Money Flow and "Following Smart Money": Why a High Hit Rate Is Not Positive Expected Value

A mathematical test of pari-mutuel racing pools | Sigma Quant Research | September 2026

Abstract

This paper tests a popular claim: if you can read money flow and follow the big bettors, you can profit from horse racing in the long run. We deliberately adopt three strong assumptions (informed bettors know the winner, know the exact probabilities, or know only approximate probabilities) and prove each in turn that a follower's expected return cannot exceed zero, and falls below the field-wide average of $-\tau$ once the signal is noisy. A synthetic simulation (100,000 races, 12 runners each) shows that "backing the horse with the most money" wins about 25.5% of the time versus 8.3% at random, yet returns about −22.6% on average. The sensible use of money flow is to blend it with your own probability estimate by precision weighting, not to replace that estimate.

Three takeaways

1. A high hit rate is not the same as value.
2. Money flow is information, not a price advantage: by the time you see the signal, the price has usually absorbed it.
3. With your own probability model, use money flow to calibrate it; without one, do not bet on flow alone.

1 Setup and Notation

In a pari-mutuel pool, horse i wins with probability p_i. Its win-pool money is W_i, the total pool is P, and its money share is $s_i = W_i/P$. The takeout rate is τ (we use 17.5% as an illustrative parameter; check the latest Hong Kong Jockey Club betting rules for actual figures). Dividends come from the pool after takeout, shared among winning stakes, so they are not fixed in advance. To stay focused we ignore rounding, dead heats and the impact of our own bets on the odds.

$$D_i = \frac{1-\tau}{s_i}, \qquad s_i = \frac{W_i}{P} \tag{1}$$

A \$1 stake returns D_i only if the horse wins and zero otherwise. Hence:

$$E[R_i] = p_i D_i = \frac{(1-\tau)\, p_i}{s_i} \tag{2}$$

Theorem 1 (identity): the money-weighted average return over the whole field is always $1 - \tau$.

$$\sum_i s_i E[R_i] = (1-\tau)\sum_i p_i = 1-\tau \tag{3}$$

Proof: multiply Eq. (2) by s_i; the shares cancel, leaving $(1-\tau)\Sigma p_i$, and all win probabilities sum to 1. ■

In other words, following the money as a whole loses τ on average. A positive expectation exists only for individual horses the market underweights (money share below the threshold below):

$$E[R_i] \geq 1 \iff s_i \leq (1-\tau)\, p_i \tag{4}$$

Example 1: A Four-Horse Mini Pool

Horse	Money (\$10k)	Share s	Dividend D	True p	Expected net return
A	40	40%	2.06	30%	−38.1%
B	30	30%	2.75	35%	−3.8%
C	20	20%	4.13	22%	−9.3%

© Sigma Quant. All rights reserved. No reproduction or modification without permission. | sigmaquanthk.com

Horse	Money ($10k)	Share s	Dividend D	True p	Expected net return
D	10	10%	8.25	13%	+7.3%
Weighted avg.	100	100%	—	100%	−17.5%

Horse A, with the most money, has the worst expected return: its 40% share is far above the threshold $(1-\tau)p = 24.75\%$. Only D, the lightest, sits below its threshold. The weighted average equals $-\tau$, as Theorem 1 says.

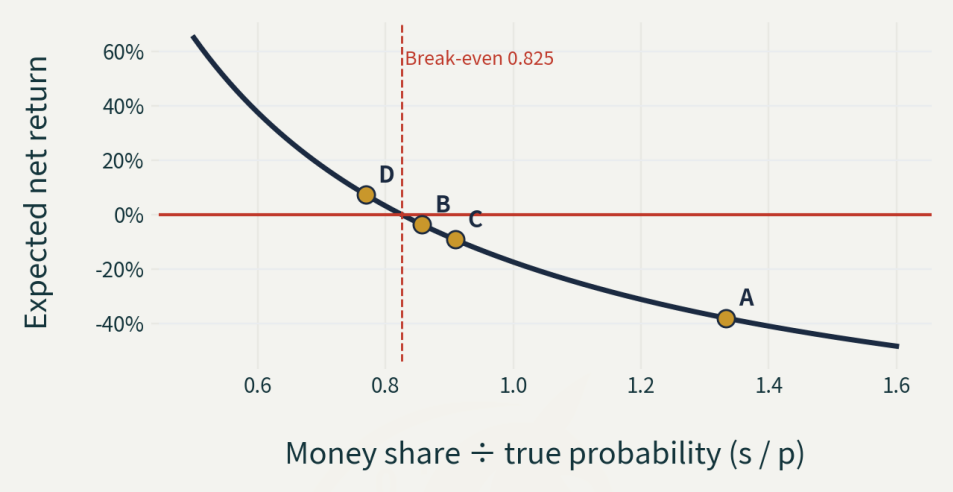


Figure 1. Expected net return depends only on "money share ÷ true probability". The four points come from Example 1; the red line marks break-even (ratio 0.825).

2 Three Bold Assumptions, Tested

Assumption A: informed bettors know the winner, with unlimited, competing capital. Knowing the winner pays whenever $s_w \le 1-\tau$, so competition pushes that horse's share up to $1-\tau$ and its dividend towards 1.00:

$$D_w = \frac{1-\tau}{s_w} \ge 1 \iff s_w \le 1-\tau \Rightarrow D_w \to 1.00 \tag{5}$$

In practice most winners pay clearly more than 1.00, so Assumption A fails. Even if it held, a follower would at best break even.

Assumption B: informed bettors know each horse's exact probability p. Let a horse have uninformed money W in a total pool P, and let the informed side add x. Because $W < P$, the odds fall strictly as x grows, and competition drives the informed side to add money until expected value hits zero:

$$o(x) = \frac{(1-\tau)(P+x)}{W+x}, \qquad o'(x) = \frac{(1-\tau)(W-P)}{(W+x)^2} < 0 \tag{6}$$

$$x^* = \frac{p(1-\tau)P - W}{1-p(1-\tau)} \Rightarrow p \cdot o(x^*) = 1 \tag{7}$$

Proof: the derivative in Eq. (6) is always negative, so $p \cdot o(x) - 1$ decreases monotonically in x; setting it to zero gives Eq. (7). Any follower entering after x^* has a negative expected return. ■

Example 2: A Horse With a 50% Chance

Let p = 50%, P = 1,000,000, uninformed money W = 350,000 and τ = 17.5%. The starting odds are 2.36, an expected net return of +17.9%. When the informed side has added $x^* \approx 106{,}000$, the odds fall to 2.00 and the edge is gone. If you then follow with 10,000, the odds become 1.97 and your expected net return is −1.3%; with 50,000 they become 1.88 and the return is −5.8%. Even if the big bettors are perfectly right, the price you get is no longer a value price.

© Sigma Quant. All rights reserved. No reproduction or modification without permission. | sigmaquanthk.com

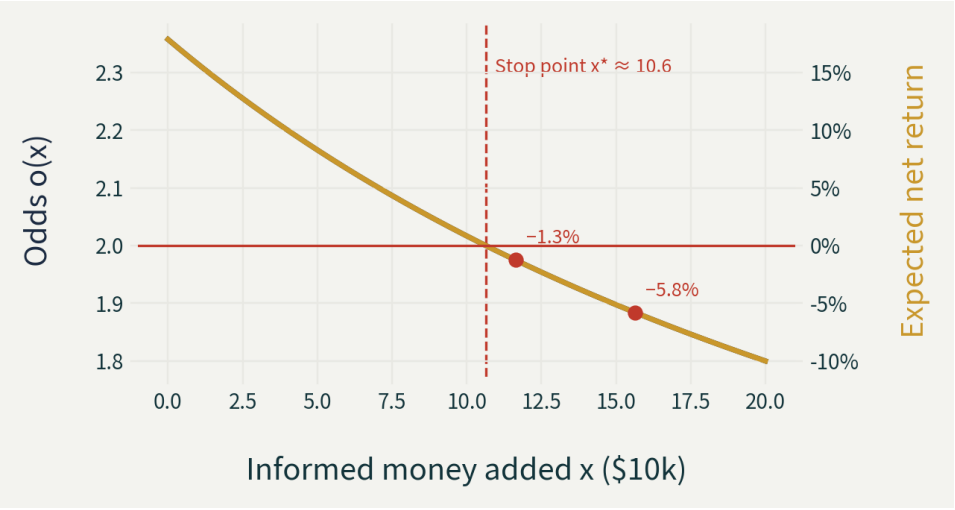


Figure 2. Odds and expected net return path for Example 2. The dashed line is the informed side's stopping point x*; anyone entering later has negative expected value.

Assumption C: informed bettors know only approximate probabilities. Treat money share as "true probability plus noise" and use a log-normal approximation:

$$\ln s_i = \ln p_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_\varepsilon^2), \quad \ln p_i \sim N(m, \sigma_u^2) \tag{8}$$

Theorem 2 (the heavier the money, the more the return is overstated):

$$E[p \mid s] \propto s^{\kappa}, \quad \kappa = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} \Rightarrow E[R \mid s] = C\, s^{\kappa - 1} \tag{9}$$

Proof: in the Gaussian model, $E[\ln p \mid \ln s] = m + \kappa(\ln s - m)$ with $\kappa < 1$, so the conditional expectation shrinks towards the mean. Exponentiating gives $E[p \mid s] \propto s^{\kappa}$, and dividing by s gives Eq. (9). ■ Intuitively, part of the highest money share is just positive noise, so following it means paying the winner's curse.

Example 3: The Cost of Doubling the Money Share

Noise weight 1 − κ	Return factor when share ×2	Return factor when share ×5
0.07	0.953	0.893
0.15	0.901	0.786
0.30	0.812	0.617

For $1 - \kappa = 0.15$, raising the money share from 10% to 20% multiplies the expected return by 0.901, about 9.9% lower again; going to 50% multiplies it by 0.786. This factor comes before takeout is considered.

3 Simulation and Empirical Evidence

What follows is a **synthetic simulation**, not Hong Kong Jockey Club history; its purpose is to verify the mechanism. Setup: 100,000 races of 12 runners; ln strength ~ $N(0, 0.9^2)$, normalised to p; money share $\ln s = \ln p + \varepsilon$ with $\sigma_\varepsilon = 0.25$; own model $\ln p_m = \ln p + \delta$ with $\sigma_\delta = 0.35$; $\tau = 17.5\%$; flat $1 bets.

Strategy	Hit rate per bet	Mean return ± s.e.
Top money-share horse	25.5%	−22.6% ± 0.5%
Top three by money share	17.8%	−20.4% ± 0.3%
Own model's top pick	24.6%	−13.7% ± 0.6%
└ agrees with top money (62.6% of races)	28.8%	−19.9% ± 0.5%
└ conflicts with top money (37.4% of races)	17.6%	−3.4% ± 1.2%
Own model, bet only if edge > 5%	7.7%	+2.5% ± 0.9%

© Sigma Quant. All rights reserved. No reproduction or modification without permission. | sigmaquanthk.com

Strategy	Hit rate per bet	Mean return ± s.e.
Blended (w = 0.34), bet only if edge > 5%	6.6%	+14.7% ± 2.4%

A random pick wins 1/12 ≈ 8.3% of the time, so following money really does have a high hit rate; but the return of −22.6% is worse than the field average of −17.5%, as Theorem 2 predicts. We also check the exponent using twenty quantile bins: the fitted slope of ln(win rate / s) on ln s is −0.0713, against a theoretical κ − 1 = −0.0716.

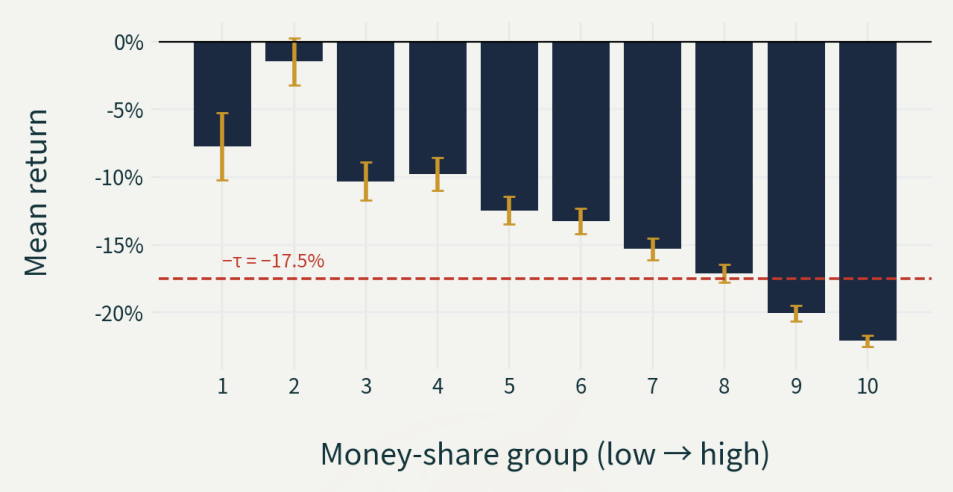


Figure 3. Simulated mean return by money-share decile. Overall, the heavier the money, the lower the return (low-share bins are noisier); all bins are below break-even, and the dashed line is −τ.

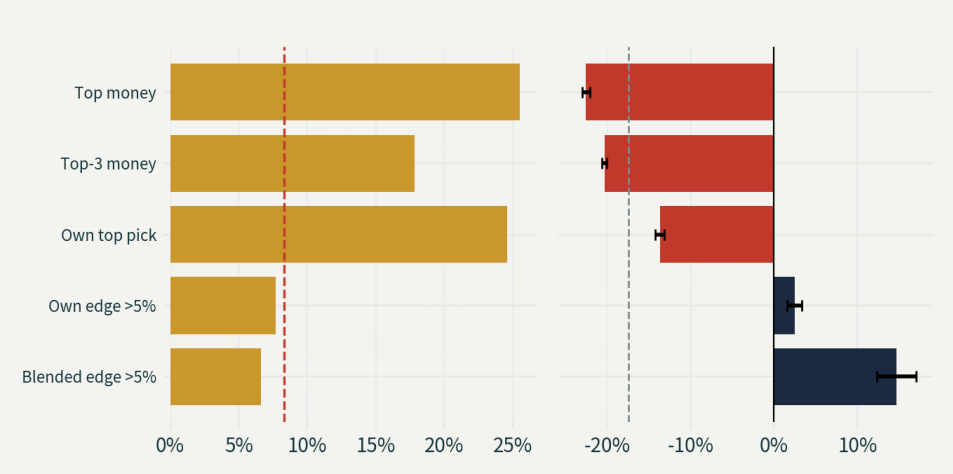


Figure 4. Hit rate per bet and mean return by strategy (error bars = ±1 standard error). A high hit rate and a positive return can be unrelated.

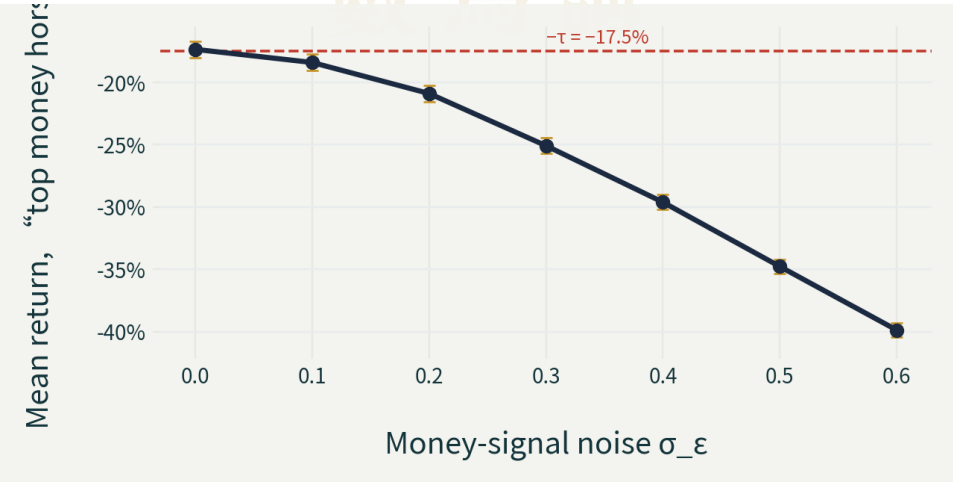


Figure 5. The noisier the money signal (larger σ_ε), the worse "back the top money horse" performs; at σ_ε = 0 it returns to −τ.

© Sigma Quant. All rights reserved. No reproduction or modification without permission. | sigmaquanthk.com

On the literature side, the favourite-longshot bias (longshots have worse expected returns) is widely documented in betting markets, and Ali (1977) explained it through disagreement in beliefs. Earlier Sigma Quant research also found that the price level remains the main signal, that odds paths add only a modest lift, and that no betting edge has been proven.

4 Using Money Flow Correctly

If you hold your own probability estimate p_m (error σ_δ) and the money share s is a second, independent error source (σ_ε), the best approach is precision-weighted blending, and betting only when the edge exceeds a safety margin m:

$$\ln p_b = w \ln p_m + (1-w)\ln s, \quad w = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2 + \sigma_\varepsilon^2}, \quad \mathrm{Var} = \frac{\sigma_\delta^2 \sigma_\varepsilon^2}{\sigma_\delta^2 + \sigma_\varepsilon^2} \tag{10}$$

$$\text{Bet if } \frac{(1-\tau)\, p_b}{s} > 1 + m, \qquad f^* = \frac{p_b\, o - 1}{o - 1} \tag{11}$$

Example 4: When Your Model and the Money Disagree

Scenario	Own p_m	Money s	Blended p_b	Trust self	Blended
Model bullish, money cold	20%	10%	12.7%	+65.0%	+4.4%
Roughly agree	12%	10%	10.6%	−1.0%	−12.2%
Model bearish, money hot	8%	14%	11.6%	−52.9%	−31.8%

First row: trusting yourself alone suggests a +65% edge and a Kelly stake of about 8.97%; after blending, the edge is +4.4% and the stake about 0.61%, roughly 15 times smaller. This is the rational middle path between "when money conflicts with my forecast, follow the money" and "ignore money completely": weight by precision instead of choosing one side. (The weight w = 0.34 comes from σ_δ = 0.35 and σ_ε = 0.25; table values ignore normalisation.)

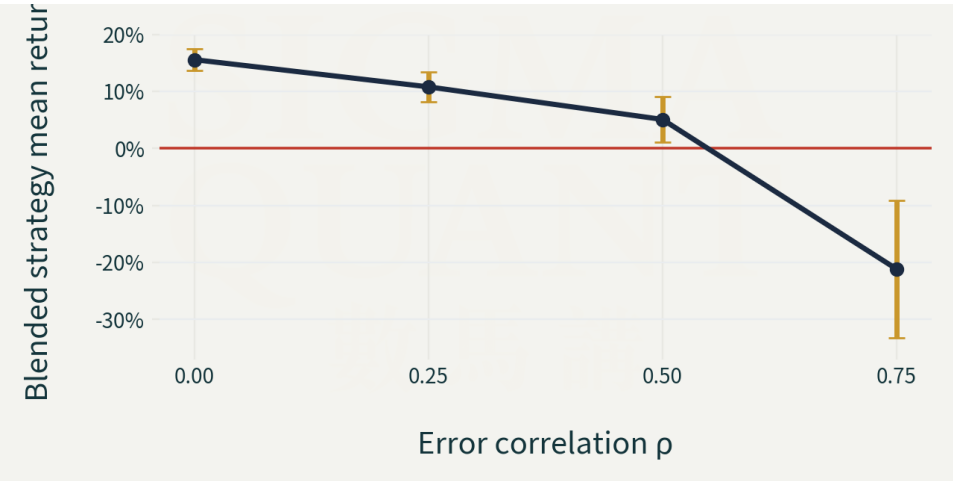


Figure 6. Simulated return of the blended strategy against the error correlation ρ between the two estimates. The more alike they are, the less extra information blending adds; at ρ = 0.75 only about 1,800 bets remain and the error is very large.

Situation	Suggested action	Reason
No own probability estimate	Do not bet on flow alone	Theorems 1 and 2
Money agrees with model	Still check the price against Eq. (11)	Agreement is mostly in the odds already
Money conflicts with model	Recompute with blended probability; cut stake or pass	Both estimates are noisy

© Sigma Quant. All rights reserved. No reproduction or modification without permission. | sigmaquanthk.com

Situation	Suggested action	Reason
Money carries information the model lacks	Use as a supplement; raise its weight	Precision weighting

5 Limitations and a Validation Plan

First, the simulation is synthetic; it proves a mechanism, not actual returns. Second, it assumes log-normality and independent noise; if money and model errors are highly correlated, the blending advantage disappears (Figure 6). Third, it ignores our own market impact, rounding, dead heats and takeout differences between pools.

We recommend validating on real data: (1) take the win-pool share s near T−2 and regress ln(win rate/s) on ln s, testing whether the slope is negative; (2) add the money movement from T−5 to T−2, use season-blocked cross-validation and compare log loss; (3) report confidence intervals for ROI rather than a single number.

References: Ali, M. (1977), Journal of Political Economy; Whelan, K., Economics of Betting Markets lecture notes; Hong Kong Jockey Club betting rules and betting guide; Sigma Quant Research 01–04 (2026).

This paper is for data analysis and education only. It is not betting advice and does not guarantee returns. Racing involves risk.

SIGMA QUANT 數馬講