly described as filters: success opens a later stage and failure ends the attempt. That language invites a simple reading: the evaluation separates capable traders from incapable ones. But a threshold selects realized paths, and every realized path contains both signal and noise.

A profitable trader can fail because a bad sequence arrives before the target. A trader with zero or negative expected market profit can pass because a favorable sequence arrives first. Increasing risk can make either boundary arrive sooner. A finite deadline adds a third outcome: neither boundary is reached in time. These features mean that pass probability, latent skill, and participant contract value are different objects.

Market Edge versus Contract Edge: Negative-Expectancy Trading under an Idealized Evaluation Contract (Matilla Serrano, 2026b) treated the evaluation success probability p and funded-stage success probability q as parameters. That abstraction establishes an existence result: a negative-expectancy market strategy can still have positive participant contract expectancy because nominal trading losses and personal contract losses are not the same cash flow. It does not identify which traders generate p and q , how volatility affects those probabilities, or whether passing identifies positive expected market P&L. This paper takes that next step by making the probabilities endogenous to a path model.

The model assigns each trader type a daily drift μ , daily volatility σ , and risk multiplier ℓ . Profit and loss follow an arithmetic Brownian benchmark monitored on a discrete intraday grid. The evaluation ends when cumulative P&L first reaches the upper target U , first reaches the lower drawdown boundary $-D$, or remains unresolved at horizon H . A successful trader then faces an independent stage with the same latent type and horizon. This makes p and q outcomes of a path model rather than free inputs.

The paper asks four questions.

1. How do drift, variance, and risk scale change evaluation and first-payout probabilities?
2. How accurately does a pass classify positive expected market P&L?
3. How do false positives and false negatives change with population composition?
4. Under what prices, payouts, and population mixes can the provider finance the participant side of the contract?

The main results challenge two shortcuts. The first is that a passed evaluation proves skill. In the baseline type grid, a zero-drift, high-variance trader passes 40.39% of evaluations, while a positive-drift, lower-variance trader passes 48.75%. The two pass rates are reasonably close even though their expected 20-day market P&L differs by \$2,000. The negative-drift, high-variance trader passes 33.84% of evaluations and receives a first payout on 11.40% of purchased attempts. Selection improves after a second successful stage, but variance remains powerful.

The second shortcut is that a large population of losing traders automatically makes the provider profitable. It does not. Provider sustainability depends on fees, activation revenue, payout incidence, payout size, payout share, operating costs, and any revenue or loss from real-market activity. A 95/5 population split supplies none of those quantities by itself. Under the closed-transfer model used here, every population mix from 5% to 50% positive-drift traders produces negative provider gross margin at the illustrative price and payout schedule. Extending the grid to 0% and 1% positive-drift entrants does not change that conclusion under equal within-group weights. Break-even appears only after the composition of the nonpositive group is changed sharply toward low-variance types. That is not a claim about an observed company. It shows that the arithmetic of cross-subsidy must be demonstrated, not assumed.

The analysis joins two questions that are usually discussed separately. The first is statistical: what kind of trader reaches a barrier before a loss limit and a deadline? The second is economic: what do those selection outcomes imply for participant and provider cash flow? Combining the two makes it possible to show when variance mimics skill, measure false selection, and test whether a stated population mix can finance the contract. First-passage probabilities and noisy selection are established subjects. Applying them to the complete two-stage cash-flow problem shows how a contract can reward variance even when variance conveys little information about expected market profit.

The rest of the paper proceeds as follows. Section 2 positions the question in the relevant literature. Section 3 defines skill, variance, and selection. Section 4 presents the stochastic model and contract cash flows. Section 5 describes the simulation design. Sections 6 and 7 report trader-level and population-level results. Section 8 examines provider sustainability. Section 9 discusses interpretation and contract design. Section 10 states the limitations, and Section 11 concludes. Derivations and sensitivity checks appear in the Appendix; code and numerical outputs are supplied separately in the accompanying research package.

2. Related Literature

2.1 Performance, luck, and selection

The distinction between observed performance and latent ability is central to the mutual-fund literature. Fama and French (2010) separate skill from luck in the cross-section of fund returns. Barras, Scaillet, and Wermers (2010) adapt false-discovery methods to estimate the proportions of skilled, unskilled, and zero-alpha funds. Carpenter and Lynch (1999) study survivorship bias and attrition effects in performance measurement. The present setting differs: the outcome is a path-dependent threshold event rather than a conventional return estimate, and participants repeatedly purchase access. The common lesson is that selection on realized outcomes changes the composition of the observed winner group.

Fama and French (2010) also show why statistical power matters when the cross-section contains only a small skilled minority. Harvey and Liu (2022) demonstrate that undersampling can make true skill difficult to detect and can distort estimates of the fraction of outperformers. The present model does not estimate trader skill from returns, but the same warning applies to any empirical use of pass/fail records: a short and selected history is a weak basis for separating skill from noise.

This paper uses the terms *false positive* and *false negative* in an economic classification sense, not as formal hypothesis-test errors. A false positive is a zero- or negative-drift type that passes. A false negative is a positive-drift type that does not pass within the evaluation horizon. These definitions require a declared type model. They cannot be inferred from observed pass/fail data alone.

2.2 Evaluation incentives and risk choice

Evaluation systems can alter risk-taking incentives. Brown, Harlow, and Starks (1996) and Chevalier and Ellison (1997) show how relative performance and flows can influence mutual-fund risk choices. Carpenter (2000), Ross (2004), Basak, Pavlova, and Shapiro (2007), and Hodder and Jackwerth (2007) show that nonlinear compensation and barriers create state-dependent incentives whose effects are more nuanced than “convex pay always increases risk.”

A purchased trading evaluation has its own nonlinear payoff. The participant pays a bounded fee, receives no refund for nominal trading losses, and may receive a discrete later payout after crossing one or more thresholds. The contract therefore changes the personal payoff to variance. It may reward a risk scale that would not maximize expected market wealth. This paper does not solve a fully dynamic optimal-control problem; it evaluates a transparent grid of risk multipliers to reveal the selection mechanism.

Entry-fee and submission-cost contests offer another parallel. Duffy, Matros, and Valencia (2023) study contests with entry fees, while Whitmeyer (2023) studies submission costs in risk-taking contests. A trading evaluation is not a rank-order contest, but each purchased attempt combines a cost of entry with a threshold-contingent reward. Re-entry may therefore shape both who participates and how much risk they choose.

2.3 Drawdown, barriers, and finite monitoring

Barrier order is intrinsically path dependent. Magdon-Ismail et al. (2004) analyze maximum drawdown for Brownian motion, and Kaminski and Lo (2014) examine conditions under which stop-loss rules change outcomes. Closed-form first-passage results provide a useful infinite-horizon benchmark for the static two-boundary problem. Finite horizons and discrete monitoring require numerical treatment. Broadie, Glasserman, and Kou (1997) demonstrate the importance of monitoring frequency for barrier problems, and Glasserman (2003) gives the broader Monte Carlo foundation.

The simulation uses common random numbers and antithetic paths to make comparisons less noisy. Common random numbers are a standard variance-reduction device when alternative systems are evaluated on the same stochastic inputs (Wright and Ramsay, 1979). Its effectiveness depends on the response structure, so variance reduction is not guaranteed for every comparison. In this application, common paths are used primarily to identify which outcome changes when the contract changes, while paired uncertainty estimates report the remaining simulation noise.

2.4 Direct funded-trader literature

Recent preprints already value funded-trader contracts. Lim’s *The Price of a Funded Account* (2026b) develops a single-contract actuarial framework, *Phantom Generosity* (2026a) compares 31 contracts across 21 firms on a common benchmark population, and Tomàs Fernández

(2026) derives a closed-form first-passage valuation with cross-firm break-even conditions. Maile (2026) discusses the industry and its possible policy relevance, while Ng (2026) studies a deferred-fee design in which payment follows a successful evaluation. The present analysis asks a different question: how does a declared mixture of latent trader types determine who passes, who reaches a first payout, and whether the resulting transfer schedule balances for the provider?

3. What Does the Evaluation Select?

3.1 Latent type

Let trader type i be represented by

$$\theta_i = (\mu_i, \sigma_i, \ell_i),$$

where μ_i is expected daily P&L before the risk multiplier, σ_i is daily P&L volatility, and $\ell_i > 0$ is the chosen risk scale. In the baseline arithmetic model, effective drift and volatility are

$$\mu_i(\ell_i) = \ell_i \mu_i, \sigma_i(\ell_i) = \ell_i \sigma_i.$$

Daily drift is an average over repeated paths, not a promised daily profit. A value of $\mu_i = \$100$, for example, means that the model adds \$100 per trading day in expectation before random variation. Scaling both quantities leaves the drift-to-volatility ratio unchanged, but it does not leave barrier outcomes unchanged because the target, drawdown, and horizon remain fixed in dollars and days. Risk scale therefore changes how quickly the process encounters the contract.

For the present paper, *positive edge* means $\mu_i > 0$, *zero edge* means $\mu_i = 0$, and *negative edge* means $\mu_i < 0$. This is a deliberately narrow definition. It does not include estimation error, skewness, serial dependence, execution costs, changing opportunity sets, or behavioral discipline.

3.2 Realized classification

Let $Y_i = 1$ if type i passes the evaluation and $Y_i = 0$ otherwise. Passing is a noisy observation of θ_i . Define:

True positive: $\mu_i > 0, Y_i = 1,$

False positive: $\mu_i \leq 0, Y_i = 1,$

False negative: $\mu_i > 0, Y_i = 0.$

A trader can also remain unresolved at the horizon. For contract purposes, the simulation treats an unresolved evaluation as not passed. Economically, it is useful to preserve the unresolved category because low risk can fail through insufficient movement rather than a drawdown breach.

The evaluation does not observe μ_i . It observes a path filtered through targets, deadlines, and rules. As a result, selectivity and accuracy depend on contract design as well as on the population.

3.3 Population composition

Let w_i be the population weight of type i , with $\sum_i w_i = 1$. If p_i is type i 's pass probability, the population pass probability is

$$\dot{p} = \sum_i w_i p_i.$$

The positive-edge share among passers is a Bayes calculation:

$$Pr(\mu > 0 \mid \text{pass}) = \frac{\sum_{i: \mu_i > 0} w_i p_i}{\sum_i w_i p_i}.$$

The nonpositive share among passers is one minus this quantity. It is labeled the *false-discovery share* for compactness, with the earlier warning that the model is not performing a statistical hypothesis test.

If q_i is the conditional probability that type i completes the funded stage after passing, the first-payout event rate is

$$\dot{r} = \sum_i w_i p_i q_i.$$

The positive-edge share among payout events is

$$Pr(\mu > 0 \mid \text{first payout}) = \frac{\sum_{i: \mu_i > 0} w_i p_i q_i}{\dot{r}}.$$

Requiring two successful stages should generally improve selection quality, but it also rejects some positive-edge traders twice.

4. Stochastic and Contract Model

4.1 Profit-and-loss process

The main notation is collected here because similar symbols refer to different levels of the model.

Table 1. Notation used in the stochastic and population model

Symbol	Meaning
$X_{i,t}$	cumulative P&L of type i at time t
μ_i	unconstrained expected P&L per trading day

Symbol	Meaning
σ_i	daily P&L volatility, measured per square-root trading day
ℓ_i	risk-scale multiplier applied to both drift and volatility
B_t	standard Brownian motion, the cumulative random-shock process
p_i, q_i	evaluation and funded-stage success probabilities
$\dot{p}, \dot{r}$	population-weighted pass and first-payout probabilities
F^*, W^*	fee or payout value that sets the stated provider margin to zero

Note: The subscript i indexes trader type. Bars denote population-weighted quantities, and stars denote break-even values.

For trader type i , cumulative P&L follows

$$dX_{i,t} = \ell_i \mu_i dt + \ell_i \sigma_i dB_t, X_{i,0} = 0 \quad (1)$$

where B_t is standard Brownian motion. In plain language, the model adds a predictable component, $\ell_i \mu_i dt$, to a random component, $\ell_i \sigma_i dB_t$. Over a short interval dt , the Brownian increment has mean zero and variance dt . It is an external source of standardized randomness, not an extra economic parameter or a balance variable. The multiplier ℓ_i changes the dollar speed of both components while leaving their ratio unchanged.

Define the first upper- and lower-boundary times

$$\tau_U = \inf \{ t \geq 0 : X_{i,t} \geq U \}, \tau_D = \inf \{ t \geq 0 : X_{i,t} \leq -D \}.$$

The finite-horizon pass probability is

$$p_i(H) = \Pr(\tau_U < \tau_D, \tau_U \leq H). \quad (2)$$

Failure by drawdown has probability $\Pr(\tau_D < \tau_U, \tau_D \leq H)$. The remaining probability is unresolved at H .

For $\mu_i \neq 0$, the infinite-horizon probability of reaching U before $-D$ is

$$p_i(\infty) = \frac{1 - \exp(-2\ell_i \mu_i D / \ell_i^2)}{1 - \exp(-2\ell_i \mu_i (U + D) / \ell_i^2)} \quad (3)$$

For zero drift, the continuous-time identity simplifies to

$$p_i(\infty) = \frac{D}{U + D}.$$

Thus, with $U = \$3,000$ and $D = \$2,000$, every zero-drift Brownian type has an infinite-horizon upper-first probability of 40%, regardless of volatility. Volatility still matters over a finite horizon because it changes the probability that either boundary is reached before the deadline.

4.2 Two-stage contract

The contract is the less-generous specification introduced in *Market Edge versus Contract Edge: Negative-Expectancy Trading under an Idealized Evaluation Contract* (Matilla Serrano, 2026b). The account is labeled “50K,” but the label is nominal and is not participant-deposited capital. A purchased evaluation costs F . A passer pays activation fee A , enters an independent funded stage, and receives participant share s of gross first payout W if that stage succeeds. For type i ,

$$V_i = -F - p_i A + p_i q_i s W. \quad (4)$$

The provider’s closed-transfer gross margin is

$$M_i = F + p_i A - p_i q_i s W = -V_i.$$

This equality is a bookkeeping benchmark, not a full business model. It excludes operating costs, commissions, market execution, hedging, copied trading profits or losses, financing income, taxes, payment defaults, marketing costs, and subsequent payouts.

When the evaluation and funded stages have the same rules and latent type, p_i and q_i should be close but need not be numerically identical because independent Monte Carlo samples are used. The first-payout probability is $p_i q_i$.

4.3 Provider break-even identities

For a population with $\dot{p} = \sum_i w_i p_i$ and $\dot{r} = \sum_i w_i p_i q_i$, participant expected contract cash flow per attempt is

$$\dot{V} = -F - \dot{p} A + \dot{r} s W.$$

The fee that sets closed-transfer provider margin to zero is

$$F^* = \dot{r} s W - \dot{p} A. \quad (5)$$

Holding the fee fixed, the maximum gross payout consistent with zero margin is

$$W^* = \frac{F + \dot{p} A}{\dot{r} s}.$$

These identities reveal why a population percentage is insufficient. A statement such as “95% lose” does not specify $\dot{p}$, $\dot{r}$, F , A , s , or W .

5. Numerical Design

5.1 Contract inputs

Table 2 states the contract benchmark.

Table 2. Baseline fictitious 50K contract inputs

Parameter	Symbol	Baseline
Nominal account label	not applicable	50K

Parameter	Symbol	Baseline
Evaluation target	U	\$3,000
Drawdown	D	\$2,000
Evaluation horizon	H	20 trading days
Funded-stage horizon	H	20 trading days
Evaluation fee	F	\$90
Activation fee	A	\$100
Participant payout share	s	90%
Gross first payout	W	\$3,000
Monitoring	not applicable	10 intraday steps per day

Note: The contract is fictitious and is not intended to reproduce any named provider. Dollar values are participant-facing contractual amounts.

No commercial provider is represented. The parameters are stylized and are chosen to preserve continuity with *Market Edge versus Contract Edge: Negative-Expectancy Trading under an Idealized Evaluation Contract* (Matilla Serrano, 2026b). The timing convention is one trading month for the evaluation and, conditional on passing, a second trading month for the first-payout stage.

5.2 Trader types

The baseline grid contains six trader types.

Table 3. Trader-type scenarios and unconstrained market P&L

Code	Latent edge	Daily drift	Daily volatility	Unconstrained expected 20-day market P&L
P-LV	Positive, lower variance	\$100	\$600	\$2,000
P-HV	Positive, higher variance	\$100	\$1,200	\$2,000
Z-LV	Zero, lower variance	\$0	\$600	\$0
Z-HV	Zero, higher variance	\$0	\$1,200	\$0
N-LV	Negative, lower variance	-\$75	\$600	-\$1,500
N-HV	Negative, higher variance	-\$75	\$1,200	-\$1,500

Note: Trader types are stylized scenarios. Expected 20-day market P&L equals daily drift multiplied by 20 before evaluation stopping rules are applied.

Within the positive group and within the nonpositive group, baseline population weight is divided equally across the relevant types. If the positive-edge population share is 5%, each positive type receives 2.5%; the remaining 95% is divided equally across the four nonpositive types.

The risk multiplier grid is $\ell \in \{0.5, 0.75, 1, 1.25, 1.5, 2\}$. The population analysis uses $\ell = 1$.

5.3 Monte Carlo method

The simulation uses 40,000 paths per type and risk scale, antithetic normal increments, and common random numbers across comparable specifications. A fixed seed of 20260727 is used to reproduce the numerical results. Each trading day contains ten monitoring steps. The evaluation and funded stages use independent path panels drawn from the same type distribution. Each antithetic pair consists of one Gaussian shock path and its sign-reversed counterpart.

At every monitoring point, cumulative P&L is checked against U and $-D$. A path that reaches neither boundary by day 20 is unresolved and does not pass. Because antithetic observations are dependent within pairs, the reported standard error is the standard error of the 20,000 pair means, not the independent-binomial shortcut. Ninety-five percent Monte Carlo confidence intervals use the paired standard error. For the product $p_i q_i$, the reported interval uses a first-order delta approximation because the evaluation and funded panels are independent. These intervals measure simulation noise only. They are not confidence intervals for an empirical trader population, which the paper does not observe.

The time grid cannot reproduce continuously monitored barriers exactly. Finer monitoring detects some between-grid boundary events earlier. A convergence check increases monitoring from 5 to 40 steps per day for the positive-drift, high-variance type. The estimated finite-horizon pass probability moves from 49.89% to 48.81%; the infinite-horizon analytical benchmark is 48.44%. The finite-horizon and infinite-horizon quantities are not required to coincide, but the direction and scale provide a useful diagnostic.

6. Trader-Level Results

6.1 The drift-volatility surface

Figure 1 maps finite-horizon pass probability across daily drift and volatility. Positive drift raises pass probability, but the effect of volatility depends on both drift and the deadline. At low volatility, many paths remain unresolved. More volatility can therefore increase upper-target arrivals even when drift is zero or negative. At sufficiently high activity, the lower boundary also arrives more often.

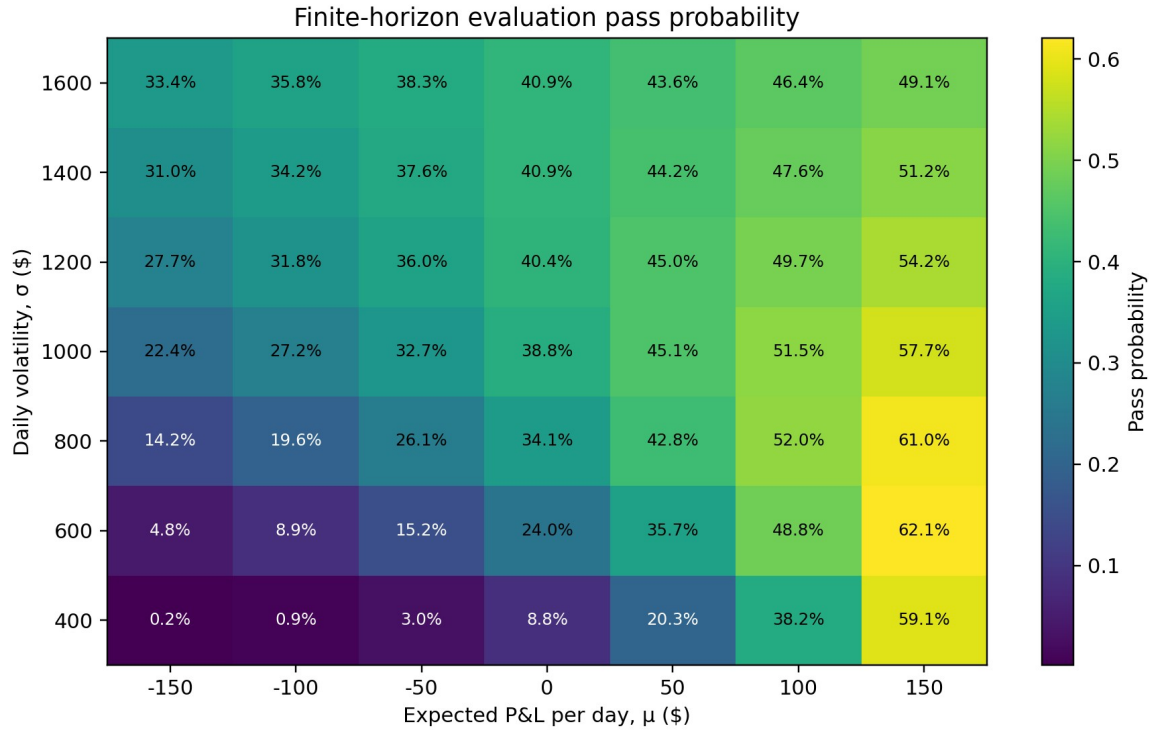


Figure 1. Finite-horizon pass probability across drift and volatility. Estimates use the 20-day fictitious contract and the baseline monitoring grid.

The infinite-horizon benchmark helps separate two effects. With zero drift, volatility does not alter the eventual upper-first probability of 40%; it alters how quickly the contract resolves. With nonzero drift, volatility also changes the balance between drift and boundary risk. The finite deadline makes both effects economically relevant.

Figure 2 shows the analytical benchmark directly. At zero drift, all volatility curves meet at 40%. With positive drift, lower volatility allows the directional component to dominate; with negative drift, additional variance can raise the chance that noise reaches the upper target before the adverse drift reaches the drawdown.

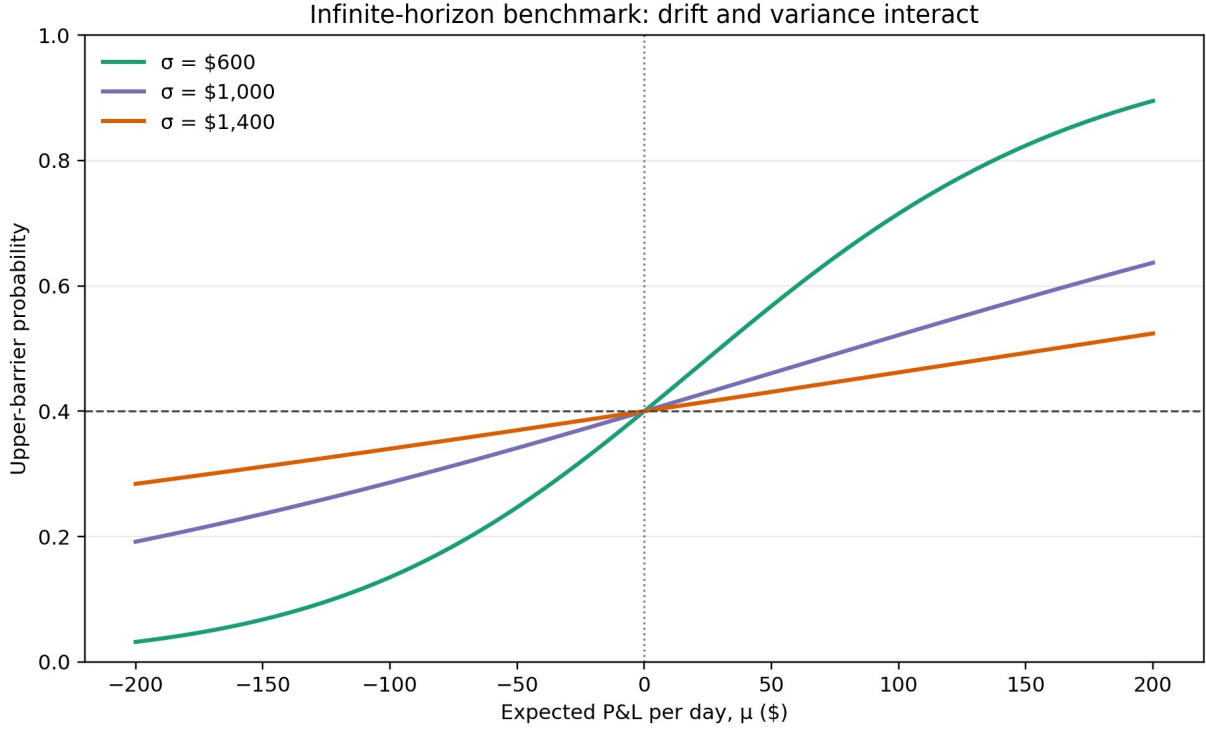


Figure 2. Infinite-horizon first-passage benchmark. The surface is analytical and omits the finite deadline.

6.2 Horizon and drift sensitivity

The 20-day assumption is retained for the main model, but it is not innocuous. Figure 3 varies the deadline from 10 to 60 trading days in a separate 20,000-path antithetic panel.

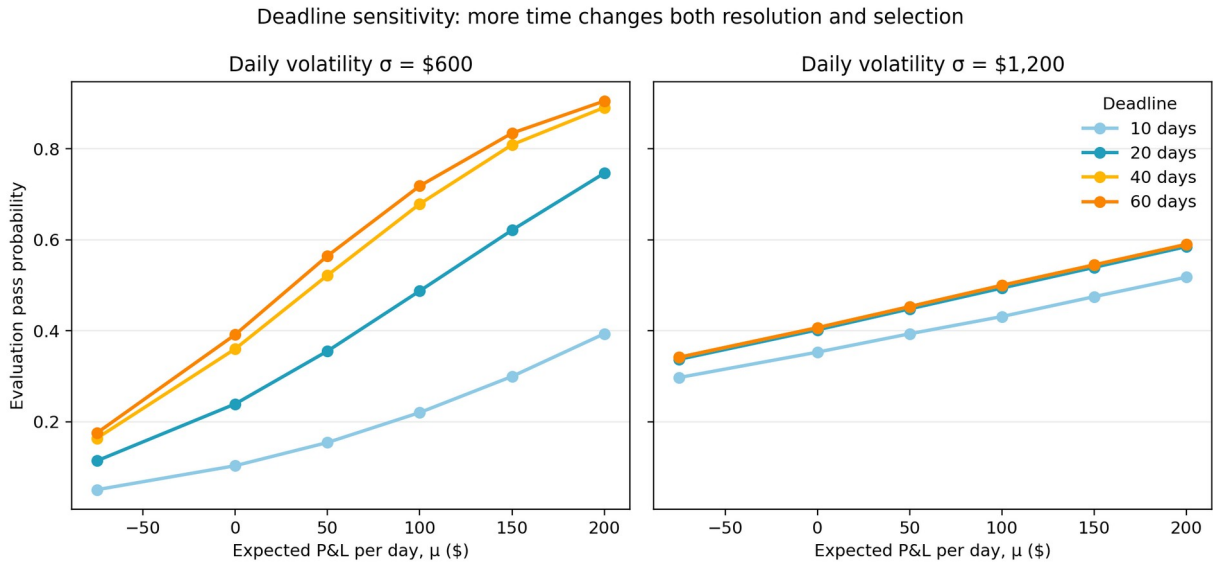


Figure 3. Deadline sensitivity across drift and volatility. Each curve changes the horizon while preserving the stated barriers and process parameters.

Table 4. Pass-probability sensitivity to drift, volatility, and horizon

Daily drift	Volatility	10 days	20 days	40 days	60 days
-\$75	\$600	5.08%	11.43%	16.40%	17.55%
\$0	\$600	10.36%	23.94%	36.05%	39.18%
\$100	\$600	22.05%	48.77%	67.88%	71.85%
\$200	\$600	39.36%	74.64%	89.08%	90.50%
-\$75	\$1,200	29.73%	33.73%	34.18%	34.19%
\$0	\$1,200	35.32%	40.22%	40.69%	40.69%
\$100	\$1,200	43.14%	49.40%	49.99%	50.01%
\$200	\$1,200	51.80%	58.51%	58.99%	58.99%

Note: Probabilities are finite-horizon Monte Carlo estimates under the stated drift, volatility, and deadline; they are not empirical trader frequencies.

More time helps the lower-variance positive types because their drift has time to accumulate. High-variance types resolve much earlier, so extending the deadline beyond 20 days changes their pass rates comparatively little. The deadline therefore changes not only the overall pass rate but also which types are favored by the filter.

6.3 Baseline type comparison

Table 5 reports the main outcomes at $\ell = 1$.

Table 5. Baseline results by trader type

Type	Pass p	95% MC CI	Funded success q	First payout $p q$	95% MC CI	Participant EV	Unconstrained E[20-day market P&L]
P-LV	48.75%	48.53–48.97%	48.57%	23.68%	23.52–23.83%	\$500.55	\$2,000
P-HV	49.72%	49.51–49.92%	49.28%	24.50%	24.36–24.64%	\$521.80	\$2,000
Z-LV	24.03%	23.69–24.38%	23.99%	5.77%	5.65–5.88%	\$41.65	\$0
Z-HV	40.39%	40.12–40.66%	40.46%	16.34%	16.19–16.50%	\$310.84	\$0
N-LV	11.71%	11.41–12.00%	11.43%	1.34%	1.29–1.39%	-\$65.58	-\$1,500
N-HV	33.84%	33.52–34.17%	33.70%	11.40%	11.25–11.56%	\$184.07	-\$1,500

Note: Evaluation and funded stages use independent path panels. Confidence intervals measure Monte Carlo uncertainty only.

Three comparisons are revealing.

First, positive drift matters. Holding volatility at \$600, the pass probability falls from 48.75% for P-LV to 24.03% for Z-LV and 11.71% for N-LV. The evaluation is not devoid of signal.

Second, variance can be an effective substitute for drift in the pass event. Z-HV passes 40.39% of evaluations, only 8.36 percentage points below P-LV. N-HV passes almost three times as often as N-LV. Those outcomes do not make N-HV a profitable market strategy; its expected 20-day market P&L remains $-\$1,500$.

Third, the contract and market rankings differ. N-HV has positive participant contract expectancy of \$184.07 per attempt under the deliberately permissive two-stage contract. This is the endogenous-probability counterpart to the contract-edge result in *Market Edge versus Contract Edge* (Matilla Serrano, 2026b). It is not an arbitrage, and it would be highly sensitive to path-dependent rules, execution costs, payout restrictions, and parameter stability.

6.4 Risk scale

Figure 4 shows pass probability across the risk multiplier. Increasing risk reduces unresolved outcomes and can initially improve the chance of reaching the target. The effect is not monotonic for every type.

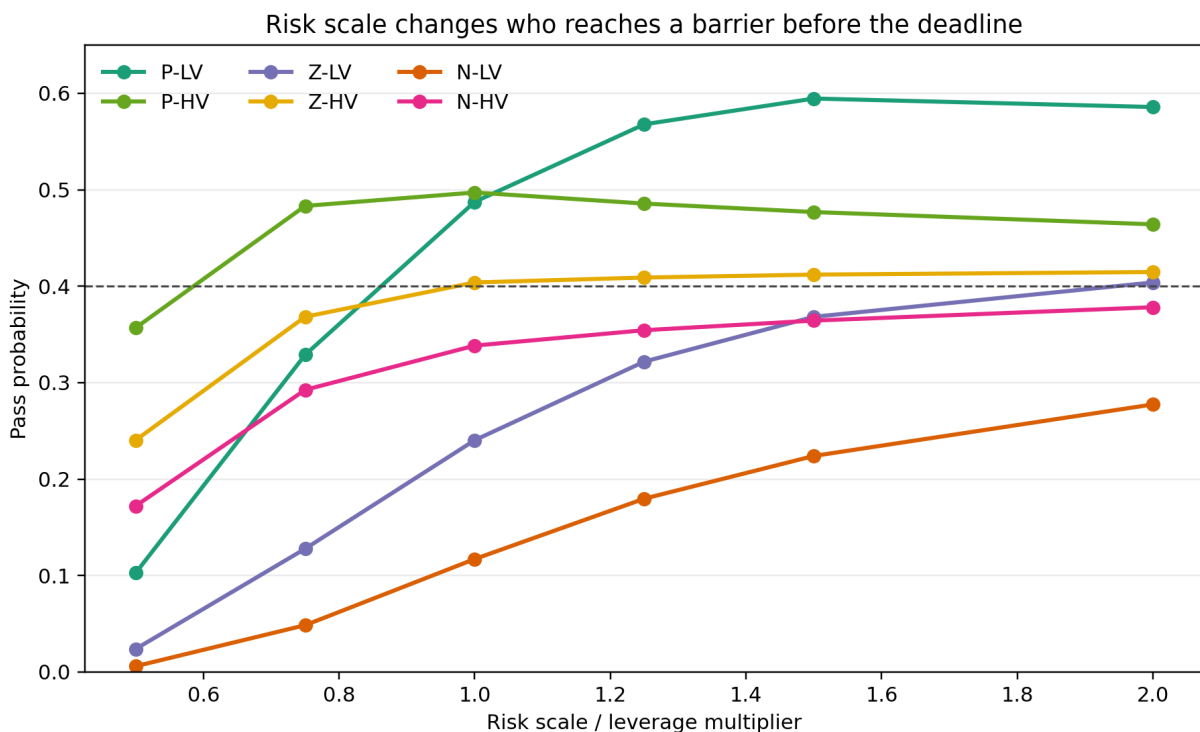


Figure 4. Risk scale, pass probability, and selection. Risk scale multiplies both daily drift and volatility while contractual dollar barriers remain fixed.

P-LV rises from a 10.26% pass probability at $\ell = 0.5$ to 59.46% at $\ell = 1.5$, then falls to 58.58% at $\ell = 2$. P-HV peaks earlier: 49.72% at $\ell = 1$, then declines to 46.42% at $\ell = 2$. For zero- and negative-drift types, higher risk mainly converts unresolved paths into barrier outcomes and continues to raise upper-target frequency over the reported grid.

The non-monotonicity follows from two opposing effects. At low ℓ , the deadline dominates: the path may simply move too slowly to reach either boundary, so more risk raises the chance of a decision. At higher ℓ , the variance term grows faster relative to the directional advantage in the barrier formula. For a positive-drift type, this weakens the protection provided by drift and brings the lower boundary forward as well as the upper one. Its pass probability can therefore peak at an interior risk scale. For a zero-drift type, the eventual upper-first benchmark stays at 40%, while more risk mainly removes unresolved paths. For a negative-drift type, more risk both removes unresolved paths and makes adverse drift less dominant relative to random movement; over the reported grid those two effects raise the pass rate toward the zero-drift benchmark. A wider leverage grid could eventually reveal a plateau or reversal, so monotonicity should not be assumed outside the simulated range.

This is a key incentive result. The risk scale that maximizes a pass probability need not maximize expected market wealth, long-run survival, or risk-adjusted performance. A threshold contract can reward “getting a result” within the deadline even when the underlying signal-to-noise ratio is unchanged.

6.5 Outcome decomposition

Figure 5 separates success, drawdown failure, and unresolved paths. Low-risk types can appear unsuccessful because they fail to resolve. High-risk types mostly resolve, but resolution includes both upper and lower barriers.

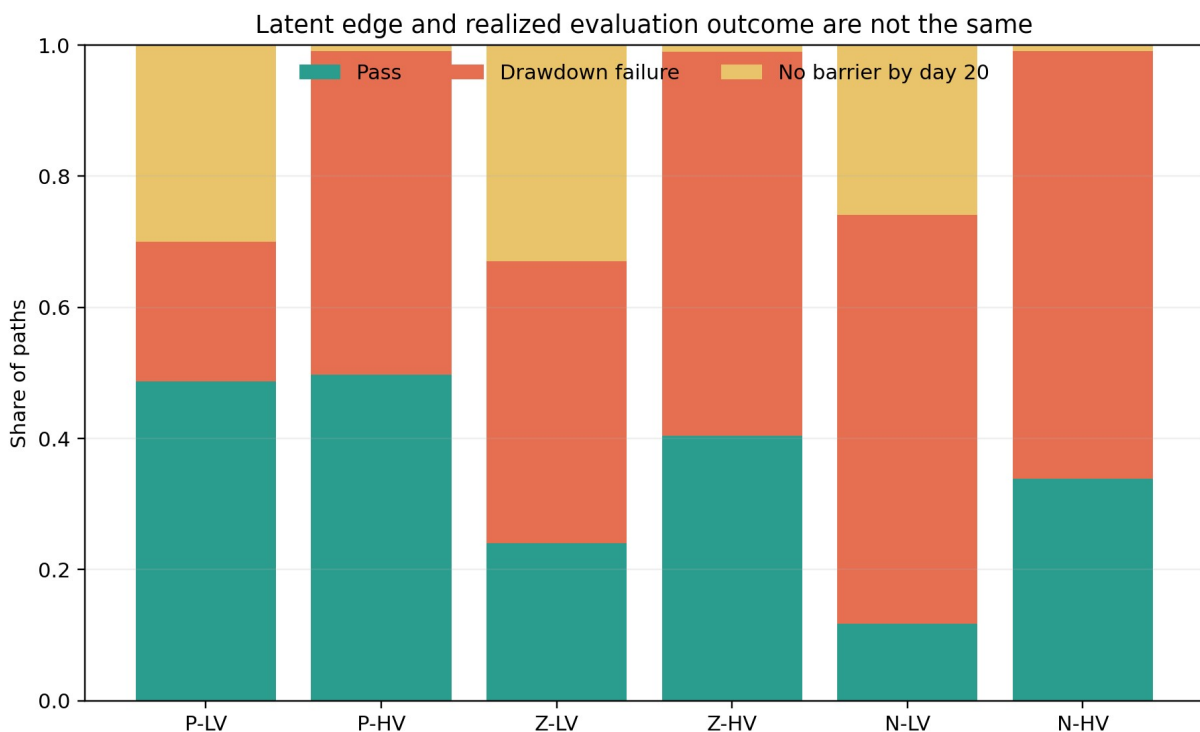


Figure 5. Evaluation outcome decomposition by trader type. Pass, drawdown failure, and unresolved outcomes sum to 100% within each type.

This distinction matters for interpretation. A deadline is not neutral. It penalizes slow strategies even if their expected return is positive, and it can make variance look like ability because variance produces faster decisions.

7. Population Selection

7.1 Population scenarios

Table 6 varies the positive-edge share from 5% to 50%. It does not claim that any row describes the industry.

Table 6. Population-composition scenarios

Positive-edge population	Population pass rate	First-payout rate	Positive among passers	Nonpositive among passers	Positive among payouts	Failed evaluations per payout
5%	28.58%	9.48%	8.61%	91.39%	12.70%	7.53
10%	29.67%	10.25%	16.60%	83.40%	23.50%	6.86
20%	31.84%	11.79%	30.92%	69.08%	40.87%	5.78
30%	34.02%	13.33%	43.42%	56.58%	54.23%	4.95
50%	38.36%	16.40%	64.17%	35.83%	73.44%	3.76

Note: Positive-edge and nonpositive population shares are scenario weights. Within each group, weight is divided equally across its baseline trader types.

When positive-edge traders are rare, most passers are nonpositive even though positive types have higher average pass probability. This is the base-rate effect. In the 5% scenario, positive types represent 8.61% of passers, an improvement over their 5% population share, but 91.39% of passers still come from the larger nonpositive group.

The second stage improves the composition. Positive types represent 12.70% of first-payout events in the 5% population and 40.87% in the 20% population. Selection is still imperfect. A two-stage threshold is stronger evidence than one stage, but it is not proof of positive drift.

The positive-type false-negative rate is 50.77% at the baseline risk scale: roughly half of the positive-drift evaluation paths do not pass within 20 days. That rate reflects this contract and these type parameters. It should not be generalized beyond them.

7.2 What the familiar 95/5 statement misses

The phrase “95% of traders lose and finance the 5% who win” combines at least three different groups:

1. participants who fail the evaluation;
2. participants whose latent expected market P&L is nonpositive; and
3. participants who never receive a payout.

They are not equivalent. Some positive-drift traders fail. Some negative-drift traders pass and receive a payout. Some passers fail the second stage. A population statistic about latent types does not directly reveal the provider’s cash flows.

In the 5% positive-edge scenario, the model produces 7.53 failed evaluations per first payout and 9.55 attempts without a payout per payout. Those counts sound favorable to the provider until the dollar amounts are included. A \$90 evaluation fee and a \$100 activation fee do not necessarily finance a 90% share of a \$3,000 first payout at the simulated event rate.

8. Provider Economics

8.1 Closed-transfer results

Figure 6 reports participant expected value and provider gross margin across population compositions.

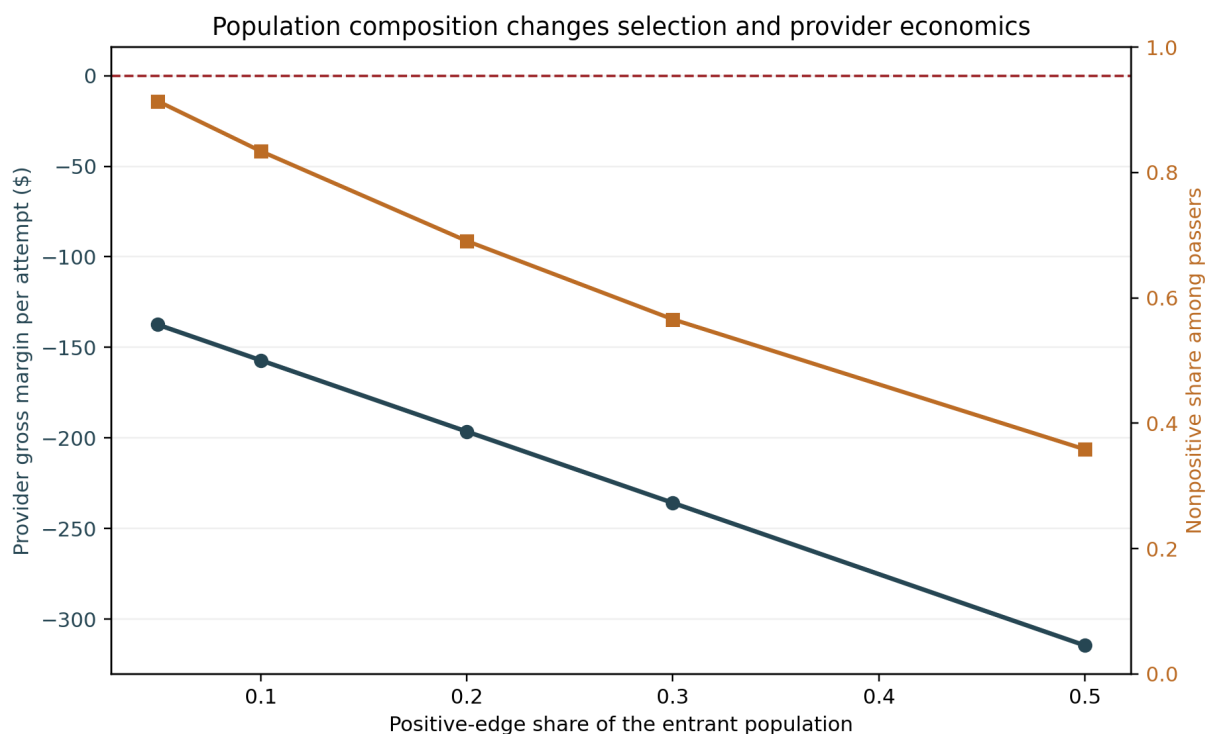


Figure 6. Population composition and provider gross margin. Margin is the closed-transfer benchmark before costs and other revenue.

Table 7. Participant and provider transfers by positive-edge share

Positive-edge population	Participant EV per attempt	Provider gross margin	Break-even evaluation fee	Maximum sustainable gross payout
5%	\$137.41	-\$137.41	\$227.41	\$1,389.64
10%	\$157.09	-\$157.09	\$247.09	\$1,297.19
20%	\$196.43	-\$196.43	\$286.43	\$1,148.47
30%	\$235.77	-\$235.77	\$325.77	\$1,034.07
50%	\$314.46	-\$314.46	\$404.46	\$869.62

Note: Provider margin is a closed-transfer benchmark equal to the negative of participant expected contract cash flow before operating costs and other revenues.

The 5% scenario gives expected fee revenue of \$90 per attempt and expected activation revenue of \$28.58. Expected participant payout is larger: $0.094813 \times 0.9 \times \$3,000 = \$256.00$, producing provider margin of approximately $-\$137.41$. Closed-transfer break-even therefore requires an evaluation fee of \$227.41, all else equal.

As the positive population share rises, pass and payout incidence increase. That improves participant expected value and worsens provider gross margin in the closed-transfer model. The result does not mean that a real provider prefers unskilled customers in every sense. Skilled traders may create market profits, marketing value, retention, or information; unskilled traders create payment-processing and acquisition costs. The result is only the transfer arithmetic of the stylized contract.

8.2 Changing the positive share is not enough

The first sensitivity asks whether a still more extreme 99/1 population would restore provider profitability. It does not under the original equal-within-group convention.

Table 8. Provider margin under extreme positive/nonpositive population splits

Positive-edge share	Nonpositive share	Population pass rate	First-payout rate	Provider margin
0%	100%	27.49%	8.71%	-\$117.74
1%	99%	27.71%	8.87%	-\$121.68
2%	98%	27.93%	9.02%	-\$125.61
5%	95%	28.58%	9.48%	-\$137.41

Note: Each row holds the within-group type weights and contract terms fixed while changing only the positive-edge population share.

The reason is visible at the type level. The provider earns \$65.58 per N-LV attempt in the closed-transfer benchmark because that type rarely reaches the first payout. The provider loses \$184.07 per N-HV attempt, \$310.84 per Z-HV attempt, and \$41.65 per Z-LV attempt. A label such as “nonpositive drift” therefore does not determine the transfer outcome. Variance and payout incidence matter inside the group.

8.3 Provider-profitability population frontier

Figure 7 varies two population dimensions at once. Let α be the positive-edge share. Let β be the high-variance share among nonpositive entrants. Positive mass is split equally between P-LV and P-HV. Within the nonpositive group, low-variance mass is split equally between Z-LV and N-LV, and high-variance mass is split equally between Z-HV and N-HV.

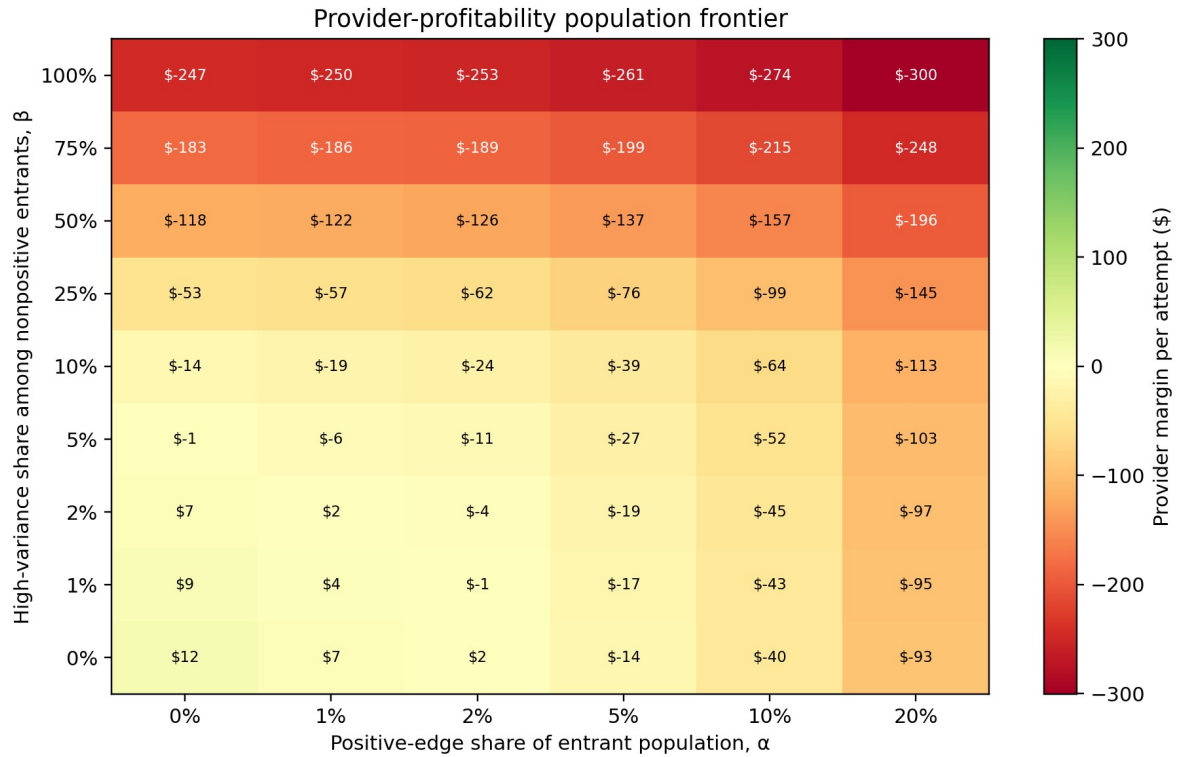


Figure 7. Provider-profitability frontier across population composition. The surface varies latent-edge and nonpositive variance composition under fixed contract terms.

Table 9. Provider-profitability frontier by population composition

Positive share α	Margin if $\beta = 0$	Margin if $\beta = 100\%$	Maximum β at break-even
0%	\$11.97	-\$247.45	4.61%
1%	\$6.74	-\$250.09	2.62%
2%	\$1.50	-\$252.73	0.59%
5%	-\$14.19	-\$260.64	No feasible value

Note: The frontier varies both latent-edge composition and the variance composition of the nonpositive group; it is a model sensitivity, not a market estimate.

The frontier answers the population question more precisely than a single 95/5 statement. Under this contract, a provider can break even with no positive-edge entrants only if fewer than 4.61% of nonpositive entrants are high-variance. At a 1% positive share, the limit falls to 2.62%. At 5%, margin is negative even if every nonpositive entrant is low-variance. This is not a universal “ideal population.” It is the break-even frontier generated by the six stylized types and the stated contract.

8.4 Contract-parameter sensitivity

Population is only one side of the arithmetic. Figure 8 varies the evaluation fee and gross first payout while holding the 5/95 path probabilities fixed.

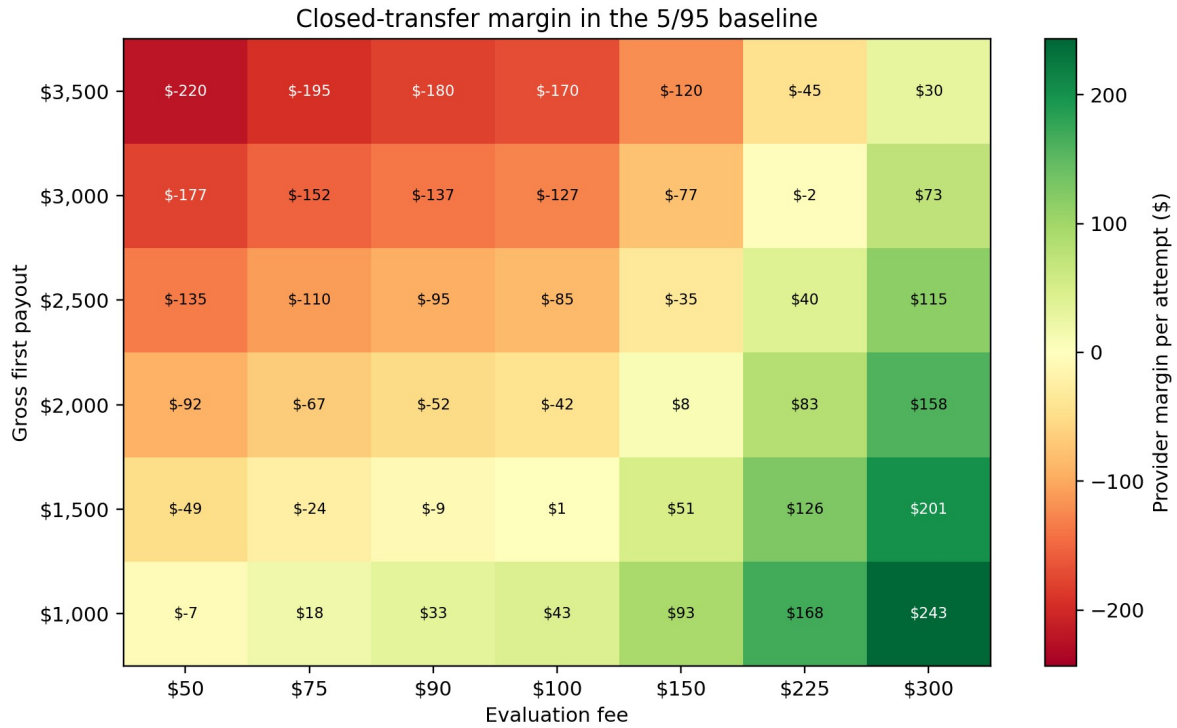


Figure 8. Fee and payout sensitivity in the 5/95 population. Path probabilities are held fixed, so the figure shows mechanical cash-flow sensitivity.

Table 10. One-way contract-parameter sensitivities

One-way change	Value	Provider margin
Evaluation fee	\$90	-\$137.41
Evaluation fee	\$225	-\$2.41
Evaluation fee	\$300	\$72.59
Activation fee	\$0	-\$165.99
Activation fee	\$200	-\$108.83
Participant share	70%	-\$80.53
Participant share	90%	-\$137.41
Participant share	100%	-\$165.86
Gross first payout	\$1,000	\$33.25
Gross first payout	\$1,500	-\$9.42
Gross first payout	\$3,000	-\$137.41

Note: Each row changes one contractual input while holding path probabilities and all other inputs at their 5/95 baseline values.

The one-way results recover the break-even identities and show their scale. Activation revenue helps, but even a \$300 activation fee leaves margin negative at the baseline event rates. Lower payout shares or payouts improve provider margin; higher evaluation fees do the same. These are mechanical sensitivities, not pricing recommendations. They deliberately hold entry, risk choice, retry behavior, and reputation fixed.

8.5 External funded-to-payout benchmark

The symbol q is conditional: it is the probability of reaching a payout after the evaluation has already been passed. That denominator matters. A September 2024 *Finance Magnates* article, *Exclusive: Only 7% of 300,000 Prop Trading Accounts Achieved Payouts*, reported figures supplied by a prop-technology vendor covering more than 300,000 accounts, 100,000 traders, and ten firms. According to the article, 14% of traders obtained a funded account and about 45% of those funded traders received a payout (Chmiel, 2024). Multiplication gives $0.14 \times 0.45 = 6.3\%$, which the source describes in rounded form as about 7% of all traders. Therefore 45%, not 7%, is the comparable conditional funded-to-payout statistic.

The evidence is useful but not interchangeable with the model. It is vendor-reported rather than a peer-reviewed cohort study; its unit is traders, whereas several model tables use purchased attempts; rule versions, retries, and observation windows are not aligned with the 20-day synthetic contract. No public account-level dataset or accompanying data dictionary was located for the reported figures, so the aggregate funnel cannot be independently reproduced from the article alone. Table 11 therefore treats 45% as a calibration sensitivity, not as the true value of q_i for every trader type.

Table 11. Conditional-funded-success calibration in the 5/95 scenario

Calibration	Evaluation pass $\dot{p}$	Conditional payout $\dot{q} = \dot{r} / \dot{p}$	First-payout rate $\dot{r}$	Participant EV	Provider margin
Endogenous path model	28.58%	33.17%	9.48%	\$137.41	-\$137.41
Model $\dot{p}$, external $q = 45\%$	28.58%	45.00%	12.86%	\$228.67	-\$228.67
Reported funnel $p = 14\%$, $q = 45\%$ under model prices	14.00%	45.00%	6.30%	\$66.10	-\$66.10

Note: The reported funnel is vendor-supplied aggregate evidence. The last row combines those reported probabilities with the paper's synthetic prices and must not be read as an estimate of the firms in the report.

The last row combines the reported probabilities with this paper's synthetic \$90 fee, \$100 activation charge, 90% share, and \$3,000 gross payout; it is not an estimate of the firms in the report. The sensitivity also clarifies the direction of the economics. Holding the model pass rate fixed, raising the conditional payout rate from 33.17% to 45% benefits the participant and makes the closed-transfer provider margin more negative.

8.6 Sources outside the closed-transfer benchmark

A complete provider model could add:

$$M^{\text{full}} = F + \dot{p} A - \dot{r} s W + R_{\text{market}} + R_{\text{other}} - C_{\text{operating}} - C_{\text{acquisition}} - C_{\text{default}},$$

where the additional terms represent market activity, other revenue, operating expenses, customer acquisition, and payment or counterparty losses. The signs and magnitudes are empirical questions.

If a provider copies or hedges selected trading activity, participant skill can become an asset rather than only a payout liability. If all accounts are simulated and no activity is monetized in external markets, fees and restrictions must carry more of the financing burden. Public legal and regulatory materials show that simulated-stage representations can be central to some offerings, but the present paper does not infer an industry-wide operating model from an individual case.

8.7 Sustainability is jointly determined

The simple margin identities hold population composition fixed. In reality, raising fees or tightening rules can change who enters, how they trade, how often they retry, and how the provider is perceived. A sustainable contract must satisfy a joint problem:

- prices and rules must finance payouts and costs;
- the product must remain attractive enough to generate entry;
- the rules must control adverse selection and participant risk shifting;
- the provider must comply with applicable legal and payment constraints; and
- any market replication process must distinguish skill from luck sufficiently well to create value.

Comparing complete rule packages and estimating behavioral responses are outside the scope of the present analysis.

9. Interpretation and Contract Design

9.1 Passing is evidence, not proof

A pass is informative when positive types are more likely to pass than nonpositive types. It is conclusive only under implausibly strong assumptions. The evidentiary value depends on:

- the prior population composition;
- target and drawdown geometry;
- the deadline;
- monitoring frequency;
- the trader's available risk scale;
- whether the second stage repeats the same test; and
- additional rules that penalize concentrated or volatile paths.

Reporting only the pass rate discards most of this information. A useful evaluation study should report the type-conditioned pass rates and the composition of passers.

9.2 Variance is not always selected in the same direction

The phrase “the evaluation selects variance” is also too broad. Variance can help by resolving a short-horizon contract, but it can hurt by bringing the drawdown boundary forward. The direction depends on drift, current state, barrier geometry, time remaining, and rule design. The baseline results show both patterns: N-HV benefits strongly from greater variance relative to N-LV, while P-HV reaches its best pass probability at a lower risk multiplier than P-LV.

9.3 The target-to-drawdown ratio is not a selection statistic by itself

The Target-to-Drawdown Ratio,

$$TDR = \frac{U}{D},$$

summarizes barrier geometry but does not rank evaluation difficulty without a path model and deadline. A larger TDR places the target farther from the starting point relative to the drawdown and therefore reduces the zero-drift, infinite-horizon upper-first probability:

$$p(\infty) = \frac{1}{1 + TDR}.$$

For positive drift, variance, and a finite horizon, the full expression is needed. Minimum days, daily-loss rules, trailing boundaries, and payout rules cannot be inferred from TDR at all. TDR is a useful descriptive ratio, not an original contribution or a sufficient measure of pass probability.

9.4 Repeated attempts complicate inference

The model evaluates one purchased attempt. If a participant can retry, the probability of at least one pass rises even when the latent process is unchanged. Observing one eventual pass after many undisclosed attempts is weaker evidence than observing a pass on the first attempt. Dependence also matters: strategy regimes, learning, fatigue, and changing risk choices make attempts non-identical.

An empirical implementation should record the complete attempt history, not only the first visible success. It should also distinguish genuine learning from mechanical selection of a favorable path.

10. Limitations and Extensions

The results are conditional on a deliberately simple model.

First, arithmetic Brownian P&L is not a realistic universal description of trading. It allows unbounded increments, constant drift and volatility, and independent Gaussian shocks. Real strategies can have skewness, fat tails, serial dependence, volatility clustering, regime changes, and discontinuous losses.

Second, drift is treated as known latent type. In practice, it must be estimated from noisy data, may change over time, and may vanish after costs. The values in Table 3 are scenarios, not calibrated estimates of retail trader ability.

Third, the type population is hypothetical. The 5/95 split is included because it is frequently asserted in informal discussion, not because this paper validates it. Credible population

estimation would require account-level data covering entries, retries, failures, funded transitions, payouts, rule versions, and ideally external-market returns.

Fourth, the funded stage is an independent repetition of the evaluation stage. Real funded rules may differ, and the same trader's stages may be dependent. Selection, learning, changing leverage, and psychological responses can make q_i differ materially from p_i .

Fifth, the provider model is a closed-transfer benchmark. It is not a firm valuation or a statement about industry profitability. Subsequent payouts, subscriptions, refunds, discounts, taxes, operating expenses, market replication, and participant acquisition are excluded.

Sixth, the monitoring grid is discrete. Although sensitivity checks use finer steps, a continuous barrier can be crossed between observations. Exact or continuity-corrected methods would improve numerical precision.

Seventh, the model holds risk scale fixed through each attempt. An optimizing participant would choose risk dynamically as P&L and time remaining change. That problem belongs to stochastic control and may alter both selection and provider economics.

The main extensions are to add minimum-day, consistency, daily-loss, trailing-drawdown, and payout rules; estimate trader-type mixtures from data; model repeated entry; allow dynamic risk choice; and compare simulated selection with subsequent market performance.

11. Conclusion

Trading evaluations select paths before they select people. Drift matters, but so do variance, deadlines, drawdown geometry, and risk choice. In the baseline simulation, a zero-drift, high-variance type passes 40.39% of evaluations, and a negative-drift, high-variance type passes 33.84%. Some nonpositive types therefore generate positive participant contract expectancy under the permissive contract even though their expected market P&L is not positive.

The population results make the inference problem explicit. When positive types are only 5% of entrants, most passers are still nonpositive types. A second stage improves the composition of observed payout recipients but does not eliminate luck. Conversely, roughly half of positive-type evaluation paths fail to pass within the deadline.

The provider result is equally important. A large majority of nonpositive or unsuccessful participants does not automatically finance a contract. Under the stated prices, payouts, and path probabilities, the closed-transfer provider margin is negative even in the 5/95 scenario. The population frontier shows why: both latent-edge composition and variance composition govern payout incidence. Sustainability depends on the entire contract and business system.

The external calibration reinforces a separate point about denominators. A reported 45% funded-to-payout conversion is a candidate benchmark for the conditional parameter q ; the roughly 7% headline is an unconditional funnel outcome. Substituting one for the other would change both participant value and provider margin for the wrong reason.

The appropriate conclusion is neither that evaluations are meaningless nor that they are clean tests of skill. They are noisy, designed selection mechanisms. To understand what they select,

one must model the process, the population, the participant payoff, and the provider payoff together.

References

- Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance*, 65(1), 179–216. <https://doi.org/10.1111/j.1540-6261.2009.01527.x>
- Basak, S., Pavlova, A., & Shapiro, A. (2007). Optimal asset allocation and risk shifting in money management. *Review of Financial Studies*, 20(5), 1583–1621. <https://doi.org/10.1093/rfs/hhm026>
- Broadie, M., Glasserman, P., & Kou, S. G. (1997). A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4), 325–349. <https://doi.org/10.1111/1467-9965.00035>
- Brown, K. C., Harlow, W. V., & Starks, L. T. (1996). Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance*, 51(1), 85–110. <https://doi.org/10.1111/j.1540-6261.1996.tb05203.x>
- Carpenter, J. N. (2000). Does option compensation increase managerial risk appetite? *Journal of Finance*, 55(5), 2311–2331. <https://doi.org/10.1111/0022-1082.00288>
- Carpenter, J. N., & Lynch, A. W. (1999). Survivorship bias and attrition effects in measures of performance persistence. *Journal of Financial Economics*, 54(3), 337–374. [https://doi.org/10.1016/S0304-405X\(99\)00040-9](https://doi.org/10.1016/S0304-405X(99)00040-9)
- Chevalier, J., & Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105(6), 1167–1200. <https://doi.org/10.1086/516389>
- Chmiel, D. (2024, September 18). Exclusive: Only 7% of 300,000 prop trading accounts achieved payouts. *Finance Magnates*. <https://www.financemagnates.com/forex/analysis/exclusive-only-7-of-300000-prop-trading-accounts-achieved-payouts/>
- Duffy, J., Matros, A., & Valencia, Z. (2023). Contests with entry fees: Theory and evidence. *Review of Economic Design*, 27(4), 725–761. <https://doi.org/10.1007/s10058-022-00318-2>
- Fama, E. F., & French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, 65(5), 1915–1947. <https://doi.org/10.1111/j.1540-6261.2010.01598.x>
- Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer. <https://doi.org/10.1007/978-0-387-21617-1>
- Harvey, C. R., & Liu, Y. (2022). Luck versus skill in the cross section of mutual fund returns: Reexamining the evidence. *Journal of Finance*, 77(3), 1921–1966. <https://doi.org/10.1111/jofi.13123>
- Hodder, J. E., & Jackwerth, J. C. (2007). Incentive contracts and hedge fund management. *Journal of Financial and Quantitative Analysis*, 42(4), 811–826. <https://doi.org/10.1017/S0022109000003409>
- Kaminski, K. M., & Lo, A. W. (2014). When do stop-loss rules stop losses? *Journal of Financial Markets*, 18, 234–254. <https://doi.org/10.1016/j.finmar.2013.07.001>
- Lim, B. C. (2026a). *Phantom Generosity: Contract Design and Value in the Market for Proprietary-Trading Evaluations*. SSRN preprint, posted 5 August 2026. <https://doi.org/10.2139/ssrn.7184138>
- Lim, B. C. (2026b). *The Price of a Funded Account: An Actuarial Analysis of Proprietary Trading Firm Challenges*. SSRN preprint, posted 5 August 2026. <https://doi.org/10.2139/ssrn.7178078>

- Magdon-Ismail, M., Atiya, A. F., Pratap, A., & Abu-Mostafa, Y. S. (2004). On the maximum drawdown of a Brownian motion. *Journal of Applied Probability*, 41(1), 147–161. <https://doi.org/10.1239/jap/1077134674>
- Maile, T. (2026). Proprietary trading firms: A strategic economic lever for Lesotho: Industry analysis, developmental opportunities, and policy recommendations. Preprint available on SSRN, abstract 6785399. <https://doi.org/10.2139/ssrn.6785399>
- Matilla Serrano, F. (2026b). *Market Edge versus Contract Edge: Negative-Expectancy Trading under an Idealized Evaluation Contract*. Working paper.
- Ng, L. (2026). Pass-first-pay-later: A structural and comparative analysis of the deferred-fee prop trading model. Preprint available on SSRN, abstract 6672818. <https://doi.org/10.2139/ssrn.6672818>
- Tomàs Fernández, M. (2026). *Valuing Proprietary Trading Firm Evaluation Contracts: Closed Form, Cross-Firm Dispersion and the Source of the Margin*. SSRN preprint, posted 12 August 2026. <https://doi.org/10.2139/ssrn.7260819>
- Ross, S. A. (2004). Compensation, incentives, and the duality of risk aversion and riskiness. *Journal of Finance*, 59(1), 207–225. <https://doi.org/10.1111/j.1540-6261.2004.00631.x>
- Whitmeyer, M. (2023). Submission costs in risk-taking contests. *Games and Economic Behavior*, 142, 101–112. <https://doi.org/10.1016/j.geb.2023.07.014>
- Wright, R. D., & Ramsay, J. B. (1979). On the effectiveness of common random numbers. *Management Science*, 25(7), 649–656. <https://doi.org/10.1287/mnsc.25.7.649>

Appendix A. Analytical Benchmark

Shift the process from interval $[-D, U]$ to $[0, U + D]$ by defining $Z_t = X_t + D$. The starting point is $Z_0 = D$. For $dZ_t = \mu dt + \sigma dB_t$, let $h(z)$ be the probability of reaching $U + D$ before zero. It solves

$$\mu h'(z) + \frac{1}{2} \sigma^2 h''(z) = 0,$$

with boundary conditions $h(0) = 0$ and $h(U + D) = 1$. For $\mu \neq 0$,

$$h(z) = \frac{1 - \exp(-2\mu z / \sigma^2)}{1 - \exp(-2\mu(U + D) / \sigma^2)}.$$

Evaluating at $z = D$ gives the formula in Section 4. When $\mu \rightarrow 0$, L'Hôpital's rule gives $h(D) = D / (U + D)$.

The formula has no deadline. It is used as an analytical benchmark, not as a substitute for the finite-horizon simulation.

Appendix B. Population Identities

Let G_+ denote positive-drift types and G_0 the remaining types. The joint probability of positive drift and passing is

$$Pr(G_+, \text{pass}) = \sum_{i \in G_+} w_i p_i.$$

Dividing by the population pass probability gives the positive share among passers. The same calculation with $p_i q_i$ gives the positive share among first-payout events.

The expected number of failed evaluations per payout event is

$$\frac{\sum_i w_i (1 - p_i)}{\sum_i w_i p_i q_i} = \frac{1 - \dot{p}}{\dot{r}}.$$

The expected number of attempts without a payout per payout event is

$$\frac{1 - \dot{r}}{\dot{r}}.$$

Neither ratio is a profit measure until cash amounts are applied.

Appendix C. Additional Sensitivity Results

C.1 Infinite-horizon geometry

Figure 2, reported in Section 6.1, compares the analytical probability across drift and volatility.

At zero drift, all volatility curves meet at 40%. With positive drift, lower volatility allows drift to dominate and raises the upper-first probability. With negative drift, higher volatility can raise the chance that noise reaches the upper target before the adverse drift reaches the drawdown.

C.2 Monitoring convergence

Table 12. Time-discretization validation

Steps per day	Paths	P-HV finite-horizon pass	Paired MC SE	95% MC CI
5	20,000	49.89%	0.15%	49.60–50.17%
10	20,000	49.55%	0.15%	49.25–49.84%
20	20,000	49.21%	0.15%	48.91–49.50%
40	20,000	48.81%	0.15%	48.51–49.10%

Note: The validation holds the trader type and path count fixed while changing monitoring frequency. Reported intervals use antithetic pair means.

Finer monitoring modestly reduces the measured pass probability in this case because lower-barrier crossings that occur between coarser observations are captured sooner.

The independent-binomial formula would give approximately 0.35% for these 20,000-path runs. It is not the reported uncertainty measure because the antithetic observations are deliberately dependent within pairs.

C.3 Alternative population weights

The equal-within-group weighting convention is transparent but not uniquely natural. Section 8.3 replaces that single convention with a two-dimensional population frontier. If low-variance types are more common, the population pass rate and payout burden fall. If high-variance nonpositive

types are more common, the false-discovery share and payout burden rise. The workbook reports the complete grid so that the break-even boundary can be audited.

Declarations

Funding. No external funding was received for this research.

Conflicts of interest. No financial or commercial ties exist with any trading-evaluation provider. The contract and population scenarios are fictitious.

Supplementary materials. The simulation code, numerical inputs, simulated aggregate outputs, workbook, and figure-generation materials are available on Zenodo: <https://doi.org/10.5281/zenodo.22770349>.

Use of synthetic data. No customer-level or provider-confidential data are used. All trader types, populations, contracts, and paths are simulated.