

# Referee Report — Round 2

## Tennis Match Simulator

Referee 2

2026-02-04

## All Major Concerns Resolved

- ✓ Reproducibility: Random seeds now set throughout codebase
- ✓ Tour averages: Calculated dynamically from data
- ✓ Documentation: Master script and README created
- ✓ Uncertainty: Bootstrap 95% CIs on ROI estimates

**Minor issues remaining:** renv.lock not generated, Python replication uses outdated constants

## Reproducibility Verified: Identical Results with Same Seed

Test	Seed	P1 Win Prob	Identical?
Run 1	42	0.5683	—
Run 2	42	0.5683	✓
Run 3	123	0.5801	✗ (expected)

### Implementation:

- ▶ `RANDOM_SEED <- 20260204` in `06_backtest.R`
- ▶ `set.seed(seed)` called at start of `backtest_period()`
- ▶ Seed recorded in output for audit trail

# Tour Averages Now Calculated from Data

## Before (Hardcoded):

- ▶ Return vs 1st: 35.0%
- ▶ Return vs 2nd: 50.0%

Source: "Magic numbers" in code

## After (Dynamic):

- ▶ Return vs 1st: 27.5%
- ▶ Return vs 2nd: 49.3%

Source: Calculated from 26,587 ATP matches

**Difference of 7.5pp in return vs first serve — this matters!**

## Cross-Language Comparison: R vs Python

	R (seed=42)	Python (seed=42)	Status
P1 Win Probability	0.5683	0.5815	✓
Difference		0.0132	
Expected MC Error		~0.010	

**Note:** Exact match impossible due to different RNG algorithms (Mersenne Twister implementations differ between R and Python).

**Verdict:** Difference within  $2\sigma$  of Monte Carlo sampling error. Implementations are statistically equivalent.

## Bootstrap Confidence Intervals Verified

Metric	Seed 123	Seed 456
ROI 95% CI Lower	-8.68%	-9.24%
ROI 95% CI Upper	+37.14%	+35.07%
Standard Error	11.35%	—

**Implementation:** Percentile bootstrap with 1,000 resamples

**Reproducibility:** Same seed → identical CIs ✓

**Sample output:**

*ROI: +5.2% (95% CI: -2.1% to +12.3%)*

## Replication Readiness: 7/10 (up from 4/10)



7/10

- ✓ Folder structure
- ✓ Relative paths
- ✓ Variable naming
- ✓ Script naming
- ✓ Master script (**NEW**)
- ✓ README (**NEW**)
- ✓ Random seeds (**NEW**)
- `renv.lock` (script exists, not run)
  - Automated figures (low priority)
  - In-text stats automation (low priority)

## New Infrastructure Created

File	Purpose	Lines
code/run_analysis.R	Master script	148
code/README.md	Replication instructions	149
code/setup_renv.R	Dependency management	48

### Master script features:

- ▶ Configurable date range, tour, simulation count
- ▶ Sets random seed for reproducibility
- ▶ Saves results to data/processed/
- ▶ Prints summary statistics

# Minor Issues Remaining

## 1. **renv.lock not generated**

- ▶ Fix: Run `source("code/setup_renv.R")`
- ▶ Risk: Package version drift across machines

## 2. **In-text statistics in CLAUDE.md**

- ▶ Manually entered, not pulled from saved results
- ▶ Low priority

**Resolved during Round 2:** Python replication now accepts tour averages as parameters, matching R implementation.

Verdict: Accept with Minor Revisions

**All Major Concerns Resolved**

**To complete acceptance:**

1. Run `source("code/setup_renv.R")` to generate `renv.lock`

**No re-review required** for this minor fix.