

Data-Driven Evaluation of Strategy & Performance in Formula 1

Quantifying upgrade effectiveness, race-strategy modelling, and the impact of the cost cap and logistical constraints

Bálint Csermák-Scharbert

Internal advisors: Yangzhuoran Yang & Renaud Jolivet | Internship supervisor: Inga Stracke | Maastricht University | June 2026



An engineering sport, decided behind the scenes



~50

ENGINEERS

Work on both cars, every single race weekend



1 TB+

OF DATA

Generated by one car over a single race weekend



\$135-215M

COST CAP

Annual limit in force since 2021

The takeaway: races are won on track, but **decided in the factory and on the pit wall**, under a budget that now governs every decision.

Where the Literature and Results Diverge



Lap-time prediction - modelled on its own



Tyre degradation & strategy - studied separately



Upgrade effectiveness - rarely quantified at all



Cost cap & logistics - described, not linked to pace



The gap

No existing work integrates **performance, strategy, upgrades, and the cost cap** into **one** framework for recent regulation eras.

To what extent can **data-driven methods** quantify the impact of F1 **upgrades** and **race strategies** on on-track performance, under **cost-cap** and **logistics constraints**?

SQ1

Part A



How reliably can tyre degradation & stint length predict pit windows?

SQ2

Part A



How well can a combined predictive + simulation policy recommend strategy?

SQ3

Part B



Do in-season upgrades produce measurable on-track gains (via difference-in-differences)?

SQ4

Part C



What do cost-cap & logistics reveal about how teams prioritise development?

One pipeline, three connected parts

Not just three studies stitched together, they share one canonical lap table and one set of feature definitions.

PART A



Race strategy

Lap-time, tyre degradation,
pit-window & value models
→ a ranking policy

SQ1 & SQ2

PART B



Upgrade effect

Difference-in-differences
on field-relative pace
around each upgrade

SQ3

PART C



Cost & logistics

Qualitative analysis of FIA
regs, team filings & expert
interviews

SQ4

Unified lens: all three parts are read under the same cost-cap & logistics constraint,
oriented toward **live decision support**.

From public timing data to a shared feature layer



Open-source APIs

FastF1 (primary)



Canonical lap table

Frozen 19-col schema



Feature engineering

Online-safe only



Shared feature layer

Feeds A - B - C



Online-safe features

Every feature uses only information available up to the current lap – no future-lap leakage, so the models can run live mid-race.



Relative, not absolute

Pace is measured against the field median, because teams' internal absolute benchmarks are classified. What matters is standing vs. rivals.

Six models, one strategy policy



Lap-time model

Gradient boosting:
predicts lap time from
tyre, position, pace &
weather



Tyre degradation

86 polynomial fits, one
per *circuit*×*compound*.
Assesses pace loss vs. a
fresh tyre



Pit classifier

Logistic regression:
probability of pitting
within 3 laps



Value regressor

Ridge: net time gained
by pitting now vs.
staying out (6-lap
window)



Strategy policy

Three-gate ranking: a
watchlist of the best
upcoming pit laps



Monte-Carlo sim

2,000 runs/scenario to
stress-test decisions
under uncertainty

Headline metric is the **top 3 hit rate**: does the lap the driver actually pitted on appear in the model's top-3 ranked laps?

Simpler models generalised better across seasons



Gradient boosting

Flexible, memorises patterns

0.297

top-3 hit rate (2025)



Logistic regression (L2)

Regularised, transfers across seasons

0.503

top-3 hit rate (2025)

Why? The boosted model learned **season-specific event-compound patterns that don't transfer**. A regularised linear model is forced onto continuous, generalisable signals.

Lap performance is highly predictable

0.78 s

validation MAE

On the unseen 2025 season

< 1% of a 70-90 s lap



Strong real-world accuracy

0.78 s sits comfortably inside 1% of typical 70–110 s lap times.



Generalises, doesn't memorise

Train MAE 0.47 s vs. validation 0.78 s – same order of magnitude, an honest cross-season gap.



Errors concentrate, not spread

RMSE 1.73 s > MAE: most laps tight, a minority (safety cars, traffic, weather) missed by more.

A useful aid bounded by safety-car randomness

50%

top-3 hit rate

The real pit lap is flagged in the model's top 3 about half the time.

~2nd

MRR $\approx$ 0.47

On average the true pit lap ranks around second on the watchlist.



What this means operationally

A strategist would have the actual pit lap on a 3-lap shortlist about half the time, being useful as a **decision aid**.

But there is a real ceiling. **Safety-car timing is close to random.**

Driven by crashes and failures no lap-feature can foresee. This caps accuracy below 1, and no extra feature fixes it.

Busting the intuitive hypothesis

The intuitive hypothesis

"Safety-car races wreck the policy. Chaos, unscheduled stops, unpredictable timing."

What the data showed

Near-identical medians: **0.200 clean vs 0.211 SC/VSC**. Safety cars aren't inherently harder just more spread out.



The real driver of difficulty

It isn't whether a safety car appears, it's **pit-window concentration**.

When a safety car creates **one clear window**, the model nails it. When the whole field pits reactively within 2–3 laps, the decision is driven by **opponent behaviour**, which lap features cannot see.

Isolating an upgrade from a developing field



Difference-in-differences

Problem: the whole field develops together, so a naïve before/after blames everything on the upgrade.

DiD subtracts **the field's change** from **the team's change** over the same window, netting out general development.

3 races before, 4 races after, 95% CI from 500 bootstraps.



The honest verdict

4 of 12 packages reached significance, while most show **no measurable relative effect**.

That's expected: in a developing field, **keeping pace is the baseline**.

And as Steve Clark asked: **"How do you even define an upgrade?"** Many parts improve balance, not raw pace.

When an upgrade is real, the model sees it

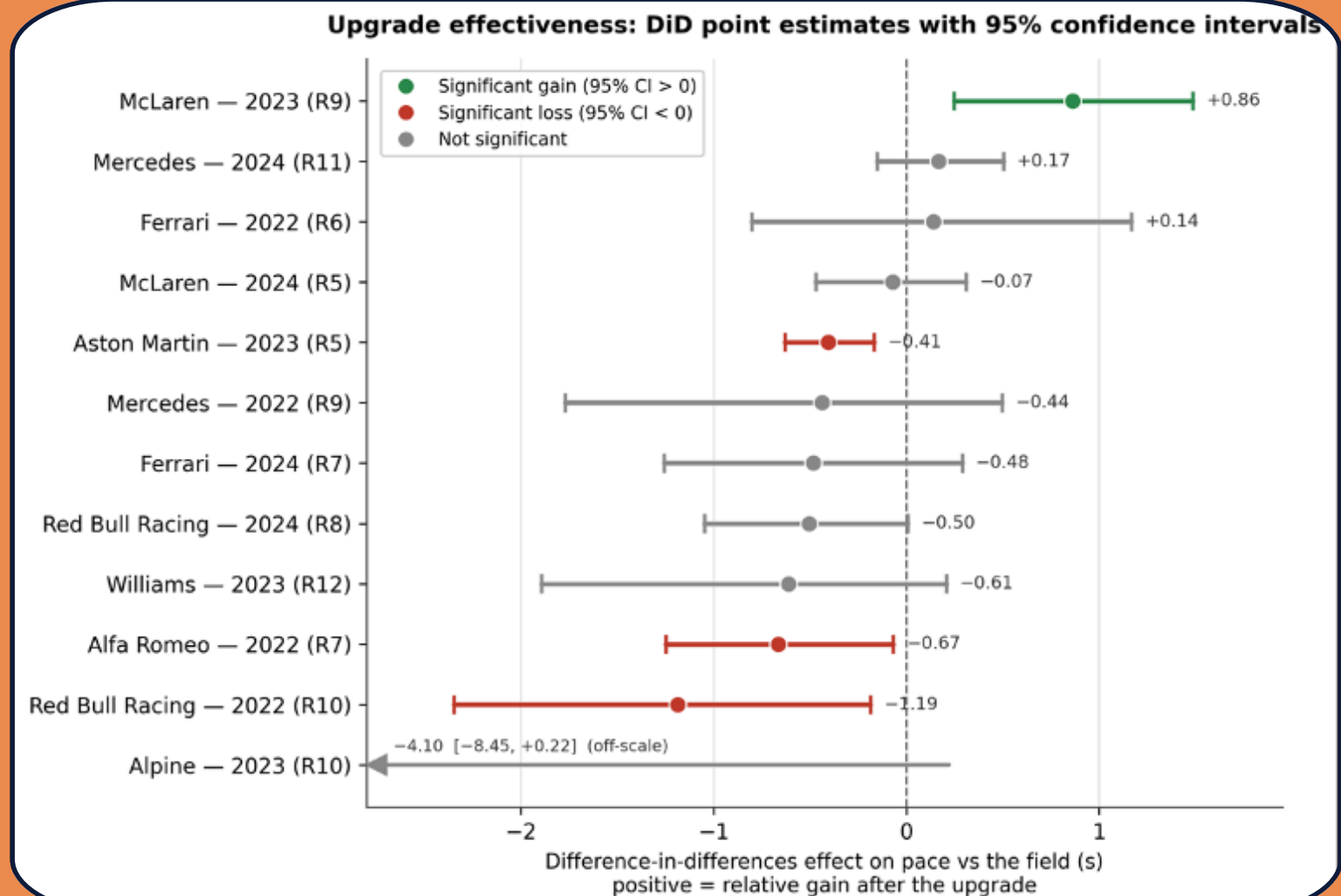


+0.86 s

relative pace gain vs. the field

95% CI [+0.25, +1.49] · significant

McLaren B-spec - 2023 - Austria



This package lifted McLaren from the lower midfield to podium contention, the first step toward their 2024 and 2025 titles.

Why so qualitative

The quantities that matter: R&D allocation, CFD/wind-tunnel limits, freight, personnel are simply not available from public sources.



FIA regulations

Defines the cap & what's in or out. Notes the boundary of what's permitted.



Team filings

Year-end accounts give scale, but never line-by-line allocation.



Expert interviews

DHL & engineers supply the operational reality public data cannot fully confirm.



The directional finding

Development is the binding constraint on competitiveness. R&D, wind-tunnel time and logistics consume the budget. **In-race strategy is near-costless.** Logistics lead-times shape when development reaches the track.

F1 under the cost cap is an allocation problem



Performance is predictable

Lap pace models well; strategy partially, is bounded by safety-car randomness.



Upgrades are situational

Hard to quantify per-package, but skipping development means falling behind fast.



Strategy is near-costless

The budget is consumed by development & logistics, not by race-day calls.



The binding constraint is conversion: how effectively a team turns **capped development budget and time into pace**. Prioritise expensive, hard-to-reverse decisions (upgrades, logistics, people) over cheap, repeatable ones.

What bounds these results



Public data only

No internal telemetry or cost breakdowns. Absolute benchmarks & line-by-line allocation stay out of reach.



Regulation boundaries

Learned relationships don't cross a rules cycle; the 2026 reset needs retraining on thin early data.



Upgrades resist isolation

Packages arrive in stages on unique tracks raced once a year. Clean causal measurement is rarely possible.



Noisy value regressor

10.3 s validation MAE reflects inherent counterfactual variance (a 22–25 s pit-loss trade-off).



Structural strategy ceiling

Top-3 hit rate is capped below 1 by random safety-car timing, not fixable with better features.



Small expert sample

Three interviews: invaluable for validation, but used as support, not standalone proof.

Possible, to a meaningful but bounded extent



Part A

Lap performance**Highly predictable***0.78 s MAE - under 1% of a lap.*

Part A

Race strategy**Useful but bounded***50% top-3 hit rate, safety-car ceiling.*

Part B

Upgrade effect**Partially quantifiable***4 of 12 significant, noisy by nature.*

Part C

Cost & logistics**Best read qualitatively***Development is the binding constraint.*

Where this goes next



v2 opponent model

A smarter strategic-opponent model to capture under/overcut dynamics the current version underestimates.



Live OpenF1 integration

Near-real-time data for live decision support, pending the licensing restrictions on live F1 data.



2026 regulation era

Re-fit and stress-test the same pipeline as the new rules accumulate enough sample size.



Cost-cap transparency

Evolving budget disclosure may push Part C from estimation toward genuine measurement.

Designed as a foundation for **live decision support**. Richer telemetry and longer samples will sharpen it beyond today's bounded ceiling.

Thank you for your attention

I'd be glad to take your questions.

Bálint Csermák-Scharbert | BSc Business Engineering |
Maastricht University



List of References

Bálint Csermák-Scharbert | BSc Business Engineering |
Maastricht University

