



Navigation

- Home
- General Information
- Call For Papers
- Editorial Board
- Current Issue
 - Volume 12 Number 3, 2009
- Archives
- The Sport Digest

Online Degree Programs

- Doctoral Degree
- Master's Degree
- Bachelor's Degree
- Continuing Education
 - Certificates
 - Certification Programs

[Home](#)

The Role of Driver Experience in Predicting the Outcome of NASCAR Races: An Empirical Analysis

ISSN: 1543-9518

Submitted by: Mary Allender - Pamplin School of Business - University of Portland

Abstract

As national interest in NASCAR grows, the field of sports economics is increasingly addressing various aspects of this sporting contest. The outcome of NASCAR races are of particular interest to fans, and, thus, models describing and predicting the outcome of NASCAR races are beginning to emerge. This paper builds a model predicting the outcome of NASCAR races using NASCAR data. Various forms of regression analysis were used as the methodology for this research. The outcome was hypothesized to depend on a set of variables and focused, in particular, on the importance of driver experience. The findings of this paper conclude that a driver's years of experience do in fact play a significant role in predicting the outcome of NASCAR races.

Introduction

NASCAR is one of the fastest growing sports in the world. It generates 3 billion dollars a year in GDP and adds new fans to its loyal fan base each year. The academic study of NASCAR is in its infancy, and this paper seeks to add to that small but growing body of literature. The origins of NASCAR reach back to the days of prohibition, when the cars used by bootleggers needed speed while making delivery runs to avoid the authorities in pursuit. More horsepower was needed, and so began the quest to modify cars for more horsepower and reliability. Simultaneously, auto racing became a sport. The inaugural auto race at Daytona Beach took place on March 8, 1936 (Felden, 2005).

These early races, however, were not officially organized, and so races were haphazard and drivers tended to show up randomly. The original tracks often consisted of dirt or sand. Fans were few in numbers, thus driving stock cars remained a hobby, since it didn't generate enough income to qualify as a job.

Over the next ten years, fan interest increased considerably, and stock car racing evolved from an occasional, hastily organized race on sand and dirt tracks to the stadiums and paved tracks we know today. In December 1947, Bill France Sr., both a driver and race promoter, developed the idea of NASCAR as organized stock car racing subject to specific rules. On February 15, 1948, NASCAR ran its first race at the Daytona Beach road course. The Daytona 500 remains the premier NASCAR race today. This paper proceeds in section II to discuss current research. Section III discusses data and methodology, while section IV discusses our empirical models and estimation methods. Section V discusses the findings of our analysis, and section VI offers concluding remarks.

Current Research

Scholarly research on NASCAR as a sport is relatively new and has taken many different directions. One avenue of research focuses on the reliability of NASCAR vehicles and explores the reasons behind part failure and the extent to which these critical part failures can be reduced. Majety, Dawande, and Rajgopal (1999) show that in general, the typical reliability allocation problem maximizes system reliability subject to a budget constraint. They note that cost is an increasing function of reliability and hence the tradeoff between dollars spent and system reliability. Although the media would have us believe that NASCAR owners are willing to spend virtually unlimited amounts of money to earn a spot in Victory Lane (New York Times, 2/13/06; CBS News, 10/6/05), NASCAR teams themselves acknowledge that in fact, a budget constraint does exist both in the form of willingness to spend money and the rules imposed on the construction of the vehicles themselves; although budgets in NASCAR racing are far more substantial than those common to commercially produced vehicles (Wachtel, 2006. Allender (2007), there continues with the reliability question, asking whether or not critical part failures in NASCAR vehicles are higher than what are expected and exploring some reasons as to why in fact they are.

Other lines of research focus on the type of tournament NASCAR represents and the most efficient type of reward structure for rank order tournaments (ROT), where finish position is all that matters to getting a prize. Becker and Harold (1992), Lynch and Zax (2000), and Maloney and McCormick (2000) use ROT theory to investigate the effect of different types of payment structures on the performance of contestants. Along similar lines, Lazear and Rosen (1981), Nabeluff and Stiglitz (1983), and O'Keefe, Viscusi, and Zeckhauser (1984) began to look seriously at a payoff structure that was preferable for the contest organizer. In fact, it was this line of research that began to take the field of sports economics into the realm of serious economic literature Fizel (2006).

Fans of NASCAR are ultimately interested in the outcome of each contest or race. The Nextel Cup Champion for the year, in essence, wins the majority of the points associated with the 38 races NASCAR holds each year at different tracks. Before the season and before each race, popular media focuses much attention on predicting the winner of each race. However, there is little in the sports economics literature that attempts to develop models that help predict the outcome of a NASCAR race. Pfitzner and Rishel (2005) develop a model predicting order of finish in NASCAR races based on variables such as car speed, driver characteristics, and the like. Allender (2008) develops a one season multivariate model showing that driver experience, along with other variables, is a statistically significant variable in determining the winner of NASCAR races. This paper seeks to add to that burgeoning body of literature by developing an empirical model that identifies the most important variables contributing to a driver's success in a race. Thus, the model can be used as a tool in predicting the outcome of NASCAR races.

Data and Methodology

The pooled time series-cross sectional data set for the study spans the period 1990-2006. Each season consists of forty three cars and thirty eight races. The data were obtained from the NASCAR website. Our methodology utilizes regression analysis by estimating two slightly different models using weighted least squares. Our third model is a logistic regression model, which essentially converts the least squares model into a probabilistic regression model (Gujarati, 1992).

Empirical Models and Estimation Methods

The basic model to be estimated is described in equation (1). FP represents finish position and is the dependent variable. SP represents

starting position or pole position as determined during qualifying runs. We expect the sign on this explanatory variable to be positive. That is, the closer to the front the driver starts the race, the closer to the front he should be expected to finish. DY*SP represents the interaction between DY which is driver years of experience and starting position. We include this variable based on the theory that driver experience enhances the positive impact of starting position on finish position. Thus, the sign on this variable is expected to be positive. PC represents the percentage of laps under caution. Since caution laps freeze car position, we expect the sign on this explanatory variable to be positive since the more the caution flag comes out, the harder it is for cars coming from behind to make up laps. DY*TL represents the interaction term between driver years of experience and track length. We expect the sign on this variable to be negative. As the track length extends and works together with driver years of experience, we expect the driver, able to negotiate the various track lengths to move further toward the front.

Empirical Results

Initially, we test the model by estimating equation (1) using weighted least squares with driver years of experience used for weighting purposes. Table I reports these findings. Based on the t-statistics all model variables are statistically significant at the 1 percent level. Starting or pole position achieved during qualifying runs positively affects winning first place, which is what was expected. The interaction of variables SP and DY also show the right sign. The more experience in years a driver has, the higher his likelihood of winning. Therefore, as expected, the sign on this interaction variable is negative. The variable designated PC or percentage of laps under caution is showing a positive correlation to winning because while all drivers are affected by caution laps, our results show that more experienced drivers take advantage of this circumstance to take the lead.

Finally, the interaction of the variables DY and TL does not help a driver to advance to the top position. A possible explanation may be that on short tracks, more wrecks occur because more passing attempts are made on the curves, which are likely to eliminate, on a random basis the wrong driver, at the wrong time regardless of experience. More specifically, "bump drafting" as a strategy for passing on curves can be successful but depends not only on the experience of the driver attempting it, but also on the condition of the car being bumped which the driver attempting the maneuver has limited knowledge of.

Hence, we expect more randomness on short tracks.

Table 1

Dependant Variable: FP
 Method: Least Squares
 Date: 01/02/08
 Sample (adjusted): 1 21698
 Included observations: 21607 after adjustments
 Weighting series: DY

Variable	Coefficient	Std. Error	t-Statistic	Prob,
C	12.04722	0.278278	43.29199	0.0000
SP	0.386505	0.012668-	30.51051	0.0000
DY*SP	-0.002630	0.000578	-4.551522	0.0000
PC	3.433934	1.284601	2.673152	0.0075
DY*TL	0.001216	0.005525	0.219188	0.8290

Weighted Statistics			
R-squared	0.568955	Mean dependent var	20.73319
Adjusted R-squared	0.568875	S.D. dependent var	21.54092
S.E. of regression	14.14379	Akaike info criterion	8.136660
Sum squared resid	4321413.	Schwarz criterion	8.138507
Log likelihood	-87899.41	F-statistic	725.3306
Durbin-Watson stat	0.988600	Prob(F-statistic)	0.000000

Unweighted Statistics			
R-squared	0.113075	Mean dependent var	21.36696
Adjusted R-squared	0.112911	S.D. dependent var	12.13023
S.E. of regression	11.42491	Sum squared resid	2819678.
Durbin-Watson stat	0.470986		

Table II reports findings of weighted least squares for equation (1) with the added variable total life time winnings of a driver in dollars. This variable is designated W. The rationale for adding driver winnings in dollars as an explanatory variable is that the winning teams and drivers enjoy added resources which improve the quality of equipment, team members, and cars. All of these factors are expected to push a driver to a winning position in future races. Therefore, one expects a negative coefficient sign for this variable. Table II suggests that this hypothesis is correct and statistically statically significant. The remaining findings in Table II are qualitatively identical to those of Table I and in the interest of brevity we do not replicate that analysis. The R squared statistic for the two variations on equation (1) hovers under 60 percent which isn't bad but suggests further research.

Table 2

Dependent Variable: FP
 Method: Least Squares
 Date: 01/02/08 Time: 16:07
 Sample (adjusted): 1 21698
 Included observations: 21607 after adjustments
 Weighting series: DY

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.46957	0.273212	49.30084	0.0000
SP	0.339250	0.012376	27.41171	0.0000
DY*SP	-0.001703	0.000562	-3.031302	0.0024

PC	11.44432	1.267665	9.027874	0.0000
DY*TL	0.071924	0.005486	13.11113	0.0000
WI	-4.63E-05	1.29E-06	-35.92230	0.0000

Weighted Statistics			
R-squared	0.593253	Mean dependent var	20.73319
Adjusted R-squared	0.593159	S.D. dependent var	21.54092
S.E. of regression	13.73968	Akaike info criterion	8.078731
Sum squared resid	4077810.	Schwarz criterion	8.080947
Log likelihood	-87272.57	F-statistic	872.9825
Durbin-Watson stat	0.950143	Prob(F-statistic)	0.000000

Unweighted Statistics			
R-squared	0.157913	Mean dependent var	21.36696
Adjusted R-squared	0.157718	S.D. dependent var	12.13023
S.E. of regression	11.13263	Sum squared resid	2677131.
Durbin-Watson stat	0.409164		

Table III reports estimation results of equation (2), that is, the logistic model. Based on the t-statistics, all variables in the model with the exception of the interaction variable TL*DY are statistically significant at the 1 per cent level. Variable SP shows that when its value is lower, the driver is starting further to the front, the higher the log of the odds of winning. This is as expected. Similarly, as before, the interaction of the variables PC and TL raises the log of the odds of winning. In contrast to our results of the weighted least squares estimation reported in Table I, the interaction of variables TL and DY turns out to be statistically insignificant in the logit model. This is unexpected and requires further investigation.

There is one possible explanation here, however. The tracks used in NASCAR range from three-quarter miles to two and a half miles in length with the vast majority being between 1 and 2 miles. In other words, there is so little variation in track length that the standard error on this explanatory variable is large. If you run track length as a stand alone explanatory variable, the t-statistic is low and makes track length an insignificant explanatory variable. In addition, the results show that more winnings in dollars for a driver, increases the log of the odds of winning races. However, the log likelihood number is a large negative number indicating that the model is a good overall fit.

Table 3

Dependent Variable: DUM1
Method: ML - Binary Logit (Quadratic hill climbing)
Date: 01/09/08 Time: 16:00

Date: 01/02/00 Time: 18:02
 Sample (adjusted): 1 21698
 Included observations: 21607 after adjustments
 Convergence achieved after 9 iterations
 Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.974485	0.116648	-25.49957	0.0000
SP	-0.089820	0.005735	-15.66153	0.0000
PC*TL	-4.204015	0.489954	-8.580419	0.0000
TL*DY	0.000310	0.003908	0.079394	0.9367
WI	1.51E-05	6.14E-07	24.54201	0.0000

Mean dependent var	0.024020	S.D. dependent var	0.153115
S.E. of regression	0.139077	Akaike info criterion	0.160929
Sum squared resid	417.8377	Schwarz criterion	0.162776
Log likelihood	-1733.598	Hannan-Quinn criter.	0.161531
Restr. log likelihood	-2447.999	Avg. log likelihood	0.080233
LR statistic (4 df)	1428.803	McFadden R-squared	0.291831
Probability(LR stat)	0.000000		

Obs with Dep=0	21088	Total obs	21607
Obs with Dep=1	519		

Conclusion

This paper set out to develop an empirical model based on theoretical hypotheses to explain the finish position of drivers in NASCAR races. The model clearly identifies the most important variables that explain the finish position of each driver. This paper utilizes both a weighted least squares model and a logistic model to test our hypotheses regarding the variables most likely to influence the finish position of drivers in NASCAR races. These models produce promising results as demonstrated by the t-statistics and the R squared statistics.

This paper offers suggestions for further research. In order to improve R^2 , it may be advisable to explore the option of including additional explanatory variables. Another avenue worth exploring is how best to frame and utilize the variable associated with caution laps. Theoretically, the number of laps under caution is totally unpredictable prior to each race. Or is it? Are there some races that involve more crashes and hence caution laps than others? If that is not the case, then the randomness of caution laps would be picked up in the error term and contribute to a lower R^2 . On the other hand, again theoretically, the number of caution laps that occur during a race should have a significant effect on the outcome because caution laps allow for pit stops that give the crew time to make adjustments, add gasoline, and change tires, all of which should affect finish position. The broader question here is that the randomness

factor plays a great role in NASCAR as a rank order tournament than it does in other rank order tournaments such as track and field.

References

Allender, Mary (2007). Are there a higher than expected number of early life critical part failures in NASCAR vehicles? A reliability Study. *The Sport Journal*, 25(1).

Allender, Mary (2008, May). Predicting the outcome of NASCAR races: The role of driver experience, *Journal of Business and Economic Research*.

Becker, B. E., & Harold, M. A. (1992). The incentive effects of tournament compensation schemes. *Administrative Science Quarterly*, 37, 336-350.

Depken, C. A., & Wilson, D. P. (2004). The efficiency of the NASCAR racing system: Initial empirical evidence. *Journal of Sports Economics*, 5(4), 371-386.

Gujarati, D. (2006). Essentials of econometrics. In Fizek, John (Ed.), *Handbook of Sports Economics Research*. London: M. E. Sharpe.

Lazear, E. P. & Rosen, S. (1981). Rank order tournaments as optimal labor contracts. *Journal of Political Economy*, 89(5), 841-864.

Majety, S. R., Dawande, M., & Rajgopal, J. (1999). Optimal reliability allocation with discrete cost-reliability data for component." *Operations Research*, 47.6.

Maloney, M. T., & McCormick, R. E. (2000). The response of workers to wages in tournaments. *Journal of Sports Economics*, 1(2), 99-123.

Martin, M. (2005). *NASCAR for dummies*. (2nd ed.). Hoboken, NJ: Wiley & Sons.

Nalebuff, B. J., & Stiglitz, J. E. (1983, Spring). Prizes and incentives: Towards a general theory of compensation and competition. *Bell Journal of Economics*, 14, 21-34.

O'Keefe, M., Viscusi, K., & Zeckhauser, R. (1984). Economic contests: Comparative reward schemes. *Journal of Labor Economics*, 2(1): 27-56.

Pfzner, B., & Rishel, T. (2005). Do reliable predictors exist for the outcomes of NASCAR races? *The Sport Journal*, 8(2).

Von Allmen, P. (2001). Is the reward system in NASCAR efficient? *Journal of Sports Economics*, 2(1), 62-79.

Author's Note

Mary Allender, Pamplin School of Business, University of Portland, Oregon.

Correspondence concerning this article should be addressed to Mary Allender, University of Portland, 5000 N. Willamette Boulevard, Portland, Oregon 97203. Email: allender@up.edu

Bookmark / Submit this article with:



Related content: [2009](#) [spring](#) [volume 12 number 2](#)

