

The Relationship between Qualifying Position and Final Position in the 2021 Fanatec GT World Challenge Europe

Samuel Jens

10/30/2021

Introduction

Over the last year and a half, I have become interested in GT3/4 racing. It started at the onset of the pandemic when I happened upon YouTube channels (e.g., Jimmy Broadbent and Jardier) showcasing the sim(ulated) racing title *Assetto Corsa Competizione* by Kunos Simulazioni, an Italian gaming studio. I was amazed with how realistic the graphics looked, how visceral the sounds were, and how detailed the overall experience appeared. After learning more about the game, I also discovered the GTWorld YouTube channel, the official channel for the SRO Motorsport Group's racing series, which live-streamed real races from various regions around the world. The racing was exciting but so too was the chat that corresponded with the stream. It was a little community of racing fans during what was an isolating time.

A common theme, both explicitly mentioned and implicitly hinted at during the broadcasts, was how important qualifying was for a driver's/team's overall position. This sparked a research question: "How important is qualifying?" Clearly cars that start at the front of the grid have a better chance of finishing well than cars that start at the back. But how "determinant" is qualifying? What are the probabilities of finishing better than one's starting position? Is the relationship different for shorter sprint races compared to long, multi-hour endurance races?

I thought this would be an exciting extracurricular research project that I could develop into a blog post for my personal website. I present the data and my analyses in the sections below. There are many analyses that could be run with the collected data. As such, this blog post may be updated as I (re)consider alternatives and new questions. Thanks for reading.

Data

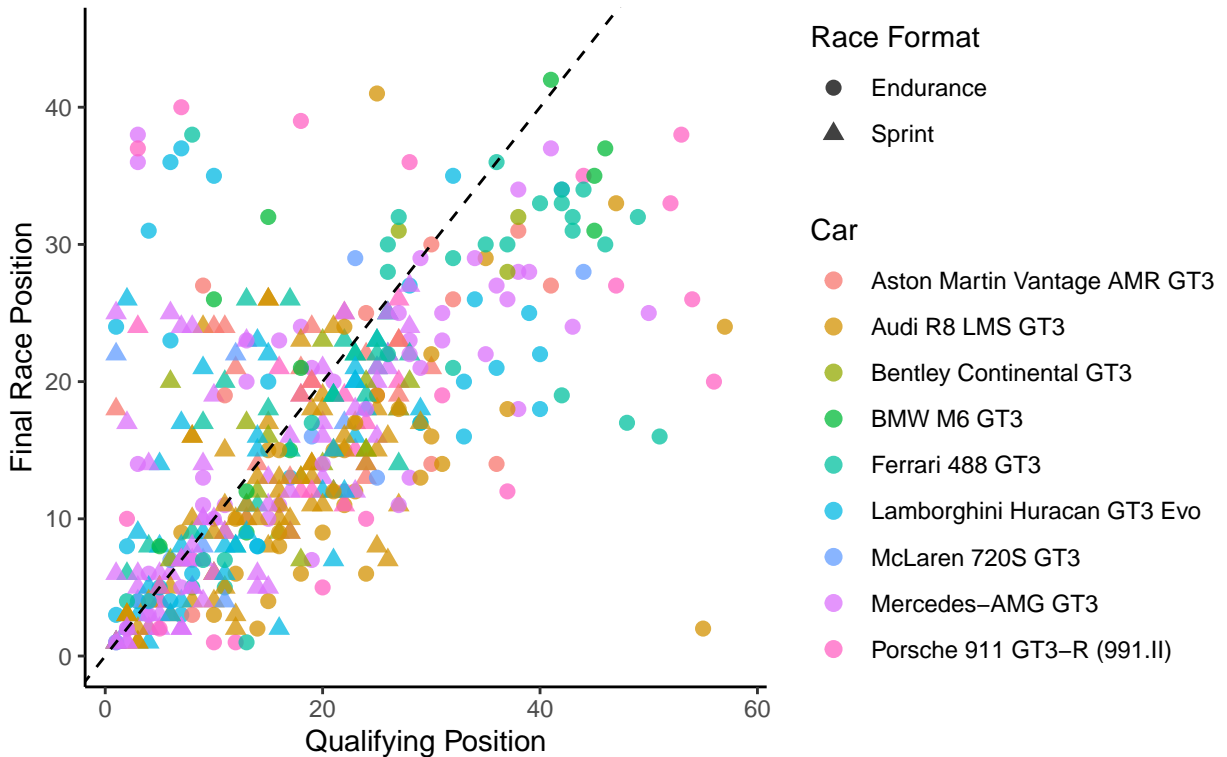
The data for this project may be found on the SRO Motorsport Group's website (see <https://www.gt-world-challenge-europe.com/results>). I collect the qualifying and final results for the ten rounds held during 2021. A few of the races did not have qualifying information, so I gather the data from GTWorld's YouTube videos (which stream practice and qualifying sessions in addition to the full race). The sprint race weekends feature two shorter 1-hour races, while the endurance weekends feature one longer race (ranging 3-24 hours). The dataset contains the qualifying position of each driver lineup – all races feature at least two drivers and a mandatory pit stop to change drivers. I also have the driver names, driver lineup category (i.e., Pro Cup, Silver Cup, or Pro-AM Cup), team names, car manufacturer and model, track, race format (i.e., sprint or endurance), race session (i.e., 1 or 2), and track.

There are a total of nine manufacturers during the 2021 season. Four of the nine manufacturers' vehicles are front engine: Aston Martin, Bentley, BMW, and Mercedes-AMG. Four are mid-engine: Audi, Ferrari, Lamborghini, and McLaren. Finally, one is rear engine: Porsche. Even with vast differences in appearance, all GT3 specification cars are subject to the SRO's balance of performance – cars are restricted in certain ways to balance them across various tracks.

Overall, there are 512 observations across the ten race weekends (though many race weekends have two races so each team appears twice for a given track). However, given that I analyze motorsports, there is attrition across the 2021 season. Most of the models presented below examine the 436 complete observations (or 85.16% of the total observations) in which the car/team not only qualifies but also finishes the race.

The figure below reveals the relationship between qualifying position (x-axis) and final race position (y-axis). The dashed 45° line indicates a perfect relationship. In other words, if all observations were on the line, then qualifying would perfectly predict (or determine) final race position. The different shapes denote whether the race was an endurance (circle) or sprint (triangle) race. The colors represent a specific manufacturer's car. I excluded the two new updated vehicles for the 2022 season which appeared in the final endurance race of the season (e.g., the BMW M4 GT3 and Audi R8 LMS GT3 Evo 2).

Fanatec GT World Challenge Europe (2021): Relationship between Qualifying and Final Positions



There appears to be a positive relationship that follows the dashed line, but there is also considerable variability suggesting qualifying is not as important as some may contend. Furthermore, most observations are below the 45* dashed line. This means most cars finished *better* than their starting (qualifying) position. This finding is in part due to cars/teams dropping out of the races due to mechanical problems and/or crash damage. As such, 65.37% percent of the complete observations finish better than (or equivalent to) their qualifying position. I turn to statistical models to evaluate this relationship in the next section.

Regression Models

The table below reveals the Ordinary Least Squares (OLS) bivariate regression when final race position is regressed on qualifying position. As one would expect, there is a highly significant relationship between the two ($p < 0.01$). However, the magnitude of the coefficient is striking. One might expect a coefficient closer to 1 rather than 0.47. This result illustrates the finding from the figure above. A one place increase in qualifying position only increases the final race position by 0.47 “places.” In other words, since the coefficient is less than 1, teams tend to do better in the final race compared to their qualifying position.

Furthermore, the R-squared value is also intriguing. Only 35.8% of the variance in final race position can be explained by the qualifying position. I find this to be rather low – 64% of the variance in final race

position remains unexplained. Qualifying clearly does not “determine” final position. There are many other important factors that influence the relationship between qualifying and final position.

Table 1: Bivariate OLS Regression with Robust Standard Errors

	<i>Dependent variable:</i>
	Final Position
Qualifying Position	0.470*** (0.037)
Constant	7.141*** (0.813)
Observations	436
R ²	0.358
Adjusted R ²	0.356
Residual Std. Error	7.727 (df = 434)
F Statistic	241.626*** (df = 1; 434)

Note: Robust SEs in parentheses *p<0.1; **p<0.05; ***p<0.01

The bivariate results from Table 1 above suggest other variables are necessary for explaining the finishing positions during the 2021 European season. While qualifying does account for final position, a large amount of variance remains unexplained. Table 2 presents three new models with new independent variables. These variables are coded as the following:

- Endurance: 1 for an endurance race and 0 for a sprint race.
- Front Engine: 1 for front engine race cars (e.g., Aston Martin, Bentley, BMW, and Mercedes-AMG) and 0 otherwise (e.g., Audi, Ferrari, Lamborghini, McLaren, and Porsche). Of the 512 total observations, 204 (39.84%) are front engine cars while 308 (60.16%) are not.

The second column of Table 2 shows endurance races are statistically different from sprint races ($p < 0.01$) in accounting for final position when controlling for qualifying position. The positive coefficient reveals final positions in endurance races are approximately two positions *higher* (i.e., worse – higher meaning further from first place) when compared to sprint races and controlling for qualifying position. In other words, the results suggest that, on average in 2021, a car qualifying 10th overall would finish a sprint race in 11th position but finish an endurance race in 13th position (when including the constant/y-intercept value in the calculations). If I disregard the constant value (which is meaningless – one cannot qualify in “zero”-th place), then qualifying in 10th would produce, on average, a 6th position finish in an endurance race and a 4th position finish in a sprint race. Furthermore, the striking finding is that even with the inclusion of the binary “Endurance” variable only improves the R-squared value by one percentage point – only 36.8% of the

Table 2: OLS Regression Results with Robust Standard Errors

	<i>Dependent variable:</i>		
	Final Position		
	(1)	(2)	(3)
Qualifying Position	0.437*** (0.042)	0.473*** (0.036)	0.440*** (0.042)
Endurance	2.116** (0.949)		2.128** (0.947)
Front Engine		1.369* (0.730)	1.385* (0.722)
Constant	6.848*** (0.756)	6.535*** (0.886)	6.232*** (0.822)
Observations	436	436	436
R ²	0.368	0.362	0.373
Adjusted R ²	0.365	0.359	0.368
Residual Std. Error	7.675 (df = 433)	7.707 (df = 433)	7.654 (df = 432)
F Statistic	125.909*** (df = 2; 433)	123.075*** (df = 2; 433)	85.531*** (df = 3; 432)

Note: Robust SEs in parentheses

*p<0.1; **p<0.05; ***p<0.01

variance in final position is explained by qualifying position and race duration.

The third column in Table 2 removes the endurance variable and replaces it with whether or not a car has a front engine. Perhaps mid-engine cars are advantaged in some way and have better final positions compared to front engine cars. (This should not occur given the SRO's balance of performance criteria, but one could posit this idea.) The results show front engine cars do *marginally* worse by approximately one position ($p < 0.1$) than non-front engine cars when controlling for qualifying position. However, like the finding above, the addition of this binary variable only improves the R-squared by 0.4 percentage points.

The final column in Table 2 presents the OLS regression results when including all three variables: qualifying position, endurance vs. sprint, and front engine vs. other. All three variables have comparable coefficient sizes to the other models and their statistical significance remains at the same levels. The R-squared value reveals an increase to 37.3% of the variance in final position being explained by the three independent variables. However, nearly 63% of the variance remains unexplained.

These initial results are intriguing. Qualifying position, as expected, does significantly account for drivers' final positions. Yet, the relationship is not as 1:1 as one might expect. The central (and, perhaps, obvious) finding is that merely finishing a race tends to provide a team with a higher finishing position than where they started. This is in large part due to the attrition associated with motorsports.

However, endurance races, which produce greater attrition due to their duration, surprisingly lead to worse finishing results for entrants than sprint races when controlling for qualifying position. Another important finding is that front engine cars do marginally worse compared to mid- and rear engine cars. This evidence suggests the balance of performance may need adjustment; however, there are other factors that I would like to explore in future blog posts before I stake this claim. (An interaction model not shown reveals no statistically significant results when interacting engine type and race format.) Perhaps front engine cars are less likely to finish races compared to other car designs. Or, there are more Pro-Cup drivers in mid-engine cars compared to front engine cars. These are ideas I will explore further.

Predicted Probabilities of Finishing Better than Qualifying Position

As a final exploration of the data, I look at the predicted probabilities that a car/team finishes better than they started. I create a new variable that is the qualifying position minus final position. Positive numbers reveal observations that finish better than their starting position. I recode this variable as a binary variable: 1 for any positive or zero value (i.e., finishing same position as start) or 0 for all negative values. The switch to a binary outcome variable means I model the data using logistic regression instead of a linear probability model (i.e., an OLS model with a binary dependent variable). The independent variables consist of the following: front engine, endurance, and driver category. Driver category is a three level variable that characterizes the experience level of a driving lineup. It takes on the value 3 for “Pro Cup” entrants (the professional drivers), 2 for “Silver Cup” (aspiring professionals) lineups, and 1 for “Pro-AM Cup” (professional and amateur combination) entrants. Only two observations were in the “AM Cup” category, so I dropped these observations going in the following analyses.

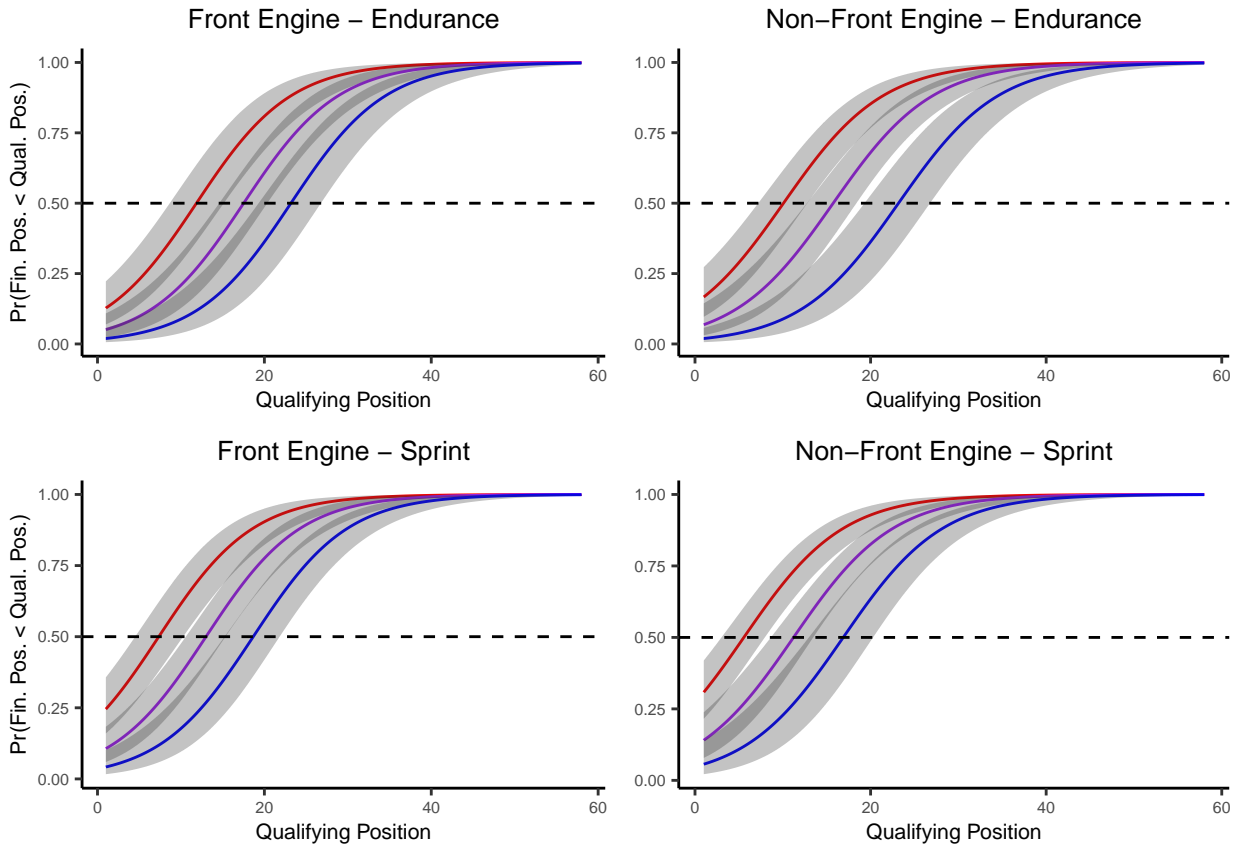
The figures below show the predicted probabilities of finishing stronger than a car/team started. The addition to this model is that I plot the probabilities by driver category. The “Pro Cup” entrants are always the red line, the “Silver Cup” drivers are always purple, and the “Pro-AM” lineups are always blue. I complete the predicted probabilities across four models:

- Top left: Front engine cars in endurance races
- Top right: Non-front engine cars in endurance races
- Bottom left: Front engine cars in sprint races
- Bottom right: Non-front engine cars in sprint races

The x-axis shows the range of possible starting (qualifying) positions and plots the probability of a 1 (i.e., doing better) on the y-axis. The 95% confidence intervals are shown in light gray. I discuss the four figures below.

“Pro Cup” drivers (red line), as expected, qualify better than the other two categories. And “Silver Cup” drivers tend to qualify better than “Pro-AM” lineups. Intriguingly, across all the configurations, if a car/team starts around 40th position, there is close to a guarantee they finish better than their starting position – regardless of engine layout or race format. Looking at the top left figure, a “Pro-AM” lineup in a front engine car has a ~50% chance of finishing better than their qualifying position if they start in 22nd position during an endurance race. This predicted probability drastically decreases across all the figures as a team qualifies closer to pole which makes sense given there are only so many places one can move up when starting further to the front of the grid.

Another illustrative example can be done with the top right plot (but extends to all plots). If we take 20th as the starting position, then “Pro-AM” teams have a ~35% probability of improving on their position. “Silver Cup” lineups, on the other hand, have a ~60% probability of improving. Finally, “Pro Cup” contenders have an ~80% probability of improving. The same type of relationship can be extended to all the other plots. For a given qualifying position, more experienced driver lineups tend to do better than less experienced teams.



Finally, the lines are shifted left along the x-axis in the bottom plots compared to the top plots. This

is evidence that sprint races are more favorable for moving up - regardless of driver category - compared to endurance races. This is, admittedly, a surprising finding since one expects longer races to allow teams to improve on their initial position. I will explore this finding in more detail in a future blog post.

Conclusion

I have many other interesting research questions on this topic. My next steps will be to look at how likely a team is to reach the Top 10 and the Top 3 if they start outside of those thresholds. I will also model these data with a Heckman selection model to account for the fact that nearly 15% of the observations are dropped due to attrition. Perhaps certain factors like race format and car manufacturer affects the likelihood a team finishes a race. I would also like to collect data from the previous years of the GT World Challenge Europe as well as its other regions (America, Asia, Australia, and Intercontinental). There are many exciting possibilities.

The results from this project follow my preconceived notions to some degree but go against them in others. Yes, qualifying matters. However, it is from determining final race position. Contrary to my expectations, though, is the finding that endurance races do not provide teams with a better finishing position compared to sprint races. The panel of plots shows that driver category matters for final positions, but qualifying at the end of the grid is not necessarily where a team will finish if they are consistent and finish the race.

Thank you for reading, and thank you for any feedback and opinions. I look forward to my next blog post. :)

Links

- GT World Challenge Europe: <https://www.gt-world-challenge-europe.com/results>
- Kunos Simulazioni: <http://www.kunos-simulazioni.com/main/>
- Assetto Corsa Competizione: <https://www.assettocorsa.it/competizione/>
- GTWorld: <https://www.youtube.com/gtworld>