# Referee Report — Round 8

## Tennis Match Simulator: ROI Calculations & Vig Handling

Referee 2

2026-02-18

# Verdict: Accept with Minor Revisions

**Focus:** Author-identified vig calculation error — independently audited and confirmed.

**Key findings:**

- **Main pipeline (`06_backtest.R`): CORRECT.** ROI formula uses actual betting odds throughout.
- **Trajectory scripts: CONFIRMED BUG.** Three scripts use vig-removed "fair" odds instead of actual Pinnacle odds to compute profit. This inflated ROI by $\sim$11pp ($+10.7\%$ apparent $\rightarrow -0.7\%$ actual).
- **Hardcoded OOS value not corrected.** `trajectory_ci.R` hardcodes the erroneous $+21.0\%$ 2025 OOS ROI. Corrected OOS result is unknown.
- **Edge definition inconsistency.** Two different "edge" definitions in use across the codebase — not a bug in ROI, but thresholds are not comparable across scripts.

**No new major methodological flaws.** Infrastructure is sound.

# Raw Data: Pinnacle Vig Is 2.8–3.0%, B365 Is 5.5–6.0%

**Verified from** `2024.xlsx`**:**

| Match | PSW | PSL | Vig |
|---|---|---|---|
| Popyrin | 1.72 | 2.23 | 3.0% |
| Shevchenko | 1.78 | 2.14 | 2.9% |
| Safiullin | 2.31 | 1.68 | 2.8% |

For same matches (B365):

| Match | B365W | B365L | Vig |
|---|---|---|---|
| Popyrin | 1.62 | 2.30 | 5.6% |
| Shevchenko | 1.62 | 2.30 | 5.6% |
| Safiullin | 2.30 | 1.62 | 5.6% |

**Implication:**
Falling back from Pinnacle to B365 **doubles** the vig hurdle:

- Pinnacle: 3% vig $\rightarrow$ each side pays $\approx 1.5\%$
- B365: 6% vig $\rightarrow$ each side pays $\approx 3\%$

**Unresolved:** What fraction of H1 2024 matches use Pinnacle vs. B365? If a material fraction uses B365, ROI claims are understated for a Pinnacle-only comparison.

# Main Backtest ROI Formula Is Correctly Implemented

`06_backtest.R`: `simulate_betting()` lines 578–586:

```
flat_profit = if_else(bet_won, (bet_odds-1)*stake, -stake)
roi_flat = sum(flat_profit) / (n_bets * stake)
```

**Why raw implied prob is correct for EV:**
A bet has positive EV if and only if:

$$\text{model\_prob} > \frac{1}{\text{odds}}$$

The vig is already embedded in the offered odds. "Raw implied prob" $= 1/\text{odds}$ is exactly the breakeven probability. No further vig subtraction is needed.

**Verified formula:**

| Component | Status |
|---|---|
| Profit on win | $(o-1) \times \text{stake}$ ✓ |
| Loss on loss | $-\text{stake}$ ✓ |
| ROI denominator | $n \times \text{stake}$ ✓ |
| Kelly formula | $(bp-q)/b$ ✓ |
| Bootstrap CI | Percentile method ✓ |

# The Trajectory Bug: Fair Odds Inflate Each Winning Bet by ∼3%

**What the code does (wrong):**

```
mkt_prob_w <- (1/PSW)/(1/PSW+1/PSL)
elo_opp_odds <- 1/(1 - mkt_prob_w)
profit <- elo_opp_odds - 1
```

**What it should do:**

```
elo_opp_odds <- PSL   # actual
profit <- elo_opp_odds - 1
```

**Numerical example (PSL = 3.50):**

| | Bugged | Correct |
|---|---|---|
| Raw prob | 1/3.50=0.286 | — |
| Vig-removed | 0.286/1.03 | — |
| Odds used | 3.606 | **3.500** |
| Profit if won | 2.606 | 2.500 |

Each winning bet reports **0.106 units too much**.

# 11pp ROI Swing Comes From Two Compounding Effects

**Effect 1: Odds inflation (≈4–5pp)**
Fair odds ≈ actual odds × (1 + vig). For the 3.0–5.0 range:

- Absolute inflation ≈ 0.09–0.15 per unit
- At 30% win rate: ∼3–5% ROI inflation

**Effect 2: Sample composition shift (≈6–7pp)**
The odds filter `elo_opp_odds >= 3.0` was applied to **fair** odds, not actual.

- PSL 2.85–2.99 matches: fair odds $\geq 3.0 \rightarrow$ *wrongly included*
- These matches have 33–35% win prob vs. 27% average
- Sample changed from N=300 (actual) to N=272 (bugged)

**Summary of impact:**

| Metric | Bugged | Corrected |
|---|---|---|
| N | 272 | **300** |
| Win rate | 30.2% | **27.3%** |
| Avg odds | 3.76 | **3.74** |
| **ROI** | +10.7% | −0.7% |
| Years pos | 4/4 | **1/4** |

**No edge exists.** The win rate barely exceeds breakeven (27.3% vs. 26.7%), and the CI is $[-19.3\%, +17.9\%]$.

# Three Scripts Infected; Main Pipeline and Others Are Clean

**Scripts with vig bug (profit inflated):**

| Script | Status |
|---|---|
| `trajectory_contrarian.R:25--27` | × |
| `trajectory_best_subset.R:19--21` | × |
| `trajectory_ci.R:11, 47` | × |

**Scripts with correct ROI:**

| Script | Status |
|---|---|
| `06_backtest.R` | ✓ |
| `validate_elo_betting.R` | ✓ |
| `find_betting_edge_subsets.R` | ✓ |
| `calibration_edge_deep.R` | ✓ |
| `test_2025_oos.R` | ✓ |
| `prospective_paper_trade.R` | ✓ |

**Unresolved:** `trajectory_ci.R` line 47 hardcodes the erroneous 2025 OOS ROI of +21.0%. The corrected 2025 trajectory OOS result has **not been computed**.

# Two Different "Edge" Definitions Create Non-Comparable Thresholds

**Definition A: Raw edge** (main backtest)

```
edge = model_prob - 1/odds
```

*Interpretation:* how much model exceeds the exact breakeven probability. Edge $> 0$ guarantees positive EV.

**Definition B: Fair edge** (analysis scripts)

```
mkt_prob = (1/w)/(1/w + 1/l)
edge = model_prob - mkt_prob
```

*Interpretation:* how much model exceeds the vig-removed market belief. Edge $> 0$ does **not** guarantee positive EV.

**Numerical comparison (3% vig, 50/50 match):**

| Model prob | Raw edge | Fair edge |
|---|---|---|
| 0.45 | $-6.5\%$ | $-5.0\%$ |
| 0.515 | $0.0\%$ | $+1.5\%$ |
| 0.53 | $+1.5\%$ | $+3.0\%$ |
| 0.55 | $+3.5\%$ | $+5.0\%$ |

"Fair edge $> 0$" can mean **negative EV** (when fair edge $<$ fair_prob $\times$ vig).

# Calibration Edge Analysis May Admit Negative-EV Bets

`calibration_edge_deep.R` computes `dog_edge` against **vig-removed** market probabilities, then bets at **actual odds**.

**Scenario where this matters:**

Actual underdog odds = 2.20, Pinnacle vig = 3%

- Raw implied prob = $1/2.20 =$ **0.455**
- Fair implied prob = $0.455/1.03 =$ **0.441**
- Model says: 0.448

`dog_edge` = $0.448 - 0.441 = +0.7\%$ (passes filter)
Raw edge = $0.448 - 0.455 = -0.7\%$ (negative EV)

**Consequence:**
The "Odds 2.0–2.5, Edge 0–10%" subset includes bets that are genuinely negative EV. The 2025 OOS failure ($-2.29\%$) may partly reflect this: some included bets never had positive EV.

**Recommendation:** Recompute using `dog_edge = model - (1/odds)` and compare. If the in-sample ROI drops, the vig-adjusted filter was doing real work.

# Recommendations (Prioritized)

1. **Fix `trajectory_ci.R`.** Line 47 hardcodes +21.0% OOS ROI from bugged calculation. Rerun 2025 trajectory analysis with actual PSL/PSW odds and update the CI file. Report the corrected OOS result.
2. **Report Pinnacle vs. B365 match breakdown.** The vig difference (3% vs. 6%) is material. Tabulate N and ROI by bookmaker source.
3. **Document edge definition in each script.** Add a comment at the top of analysis scripts noting whether edge is vs. raw or vig-adjusted implied probability.
4. **Recompute calibration edge with raw-edge filter.** Test whether the in-sample +5.16% ROI survives when only genuinely positive-EV bets are included.
5. **Clarify paper trade Elo update.** `prospective_paper_trade.R` updates Elo during the OOS period. Document whether this is intentional and how it interacts with the "frozen model" claim.