

What Explains Constructor Performance in Formula 1's Hybrid Turbo Era?*

Suneel Chandra Vanamu

2025-12-03

This paper examines the joint role of engine supplier and constructor experience in determining team performance during Formula 1's 2014–2020 Hybrid Turbo Era. Using season-level data from the F1DB dataset, I estimate a model that has log points as a function of engine group and years since debut and compare those results to a LASSO specification and a train–test prediction framework. The engine supplier explains the vast majority of predictable variation in performance, while constructor experience explains very little additional variation once engine choice is conditioned on. These results formalize the strength of the so-called “engine gap” characteristic of this era and show that adding experience does little to improve predictive accuracy.

1 Introduction

Performance in Formula 1 reflects a mix of technical and organizational factors, and the introduction of the V6 turbo-hybrid power units in 2014 made engine programs a central component of competitive differences across teams. Engineering studies have documented large performance gaps across power-unit designs and energy-recovery systems during this era (Cavallaro 2018; Scarf 2019), while competitive analyses highlight how structural factors such as budget, experience, and continuity influence results (Jenkins and Floyd 2010). However, these discussions rarely quantify how engine supplier performance compares to team longevity, even though both factors are often cited in commentary about competitive balance.

This paper asks a specific version of that question: **To what extent can constructor performance in the 2014–2020 Hybrid Turbo Era be explained by engine supplier after accounting for years since debut?** Engine programs varied a great deal in reliability and efficiency as well as in hybrid energy use, whereas teams varied in terms of organizational maturity. By measuring the two visible factors, we can see how much of the outcomes in this period come from mechanical advantages versus accumulated experience.

To address this, I use season-level data from the public Formula 1 Database (F1DB) linking constructor points, engine assignments, and debut years. I model log-transformed points as a function of engine supplier and constructor experience, compare a full multiple regression with a reduced engine-only

*Project repository: <https://github.com/Suneel1508/MATH261A-project2>

specification, and use LASSO regularization and a train–test split to evaluate predictive performance. The analysis is descriptive rather than causal: the goal is to measure how strongly each factor is associated with observed performance rather than to isolate structural effects.

The remainder of the paper is organized as follows. Section 2 describes the dataset and variables. Section 3 outlines the modeling framework and estimation procedures. Section 4 presents the empirical and predictive results. Section 5 discusses implications for interpreting competitive structure in the Hybrid Era and outlines limitations of the approach.

2 Data and Variables

This analysis uses the public Formula 1 Database (F1DB contributors 2025), covering championship races from 2014 through 2020, corresponding to the Hybrid Turbo Era. All officially classified results are included. Race-level constructor standings were combined with engine assignment and constructor chronology records to create a season-level dataset, where the observational unit is a constructor–season. Data preparation, modeling, and visualization were conducted in R (R Core Team 2024) using `readr` (Wickham, Hester, and François 2024), `dplyr` (Wickham et al. 2023), `tidyverse` (Wickham and Girlich 2024), and `janitor` (Firke 2023) for data cleaning; `ggplot2` (Wickham 2016) for graphics; `caret` (Kuhn 2023) for stratified train–test splitting; and `glmnet` (Friedman et al. 2010) to fit the LASSO models. For each constructor in each season, I use the following variables:

- **Constructor points (response):** Total points scored in that season, summarizing overall competitive performance.
- **Engine supplier (predictor A):** A categorical variable indicating the power unit used that year (e.g., Mercedes, Ferrari, Renault, Honda, or customer rebrandings such as Tag Heuer).
- **Constructor experience (predictor B):** Years since the constructor’s debut, capturing organizational maturity and long-term technical continuity.

After filtering to the 2014–2020 seasons and dropping entries with missing debut year or unclear engine mapping, the final dataset contains **924 constructor–season rows across 17 constructors**. This coverage includes all major engine programs and long-running teams.

2.1 Data Limitations

There are two limitations that bear on interpretation. First, the debut year for constructors was not available for all teams, so observations missing that value were left out. This cuts the sample size for the models using experience but it keeps the overall mix of teams. Second, some of the engine labels reflect customer branding (such as Tag Heuer). These are treated as separate engine groupings, but they can be combined in broader groupings where necessary to aid in interpretation.

2.2 Summary of Key Variables

2.2.1 Distribution of Constructor Points

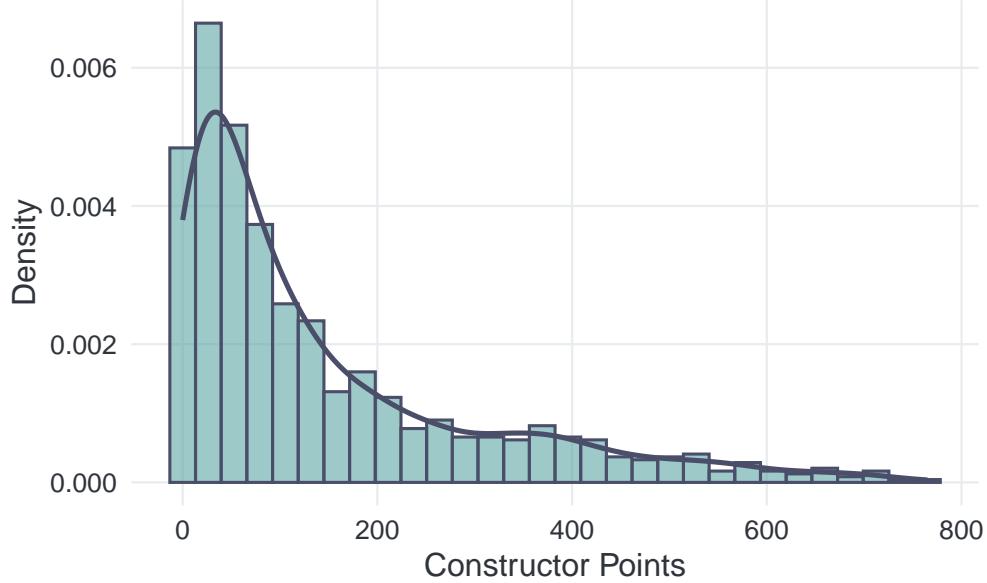


Figure 1: Distribution of constructor points in the 2014–2020 Hybrid Turbo Era.

The raw distribution of constructor points is highly right-skewed, reflecting the dominance of a few teams. This motivates the log transformation used in the regression models.

2.2.2 Distribution of Log-Transformed Constructor Points

Given the extreme right-skew in constructor points, driven by a few dominant teams, I apply a log transformation to stabilize variance before modeling; the resulting distribution is shown below.

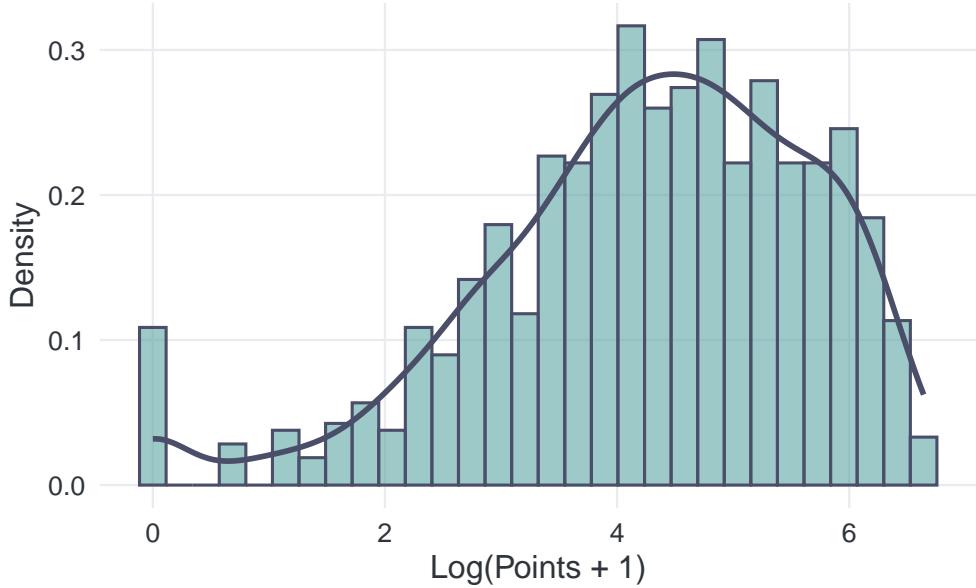


Figure 2: Distribution of log-transformed constructor points, reducing skew for modeling.

The transformed distribution is notably more symmetric than the raw points distribution in Section 2.2.1. This motivates modeling performance on the log scale and supports the regression assumptions examined later in the paper.

2.2.3 Points by Engine Supplier

Table 1: Summary of constructor points by engine supplier (2014–2020).

engine	avg_points	sd_points	n
tag-heuer	215.62712	125.15410	59
mercedes	199.20732	195.38357	328
ferrari	137.52212	148.15887	226
honda	88.83495	103.45901	103
renault	75.66071	77.49683	168
bwt-mercedes	67.56667	52.57814	30
toro-rosso	29.80000	16.57843	10

Table 1 provides a numerical comparison of performance across engine suppliers and highlights the clear separation between groups. Teams powered by Tag Heuer-branded Renault units or Mercedes engines achieve the highest average points, while Renault, Honda, and customer-Mercedes-powered teams tend to score fewer points. These differences motivate including engine supplier as a categorical predictor in the regression analysis.

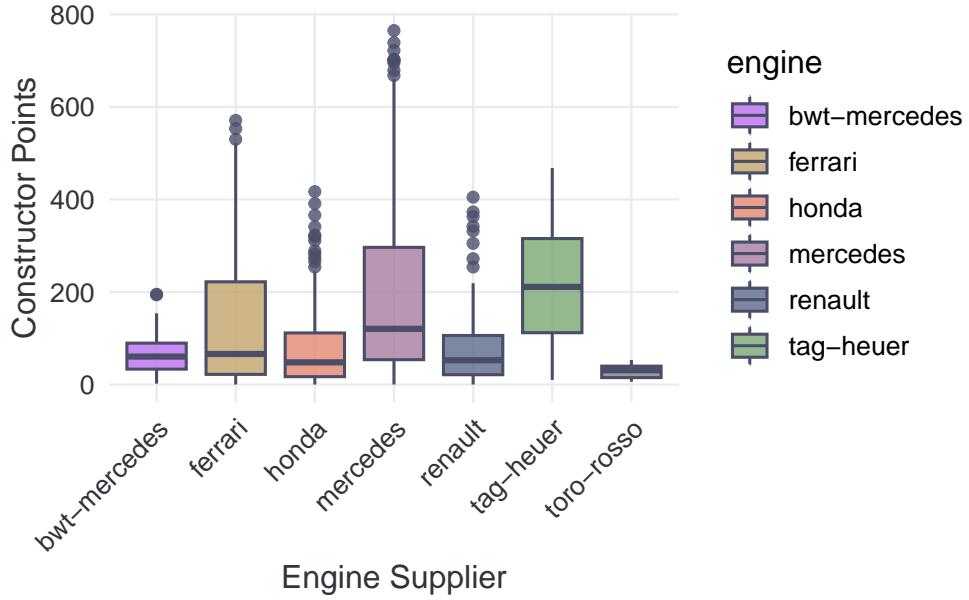


Figure 3: Season-level points by engine supplier, showing large performance differences across engine groups.

Engine suppliers differ substantially in performance: Mercedes-powered teams have the highest and most variable results, Ferrari teams cluster lower with wide spread, and Honda and Renault engines tend to yield lower scores. These differences justify modeling engine supplier as a categorical predictor.

2.3 Bivariate Relationships

2.3.1 Experience vs Constructor Points

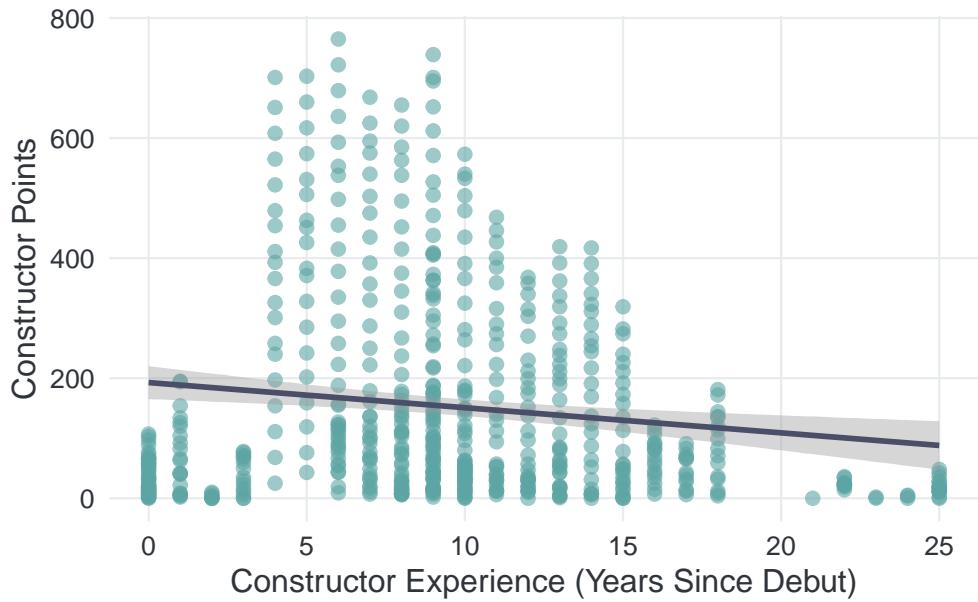


Figure 4: Relationship between constructor experience and season points.

Experience shows only a weak negative association with points, suggesting that team age alone is not a strong indicator of performance. This strengthens the use of experience as a secondary predictor in the regression models.

2.3.2 Constructor Experience by Engine Supplier

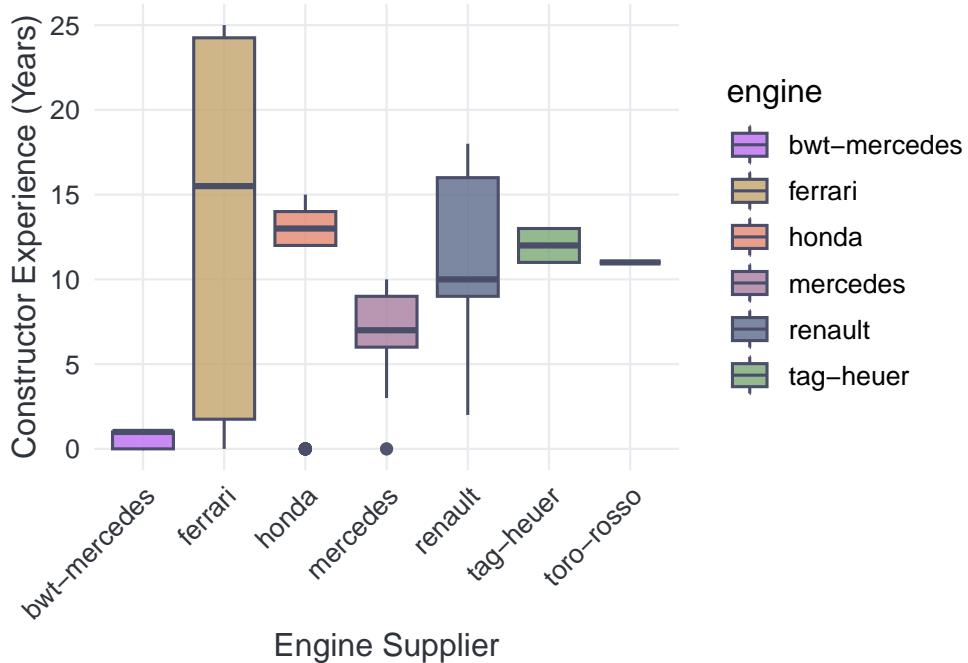


Figure 5: Distribution of constructor experience across engine suppliers.

Experience levels differ across engine groups, but no supplier is exclusively associated with older or younger teams. This supports including both engine supplier and experience in the models.

3 Methods

This section describes the statistical models used to examine how engine supplier and constructor experience relate to log-transformed points during the 2014–2020 Hybrid Turbo Era. The analysis includes a multiple linear regression model, a reduced engine-only model for comparison, a LASSO model for variable selection, and a train–test evaluation of predictive accuracy. All results are descriptive and measure associations rather than causal effects.

3.1 Response Transformation and Rationale

The dominance of a few teams creates substantial right-skew in constructor points, so I apply a log transformation to stabilize variance before modeling:

$$Y_i^* = \log(\text{points}_i + 1),$$

where Y_i^* is the transformed response for constructor–season i . Adding 1 ensures that seasons with zero points are included. This transformation produces a more symmetric distribution and supports the linear-model assumptions evaluated later.

3.2 Multiple Linear Regression Model

Let Y_i^* denote the log-transformed points scored by constructor i in season i . Let Engine_i indicate the engine supplier, and let Exp_i denote the constructor's experience measured as years since debut. The main regression model is

$$Y_i^* = \beta_0 + \sum_{k=1}^{K-1} \beta_k \mathbf{1}(\text{Engine}_i = k) + \beta_{\text{exp}} \text{Exp}_i + \varepsilon_i,$$

where the indicator terms capture differences across engine suppliers, and β_{exp} represents the association between experience and log points. The error term ε_i absorbs unexplained season-level variation.

To assess whether constructor experience contributes meaningfully beyond engine group, I also fit a reduced model:

$$Y_i^* = \alpha_0 + \sum_{k=1}^{K-1} \alpha_k \mathbf{1}(\text{Engine}_i = k) + \eta_i$$

Comparing the two specifications tests whether experience improves explanatory power once engine differences are accounted for.

3.3 Hypothesis Tests

Two nested-model F-tests evaluate the marginal contributions of the predictors.

Engine supplier:

$$H_0^{(\text{engine})} : \beta_1 = \beta_2 = \dots = \beta_{K-1} = 0, \quad H_1^{(\text{engine})} : \text{at least one } \beta_k \neq 0.$$

This test compares the full model with a model containing only experience.

Constructor experience:

$$H_0^{(\text{exp})} : \beta_{\text{exp}} = 0, \quad H_1^{(\text{exp})} : \beta_{\text{exp}} \neq 0.$$

This compares the reduced engine-only model with the full model.

These tests quantify whether each predictor group explains additional variation in log points.

3.4 LASSO Model for Variable Selection

Because engine supplier and experience may overlap in the variation they explain, I use LASSO regression to assess variable importance under coefficient shrinkage. The LASSO estimator solves:

$$\hat{\theta} = \arg \min_{\theta} \left[\sum_{i=1}^n (Y_i^* - \theta_0 - x_i^\top \theta)^2 + \lambda \sum_{j=1}^p |\theta_j| \right]$$

where x_i contains all engine indicators and experience. The penalty parameter λ controls the degree of shrinkage. I select two values using 10-fold cross-validation:

- λ_{\min} : which minimizes cross-validated error, and
- λ_{1se} : a more parsimonious choice within one standard error of the minimum.

Predictors set to zero under λ_{1se} are interpreted as contributing limited explanatory value once regularization is applied.

3.5 Predictive Evaluation

To assess generalization to new constructors or seasons, I perform an 80/20 train–test split stratified by engine supplier. All models are estimated on the training set and evaluated on the test set using:

- **RMSE**: sensitivity to large errors
- **MAE**: typical error magnitude
- **Test-set R^2** : proportion of explainable variation

Comparing predictive metrics provides an additional check on whether experience improves model performance and whether the regularized model offers meaningful benefits.

3.6 Model Assumptions

The multiple regression models rely on several assumptions about the error terms ε_i . First, the linearity assumption requires that the expected value of the log-transformed points varies linearly with the engine indicators and constructor experience. The exploratory plots in Section 2 and the use of the log scale help reduce curvature and make this approximation reasonable. The models also assume independence of errors across constructor–season observations; although seasons within the same constructor may exhibit some correlation, the goal of the analysis is descriptive rather than causal, so this dependence primarily affects interpretation rather than validity. A further requirement is that the variance of the errors remains roughly constant across fitted values. Any deviation from homoskedasticity would affect standard errors and significance tests, so the residual–fitted plot in Section 4.5 is used to assess this visually. Finally, OLS inference assumes that the errors are approximately normally distributed. The Q–Q plot in Section 4.5 evaluates this condition, and with the reasonably large sample size available here, mild departures from normality are unlikely to materially affect conclusions.

The LASSO models are not based on distributional or variance assumptions and are evaluated instead through their out-of-sample predictive accuracy. This alternative perspective also provides a means of determining how well the correlations found by OLS are robust when using penalties.

4 Results

4.1 Multiple Linear Regression Estimates

The full OLS model shows large differences across engine suppliers. Using Mercedes as the reference category, Ferrari, Honda, Renault, and Toro Rosso-powered constructors all have substantially lower expected log points. For example, the Ferrari estimate of -2.46 corresponds to roughly $e^{-2.46} \approx 0.09$ times the points of a Mercedes-powered constructor, holding experience constant.

These estimates indicate that engine supplier explains most of the systematic variation in constructor performance, while experience contributes only a modest incremental effect.

Table 2: Full OLS regression coefficients for log-transformed constructor points.

term	estimate	std.error	statistic	p.value
(Intercept)	4.735	0.115	41.085	0.000
enginebwt-mercedes	-0.911	0.256	-3.565	0.000
engineferrari	-2.456	0.199	-12.333	0.000
enginehonda	-0.939	0.173	-5.429	0.000
enginerenault	-1.320	0.152	-8.660	0.000
enginetag-heuer	0.109	0.192	0.569	0.569
enginetoro-rosso	-1.729	0.409	-4.224	0.000
experience	0.023	0.011	1.964	0.050

Full OLS: Predicted vs Observed

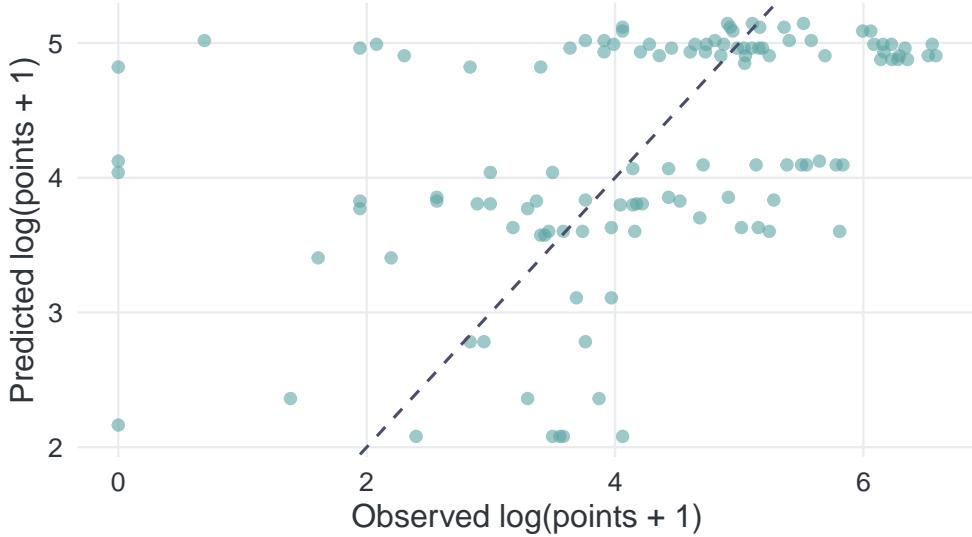


Figure 6: Predicted vs observed log points for the full OLS model on the test set.

The coefficient on constructor experience is small and borderline significant, indicating that experience plays a limited role once engine supplier is included.

4.2 Hypothesis Test Results

To assess whether engine supplier and constructor experience each provide explanatory power for performance, I use nested-model F-tests corresponding to the hypotheses defined in Section 3.3.

Table 3: Nested-model F-tests evaluating the marginal contributions of engine supplier and constructor experience.

Test	F_statistic	p_value
Engine effect (experience-only vs full model)	37.00596	0.00000
Experience effect (engine-only vs full model)	3.85560	0.05005

Nested-model F-tests confirm these patterns. The omnibus engine-effect test strongly rejects the null hypothesis ($F = 37.01, p < 0.00001$), indicating that engine supplier explains substantial variation in log-transformed points even after adjusting for constructor experience. This confirms the OLS findings in Section 4.1: engine differences are large, persistent, and statistically meaningful.

The experience-effect test yields $F = 3.86$ with $p = 0.05005$, which is borderline at the 5% level. This suggests that experience provides only a modest improvement in explanatory power beyond engine supplier.

4.3 LASSO Model Results

The LASSO model provides a complementary perspective by shrinking weak predictors toward zero. Under the one-standard-error penalty, the coefficient for constructor experience is shrunk to exactly zero, while all engine indicators remain nonzero. This indicates that experience provides little unique explanatory value once engine differences are accounted for.

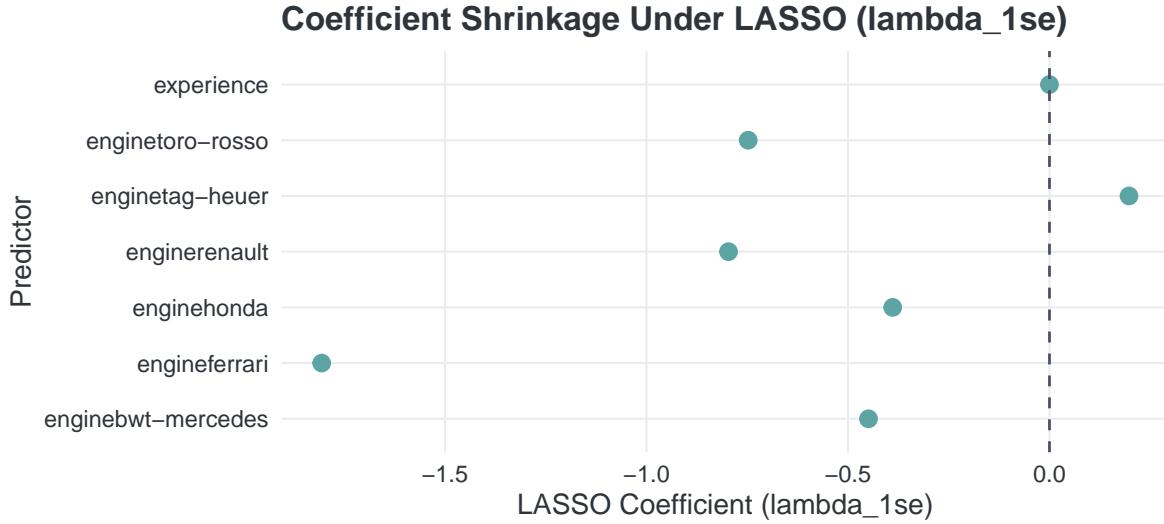


Figure 7: LASSO coefficient estimates under the lambda_1se penalty.

The retained engine coefficients follow the same ordering as in the OLS model: Ferrari, Renault, Honda, and Toro Rosso have negative effects relative to Mercedes, with broadly similar magnitudes. The LASSO predicted-versus-observed plot shows tighter shrinkage toward the mean relative to the OLS model, which is expected given the penalization. Nonetheless, the LASSO model accurately captures group-level differences across engine suppliers while reducing variability in fitted values.

LASSO: Predicted vs Observed

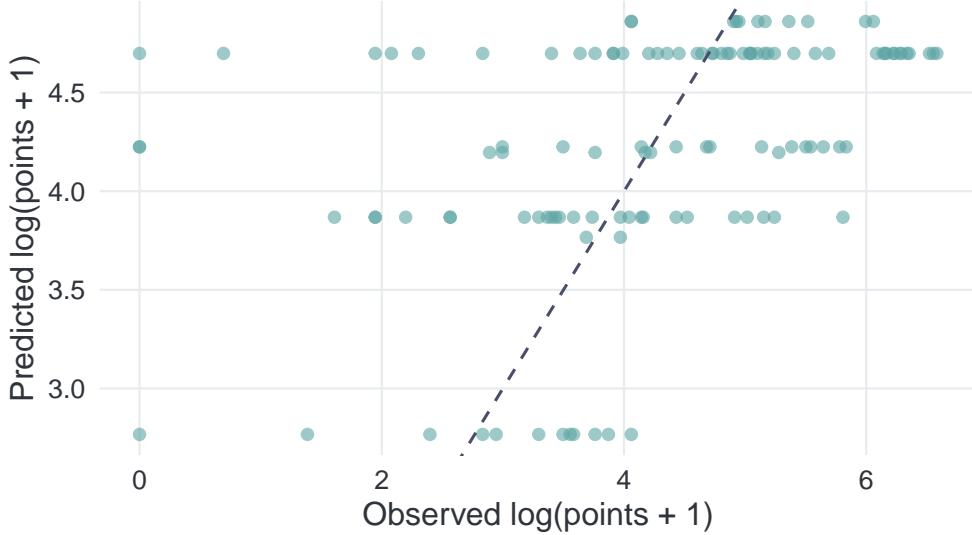


Figure 8: Predicted vs observed log points for the parsimonious LASSO model on the test set.

Overall, the LASSO results corroborate the OLS findings: engine supplier is consistently the strongest predictor, and experience does not survive penalization when the model is simplified.

4.4 Predictive Evaluation Results

The following table compares out-of-sample predictive accuracy across the four models. Test-set R^2 values range from 0.16 to 0.18, reflecting the substantial noise inherent in season-level performance. While these values are modest, they are typical for models using only structural characteristics and no race-level or driver-level information.

Table 4: Out-of-sample predictive accuracy for four models using an 80/20 stratified train–test split.

Model	RMSE	MAE	R2
Reduced OLS	1.3512	1.0108	0.1678
Full OLS	1.3579	1.0136	0.1596
LASSO (λ_{min})	1.3569	1.0128	0.1608
LASSO ($\lambda_{1\text{se}}$)	1.3409	0.9958	0.1804

Across the four models, several patterns emerge. The reduced engine-only OLS model performs slightly better than the full OLS specification, indicating that constructor experience does not strengthen predictive performance and may introduce additional noise. The LASSO model using the $\lambda_{1\text{se}}$ penalty achieves the best out-of-sample accuracy, yielding both the lowest RMSE (1.3409) and the highest test-set R^2 (0.1804). This sparse model eliminates constructor experience completely, reducing all other engine parameters down towards zero, and yet performs as well as, and in many cases better than, models with far more complex formulations. Taken in conjunction, these findings suggest that engine

supplier explains almost all of the predictable patterns in season-level point data, and that simpler models, those which preserve only engine data, fare better for generalization purposes than their more complex counterparts which also preserve experience.

4.5 Diagnostic Checks

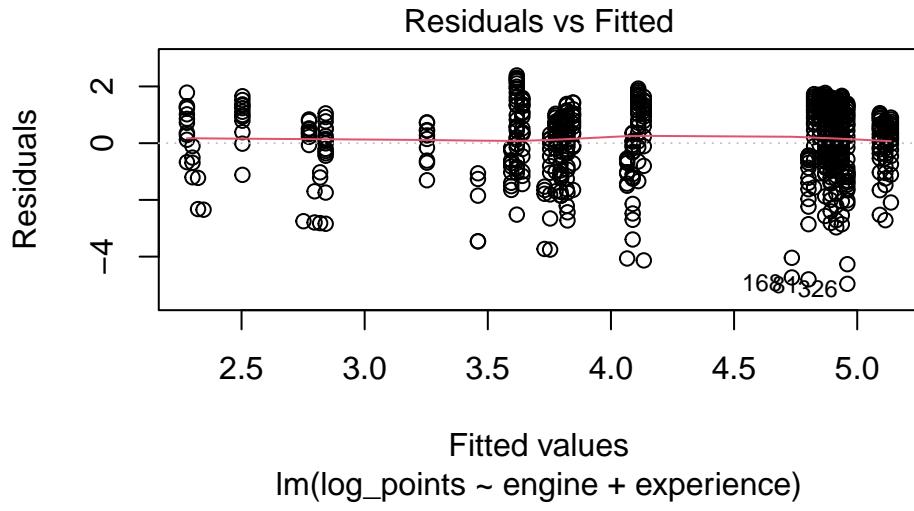


Figure 9: Residual-fitted plot for assessing linearity and variance.

Residual-fitted plots do not show strong curvature, indicating that the linear structure on the log scale is reasonable. There is mild heteroskedasticity at higher fitted values, which is expected given the concentration of high-scoring constructors in a small subset of seasons.

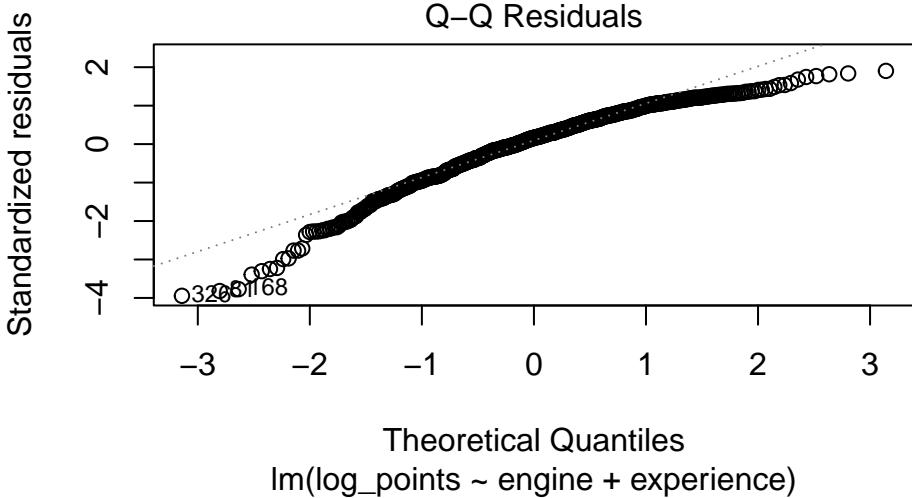


Figure 10: Normal Q–Q plot for standardized residuals.

The Q–Q plot shows heavier tails than those of a normal distribution, reflecting occasional extreme performances.

These deviations do not undermine the main conclusions: the linear mean structure is appropriate, the variance pattern is stable enough for inference, and the heavy tails explain why predictive accuracy remains limited even with strong group-level predictors.

5 Discussion

This paper examined the ways in which engine supplier and constructor experience related to team performance in Formula 1’s 2014–2020 Hybrid Turbo Era. Throughout all models, engine supplier was the strongest and most consistent predictor of log-transformed points, while constructor experience added little beyond what was captured with engine differences accounted for. The LASSO model removing the experience term further supports this: experience only explains a small amount of variation compared to engine programs.

When we interpret the coefficients on the original scale, teams powered by Ferrari, Renault, Honda, and Toro Rosso earned only a small fraction of the points that Mercedes-powered teams did. This puts numbers on the common idea of an “engine gap” in that era, and suggests that long-term participation doesn’t lead to big performance gains when the underlying mechanical differences stay large.

There are several limitations to these findings. Season-level data cannot capture race-level factors such as driver skill, teams’ speed of development, reliability, or aerodynamics that impact results. Constructor experience is only a rough stand-in for the strength of an organization, and there is limited accuracy due to variables we did not observe.

Future work might incorporate race-level performance data, financial information, or driver metrics to disentangle mechanical from organizational effects. Examining later periods of regulation would also help demonstrate if engine-driven gaps in performance remain consistent across alternative technical regimes

References

- Cavallaro, Fausto. 2018. “Improving Power-Unit Efficiency in Formula 1: Hybrid Technologies and Performance Trends.” *Energy Procedia* 148: 1165–72.
- F1DB contributors. 2025. “F1DB: Formula 1 Database (CSV Release and Documentation).” GitHub repository. <https://github.com/f1db/f1db>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Friedman, Jerome, Trevor Hastie, Robert Tibshirani, Noah Simon, and Balasubramanian Narasimhan. 2010. *Glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. <https://CRAN.R-project.org/package=glmnet>.
- Jenkins, Mark, and Steven Floyd. 2010. “Organizational Routines and Performance in High-Technology Motorsport.” *Organization Science* 21 (4): 633–53.
- Kuhn, Max. 2023. *Caret: Classification and Regression Training*. <https://CRAN.R-project.org/package=caret>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Scarf, Philip. 2019. “Competitive Balance in Formula 1: Modelling Constructors’ Performance.” *Journal of Quantitative Analysis in Sports* 15 (2): 123–38.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://doi.org/10.1007/978-3-319-24277-4>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- Wickham, Hadley, Jim Hester, and Romain François. 2024. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.