OUTDRIVE TO SURVIVE
Extracting driver skill from constructor advantage

Soumil Banerjee, Won Seok Kang, Nikhil Krishnan, Vianna Seifi

1. Introduction

In Formula 1, there are many factors that determine the outcome of a race. Of them, the two most critical factors are the individual driver's skill and the inherent advantage of their car. Not all cars are built equal: there are 10 constructors each season, each fielding two drivers with machinery that can vary in pace, performance, operating window, design, etc. Previous research has shown that most of the results of a driver can be attributed to their constructor. Thus, if driver skill matters less, are better teams actually employing better drivers? If not, the official driver's championship ranking loses its meaning and credibility.

Specifically, we want to answer the following question: **Which drivers are actually the fastest, regardless of the car they're in?** Separating the impact of driver ability from the strength and speed of the cars is challenging, especially as the cars receive changes over a season and teammates run different setups. In this project, our objective is to analyze the problem of isolating driver performance from constructor effects by analyzing race finishing positions from the hybrid era (2014-2024). Once isolated, we rank drivers' skills, agnostic of constructors.

Using a Bayesian hierarchical model, we estimate and rank drivers' abilities independently of car strength. This approach provides interpretable uncertainty estimates and avoids overfitting drivers with only a few races logged. Our results reinforce the idea that finishing race positions are significantly affected by the constructor's advantage rather than the driver's skill. However, we find that individual driver skills are still relevant in determining the outcome of a race, with isolated driver ability still aligning similarly to the official drivers' championship rankings.

2. Data

2.1. Dataset Description

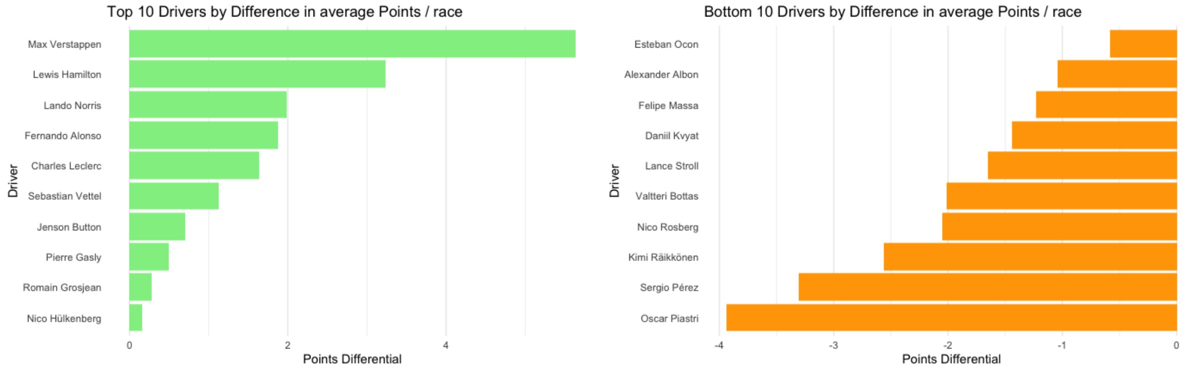
The dataset used for this project combines several tables from a public Formula 1 database on Kaggle, compiled using the Ergast API. The first table (`races`) contains information about each Grand Prix race, including race ID, year, name, and dates. The second table (`drivers`) provides demographic and identification information for each driver, including driver ID and name. The third table (`results`) includes detailed race-by-race results for each driver, including finishing position, points scored, constructor ID, and race status. The three datasets were merged using common keys (`raceId`, `driverId`, and `constructorId`) to create a dataset containing a single driver's result for a particular race for each observation. For this study, we restricted the data to races from 2014 to 2024 to focus on the modern hybrid power era.

2.2. Exploratory Data Analysis (EDA)

For our EDA, we looked at teammate comparisons since each team has two drivers. This is because a driver's teammate is the closest comparison we can get to one that is completely independent of the car. However, it is still important to note that the two drivers within a team may still have different setups and that the car may be built to favor the driving style of one driver over the other. For our comparison, we looked at the average point differential per race of a driver with their teammate and isolated the top ten and bottom ten. This is shown in Fig. 1.

Looking at the top ten drivers, we can see that the current 4-time defending world champion Max Verstappen has the highest lead of about 5 points over his teammates, with 7-time world champion Lewis Hamilton in second place with about 3 points over his teammates. Several other world champions make the top 10, such as Fernando Alonso, Sebastian Vettel, and Jenson Button. However, Max Verstappen has notably almost always been paired with teammates who are significantly worse than him, while Lewis Hamilton has had several world champions as teammates. This is not to say that one is better than the other, but rather to note that teammate comparisons contain only general trends and should not be taken as a decisive litmus test.

This is more so apparent in the bottom 10. For example, Nico Rosberg and Kimi Raikkonen are both world champions, but have had strong teammates for significant portions of their careers. Alex Albon, Valtteri Bottas and Sergio Perez are also examples of strong drivers who have had even stronger teammates. We also see Oscar Piastri in last, but he has only completed two years in F1 and has been outperformed by his teammate. Thus, it appears that teammate comparisons are better at highlighting elite drivers at the top than segregating the poor drivers at the bottom.



(a) Top 10 Drivers

(b) Bottom 10 Drivers

Fig. 1: Teammate comparisons provide a rough proxy for isolating individual driver performance.

In addition, we also looked at the performance of the current ten constructors in the past few years (2021-2024). Their performance was measured by their ranking in the Constructors' Championship after each race.

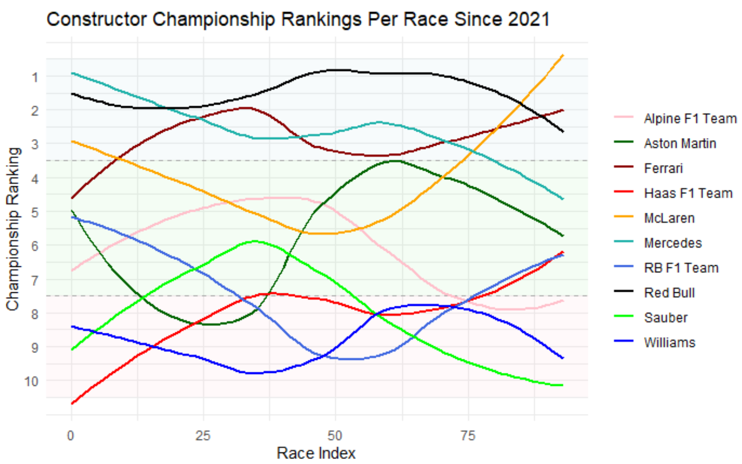


Fig. 2: Teams tend to cluster into frontrunners, midfielders, and backmarkers across seasons.

If we split the rankings into three groups as per the bands created (i.e. front-runners, midfield, and back-markers), we can see that teams generally tend to stay within their group. This supports the idea that the constructor is more important than the driver when it comes to the final outcome of a race.

3. Methods

To isolate and quantify the contributions of individual drivers and constructors to race outcomes, we employed a Bayesian hierarchical model. Using the `brms` package in R, we estimated full posterior distributions of driver and constructor effects, allowing us to separate the influence of driver ability from constructor performance on finishing positions.

Our model treats the finishing position as an ordinal outcome and includes random intercepts for both drivers and constructors. This structure captures how much of a driver's observed performance can be attributed to their own skill versus the quality of their car, while accounting for uncertainty using posterior distributions and credible intervals. The Bayesian framework also provides natural regularization via prior distributions, improving estimation, especially for drivers or constructors with fewer observations.

Posterior means and 95% credible intervals were extracted to interpret the estimated effects and quantify uncertainty. Given the large number of drivers in the dataset, we focused our interpretation on the top 10 and bottom 10 drivers ranked by their estimated random effects, to provide clearer comparisons and insights into individual performance.

This hierarchical framework supports our main goal of extracting driver skill from team advantage while appropriately modeling the ordinal nature of race outcomes.

3.1. *Feature Engineering + Covariate Selection*

We restricted our analysis to races from 2014-2024 and included only drivers who finished the race. Several additional covariates were selected to allow the model to incorporate race-level dynamics. A binary indicator for street circuits (`street_race`) was created by referencing circuit names against a list of known street tracks, controlling for the increased difficulty and variability of these tracks, where driver skill might potentially play a larger role. Starting grid position (`grid_position`) represents where the driver qualified, which accounts for how many positions a driver was able to gain over the course of the race to get to their ending position. A fast car is helpful for overtakes, but driver skill becomes very important against strong defending and for difficult passes. Finally, we include the number of pit stops per race (`pit_stop_count`) since this is typically dependent on the strategists rather than the driver, and a poor or intelligent strategy can impact the final standings.

3.2. *Model Specification*

Our aim was to model the final position of each driver in a race as an ordinal outcome, accounting for both driver ability and constructor strength while adjusting for race-specific factors. To achieve this, we employed a Bayesian cumulative logit model, which is specifically designed for ordered but non-continuous outcomes such as race finishing positions.

In Formula 1, the gap between winning and placing 2nd is more meaningful than the gap between 10th and 11th. Critically, this means that ranks are not equally spaced in significance, and that our model needs to distinguish between difficulty thresholds between finishing positions. This motivated the choice of modeling finishing position as an ordinal outcome using a cumulative logit link. By estimating separate threshold parameters between each pair of adjacent positions, the model effectively represents the non-uniformity of different race outcomes and avoids assigning equal value to all positional differences.

For each driver i in race j, the cumulative logit model is given as:

$$\log\left(\frac{\Pr(\text{Position}_{ij} \leq k)}{\Pr(\text{Position}_{ij} > k)}\right) = \theta_k - X_{ij} \quad \text{for each } k = 1, \ldots, K-1,$$

where θ_k are learned threshold parameters between ordinal categories, and X_{ij} is the linear predictor for driver i in race j. In other words, θ_k refers to the corresponding threshold for category k,

The linear predictor X_{ij} is defined as:

$$X_{ij} = \begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix}^T \begin{bmatrix}\text{street\_race}_j\\ \text{grid\_position}_{ij}\\ \text{pit\_stop\_count}_{ij}\end{bmatrix} + u_{\text{driver}^{(i)}} + v_{\text{constructor}^{(i)}},$$

where:

- β terms are fixed effect coefficients,
- $u_{\text{driver}^{(i)}} \sim \mathcal{N}(0, \sigma^2_{\text{driver}})$ are random intercepts for drivers,
- $v_{\text{constructor}^{(i)}} \sim \mathcal{N}(0, \sigma^2_{\text{constructor}})$ are random intercepts for constructors.

Posterior samples were generated through Markov Chain Monte Carlo (MCMC) sampling across four chains, with convergence diagnostics monitored to ensure proper mixing and sampling efficiency. A table of coefficient estimates and the posterior estimates for the random effect standard deviations can be found in the appendix.

3.3. *Model Evaluation*

We evaluated model fit using the posterior predictive check, which involves simulating new datasets from the posterior predictive distribution and comparing them to the observed race outcomes. If the simulated and observed distributions closely match, this provides evidence that the model captures the underlying data-generating process.

4. Results

4.1. *Relative Ability and Posterior Distributions*

We summarize our results by displaying the random effects for all the constructors and the random effects for our top 10 and bottom 10 drivers. We also include a 95% interval to quantify uncertainty. Note that we can interpret

Relative Ability as how many positions the driver can do better/worse than the average driver. The constructor relative ability can be interpreted in an analogous manner. For example, having Mick Schumacher drive will result in a drop of -0.5 positions compared to the baseline driver. Another example is that a Red Bull car will perform about 2 positions better than the baseline constructor.

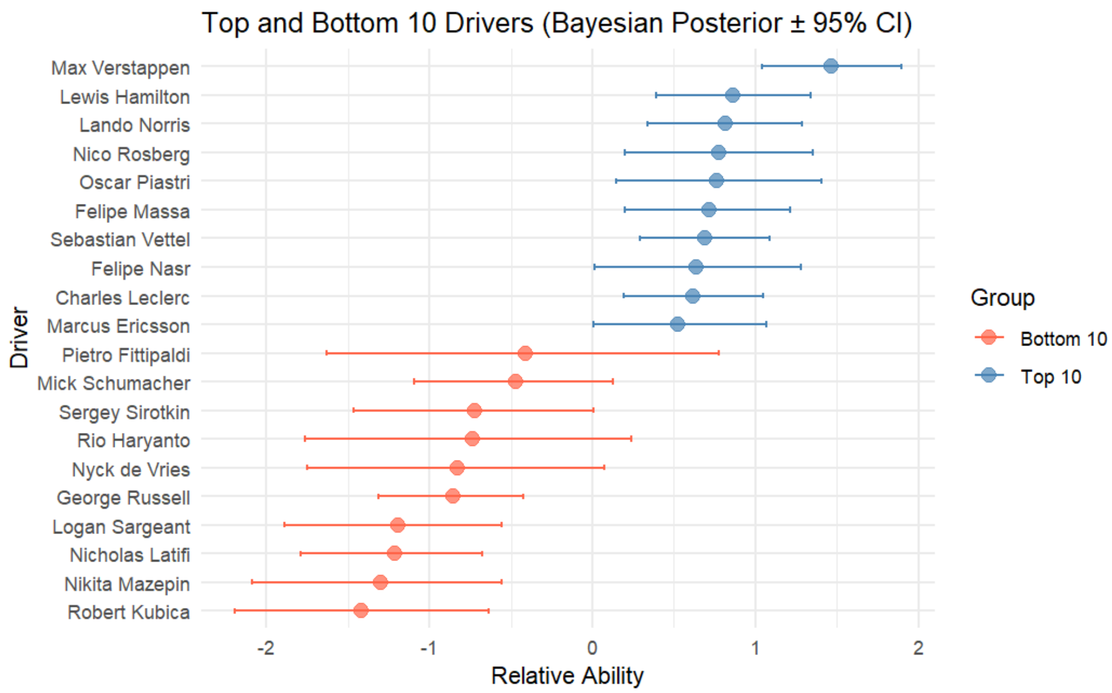


Fig. 3: Top drivers like Verstappen and Hamilton stand far above, while Mazepin and Sargeant anchor the bottom.

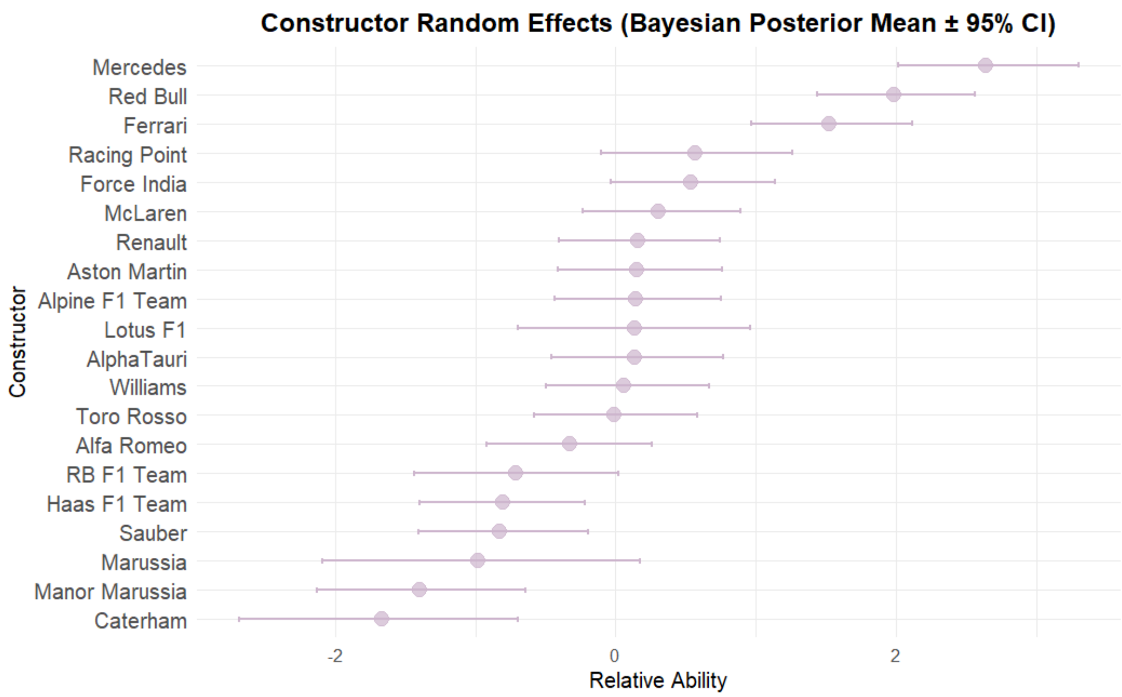


Fig. 4: Mercedes and Red Bull show the strongest positive effects, while Caterham and Marussia exhibit the largest negative impacts on finishing position.

We can see that the range of effect of the constructor is larger than the range of effect of the driver. The drivers are within ± 1.5 positions of each other, while constructors are within almost ± 3 positions of each other.

In the top ten drivers, we can see several world champions (i.e. Max Verstappen, Lewis Hamilton, Nico Rosberg, Sebastian Vettel) and world champion runner-ups (i.e. Lando Norris, Felipe Massa, Charles Leclerc), matching the top 10 from the EDA section. In the bottom 10 drivers, we can see drivers that were dropped for poor finishes or too many crashes (i.e. Mick Schumacher, Logan Sargeant, Nicholas Latifi, Nikita Mazepin). Similarly, the known frontrunner teams such as Mercedes, Red Bull, and Ferrari are at the top of the constructors. Backmarker teams like Marussia, Caterham, and Sauber show up at the bottom of the constructors.

4.2. Posterior Predictive Assessment of Model Fit

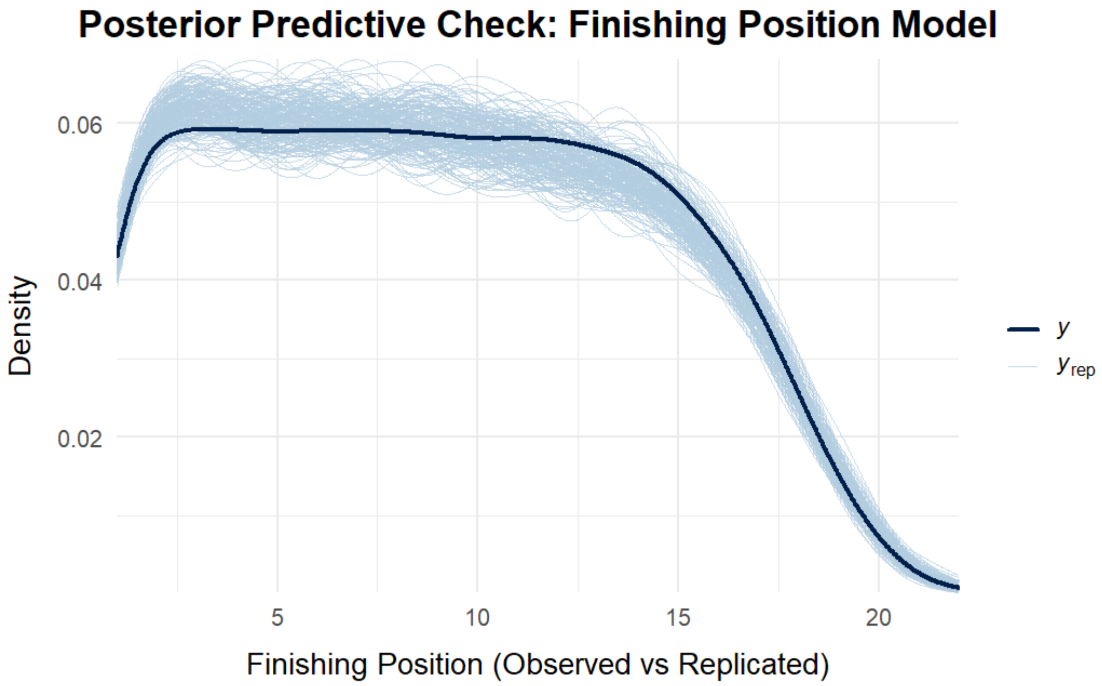


Fig. 5: Observed and simulated densities show strong alignment across finishing positions.

In the posterior predictive check, we compare the observed distribution of finishing positions to replicated datasets simulated from the fitted cumulative logit model. As can be seen from figure 5, the model successfully captures the key features of the data: a steep rise at better finishing positions (from 1st to 5th), a broad plateau between 5th and 15th place, and a sharp decline after 15th. The close overlap between the observed and simulated curves suggests that the model adequately represents the distributional structure of real race outcomes across the entire finishing spectrum, especially given the prevalence of Did Not Finish (DNF) results, leading to smaller likelihoods of the bottom 5 finishes.

5. Discussion

5.1. Understanding Results

From our modeling results, the larger random effects of the constructor over the drivers reinforces previous results that **the constructor has a larger impact on race outcome than the driver**. In terms of the actual rankings, it does appear that drivers on average tend to race for teams of similar caliber. For example, Max Verstappen, Lewis Hamilton, Nico Rosberg, and Sebastian Vettel spent most, if not all of their careers driving for at least one of the front-runner teams Mercedes, Red Bull, and Ferrari. Drivers such as Robert Kubica, Mick Schumacher, and Nikita Mazepin drove for Sauber and Haas. In this sense, we can tentatively conclude that a driver's skill does matter and they are being selected accordingly by teams. Therefore, we do not have to worry that the Drivers' Championship isn't a valid ranking of driver skill. World Champions tend to be more likely to "outdrive" their car, and those at the bottom of the rankings do not.

However, we recognize that there are some unexpected driver rankings in our results. Two cases of this are Oscar Piastri and George Russell. Oscar Piastri has only been in F1 for two years, and has been consistently beaten by his McLaren teammate Lando Norris in both of those. On the flipside, George Russell has obtained several wins at Mercedes and beaten Mercedes teammate Lewis Hamilton in head-to-head qualifying over the past three years. If we look at the performance of Piastri and Russell's constructors, we can see a common factor. Piastri joined McLaren right when the team's car became the fastest of the grid, and Russell joined Mercedes when the team started to struggle with new regulations. So while we have attempted to disentangle driver and constructor to our best efforts, it appears that there still seems to be some connection between the two if the constructor undergoes a large change in performance.

5.2. *Limitations*

This brings us to some of the limitations of our methodology and analyses. As mentioned previously, upswings or downswings in constructor performance can affect our perspective of driver ability if their stint at a team aligns with these performance shifts. In addition, certain cars are easier to drive with certain driving styles. An example of this is Red Bull, which is hyper-tuned to Max Verstappen's extremely sensitive, pointy driving style, which many of Red Bull's second drivers cannot adapt to. This also applies to car regulations. Notably, Lewis Hamilton has struggled with recent ground-effect regulation changes, and Fernando Alonso's performance took a noticeable dive once the hybrid era began. Ideally, we would like to model the effect of a driver in a specific team. Unfortunately, there is not enough data to achieve any useful results, and drivers race for only a few teams at most over their career. Finally, we excluded all DNFs from our data, but it would instead be better to include and distinguish non-finishing results that are caused by the driver versus the constructor. This could help include information about drivers who are more crash-happy and constructors whose cars aren't as reliable.

5.3. *Future Work*

For future work, we could use a more fitting model like the Plackett-Luce model, which handles list-wise preferences. This is perfectly suited for representing rankings, and it also handles partial rankings, which is useful for handling DNFs. Another avenue of interest is including more covariates in our model. Some possible features are an indicator of a wet race, pre-race championship position, number of upgrades the car has received by that race, etc. Overall, our goal for future work would be to put greater effort into separating driver and constructor so that we can find more accurate rankings of driver ability.

6. Appendix

Table 1: Posterior Estimates for Covariates and Threshold Parameters

Parameter	Estimate	95% CI Lower	95% CI Upper
street_race	-0.15	-0.28	-0.02
grid_position	0.25	0.23	0.26
pit_stop_count	0.16	0.10	0.21
θ_0	-3.16	-3.76	-2.61
θ_1	-1.98	-2.56	-1.43
θ_2	-1.11	-1.69	-0.55
θ_3	-0.38	-0.95	0.17
θ_4	0.28	-0.30	0.82
θ_5	0.89	0.33	1.44
θ_6	1.46	0.89	2.01
θ_7	2.00	1.43	2.56
θ_8	2.50	1.92	3.06
θ_9	2.99	2.42	3.55
θ_{10}	3.48	2.90	4.05
θ_{11}	3.96	3.38	4.53
θ_{12}	4.47	3.88	5.04
θ_{13}	5.02	4.43	5.60
θ_{14}	5.61	5.01	6.18
θ_{15}	6.25	5.66	6.83
θ_{16}	7.04	6.44	7.63
θ_{17}	7.93	7.30	8.54
θ_{18}	9.04	8.36	9.73
θ_{19}	10.55	9.63	11.51
θ_{20}	12.00	10.58	13.74

Table 2: Standard deviation estimates for driver and constructor effects

Group	Parameter	Estimate (95% CI)
Drivers	σ_{driver}	0.69 (0.53, 0.90)
Constructors	$\sigma_{\text{constructor}}$	1.16 (0.81, 1.65)