Supplemental appendix to *The Decline of the Follow-On in 125 Years of First-Class Cricket*[*]

Johan Fourie[†]

This appendix reports the rule history that identifies the design, the data construction, and the robustness checks referred to in the paper. All numbers are produced by the replication scripts in `scripts/`, which run end to end from `scripts/run_all.R`.

1 The follow-on rule

The margin required to enforce the follow-on is set by the Laws of Cricket and depends on the scheduled length of the match. Table A1 states the history. Two features identify the design. First, enforcement became the captain's choice in 1900; before then a side that led by the stated margin was obliged to enforce, so there is no decision to study. Second, in 1980 the margin for matches of five days or more rose from 150 to 200 runs while the margin for three- and four-day matches stayed at 150. Test cricket therefore changed cutoff and domestic first-class cricket did not.

Table A1: The statutory follow-on margin.

Period	Match length	Margin (runs)	Enforcement
1835–1853	any	100	Compulsory
1854–1893	any	80	Compulsory
1894–1899	any	120	Compulsory
1900–1979	3 or more days	150	Captain's choice
1900–1979	2 days	100	Captain's choice
1980–present	5 or more days	200	Captain's choice
1980–present	3 or 4 days	150	Captain's choice
1980–present	2 days	100	Captain's choice

Sources: MCC Laws of Cricket, 1884 code Law 53 and its 1894 revision; the 1900 revision, which introduced the captain's discretion; the 1980 code Law 13, which introduced the graduated

[*]This is the supplementary material for Fourie, Johan. 2026. "The Decline of the Follow-On in 125 Years of First-Class Cricket." Working Paper, Department of Economics, Stellenbosch University. This paper was created with the help of Anthropic's Claude Code and OpenAI's Codex, which assisted with statistical code development, reproducibility and methods audits, and language editing; I reviewed and verified all analyses, results and text and take full responsibility for the content. Cite the paper, not this supplement.

[†]Department of Economics, Stellenbosch University. Email: johanf@sun.ac.za.

scale by match length; and the current code Law 14, which retains the 1980 scale (`https://www.lords.org/mcc/the-laws/the-follow-on`).

Because the Law is written in terms of scheduled days rather than format, a small number of matches break the usual correspondence. Six Test matches scheduled for three or four days have been played since 1980, and they take the 150 margin. A number of domestic matches, mostly finals, are scheduled for five days and take the 200 margin after 1980. Assigning the margin by format rather than by scheduled days would place these matches on the wrong side of the cutoff.

The County Championship suspended the follow-on for the 1961 and 1962 seasons, which are excluded. The data confirm the suspension: enforcement counts fall from 49 in 1960 to 5 and 4, then return to 24 in 1963.

2 Data construction

2.1 Who may enforce

Under Law 14 only the side that batted first may enforce the follow-on. If the side batting second scores more in the first innings, no follow-on is possible whatever the size of the gap. Identifying the leading side by its score rather than by its batting order therefore counts matches as eligible in which no decision existed.

The DAFT database records batting order in a column that takes the value 1 for the side batting first and 2 for the side batting second. We validate it against the legal requirement that an enforcing side must have batted first and led: 99.94 per cent of recorded enforcements satisfy this. The correction is large. An order-blind construction counts 8,221 eligible matches where the correct number is 4,939, an overstatement of 66 per cent, and reports an enforcement rate of 36.7 per cent where the correct rate is 61.1 per cent. These counts precede the sample restrictions below; 4,227 eligible matches remain after applying them.

2.2 Sample restrictions

We exclude fixtures that are first-class but not competitive: matches against touring international sides and against Oxford and Cambridge universities. In university matches the enforcement rate is 7.6 per cent, against about 65 per cent in Championship matches, because the county has little interest in forcing a result. Including them would distort both the level and the trend.

A small number of matches record an enforcement that the Laws did not permit, either by a side that did not lead or by a side whose lead fell short of the margin. Eligibility is a deterministic legal condition, so these records contain an error either in the follow-on flag or in the recorded scores. We cannot tell which, so we remove them. They are 1.2 per cent of enforcements and their removal leaves every estimate unchanged to the third decimal place.

2.3 Coverage

The database contains no first-class records for India, Sri Lanka or Bangladesh; for those countries only limited-overs files exist. The Ranji Trophy is therefore not available. County Championship coverage begins in 1945, so before that year the domestic sample is limited to Australian and New Zealand first-class cricket.

3 Validity of the discontinuity design

3.1 Manipulation of the running variable

The side batting second knows the margin and bats to avoid it, an objective cricket describes as saving the follow-on. This produces excess mass just below the cutoff. Within three runs there are 195 matches below and 122 above (binomial $p < 0.001$), and a local density test rejects ($p = 0.040$).

The manipulation is one-sided and is carried out by the side that does not make the decision under study. This does not remove the problem. The strength of the opponent affects the match outcome and its ability to save the follow-on.

We therefore apply the bounds of Gerard, Rokkanen and Rothe (2020). We reverse the running variable and treatment coding so assignment means avoiding eligibility and the treatment means not enforcing. We then reverse the sign of the estimated bounds. Table A2 reports point-identified sets for the effect of enforcing. The density bandwidth implies that 31 to 38 per cent of units just below the cutoff may have avoided eligibility. Every identified set contains zero.

Table A2: Manipulation-robust fuzzy RD bounds. The outcome bandwidth is 87 runs.

Density bandwidth	Manipulated share	Lower bound	Upper bound
10 runs	0.380	−0.225	0.607
15 runs	0.339	−0.138	0.606
20 runs	0.306	−0.077	0.604

These are identified sets under the weak assumptions in Gerard, Rokkanen and Rothe (2020), rather than confidence intervals. A refinement assuming that units able to avoid eligibility are at least as likely not to enforce changes the upper bounds by less than one percentage point.

Table A3 reports estimates that exclude matches close to the cutoff. The estimate falls from 0.240 to 0.194 as the excluded region widens to five runs, and remains significantly positive throughout. A donut does not restore identification by itself; it changes the extrapolation. We report it as a bound on how much the manipulated region contributes.

Table A3: Donut estimates: fuzzy effect of enforcing on the probability that the side batting first wins, excluding matches within a given number of runs of the cutoff.

Excluded region	Matches	Estimate	Std. error
none	22,113	0.240	0.053
within 1 run	21,933	0.224	0.057
within 2 runs	21,844	0.214	0.058
within 3 runs	21,764	0.203	0.058
within 5 runs	21,582	0.194	0.059

3.2 Placebo cutoffs

If the statutory rule identifies the effect, a discontinuity in enforcement should appear only where the Law places one. Table A4 tests this. Each placebo is estimated on a window that excludes the true cutoff, so that the local polynomial is not fitted across a real discontinuity.

Table A4: Discontinuity in the probability of enforcing at true and placebo cutoffs.

Sample	Cutoff		Estimate	p
Tests before 1980	150	true	0.305	0.012
Tests before 1980	200	placebo	0.043	0.420
Tests from 1980	150	placebo	0.000	–
Tests from 1980	200	true	0.443	<0.001
Domestic before 1980	150	true	0.683	<0.001
Domestic before 1980	200	placebo	0.048	0.661
Domestic from 1980	150	true	0.473	<0.001
Domestic from 1980	200	placebo	−0.136	0.201

Placebo cutoffs in the outcome, placed 50 to 100 runs either side of the true line, produce estimates from −0.108 to +0.061, none statistically distinguishable from zero.

3.3 Honest confidence intervals

The running variable is a whole number of runs and so has mass points, in which case conventional intervals can undercover (Kolesár and Rothe, 2018). We report bias-aware intervals over a grid of curvature bounds rather than a single data-driven value, because a curvature estimated from the data is not an upper bound on curvature. For the fuzzy parameter the bound is a pair, one for the outcome and one for the first stage.

Every interval excludes zero under its stated curvature bound. These intervals use curvature-specific bandwidths and weights, so their point estimates need not equal the conventional estimate in the paper. They address discreteness and smoothness. They do not address sorting.

Table A5: Bias-aware honest intervals for the effect of enforcing on the probability that the side batting first wins.

Curvature bound M	Estimate	Honest interval
(0.001, 0.002)	0.381	[0.145, 0.617]
(0.002, 0.004)	0.391	[0.131, 0.650]
(0.004, 0.008)	0.388	[0.106, 0.670]
data-driven	0.292	[0.157, 0.426]

3.4 Bandwidth, polynomial order and balance

The estimate falls monotonically as the bandwidth widens, from 0.299 at half the optimal bandwidth to 0.187 at twice it, and is significant at every value. A quadratic local polynomial gives 0.276 against 0.240 for the linear specification. Scheduled match length does not jump at the cutoff (0.001, p = 0.910), nor does calendar year (−0.749, p = 0.818). A home-status measure was removed because venue country does not identify the home side in domestic competitions.

4 The behavioural design

4.1 Why the comparison group is other enforcements

Comparing teams that suffered a disaster with teams that did not is biased. A team can suffer this disaster only if it enforced, and it enforces when its propensity to enforce is high, so its rate would fall afterwards through regression to the mean alone. Teams that never suffered a disaster are disproportionately teams that rarely enforce.

Conditioning both groups on the same action reduces this comparability problem. What remains is selection on the realised outcome, which is not random: conditional on enforcing with a given lead, losing still depends on team and opponent strength, form, pitch and weather. The design is therefore a matched observational event study, not an experiment, and we describe it that way throughout.

Two features of the construction matter. Controls are not required to be free of later disasters, because that would select controls on outcomes occurring after the event while imposing no equivalent restriction on treated teams. Instead both groups are censored at any later disaster. Event time is measured from the event date rather than the season label, so that a match played later in the same season counts as after the event.

4.2 Inference

With 34 treated events, cluster-robust standard errors are unreliable (MacKinnon, Nielsen and Webb, 2023). We report three procedures. Clustering on team gives a standard error of 0.046. A wild cluster bootstrap with Webb weights and the null imposed gives p = 0.483. The randomisation test reassigns, within each matched set, which of the comparable enforcements ended in defeat; it

permutes a studentised statistic and uses the finite-simulation p-value $(1 + r)/(B + 1)$. It gives p = 0.053 for the first post-event year and p = 0.498 for the five-year average. Inverting the latter test gives the interval $[-0.140, +0.070]$ reported in the paper.

The minimum detectable effect of 0.135 describes what the design could have found. It is not a bound on the true effect; the inverted interval is the bound.

4.3 Exploratory comparisons

Table A6 reports two comparisons built with the same matching, censoring and studentised randomisation procedure. Control events are used at most once. Losing after declining gives an estimate near zero. A near miss is an enforcement in which the opponent erased the deficit but did not win. Its estimate is also imprecise and positive. These checks bear on the five-year average, not on the first-year coefficient in the main event study.

Table A6: Exploratory matched event studies.

Event	Events	Estimate	Std. error	Rand. p
Declined and lost	116	0.003	0.028	0.924
Near miss after enforcing	543	0.020	0.032	0.587

Only 39.4 per cent of treated post-event observations retain the captain from the event match. Separate estimates are −0.043 (0.067) under the same captain and −0.030 (0.057) after a change. The team-level design cannot isolate a captain's personal response. A stacked comparison around disasters suffered by rivals gives 0.022 (0.014), but it reuses matches and has few competition clusters. We treat it as descriptive.

4.4 Test cricket after Kolkata, 2001

In March 2001 India beat Australia at Kolkata having followed on, the most discussed instance of the disaster in the modern game. Every Test captain saw it at the same time, so there is no control group inside Test cricket and no credible comparison outside it: domestic captains watched the same match.

We therefore report a descriptive interrupted series. Test enforcement among eligible captains was 0.805 in 1981–2000 and 0.519 from 2001, a fall of 0.127 in a weighted regression with a linear trend. To judge whether a break of that size is unusual, we estimate the same break at every other feasible year and locate 2001 in that distribution. Twenty-three per cent of candidate years produce a larger fall, so the break at 2001 is not exceptional against the background variation in a series this short. We draw no causal conclusion from it.

Table A7: Event-study coefficients, effect on the probability of enforcing by year relative to the event. Reference year −1. Standard errors clustered on team.

Year relative to event	Estimate	Std. error
−5	+0.028	0.110
−4	−0.121	0.086
−3	−0.184	0.114
−2	−0.036	0.112
−1	0	(reference)
0	−0.252	0.107
1	−0.066	0.113
2	−0.044	0.140
3	−0.029	0.098
4	−0.182	0.137
5	+0.110	0.122
Joint test of pre-event coefficients: p = 0.492		

4.5 Full event-study estimates

References

Gerard, François, Miikka Rokkanen, and Christoph Rothe. 2020. "Bounds on treatment effects in regression discontinuity designs with a manipulated running variable." *Quantitative Economics*, 11(3): 839–870.

Kolesár, Michal, and Christoph Rothe. 2018. "Inference in regression discontinuity designs with a discrete running variable." *American Economic Review*, 108(8): 2277–2304.

MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb. 2023. "Cluster-robust inference: a guide to empirical practice." *Journal of Econometrics*, 232(2): 272–299.