

Referee Report: Tennis Match Simulator

Round 1 Audit — Cross-Language Replication & Code Review

Referee 2

2026-02-04

Verdict: Major Revisions Required

Overall Assessment: Sound core logic, but critical infrastructure gaps.

Key Findings:

- ✗ **No random seeds** — Results not reproducible
- ✗ **Hardcoded tour averages** — Should be computed from data
- ✗ **No master script** — Cannot replicate with single command
- ✗ **No uncertainty on ROI** — Point estimates only
- ✓ **Python replication confirms logic is correct**
- ✓ **Missing value handling is reasonable**

Replication Readiness Score: 4/10

Python Replication: Monte Carlo Engine Verified

Created: `code/replication/referee2_replicate_mc_engine.py`

Test Case	Python Result	Status
Typical ATP matchup (P1 win prob)	0.5815	✓
Same matchup, no adjustment	0.6737	✓
Opponent adjustment effect	+9.2pp	✓
Big server vs elite returner	0.4155	✓
Point-level: Aces	8.17% (exp: 8.00%)	✓
Point-level: Double faults	2.99% (exp: 3.00%)	✓

Note: Exact numerical comparison impossible — R code has no `set.seed()`.

Replication Limitation: No Random Seeds in R Code

Problem: Cannot verify Python-R equivalence to 6 decimal places.

What I Did:

- ▶ Line-by-line translation of R \rightarrow Python
- ▶ Verified all formulas match
- ▶ Point-level statistics correct
- ▶ Confidence intervals match Wilson formula

What's Missing:

- ▶ R output with fixed seed
- ▶ Direct comparison table
- ▶ Guarantee of identical RNG sequences

Resolution Required: Author must add `set.seed()` and re-run.

Code Audit: Hardcoded Tour Averages

File: 01_mc_engine.R, lines 45 and 68

```
# Line 45:  
avg_return_vs_first <- 0.35
```

```
# Line 68:  
avg_return_vs_second <- 0.50
```

Issue: These values should be computed from data, not hardcoded.

Impact:

- ▶ If data period changes, values become incorrect
- ▶ Cannot verify these match actual tour averages
- ▶ Acknowledged in CLAUDE.md as “Technical Debt #1”

Fix: Use `stats_db$tour_avg_surface$return_vs_first_pct`

Replication Readiness: 4/10

4/10



- ✓ Folder structure (data/raw, data/processed)
- ✓ Relative paths used
- ✓ Informative variable names
- ✓ Numbered script execution order

- ✗ **No master script**
- ✗ **No renv.lock**
- ✗ **No README in /code**
- ✗ **Random seeds not set**

Bottom line: A replicator cannot run the analysis without reading all code.

Econometrics: ROI Estimates Lack Uncertainty

Reported Result (from CLAUDE.md):

ROI at 5% edge: -5.1%

Problem: No confidence interval or standard error provided.

Why This Matters:

- ▶ Betting returns have high variance
- ▶ -5.1% could be consistent with true ROI of +10% or -20%
- ▶ Cannot determine if result is statistically distinguishable from zero
- ▶ Cannot assess whether model has genuine edge

Fix: Add bootstrap CIs by resampling the betting outcomes.

Output Automation: Mostly Manual

Output Type	Status	Concern Level
Backtest results to RDS	✓ Automated	—
Tables for publication	✗ Manual	Minor
Figures (calibration, ROI)	✗ Not saved	Minor
In-text statistics in CLAUDE.md	✗ Manual	Major
Reproducible outputs	✗ No seeds	Major

Key Issue: Statistics in CLAUDE.md (58.9% accuracy, 0.2336 Brier, -5.1% ROI) appear manually entered. These should be pulled programmatically.

Priority Recommendations for Resubmission

1. **Add `set.seed()` to all scripts**
Use documented seed, re-run backtests
2. **Create master script**
`code/run_analysis.R` that runs full pipeline
3. **Calculate tour averages dynamically**
Replace hardcoded 0.35/0.50 with computed values
4. **Add bootstrap CIs to ROI**
Quantify uncertainty on betting performance
5. **Set up renv**
Document package versions for reproducibility
6. **Verify Python-R equivalence**
With fixed seed, confirm 6-decimal match

Referee 2 Deliverables

Replication Scripts:

- ▶ `code/replication/referee2_replicate_mc_engine.py`

Report Files:

- ▶ `correspondence/referee2/2026-02-04_round1_report.md`
- ▶ `correspondence/referee2/2026-02-04_round1_deck.tex`
- ▶ `correspondence/referee2/2026-02-04_round1_deck.pdf`

Next Steps:

1. Author addresses Major Concerns
2. Author files response at `2026-02-XX_round1_response.md`
3. Round 2 audit verifies fixes