

Statistical Measurement of F1 Fan Attention Dynamics: Competitive Performance, Media Exposure, and Social Interaction

Lingrui Zhou

*High School Affiliated to Nanjing Normal University Jiangning Campus, Nanjing, China
lingrui_zhou2009@outlook.com*

Abstract. With the expansion of F1 racing across social media, search platforms, and streaming media, driver attention has gradually become a continuously observable data point. Existing research typically uses Google Trends, fan size, fan growth, and engagement rate to measure changes in attention, but these indicators correspond to search attention, cumulative audience, and immediate engagement, respectively, and their statistical meanings are not interchangeable. Regarding driving factors such as competitive performance, race suspense, media exposure, and social interaction, relevant literature employs methods such as linear regression, count regression, panel models, event studies, and nonparametric tests. Existing evidence shows that F1 fan attention exhibits time dependence, long-tail distribution, cardinality effects, and platform heterogeneity, and there is also endogeneity between media exposure and competitive performance. Statistical analysis requires first distinguishing between Search, Growth, and Engagement, then selecting a model based on the distribution of the dependent variable and the data structure, and controlling for individual differences and time impacts through fixed effects, event windows, and robust standard errors. This establishes an analytical path from indicator measurement and driving factors to dynamic identification, providing a statistical foundation for the quantitative comparison of F1 fan attention and subsequent driver-level panel analysis.

Keywords: Fan Attention, Mathematical Statistics, Social Media, Panel Data, Event Study

1. Introduction

The competitive performance of F1 drivers is not entirely synchronized with fan attention. Points, podium finishes, race wins, and qualifying results directly describe track performance, but sustained fan attention to drivers and the race is also influenced by factors such as social media content, documentary narratives, race controversies, team promotion, and platform usage habits. As race broadcasting shifts from television to social media, short videos, and streaming media, fan behavior leaves more continuous digital traces. Race results can be represented as discrete or continuous variables, likes, comments, and new followers are count data, Google Trends forms a standardized time series, and drivers and weeks can also form longitudinal panel data. This data transforms fan

attention from a dissemination phenomenon into a measurable and comparable statistical object, and also makes variable definition, distribution characteristics, and time structure the basis for the reliability of the analysis results.

Existing F1 literature has quantified fan behavior from different perspectives. Garcia-del-Barrio and Reade used Google Trends to examine search interest before and after the championship was decided, and the results showed that there was a statistical relationship between race uncertainty and search attention [1]; Palermo and Delaine compared the interaction rates of F1 teams on Facebook and Instagram in the 2021 season and found that there was no stable one-to-one correspondence between competitive ranking and social interaction [2]; Wood and Burkhalter's analysis of F1's digital transformation showed that social platforms, YouTube, podcasts, F1 TV and streaming content have jointly expanded the channels for fans to access races and drivers [3]. In terms of statistical measures, attention, popularity, engagement and follower growth correspond to different behavioral processes, so search popularity cannot be directly equated with fan size, and fan growth cannot be replaced by interaction rate.

Regarding this measurement difference, statistical analysis needs to first distinguish between search attention, fan growth and social interaction, and then discuss the role of competitive performance, media exposure and content dissemination on various outcome variables. Further consideration of the excessive dispersion of count data, the cardinality difference between drivers, the correlation of continuous observation of the same driver, and the endogeneity between media exposure and competitive performance are necessary for model selection to be consistent with the data structure. Therefore, this paper discusses the indicators of interest and their applicable methods, the main driving factors, dynamic statistical models and identification limitations in turn. When direct evidence from F1 is insufficient, it introduces methodological research from adjacent sports fields to clarify the scope of conclusions that different statistical methods can support.

2. Statistical measurement of F1 fan attention

2.1. Definition of indicators and data types

Fan attention can be divided into three dimensions: search attention, fan size and growth, and social interaction. Google Trends normalizes search data by time and region and scales it to 0-100 to represent relative search popularity rather than absolute search volume; when comparing across windows, the search conditions should be kept consistent. Fan size reflects the accumulated audience, and new fans are more sensitive to championships, transfers, or media events. Social interaction mainly includes likes, comments, and reposts, which can measure the participation of existing audiences. The same keyword will be renormalized in different search time ranges. If the downloaded sequences are directly spliced, they may produce inconsistent scales. Therefore, long-term analysis still needs to unify the time range or perform scale calibration. The Shanghai F1 race study also shows that service experience and perceived value affect fan participation [4], indicating that online data cannot cover all fan behaviors. Table 1 shows the main statistical indicators and method selection of F1 fan attention.

Table 1. Main statistical indicators and method selection of F1 fan attention

Measurement Dimension	Observable Indicator	Data Type	Suitable Methods	Main Risk
Search Attention	Google Trends	0–100 relative index series	Time-Series Analysis, Event Study	Comparability Across Time Windows

Table 1. (continued)

Follower Base	Number of Account Followers	Non-negative, Right-Skewed	Log-Transformation Models / Count Models	Historical Base Effect
Follower Growth	Daily/Weekly New Followers	Longitudinal Count Data	Panel Data Models, Event Study	Autocorrelation
Social Engagement	Likes, Comments, and Shares	Long-Tailed Count Data	Poisson / Negative Binomial Regression	Overdispersion
Relative Engagement	Engagement / Number of Followers	Ratio Data	Generalized Linear Models, Robust Regression	Inflation Due to Small Follower Base

2.2. Interaction rate and cross-platform comparability

Differences in account size can be controlled through relative interaction rates. Let driver i receive L likes, C comments, and S shares at time t , and have F followers. Then the interaction rate is:

$$ER_{it} = \frac{L_{it} + C_{it} + S_{it}}{F_{it}} \times 100\% \quad (1)$$

Interaction rate is convenient for cross-account comparisons, but smaller accounts are easily amplified by the denominator, and the participation costs of likes, comments, and reposts differ. Interaction rates may also exceed 1, therefore Beta regression, which only applies to continuous proportions in the (0,1) interval, cannot be directly applied; proportions containing 0 or 1 can be considered using a fractional response model, while those exceeding this interval should be handled by a transformation or generalized linear model based on the actual distribution. Search engagement, follower growth, and interaction rate should be treated as separate outcome variables. When using interaction count as the dependent variable, follower size can also be used as a control variable, or treated as exposure in a suitable counting model, thereby reducing the instability caused by simply calculating the ratio. Media mentions are exposure variables; if they are included in both the overall popularity index and explanatory variables, a mechanical correlation will occur. Cross-platform analysis should also control for platform differences, as the audiences and interaction mechanisms of Instagram, X, TikTok, and Google are not the same. The robustness of the same conclusion is generally higher if it is consistent across multiple platforms than with a single-platform result.

3. Main driving factors of F1 fan attention

3.1. Competitive performance and race suspense

Competitive factors include overall race suspense and individual driver performance. Garcia-del-Barrio and Reade used Google Trends Web data from 2004 to 2017 and News data from 2008 to 2017 to incorporate the determination of the champion, race week, and season phase into the regression. The results showed that overall search interest decreased after the suspense of the annual champion disappeared [1]. This result supports the relationship between event uncertainty and search attention, but it cannot explain the long-term fan growth of individual drivers. At the driver level, points changes, race rankings, podium finishes, and relative teammate performance can be used. Different indicators reflect different competitive information. For example, rankings emphasize single-race results, while points changes are more suitable for describing the cumulative

state of the season. Such variables are highly correlated and collinearity should be controlled by correlation matrix, variance inflation factor, or dimensionality reduction. If weekly data is used, the current period's results can be compared with the lagged period to determine whether there is a delay in the change of attention. Adjacent sports research can supplement the model design. Doyle et al. analyzed the Instagram content of professional football players and found a statistical relationship between competitive performance content and interaction rate [5]. This result suggests that actual game results and competitive content on social media should be encoded separately to avoid confusing performance with content exposure. If both types of variables are included in the model, the effect of the results themselves and the additional effect of the results being disseminated can be further compared, rather than attributing all changes in attention to competitive performance.

3.2. Media exposure and content dissemination

The digital expansion of F1 has made media exposure an independent explanatory factor. Related studies have summarized the changes after 2017 as a fan contact system composed of social media, videos, podcasts, and streaming media [3]. Exposure can be quantified as mentions on official accounts, video appearances, news reports, documentary narrative centrality, or content release frequency. Drive to Survive studies show that documentaries can help some viewers form or strengthen F1 fan connections [6], and streaming narratives can also change visibility distribution through character portrayal [7]. These qualitative evidences indicate that the exposure mechanism exists, but cannot directly give a fan growth effect. Quantitative analysis can align the number of reports, appearance duration, or mentions on official accounts with the driver's weekly fan growth, while controlling for race results, to avoid mistaking the exposure brought about by competitive performance as an independent media effect. Longitudinal studies of tennis events use fan change data and wild-cluster bootstrap standard error tests to test growth factors [8], which can provide a methodological reference for F1 to track driver fan changes weekly.

3.3. Social interaction and group differences

Social interaction usually has a long tail distribution, and top drivers or controversial events may produce extreme values. Lick et al. studied the participation of female F1 fans through online questionnaires and X platform content analysis and found that there are differences in participation methods and content interests between genders [9]. Therefore, gender, region, new and old fans and content categories can be used as sources of heterogeneity; when the data is severely skewed, nonparametric tests such as Kruskal-Wallis can be used to compare the distribution between groups. Le Clue discussed controversial narratives and online participation around #VoidLap58 after the 2021 Abu Dhabi Grand Prix [10]. Related phenomena can be further quantified as the number of posts, the proportion of sentiment or the reply network, and structural changes can be identified through the change point or interruption time series. Social platforms take active users as the main observation objects, and their results should not be directly extrapolated to all F1 fans. Text sentiment, forwarding networks, and interaction volume also represent different levels of group behavior. An increase in the proportion of extreme sentiments does not necessarily mean that the network structure has become polarized; these two should be measured separately before joint interpretation.

4. Dynamic statistical analysis of fan attention

4.1. Count regression and panel model

New fans, likes, and comments are usually non-negative integers and may exhibit excessive dispersion. The counting model for F1 drivers can be expressed as:

$$Y_{it} \sim NB(\mu_{it}, \theta), \log \mu_{it} = \beta_0 + \beta_1 \text{Perf}_{it} + \beta_2 \text{Media}_{it} + \beta_3 \text{Content}_{it} + \alpha_i + \tau_t \quad (2)$$

Y represents the number of new followers or interactions. Perf, Media, and Content represent competitive performance, media exposure, and content variables, respectively. α and τ are used to control for driver differences and shared time impact. The Poisson model requires a strong relationship between the conditional mean and variance. Social media counts often exhibit excessive dispersion due to driver differences and event impacts; in such cases, a negative binomial model is usually more suitable. A zero-inflation model should be tested when zero values are abnormally concentrated. Model evaluation should combine likelihood, AIC/BIC, residual, and excessive dispersion tests, avoiding comparisons based solely on goodness of fit. Continuous growth rates or logarithmic indicators can be analyzed using fixed-effects panel regression, with cluster robustness standard errors used to handle the correlation of consecutive observations of the same driver.

4.2. Event impact and time dependence

Winning championships, accidents, penalties, transfers, and documentary releases all have clear time points, allowing for the establishment of event windows to compare dynamic changes before and after their occurrence. For multiple drivers and seasons, driver and time fixed effects can be incorporated to control for stable individual differences and shared race impacts. Event studies should also observe trends before the impact; if a sustained increase or decrease already exists before the event, subsequent differences cannot be entirely attributed to that event. When fan engagement exhibits inertia, a lagged dependent variable can be added:

$$Y_{it} = \rho Y_{i,t-1} + \beta X_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (3)$$

When the time dimension is short, the simultaneous inclusion of lagged dependent variables and fixed effects in the model may introduce dynamic fixed effects bias. In such cases, a dynamic estimation method suitable for short panels should be used, or the stability of the results should be tested by adding a time dimension. Ordinary fixed effects estimation is not suitable for direct unbiased interpretation. For high-frequency data within the same season, the race cycle and overall platform traffic changes should also be examined to avoid time trends being included in the event coefficients.

5. Statistical identification and comprehensive analysis

The main limitations of F1 fan statistics are related to measurement error, endogeneity and cardinality effects. Each platform serves a specific audience and findings from one platform are only able to account for the sample of that platform. Covered drivers with better performance will be

more likely to get coverage and popular drivers will be more likely to get additional content resources. Thus, in ordinary regression the exposure coefficient of the media is likely to represent conditional correlation. The fullness of causal explanations can only be achieved with the presence of credible exogenous shocks, instrumental variables, or controlled designs. The cardinality of fans is also a big influence; 100,000 more fans has different implications for top accounts vs small accounts. As a result, absolute growth should be used in conjunction with relative growth. Variable selection, regularization or principal component analysis may be used to handle the case of collinearity between competitive indicators. In addition, for time series data, autocorrelation, seasonality, and policy changes on the platform must be explored. Control variables and/or confidence intervals should be well motivated, and exogenous assumptions or sources of identification should be clearly explained, if causal judgments are being made.

Given the evidence, it would be more helpful to think of this as a sequence, beginning with a distinction between Search, Growth, and Engagement, then adding explanatory variables (Performance, Media, Content, Team), and then fixed effects, event windows, lagged terms, and differences between platforms. Including different outcome variables in models to reflect the domain and distribution of each outcome variable decreases the conceptual confusion, and helps to compare the effects of competitive performances and media exposure across different attention dimensions. Search indices can be used to track fluctuations in attention; growth in followers is appropriate to look at the accumulation process; and numbers of interactions is closer to measuring immediate responses of existing users: Only when these three types of outcomes converge can there be a better basis to discuss changes in attention that are relatively stable. The representative evidence is set out in Table 2 together with its statistical role.

Table 2. Representative evidence and its statistical role

Evidence Type	Representative Literature	Main Data / Methods	Statistical Role
Direct F1 Evidence	Garcia-del-Barrio & Reade [1]	Google Trends, Regression Analysis	Relationship Between Race Uncertainty and Search Attention
Direct F1 Evidence	Palermo & Delaine [2]	Team Social Media Data, Engagement Rate	Distinguishing Competitive Performance from Fan Engagement
Direct F1 Evidence	Lick et al. [9]	Surveys, Content Analysis	Differences Across Fan Groups and Content
Direct F1 Evidence	Le Clue [10]	Controversial Events, Online Discourse	Measurement of Event Impacts
Methods from Related Sports Research	Doyle et al. [5]	ANOVA, Hierarchical Regression	Relationship Between Performance-Related Content and Engagement
Methods from Related Sports Research	Bredikhina et al. [8]	Longitudinal Regression, Wild-Cluster Bootstrap	Follower Growth During Event Periods

6. Conclusion

The problem of fan engagement can be turned into a problem of statistics with well specified variables, data structures and estimation conditions. Attention, accumulation and participation are captured in the measures of search popularity, fan base size and fan growth, respectively. There is no one metric that can capture the increase or decrease in popularity of a driver or race. Some of the fluctuations in short-term performance can be attributed to competitive performance, race suspense;

digital media, streaming narratives and social content will serve as avenues for exposure outside the racetrack. In reality, both types of factors may change together in the data and the correlation and causality should be carefully differentiated. Analysis of the effects on the driver level should be based on continuous longitudinal data, depending on the domain of support of the outcome variables, panel regression, counting models or event studies should be used, controlling for individual differences, common time impact, cardinal effects and platform heterogeneity. For impacts like controversial events, documentary releases and major achievements, pre-event trends and attention inertia should also be investigated. A more powerful analytical approach is to estimate the three components of Search, Growth and Engagement separately and then compare direction and magnitude of the effects of various driving factors within each component to make the measurement of F1 fan dynamics more comparable and statistically explanatory.

References

- [1] Garcia-del-Barrio, P., & Reade, J. J. (2022). Does certainty on the winner diminish the interest in sport competitions? The case of formula one. *Empirical Economics*, 63(2), 1059–1079.
- [2] Palermo, A., & Delaine, G. (2023). Formula One on social networks: A successful strategy. *Zer*, 28(54), 13–29.
- [3] Wood, N. T., & Burkhalter, J. N. (2023). A formula for success: How Formula One racing embraced digital and social media to engage fans. *Journal of Digital & Social Media Marketing*, 11(1), 43–59.
- [4] Jones, C. W., Byon, K. K., & Huang, H. (2019). Service quality, perceived value, and fan engagement: Case of Shanghai Formula One racing. *Sport Marketing Quarterly*, 28(2), 63–76.
- [5] Doyle, J. P., Su, Y., & Kunkel, T. (2022). Athlete branding via social media: Examining the factors influencing consumer engagement on Instagram. *European Sport Management Quarterly*, 22(4), 506–526.
- [6] Shah, D., & Williams, A. L. (2025). "Drive to survive" drives new fans to formula 1? Studying viewer experiences of a sports documentary and its influence on the sport's fandom. *Communication & Sport*, 13(6), 1213–1234.
- [7] Campbell, D. J., & Davis, L. (2026). Drive to survive and symbolic capital in Formula One. *International Review for the Sociology of Sport*, 10126902261448900.
- [8] Bredikhina, N., Gupta, K., & Kunkel, T. (2023). Superboosting the athlete social media brand: Events as an opportunity for follower growth. *European Sport Management Quarterly*, 23(6), 1819–1842.
- [9] Lick, E., Bakirov, R., & Ramjaun, T. A. (2024). Female motorsport fan engagement on social media-based brand communities. *Journal of Digital & Social Media Marketing*, 12(1), 28–51.
- [10] Le Clue, N. (2025). Controversy, social media, and formula one: Examining# VoidLap58. *Transformative Works and Culture*, 45.