 DOI 10.1515/jqas-2013-0031 — JQAS 2014; 10(2): 261–278

Andrew J. K. Phillips*

Uncovering Formula One driver performances from 1950 to 2013 by adjusting for team and competition effects

Abstract: Subjective ratings of the best drivers in the history of Formula One are common, but objective analyses are hampered by the difficulties involved in comparing drivers who raced for different teams and in different eras. Here, we present a new method for comparing performances within and between eras. Using a statistical model, we estimate driver and team contributions to performance, as well as the effects of competition with other drivers. By adjusting for team and competition effects, underlying driver performances are revealed. Using this method, we compute adjusted scoring rates for 1950–2013. Driver performances are then compared using: (i) peak performances for 1-year, 3-year, and 5-year intervals; and (ii) number of championships. Overall, these comparisons rank Clark, Stewart, Fangio, Alonso, and Schumacher as the five greatest drivers. We confirm the model's accuracy by comparing its performance predictions to 2010–2013 lap-time data. The results of the analysis are generally in good agreement with expert opinions regarding driver performances. However, the model also identifies several undervalued and overvalued driver performances, which are discussed. This is the first objective method for comparing Formula One drivers that has yielded sensible results. The model adds a valuable perspective to previous subjective analyses.

Keywords: Formula One; historical analysis; objective comparison; performance; statistical model.

***Corresponding author: Andrew J. K. Phillips,** Brigham and Women's Hospital, Harvard Medical School, Division of Sleep Medicine, 221 Longwood Ave, Boston, MA 02148, USA, Phone: +617-278-0057, e-mail: ajkphillips@gmail.com

1 Introduction

While motor racing has existed in many forms since the late 19th century, the Formula One World Drivers' Championship has represented the highest competitive level since its commencement in 1950 (Rendall 1991; Smith 2011). During this time, the sport has evolved considerably in its technology and regulations, but the basic rules of racing have remained relatively consistent over time (Jones 1995). There is frequent interest in comparing and ranking the best drivers from 1950 to the present day (F1 Racing 1997, 2004, 2008; Henry 2008; Autosport 2009). In this respect, Formula One presents an interesting challenge compared to most other sports, because drivers race against each other using unequal cars provided by different teams. Differences in performance between teams may be much larger than differences in performance between drivers (Jones 1995). Consequently, the winner of the World Drivers' Championship is not necessarily the objectively best performing driver in any given year. To accurately compare drivers within and between eras it is therefore necessary to apply statistical techniques that account for team performances.

Each Formula One season consists of an international series of races (Jones 1995). Drivers score points in each race based on their finishing positions. The driver with the most points at the end of the year is awarded the World Drivers' Championship title (Smith 2011). In each race, drivers must drive for a particular team. The team provides the car, performs pit-stops during the race, and performs maintenance and repairs (Jones 1995). While drivers are usually contracted to race for the same team throughout a season, drivers have occasionally switched teams during a season.

A driver's result in the championship relative to competitors in that year may be considered the combined result of two factors: driver performance and team performance. Simple statistics, such as number of championship titles or number of wins, fail to differentiate these two factors and are therefore unreliable metrics for comparing driver performances.

Team performance can be summarized in terms of two key factors: reliability and speed. Reliability is a measure of how frequently a team suffers failures that prevent its drivers from finishing races, e.g., mechanical faults. Speed is a relative measure of the ability of a team to finish ahead of other teams in a race where the car is not prevented from finishing the race by a failure. These two factors need not be correlated; a car can be reliable and slow, or unreliable and fast. In many cases, it has been argued that a driver's true skill has been masked by their driving an unreliable or slow car (Smith 2011; F1 Racing 2012). Our goal is to estimate and adjust for the effects of these two factors to reveal the underlying driver performances.

Separating differences in driver performance from differences in team performance is a challenging analytical task that has received very limited attention. Part of the difficulty lies in tailoring methods to address the unusual structure of motorsport data, which include scoring finishes, non-scoring finishes, and a variety of types of non-finishes. While statistical methods have been proposed for comparing performances between eras in other sports (e.g., Berry, Reese, and Larkey 1999), they are best suited to sports in which there are straightforward performance metrics and in which competitors interact on a level playing field, e.g., golf, tennis. In basketball and football, sophisticated models have been used for ranking teams (West and Lamsal 2008), predicting team performances (Kvam and Sokol 2006; West 2006; Brown and Sokol 2010), ranking individual players within teams (White and Berry 2012), and relating the contributions of individual players to team performances (Cohea and Payton 2011). Here, we develop a relatively simple model that is well suited to the data structure of Formula One.

Previous attempts to rank Formula One drivers have been largely based on qualitative metrics, such as expert opinions (F1 Racing 1997, 2004; Autosport 2009). Within these assessments, achievements in less competitive cars are often cited as a key marker of driver performance (F1 Racing 1997; Henry 2008), although they are not quantified in any way. To date, only one statistical model has been proposed for determining driver performance independent of team performance (Eichenberger and Stadelmann 2009). That analysis used a linear model of performance, including both driver and team effects, to accomplish the first objective estimates of driver performances. While this was a significant step towards objective comparisons, some of the model's predictions seem spurious. For example, the model rated Mike Hawthorn as the 5th greatest driver of all time, although he is widely considered to be one of the weakest champions and is universally ranked outside of the top 20 by experts and fans alike (F1 Racing 1997, 2004, 2008; Henry 2008; Autosport 2009). Other questionable results include: rating Erik Comas ahead of Ayrton Senna; rating Arturo Merzario ahead of Jack Brabham and Nelson Piquet; rating Harry Schell ahead of Bruce McLaren; rating Mark Blundell ahead of Niki Lauda; and rating Ukyo Katayama ahead of Nigel Mansell. We note that some of these anomalous results may be due to the model specification: finishing position was used as the performance metric rather than points scored. Consequently, the model applied a large penalty to drivers who failed to finish a race, and weighted differences between all finishing positions equally, despite their different point values. We argue that points are a more salient and reliable metric, since the ultimate objective of Formula One is to score the most points.

Here, we present a new quantitative method for estimating and comparing driver performances across the history of the Formula One World Drivers' Championship. This method uses points scoring as the comparison metric. Scoring rate is related to driver and team performances, as well as competition effects, by a nonlinear function. By fitting the parameters of this model to data from 1950 to 2013, we are able to adjust driver scoring rates to compensate for team and competition effects, providing an estimate of underlying driver performance. This allows us to compute adjusted championship results from 1950 to 2013 and compare drivers' relative performances between eras.

2 Methods

In this section, we describe the data used in this study. We then explain our data analysis methods and our statistical model. Finally, we explain how the model is applied to generate the results in this paper.

2.1 Data

For this paper, we analyzed results for every driver from every championship race in Formula One from 1950 to 2013. Data were sourced from Wikipedia, where Formula One race data are compiled in a central location from other available sources as part of the Formula One Portal (http://en.wikipedia.org/wiki/Portal:Formula_One). Results from driver pages were all checked against both race and season pages. All data were collected, scored, and reviewed by the author, who has a deep knowledge of the sport. In rare cases where data were unclear or appeared to be in error, results were checked against race reports at grandprix.com and against Smith (2011). For each driver, the following data were collected for our analysis:

1. Year and round of the championship.
2. Team for which the driver drove.
3. Finishing position if the driver finished the race, or type of non-finish if the driver did not finish the race.

We did not include races where a driver did not start, did not attend (for any reason, including a suspension), or did not qualify. In the rare event of a car being shared by two drivers, both drivers were awarded the position in which their shared car finished. In the very rare event of a driver sharing more than one car in a single race, they were awarded their highest finish.

Teams in each season were determined based on the chassis-engine combination; this is important because in many seasons the same chassis was paired with different engines (e.g., Lotus-Climax and Lotus-BRM). Examples of different teams are the 1964 Lotus-Climax, the 1964 Lotus-BRM, the 2010 Ferrari-Ferrari, and the 2010 Red Bull-Renault. Changes in the chassis number within a season and minor upgrades to car components were discounted (e.g., changes between the Wolf WR1, WR2, and WR3 within the same season). Including such changes could potentially increase model accuracy, but at the very real risk of model overspecification.

Non-finishes were classified as one of two types:

(i) *Non-driver failures*: Races in which the failure of a car component is listed as the reason for the non-finish, e.g., engine, transmission.

(ii) *Driver failures*: Races in which the driver spun off, crashed, or was involved in a collision with another driver.

While these simple definitions covered the vast majority of non-finishes, a few special cases had to be treated separately. These are described below.

Not classified: These are races in which a driver finished the race having covered <90% of the laps. In cases where a mechanical failure was listed, these were classified as non-driver failures. Otherwise, they were classified as driver failures.

Disqualified: In cases where the disqualification was due to an infringement on the part of the team (e.g., using illegal fuel), these were classified as non-driver failures. In cases where the disqualification was due to a driver error (e.g., receiving a push-start after spinning off the track), these were classified as driver failures.

Fuel: Races in which a driver ran out of fuel before the finish. In cases where the failure occurred more than five laps before the finish, these were classified as non-driver failures. In these cases, the team is considered at fault for supplying insufficient fuel or miscommunicating the rate of fuel consumption to the driver. In cases where the failure occurred within five laps of the finish, these were treated as driver failures. In these cases, the driver is considered at fault for failing to correctly manage fuel consumption. This approach was used because fuel management by the driver has been an important skill at several points in the history of Formula One, including the mid-1980s.

Handling: Races in which a driver withdrew due to the car having very poor or dangerous handling. These were classified as non-driver failures. We used this approach because, although a driver is partly responsible for car set-up, they cannot reasonably be blamed for a car that is completely undriveable. Also, in many cases the handling problem may be due to an underlying mechanical failure.

Physical: Races in which a driver withdrew due to illness or prior injury. These were classified as non-driver failures, since they were due to circumstances beyond the driver's control.

Withdrew: Races in which a driver withdrew for none of the above reasons, nor any technical problems, while other drivers continued (e.g., due to heavy rain). These were classified as driver failures, since they were purely down to the decision of the driver.

Retirement: In other very rare cases where a non-finish was simply listed as "Retirement" and no more information could be found, the result was scored as a driver failure.

2.2 Points calculations

The ultimate goal of Formula One is to score the most possible points in each season. We therefore used points as the basis for our performance metric. Across the history of Formula One, the rules for points scoring have varied slightly (Jones 1995). The number of scoring drivers has changed from five (1950–1959) to six (1960–2002) to eight (2003–2009) to ten (2010–2012). One point was once awarded for the fastest lap in each race (1950–1959). The number of races that count towards the championship has also changed from a subset (always more than 50%) of the races (1950–1990) to all of the races. However, the subset rule had no effect in most years of the championship (due to drivers not finishing many races); only in 1964 and 1988 did the subset rule determine the winner of the championship.

To enable direct comparisons between seasons, we retrospectively applied the points scoring method that was in use from 1991 to 2002 to all races, since this is a good compromise to the points systems used across the sport's history. The points allocated for finishes in each race are as follows:

1st: 10 points
2nd: 6 points

3rd: 4 points
4th: 3 points
5th: 2 points
6th: 1 point.

Retrospective scoring potentially introduces the problem of neglecting the actual state of the championship. For example, in years where only a subset of the races counted, a driver may have already won the championship before the last race of the season, meaning they are potentially not trying as hard in the remaining races. However, such cases are rare in practice. Additionally, Formula One drivers are highly competitive by nature and usually try to achieve the best possible finishing position regardless of the championship situation. We therefore do not consider this to be a major flaw in our analysis.

For finishing positions below 6th place, fractional points were awarded using the formula

$$p\text{th}: a^{p-6}.$$

A base of $a=(1/10)^{1/5}$ was chosen to match the change from 10 points for 1st place to 1 point for 6th place. This additional scoring for lower positions makes negligible difference to the scoring rates of drivers who regularly finished in the points, but allows the model to detect performance differences between drivers who scored no points (e.g., comparing drivers in the worst teams).

2.3 Team reliability correction

Our first method of correcting for team performance was to remove the effects of non-driver failures. All races that were scored as non-driver failures were therefore not counted towards a driver's career. All other races (i.e., finishes and driver failures) were counted towards a driver's career and these races are referred to as *counting races*. For each counting race, points were awarded according to the scoring system described above. For further analysis, we included only drivers who had completed at least three counting races in a single season. In total, this analysis included 342 drivers and 760 unique teams (i.e., chassis-engine combinations each pertaining to a particular season).

In making the above correction for team reliability, we assumed that all non-drivers failures are primarily due to factors beyond the drivers' control. In other words, we neglected the fact that some drivers may contribute to their own car problems through rough driving.

2.4 Statistical model

For each driver i in each team j in each season k, we computed the average scoring rate, s_{ijk}, by dividing the total number of points scored by the number of counting races. Note that a driver may in some cases compete for more than one team in the same season, with a different s_{ijk} for each.

As the basis for our statistical model, we assume that there is a performance variable y_{ijk} that is predictive of s_{ijk}. The simplest reasonable model is to assume that y_{ijk} is a linear function of driver, team, and competition effects on performance. Since s_{ijk} is theoretically bounded between 0 (never scoring points) and 10 (winning every race), we assume that s_{ijk} is a sigmoidal function of y_{ijk},

$$s_{ijk}=S(y_{ijk})=5(1+\mathrm{erf}(y_{ijk})), \tag{1}$$

where S is a sigmoidal function ranging from 0 to 10, $\mathrm{erf}(x)=(2/\sqrt{\pi})\int_0^x e^{-t^2}dt$ is the standard error function, and y_{ijk} is the linear predictor function. For each s_{ijk}, the value of y_{ijk} is computed by applying the inverse function,

$$y_{ijk}=S^{-1}(s_{ijk})=\mathrm{erf}^{-1}\left(\frac{s_{ijk}-5}{5}\right). \tag{2}$$

We propose a simple linear statistical model for the predictor,

$$y_{ijk}=\alpha_i+\beta_j-\delta_k+\varepsilon_{ijk}, \tag{3}$$

where α_i is a fixed effect representing driver performance, β_j is a fixed effect representing team performance, δ_k is a fixed effect representing difficulty scoring in season k due to competition with other drivers, and ε_{ijk} is a random effect representing variability in performance.

For each driver i in each season k, we define a weighted predictor by averaging across all teams j that the driver drove for in that season

$$y_{ik}=\sum_j w_{ijk}y_{ijk}\Big/\sum_j w_{ijk}, \tag{4}$$

where w_{ijk} is the number of counting races for that driver/team combination in that season.

2.5 Competition effects

Because there are only so many points available in each race, competition with other driver/team combinations makes it more difficult to score points. The level of competition is expected to vary from season to season, depending on both the number of other driver/team combinations and their speed. Competition effects on scoring rates are modeled by the term δ_k in Eq. (3) above. Years with higher levels of competition will have higher values of δ_k.

We can estimate the form of δ_k by considering the expected scoring rate of a population of N drivers with normally-distributed $\theta_{ij}=\alpha_i+\beta_j$, with mean μ and standard deviation σ. The expected total scoring rate for this population is

$$S_{\text{tot}}=N\int_{-\infty}^{\infty}S(y)P(y)dy=N\int_{-\infty}^{\infty}5(1+\text{erf}(y))\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y+\delta-\mu)^2}{2\sigma^2}}dy$$
$$=5N\left[\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y+\delta-\mu)^2}{2\sigma^2}}dy+\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\text{erf}(y-\delta+\mu)e^{-\frac{y^2}{2\sigma^2}}dy\right].\tag{5}$$

This can be evaluated using the identity $\int_{-\infty}^{\infty}\text{erf}(x+a)e^{-x^2/b}dx=\sqrt{\pi b}\,\text{erf}\left(a\sqrt{\frac{1}{1+b}}\right)$, to yield

$$S_{\text{tot}}=5N\left[1+\text{erf}\left((-\delta+\mu)\sqrt{\frac{1}{1+2\sigma^2}}\right)\right].\tag{6}$$

Now, we can use the fact that the total scoring rate for the whole population must be equal to the total number of points on offer per race (so long as $N\gtrsim 6$), $P_{\text{tot}}=10+6+4+3+2+1+\cdots=27.097$. Rearranging for δ yields

$$\delta=\mu-\sqrt{1+2\sigma^2}\,\text{erf}^{-1}\left(\frac{P_{\text{tot}}-5N}{5N}\right).\tag{7}$$

We use this function as our competition function, $\delta_k(\mu_k, \sigma_k, N_k)$.

The values of μ_k, σ_k, and N_k are computed by weighting each driver/team in proportion to w_{mnk}/R_k, where w_{mnk} is the number of counting races completed by that driver/team combination in season k and R_k is the total number of races in season k. Thus,

$$N_k=\sum_{m,n}\frac{w_{mnk}}{R_k}\tag{8}$$

$$\mu_k=\sum_{m,n}\frac{w_{mnk}\theta_{mnk}}{R_kN_k}\tag{9}$$

$$\sigma_k^2=\sum_{m,n}\frac{w_{mnk}(\theta_{mnk}-\mu_k)^2}{R_kN_k}.\tag{10}$$

2.6 Model fitting

The model's random effects, ε_{ijk}, are assumed to be independent and identically normally distributed with mean zero and variance σ^2. The model's maximum likelihood is thus achieved by estimating values $\hat{\alpha}_i$ and $\hat{\beta}_j$ that correspond to the least-squares fit for y_{ijk} to the data, where hats denote estimated parameter values. We achieved this using the nlinfit function in Matlab (version R2014A, Natick, MA, USA), which uses the Levenberg-Marquardt algorithm for global convergence. Each y_{ijk} value was weighted by the corresponding number of counting races w_{ijk}.

Cases where a driver scored 0 points or 10 points in every counting race in a season present potential problems for fitting, since they theoretically correspond to $y_{ijk}\to-\infty$ and $y_{ijk}\to+\infty$, respectively. We treated these cases by making conservative estimates of scoring rates (i.e., assuming that with enough races, a driver would eventually stop scoring 0 or 10 points, respectively). For cases where a driver scored 10 points in every counting race, we conservatively estimated scoring rate by supposing the existence of one additional race in which the driver came second (scored 6 points), such that $y_{ijk}=(10w_{ijk}+6)/(w_{ijk}+1)$. This method also allows us to differentiate between drivers with perfect records in counting races (e.g., Alberto Ascari in 1952 and Jochen Rindt in 1970) by giving a higher scoring rate to drivers who achieved perfect records over more counting races. This approach is sensible, as it rewards drivers for more statistically reliable perfect performances, e.g., 10 wins in 10 races is more impressive than one win in one race.

In cases where a driver scored 0 points in every counting race (i.e., a driver failure in every counting race), we conservatively estimated scoring rate by supposing the existence of one additional race in which the driver scored the average scoring rate of their teammates in the same season. In the very rare event that the driver had no teammates with counting races in that season, we estimated scoring rate by supposing the existence of one additional race in which the driver finished in the lowest position that any driver finished in that season. The results were found to be quite insensitive to this choice.

2.7 Adjusted scoring rates

Without any additional constraints, the above model is not identifiable, because any parameter set $\{\hat{\alpha}_i,\hat{\beta}_j\}$ makes identical predictions to the parameter set $\{\hat{\alpha}_i+K,\hat{\beta}_j-K\}$. It is therefore necessary to fix the value of at least one of the estimated $\hat{\alpha}_i$ or $\hat{\beta}_j$ parameters, so that other parameter values can be fitted relative to that reference value. Without loss of generality, we set the $\hat{\beta}_j$ parameter corresponding to the Red Bull team in 2013 to 0.5, since this value results in similar ranges for $\hat{\alpha}_i$, $\hat{\beta}_j$, and $\hat{\delta}_k$.

Once fitted, the parameter values can be used to infer underlying driver performances in each season from 1950 to 2013. For the purposes of comparison, we used Sebastian Vettel driving for the Red Bull team in 2013 as a baseline. Any baseline is arbitrary and will not affect the order of our rankings. We chose this baseline because it provides a modern reference point that can be easily related to, and because many Formula One pundits have speculated on how other drivers might have performed in the 2013 Red Bull. We thus defined β_{2013} as the value of $\hat{\beta}_j$ for the 2013 Red Bull, and δ_{2013} as the value of $\hat{\delta}_k$ for the strength of competition in 2013. Normalized driver performances were computed by correcting for team and competition effects,

$$\tilde{y}_{ijk} = y_{ijk} - \hat{\beta}_j + \hat{\delta}_k + \beta_{2013} - \delta_{2013}. \tag{11}$$

We then computed *adjusted scoring rates* for each driver *i* in each season *k* in which they completed at least 3 counting races, by taking a weighted sum across all teams, *j*, for which they drove in that season,

$$\tilde{s}_{ik} = S\Big(\sum_j w_j \tilde{y}_{ijk} \Big/ \sum_j w_j\Big). \tag{12}$$

In most years, a driver competed for only a single team, in which case Eq. (12) simply reduces to $\tilde{s}_{ik} = S(\tilde{y}_{ijk})$. For reference, Vettel scored $\tilde{s}_{ik} = 8.33$ points per race in 2013. The adjusted scoring rate estimates what would be an equivalent level of performance from a driver in Vettel's place at Red Bull in 2013. Note that this is not an estimation of how a driver from the past would actually perform in the Red Bull if somehow transported to the year 2013. It is not possible to estimate how the technical changes in the sport would affect the performances of past and future drivers in cars of different eras. The adjusted scoring rate should therefore only be viewed as an equivalence measure that allows *relative* performances from different eras to be compared on a common scale.

Using this method, we computed adjusted scoring rates for each driver in each year of the World Drivers' Championship from 1950 to 2013. These were then used to determine adjusted ranks in each year, which were compared to historical results. In addition, we computed metrics of peak driver performance over different timescales, as has been done in analyses of other sports (e.g., Sonas 2005). For each driver, we calculated the peak adjusted scoring rates over 1, 3, and 5 consecutive years, by averaging across consecutive seasons. A driver needed to have completed at least three counting races in each season in the consecutive sequence. These peak performance ranges were selected because different observers may value short-term vs. long-term performances differently. The choice of metric is ultimately a subjective matter; our goal was to provide the first reasonable quantitative framework for making comparisons, whatever the choice of metric.

Our model assumes that $\hat{\alpha}_i$ is constant for each driver's career and variations in performance ε_{ijk} are random uncorrelated fluctuations. In reality, there may be time-correlated fluctuations in form across a driver's career. Without modeling these dynamics explicitly, we instead approximate each driver's peak form by computing peak adjusted scoring rates. Methods that could be used for modeling changes in form across a driver's career are considered in the Discussion, as well as some of the particular difficulties associated with doing this in the context of Formula One.

2.8 Lap-time comparison

To test our model, we compared predictions of driver/team performances to lap-time data. For this, we used the fastest lap time set by each team at each race (in any session) in 2010, 2011, 2012, and 2013. The percentage difference between each team's best time and the best overall time (available at http://www.f1fanatic.co.uk/2012/12/03/2012-f1-car-performance/) was then used as a performance metric. The average value of this metric for each team in each year, T_{jk}, was regressed against the expected driver/team performance for each team's best performing driver in that year,

$$y_{jk} = \max_i(\hat{\alpha}_i + \hat{\beta}_j - \hat{\delta}_k). \tag{13}$$

The data for all 4 years were then fitted by an exponential function, $T_{jk} = a_1 e^{-a_2 y_{jk}}$, since the data followed approximately this form and because we expect $T_{jk} \to 0$ as $y_{jk} \to \infty$, since a team with very strong performance should always set the best time.

3 Results

3.1 Model fitting

Fitting our model to the 64 years of championship data, with 1921 y_{ijk} values, yielded good estimates of all 342 $\hat{\alpha}_i$

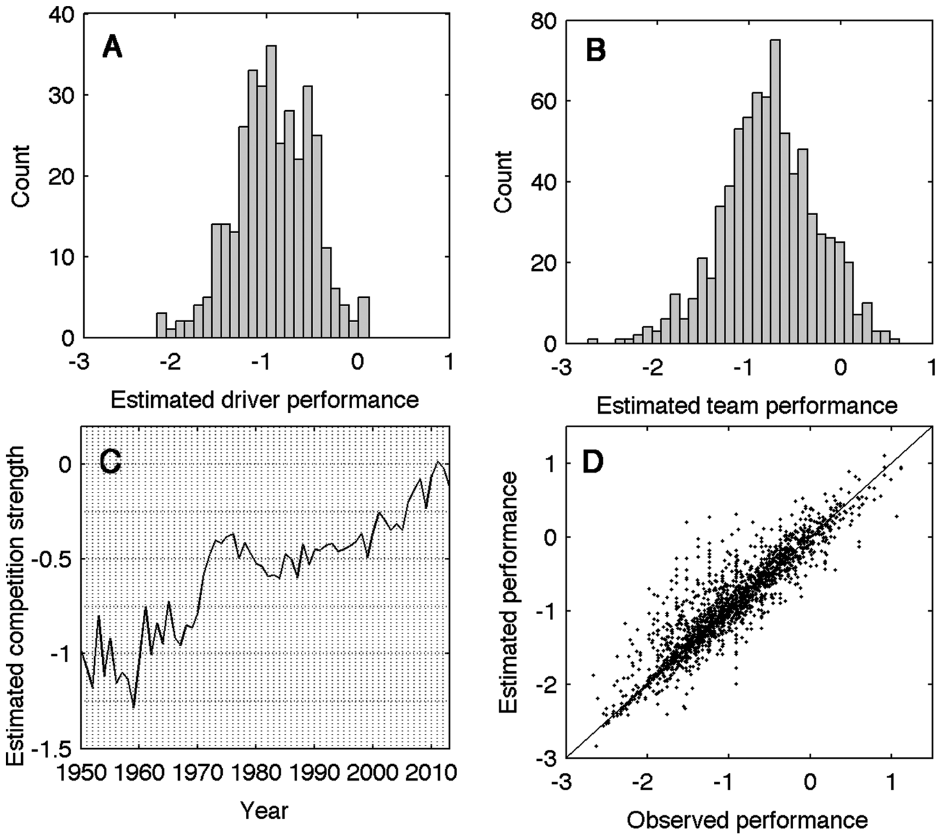

Figure 1 Histograms for estimated values of the model parameters: (A) estimated driver performance $\hat{\alpha}_i$ and (B) estimated team performance $\hat{\beta}_j$. Histograms are in bins of width 0.1. (C) The estimated competition effect $\hat{\delta}_k$ for each year from 1950 to 2013. (D) The model's estimated values of performance $\hat{y}_{ijk}$ versus the observed data values. The line $y=x$ is plotted for reference.

and 760 $\hat{\beta}_j$ parameters, representing fixed driver and team effects, respectively. Histograms for the model parameter values are shown in Figure 1A and B. Both distributions are approximately unimodal, with the mean of $\hat{\alpha}_i$ being −0.92 and the mean of $\hat{\beta}_j$ being −0.60. The distribution for $\hat{\alpha}_i$ is narrower than the distribution for $\hat{\beta}_j$, with standard deviations of 0.42 and 0.51, respectively. This finding is consistent with the supposition that differences in performance between teams are typically larger than differences in performance between drivers in Formula One.

The $\hat{\delta}_k$ parameter increases from 1950 to 2013, as shown in Figure 1C, meaning the level of competition has typically increased from one year to the next across the sport's history. However, the increase has not been monotonic. There is a decrease across the late 1970s and early 1980s. The mean value is −0.57, with a minimum value of −1.42 in 1950 and a maximum value of 0.01 in 2011. The increase in competition is partly due to an increase in the number of competitors, since Eq. (7) depends on N_k. However, it is not solely due to this change, as changes in $\hat{\delta}_k$ do not simply track changes in the number of competitors. Overall, the model fits the data well, as shown in Figure 1D. The adjusted R^2 value for the fit is 0.82. Ways in which the fit could be improved are considered in the Discussion.

3.2 World Drivers' Championship

Using adjusted scoring rates, $\tilde{s}_{ik}$, we calculated adjusted results for each year of the World Drivers' Championship from 1950 to 2013. By correcting for team and competition effects, these adjusted scoring rates put all underlying driver performances onto an equivalent scale. The top three drivers in each year are given in Tables 1–3.

In 38 of the 64 years (59%), our model awarded the championship to a driver other than the historical winner. In these cases, another driver was deemed to have performed better, having been restricted by non-driver failures and/or worse team performance. The model also awarded the championship to fewer drivers in total, with 21 champions compared to the historical 32 champions. Historical champions who no longer win based on model predictions are: Giuseppe Farina (1950), Mike Hawthorn (1958), Jack Brabham (1959, 1960, 1966), Phil Hill (1961), Graham Hill

Table 1 The top three adjusted scoring rates for each season from 1950 to 1969.

	Driver	Raw	Adjusted		Driver	Raw	Adjusted
1950				1960			
1	JM Fangio	10.00	7.69	1	S Moss	5.40	7.42
2	A Ascari	4.67	7.05	2	**J Brabham**	7.57	6.82
3	**G Farina**	6.60	4.47	3	I Ireland	2.36	5.09
1951				1961			
1	**JM Fangio**	7.04	7.39	1	S Moss	4.68	8.69
2	JF Gonzalez	6.40	6.89	2	I Ireland	4.39	8.53
3	A Ascari	6.00	6.50	3	J Clark	1.69	5.97
1952				1962			
1	**A Ascari**	10.00	8.90	1	J Clark	6.65	9.27
2	M Hawthorn	2.50	4.36	2	B McLaren	4.71	6.46
3	G Farina	4.67	3.32	3	**G Hill**	6.91	6.04
1953				1963			
1	**A Ascari**	6.67	7.54	1	**J Clark**	8.89	9.29
2	JM Fangio	5.17	7.05	2	D Gurney	3.80	6.11
3	G Farina	4.86	5.87	3	B McLaren	3.40	5.35
1954				1964			
1	**JM Fangio**	8.38	8.00	1	J Clark	10.00	8.70
2	JF Gonzalez	5.50	5.52	2	G Hill	5.87	5.93
3	M Hawthorn	5.17	5.19	3	B McLaren	2.85	5.82
1955				1965			
1	**JM Fangio**	9.20	9.07	1	**J Clark**	10.00	8.97
2	M Trintignant	4.35	5.81	2	B McLaren	2.03	7.56
3	G Farina	4.33	5.80	3	G Hill	5.44	7.25
1956				1966			
1	**JM Fangio**	7.50	6.69	1	J Surtees	7.50	8.35
2	S Moss	6.40	6.16	2	J Clark	4.25	7.92
3	P Collins	6.33	5.41	3	J Rindt	3.43	6.17
1957				1967			
1	**JM Fangio**	8.67	8.83	1	J Clark	7.50	8.00
2	L Musso	3.85	5.02	2	B McLaren	1.21	5.63
3	S Moss	5.40	4.76	3	P Rodriguez	2.68	5.49
1958				1968			
1	S Moss	9.20	7.33	1	J Stewart	4.95	7.92
2	**M Hawthorn**	5.75	5.71	2	J Ickx	4.67	6.95
3	R Salvadori	2.05	4.35	3	J Rindt	2.67	6.37
1959				1969			
1	S Moss	6.50	5.91	1	**J Stewart**	7.67	9.04
2	T Brooks	6.05	5.35	2	J Rindt	4.60	6.89
3	**J Brabham**	6.17	5.20	3	J Ickx	5.57	6.46

Also shown are raw scoring rates across all counting races, using the scoring method described in the results. Historical champions are marked by bold text.

(1962, 1968), Denny Hulme (1967), Emerson Fittipaldi (1972, 1974), Mario Andretti (1978), Jody Scheckter (1979), Nelson Piquet (1981, 1983, 1987), Keke Rosberg (1982), Nigel Mansell (1992), Damon Hill (1996), Jacques Villeneuve (1997), Mika Hakkinen (1998, 1999), Kimi Raikkonen (2007), Jenson Button (2009), and Sebastian Vettel (2010–2013).

Seven drivers who were not historical champions are now awarded titles: Stirling Moss (1958–1961), Carlos Reutemann (1974), John Watson (1981), Elio de Angelis (1982), Stefan Johansson (1989), Jean Alesi (1992), and Heinz-Harald Frentzen (1999). Moss historically came close to winning the championship in each of those 4 years and is widely regarded as the best driver to have never won a championship. In 1974, Reutemann is rewarded for driving the less competitive Brabham-Ford and suffering 5 non-driver failures (compared to 2 for Fittipaldi). In 1981, Watson is rewarded for dominating teammate Andrea de Cesaris (27 points to 1 point) and for driving a relatively uncompetitive car. In 1982, de Angelis is rewarded for greatly outperforming teammate Nigel Mansell, and suffering 6 non-driver failures (compared to 2 for Keke Rosberg). In 1989, Johansson is narrowly awarded the

Table 2 The top three adjusted scoring rates for each season from 1970 to 1989.

	Driver	Raw	Adjusted		Driver	Raw	Adjusted
1970				1980			
1	**J Rindt**	10.00	9.64	1	**A Jones**	6.34	7.33
2	J Stewart	5.20	9.15	2	E de Angelis	1.36	6.90
3	C Amon	3.00	7.85	3	J Watson	0.98	6.78
1971				1981			
1	**J Stewart**	7.56	8.79	1	J Watson	2.71	7.81
2	R Peterson	3.73	6.85	2	A Prost	4.60	7.24
3	C Amon	1.15	6.58	3	E Cheever	1.00	6.89
1972				1982			
1	J Stewart	5.51	8.59	1	E de Angelis	2.71	7.01
2	R Peterson	1.21	7.67	2	E Cheever	2.53	6.49
3	**E Fittipaldi**	7.34	7.09	3	**K Rosberg**	3.78	6.27
1973				1983			
1	J Hunt	2.93	8.13	1	A Prost	4.78	6.95
2	**J Stewart**	5.86	8.02	2	J Watson	2.14	6.54
3	C Pace	1.03	6.75	3	K Rosberg	2.40	6.36
1974				1984			
1	C Reutemann	3.59	7.95	1	A Prost	6.97	6.85
2	**E Fittipaldi**	4.52	7.46	2	E de Angelis	2.97	6.72
3	J Scheckter	4.31	7.40	3	K Rosberg	3.34	6.70
1975				1985			
1	J Hunt	3.70	8.77	1	**A Prost**	6.23	7.50
2	M Donohue	0.69	7.33	2	A Senna	4.52	7.36
3	V Brambilla	2.31	6.75	3	M Alboreto	6.11	6.44
1976				1986			
1	**J Hunt**	5.36	7.73	1	**A Prost**	5.57	7.25
2	N Lauda	5.65	6.94	2	A Senna	5.18	6.54
3	C Amon	0.41	6.85	3	A Jones	0.56	5.74
1977				1987			
1	**N Lauda**	6.25	7.94	1	A Senna	4.59	7.08
2	J Hunt	3.69	7.34	2	A Prost	4.51	6.51
3	A Jones	2.15	7.17	3	E Cheever	1.29	6.46
1978				1988			
1	N Lauda	5.11	7.98	1	A Prost	8.00	7.48
2	J Scheckter	1.73	7.03	2	I Capelli	1.94	6.22
3	C Reutemann	3.59	6.73	3	**A Senna**	6.80	6.16
1979				1989			
1	A Jones	5.95	7.12	1	S Johansson	2.13	6.77
2	J Watson	1.43	7.05	2	**A Prost**	5.67	6.75
3	E Fittipaldi	0.36	6.47	3	M Gugelmin	0.86	6.72

Also shown are raw scoring rates across all counting races, using the scoring method described in the results. Historical champions are marked by bold text.

1989 championship for his performances (including a 3rd place) in the slow and unreliable Onyx. In 1992, Alesi is rewarded for his strong performances in an uncompetitive Ferrari, dominating teammate Ivan Capelli (18 points to 3 points), while historical winner Mansell is heavily penalized for driving an extremely dominant Williams. In 1999, Frentzen is rewarded for greatly outperforming teammate Damon Hill.

The drivers awarded championships by the model are listed in Table 4, with the most championships going to Michael Schumacher (10), Fernando Alonso (9), Juan Manuel Fangio (6), and Alain Prost (6). The number of championships may be considered a quantitative metric of cumulative career performance. However, it is dependent on the quality of competitors faced during each driver's career, as well as career duration.

3.3 Peak driver performances

It is common in sports hypotheticals to compare competitors from different eras at their respective peaks of form.

Table 3 The top three adjusted scoring rates for each season from 1990 to 2013.

	Driver	Raw	Adjusted		Driver	Raw	Adjusted
1990				2002			
1	**A Senna**	7.00	7.41	1	**M Schumacher**	8.47	7.91
2	A Prost	6.00	6.82	2	N Heidfeld	0.65	6.56
3	I Capelli	1.48	6.31	3	J Button	1.25	6.30
1991				2003			
1	A Prost	3.78	7.05	1	F Alonso	3.37	8.29
2	**A Senna**	6.78	6.59	2	**M Schumacher**	4.90	7.15
3	N Mansell	6.73	6.10	3	HH Frentzen	0.77	7.05
1992				2004			
1	J Alesi	2.00	6.97	1	**M Schumacher**	7.93	7.88
2	M Schumacher	4.13	6.67	2	K Raikkonen	3.11	7.77
3	**N Mansell**	7.71	6.37	3	F Alonso	2.39	7.59
1993				2005			
1	M Schumacher	4.33	7.32	1	**F Alonso**	6.77	8.62
2	A Senna	6.08	6.96	2	K Raikkonen	6.64	7.55
3	**A Prost**	6.60	6.79	3	M Schumacher	2.47	7.18
1994				2006			
1	**M Schumacher**	7.67	8.75	1	**F Alonso**	7.25	8.66
2	D Hill	6.09	7.20	2	M Schumacher	6.50	7.96
3	G Berger	5.13	6.39	3	R Kubica	0.92	7.15
1995				2007			
1	**M Schumacher**	6.38	7.56	1	L Hamilton	5.17	7.64
2	HH Frentzen	1.28	6.45	2	F Alonso	5.07	7.56
3	M Salo	0.58	5.91	3	N Heidfeld	2.40	7.32
1996				2008			
1	M Schumacher	5.90	8.38	1	F Alonso	2.60	8.23
2	**D Hill**	6.93	6.26	2	S Vettel	1.38	7.66
3	M Salo	0.75	5.93	3	F Massa	5.23	7.13
1997				2009			
1	M Schumacher	5.20	8.06*	1	F Alonso	0.89	8.88
2	**J Villeneuve**	5.06	6.71	2	N Rosberg	1.11	7.75
3	R Barrichello	1.52	6.65	3	K Raikkonen	2.43	7.69
1998				2010			
1	M Schumacher	6.15	7.71	1	F Alonso	4.58	8.21
2	**M Hakkinen**	7.14	6.39	2	N Rosberg	1.72	8.14
3	D Hill	2.01	6.19	3	R Kubica	1.71	7.39
1999				2011			
1	HH Frentzen	4.15	8.09	1	F Alonso	3.88	8.36
2	J Trulli	0.87	6.26	2	**S Vettel**	8.17	8.08
3	E Irvine	4.93	6.25	3	M Schumacher	0.77	7.50
2000				2012			
1	**M Schumacher**	7.20	7.61	1	F Alonso	4.34	8.52
2	HH Frentzen	1.47	7.31	2	N Rosberg	1.20	7.98
3	G Fisichella	1.40	6.44	3	**S Vettel**	5.06	7.45
2001				2013			
1	**M Schumacher**	8.20	8.47	1	F Alonso	3.80	8.38
2	T Marques	0.10	7.70	2	**S Vettel**	8.33	8.33
3	F Alonso	0.06	7.10	3	N Rosberg	2.92	7.25

Also shown are raw scoring rates across all counting races, using the scoring method described in the results. Historical champions are marked by bold text. *Note that Schumacher is included in the 1997 championship here; he was not disqualified from any individual races, but was excluded from the championship for unsporting conduct.

Peak performance can be defined in different ways and on different timescales. The choice of metric for peak performance is intrinsically subjective, but the comparison itself ought to be objective. Until now, however, peak performances have not been accurately quantified for Formula One drivers.

Table 4 Number of World Drivers' Championships awarded to each driver by the model, as well as the change in number relative to the sport's actual history in parentheses.

Championships	Drivers
10	M Schumacher (+3)*
9	F Alonso (+7)
6	JM Fangio (+1), A Prost (+2)
5	J Clark (+3)
4	S Moss (+4), J Stewart (+1)
3	J Hunt (+2)
2	A Ascari (=), N Lauda (-1), A Jones (+1), A Senna (-1)
1	J Surtees (=), J Rindt (=), C Reutemann (+1), J Watson (+1), E de Angelis (+1), S Johansson (+1), J Alesi (+1), HH Frentzen (+1), L Hamilton (=)

*Note that Schumacher's 1997 championship is included here; he was not disqualified from any individual races, but was excluded from the championship for unsporting conduct.

Jochen Rindt's 1970 performance is ranked the best single-year performance, with an adjusted scoring rate of 9.64 points per race. In this year, he won all five counting races, despite strong competition from other drivers, including Jackie Stewart, Jacky Ickx, Denny Hulme, and Jack Brabham. Jim Clark scores the 2nd, 3rd, and 7th best single-year performances for 1963, 1962, and 1965, respectively. Alberto Ascari's record of six wins from six counting races in 1952 is ranked the 8th best performance of all time, penalized for the very dominant Ferrari that he drove.

In Table 5, we rank the top 40 drivers for peak performances over 1-year, 3-year, and 5-year intervals. For both 3-year and 5-year intervals, the top five drivers are Jim Clark, Jackie Stewart, Fernando Alonso, Michael Schumacher, and Juan Manuel Fangio.

In Figure 2, we plot the adjusted scoring rates from 1950 to 2013 for all drivers ranked in the top 20 for peak 3-year performance. The top 20 drivers are quite evenly distributed over time, suggesting that the model is not heavily biased towards drivers in any particular era.

3.4 Lap-time predictions

To test our model predictions, we compared the predicted driver/team performances y_{jk} to the best lap times for each team across races in 2010, 2011, 2012, and 2013. Although the model parameters are fitted only to finishing position data, the model does a satisfactory job of predicting lap-time data for each team, using the same fit for all 4 years (adjusted R^2=0.49), shown in Figure 3. The parameter values for the exponential fit are $\hat{a}_1$=1.83 and $\hat{a}_2$=1.62. The model fits the data with high confidence for midfield and front-running teams, with much lower confidence for back-markers.

4 Discussion

As is also true for other sports (Davis 2000; Aitken 2004; Albert 2006; Kvam 2011; Oberstone 2011; Piette and Jensen 2012), there is frequent interest in ranking the greatest Formula One drivers of all time. To date, however, no suitable quantitative methods have been proposed. We addressed this problem by developing a simple statistical model of driver performance over time that accounts for both team and competitor effects on scoring rates. The model predictions are sensible and can also be used to predict lap-time data. In many cases, the model provides a quantitative foundation for expert opinions, while in a few interesting cases it challenges prevailing views, as discussed below.

Our model allows comparisons within and between eras due to two facts: (i) drivers can be compared to their teammates, and (ii) drivers have multiple different teammates across their careers. The only exceptions to the latter point are drivers who have spent very little time in Formula One and therefore have minimal bearing on determining the greatest drivers of all time. Using our model, we were able to rank all drivers on a single scale: equivalent performance in the place of Sebastian Vettel in the 2013 Red Bull. This should be interpreted as an equivalence measure of performance, *not* as a prediction of how a driver from another era would actually perform in a 2013 car. While statistical 'time machine' methods similar to those employed here (Berry et al. 1999) may allow more direct hypothetical comparisons in sports where there have been negligible technical/equipment changes over time, they clearly cannot allow such strong comparisons in Formula One, where technology and required driving styles have changed significantly from 1950 to 2013 (Jones 1995; Smith 2011). Our model therefore makes *direct* comparisons of performance within eras and *relative* comparisons of performance between eras, where no direct comparison is possible.

4.1 Comparisons to expert rankings

Qualitative rankings of Formula One drivers typically list Jackie Stewart, Jim Clark, Michael Schumacher, Alain Prost, Juan Manuel Fangio, and Ayrton Senna as the top six in some permutation (F1 Racing 1997, 2004, 2008;

Table 5 The top 40 adjusted scoring rates for 1-year, 3-year, and 5-year peaks.

#		1-year peak		3-year peak		5-year peak
1	J Rindt	9.64 (1970)	J Clark	9.08 (1962–1964)	J Clark	8.83 (1962–1966)
2	J Clark	9.29 (1963)	J Stewart	8.99 (1969–1971)	J Stewart	8.72 (1969–1973)
3	J Stewart	9.15 (1970)	F Alonso	8.48 (2009–2011)	F Alonso	8.47 (2009–2013)
4	JM Fangio	9.07 (1955)	M Schumacher	8.23 (1994–1996)	M Schumacher	8.09 (1994–1998)
5	A Ascari	8.90 (1952)	JM Fangio	8.20 (1955–1957)	JM Fangio	7.93 (1953–1957)
6	F Alonso	8.88 (2009)	J Hunt	7.96 (1973–1975)	J Hunt	7.79 (1973–1977)
7	J Hunt	8.77 (1975)	S Vettel	7.95 (2011–2013)	N Rosberg	7.66 (2009–2013)
8	M Schumacher	8.75 (1994)	N Rosberg	7.78 (2010–2012)	S Vettel	7.46 (2009–2013)
9	S Moss	8.69 (1961)	A Ascari	7.65 (1951–1953)	A Prost	7.12 (1984–1988)
10	I Ireland	8.53 (1961)	J Rindt	7.63 (1968–1970)	N Lauda	7.11 (1975–1979)
11	J Surtees	8.35 (1966)	N Lauda	7.62 (1976–1978)	K Raikkonen	7.01 (2003–2007)
12	S Vettel	8.33 (2013)	S Moss	7.34 (1959–1961)	L Hamilton	6.89 (2007–2011)
13	N Rosberg	8.14 (2010)	K Raikkonen	7.31 (2003–2005)	J Watson	6.87 (1979–1983)
14	HH Frentzen	8.09 (1999)	L Hamilton	7.26 (2007–2009)	A Jones	6.86 (1976–1980)
15	N Lauda	7.98 (1978)	J Watson	7.21 (1979–1981)	S Moss	6.82 (1957–1961)
16	C Reutemann	7.95 (1974)	A Prost	7.20 (1984–1986)	R Kubica	6.82 (2006–2010)
17	C Amon	7.85 (1970)	HH Frentzen	7.06 (1998–2000)	A Senna	6.77 (1987–1991)
18	J Watson	7.81 (1981)	R Peterson	7.04 (1971–1973)	N Heidfeld	6.66 (2005–2009)
19	K Raikkonen	7.77 (2004)	A Senna	6.99 (1985–1987)	E Fittipaldi	6.65 (1974–1978)
20	T Marques	7.70 (2001)	R Kubica	6.98 (2008–2010)	J Button	6.61 (2003–2007)
21	R Peterson	7.67 (1972)	A Jones	6.94 (1978–1980)	HH Frentzen	6.59 (1999–2003)
22	L Hamilton	7.64 (2007)	J Button	6.93 (2005–2007)	R Peterson	6.57 (1971–1975)
23	B McLaren	7.56 (1965)	E Fittipaldi	6.84 (1974–1976)	J Scheckter	6.56 (1974–1978)
24	A Prost	7.50 (1985)	C Reutemann	6.84 (1973–1975)	J Trulli	6.41 (2001–2005)
25	E Fittipaldi	7.46 (1974)	N Heidfeld	6.82 (2005–2007)	C Pace	6.33 (1973–1977)
26	A Senna	7.41 (1990)	C Amon	6.70 (1970–1972)	E de Angelis	6.26 (1980–1984)
27	J Scheckter	7.40 (1974)	E de Angelis	6.63 (1980–1982)	D Hill	6.24 (1994–1998)
28	R Kubica	7.39 (2010)	J Scheckter	6.61 (1974–1976)	B McLaren	6.15 (1961–1965)
29	A Jones	7.33 (1980)	F Massa	6.54 (2008–2010)	K Rosberg	6.13 (1981–1985)
30	M Donohue	7.33 (1975)	J Trulli	6.46 (2003–2005)	C Reutemann	6.08 (1972–1976)
31	N Heidfeld	7.32 (2007)	K Rosberg	6.45 (1983–1985)	F Massa	6.06 (2008–2012)
32	J Trulli	7.31 (2004)	JP Beltoise	6.42 (1972–1974)	J Ickx	6.04 (1968–1972)
33	G Hill	7.25 (1965)	G Fisichella	6.36 (1999–2001)	G Hill	5.98 (1962–1966)
34	JP Beltoise	7.25 (1974)	D Hill	6.36 (1994–1996)	N Mansell	5.98 (1988–1992)
35	D Hill	7.20 (1994)	C Pace	6.34 (1973–1975)	D Hulme	5.93 (1968–1972)
36	F Massa	7.13 (2008)	G Hill	6.31 (1964–1966)	G Fisichella	5.93 (1999–2003)
37	J Button	7.03 (2005)	E Cheever	6.24 (1981–1983)	JP Beltoise	5.86 (1970–1974)
38	E de Angelis	7.01 (1982)	B McLaren	6.24 (1963–1965)	D Gurney	5.85 (1961–1965)
39	J Alesi	6.97 (1992)	JP Montoya	6.13 (2002–2004)	J Villeneuve	5.84 (1997–2001)
40	J Ickx	6.95 (1968)	J Villeneuve	6.12 (2000–2002)	R Barrichello	5.76 (1993–1997)

The corresponding year ranges for each driver are given in parentheses.

Autosport 2009). Our analysis supports the qualitative rankings of Stewart, Clark, Fangio, and Schumacher as the greatest drivers in terms of peak performances over 1, 3, and 5 years, adding the currently active driver Fernando Alonso to this list. Prost is rated outside the top six based on peak performance metrics and is rated 3rd by number of championship titles. Senna is rated outside the top 10 by all peak performance metrics and is rated 9th by number of championship titles.

Using adjusted scoring rates, we simulated a revised history in which team and competition effects were removed. In many cases, our model results are consistent with expert consensus. Some examples include: (i) Stirling Moss, who is widely considered the best driver to have never won a championship, is awarded four championships by our model. (ii) The 1996 championship was historically dominated by the Williams team, but Michael Schumacher's performance in the less competitive Ferrari is widely considered the best of that year. Our model awards Schumacher that championship by a large margin. (iii) Fernando Alonso was widely considered the best performing driver of 2012 (F1

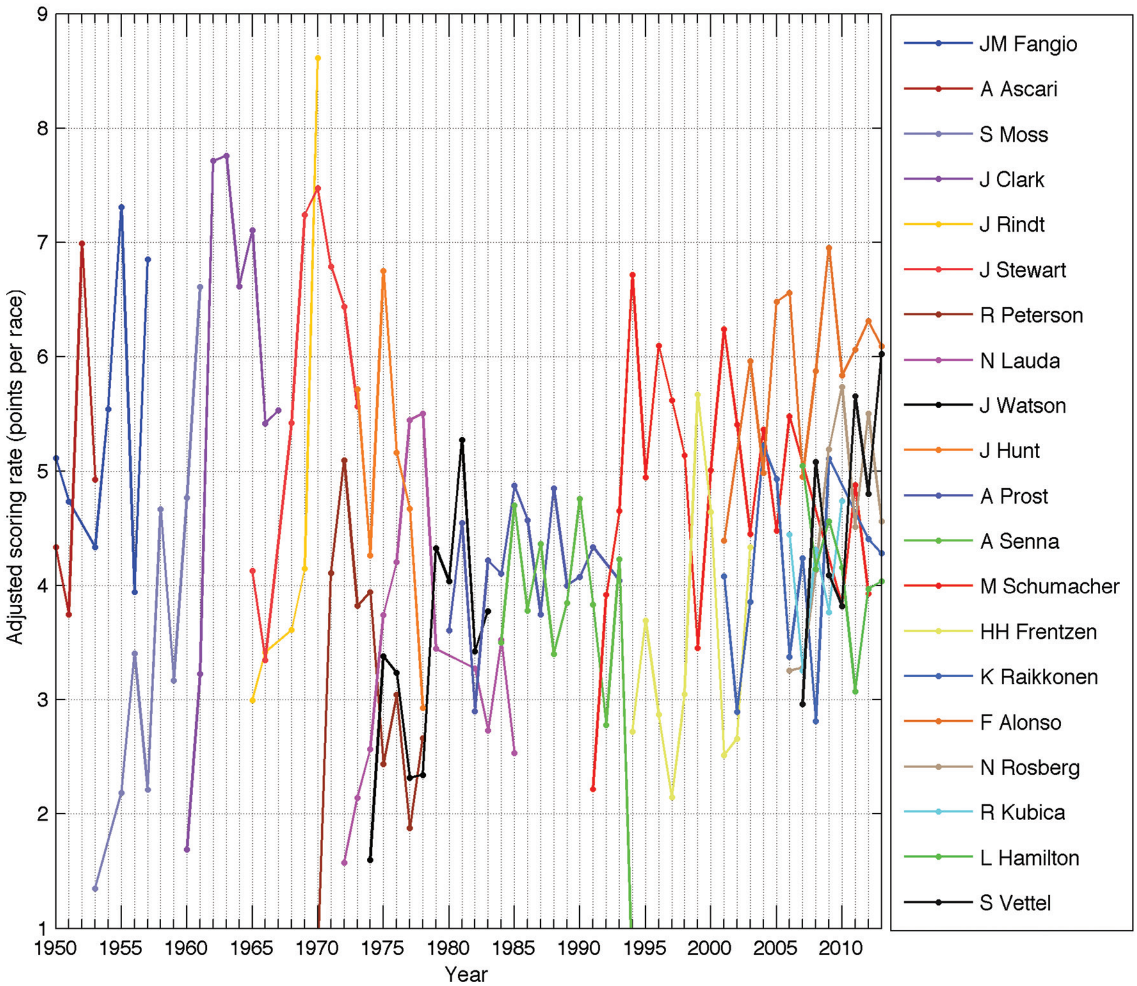

Figure 2 Adjusted scoring rates for the top 20 drivers based on 3-year peak adjusted scoring rates. Different colors are used for different drivers, as indicated in the legend (listed in chronological order).

Racing 2012). Our model awards Alonso that championship. Overall, our driver rankings based on peak performances are similar to expert rankings. Of the top 25 drivers rated by experts (Autosport 2009), 16, 16, and 15 are rated in our top 25 for 1-year, 3-year, and 5-year peak performances. Team performance predictions are also sensible, with the 2011 Red Bull, 2013 Red Bull, 1992 Williams, and 2002 Ferrari rated as the four best performing teams. While our model produces rankings that are mostly consistent with expert opinions, there are some interesting exceptions.

Once team and competition effects are corrected for, the best drivers from the 1980s are generally rated lower than the best drivers from other decades. None of the top 14 drivers by 3-year peak performance are from the 1980s. Overall, the model rankings suggest that the following drivers have been underrated by experts: James Hunt, Nico Rosberg, John Watson, and Heinz-Harald Frentzen. The model rankings also suggest that the following drivers have been overrated by experts: Jack Brabham, Graham Hill, Mario Andretti, Gilles Villeneuve, Nelson Piquet, Nigel Mansell, Ayrton Senna, and Mika Hakkinen. We consider each case in sequence below.

James Hunt: Hunt is rated between 19th and 24th by experts (F1 Racing 1997, 2004; Autosport 2009), but is rated significantly higher by our analysis, with the 6th highest 3-year peak performance. This rating is due to Hunt outperforming all teammates in his career, scoring an exceptional 2.93 points per race in 1973 in the uncompetitive March-Ford, and outperforming teammate Jochen Mass from 1976 to 1977 by a greater margin than Emerson Fittipaldi outperformed Mass in 1975. Our high ranking of Hunt is partially supported by the earlier analysis of Eichenberger and Stadelmann (2009), which placed Hunt at 14th. This suggests that Hunt's objective performances are vastly underrated by experts.

Nico Rosberg: Rosberg is a currently active driver who has not featured in previous expert rankings. His 8th

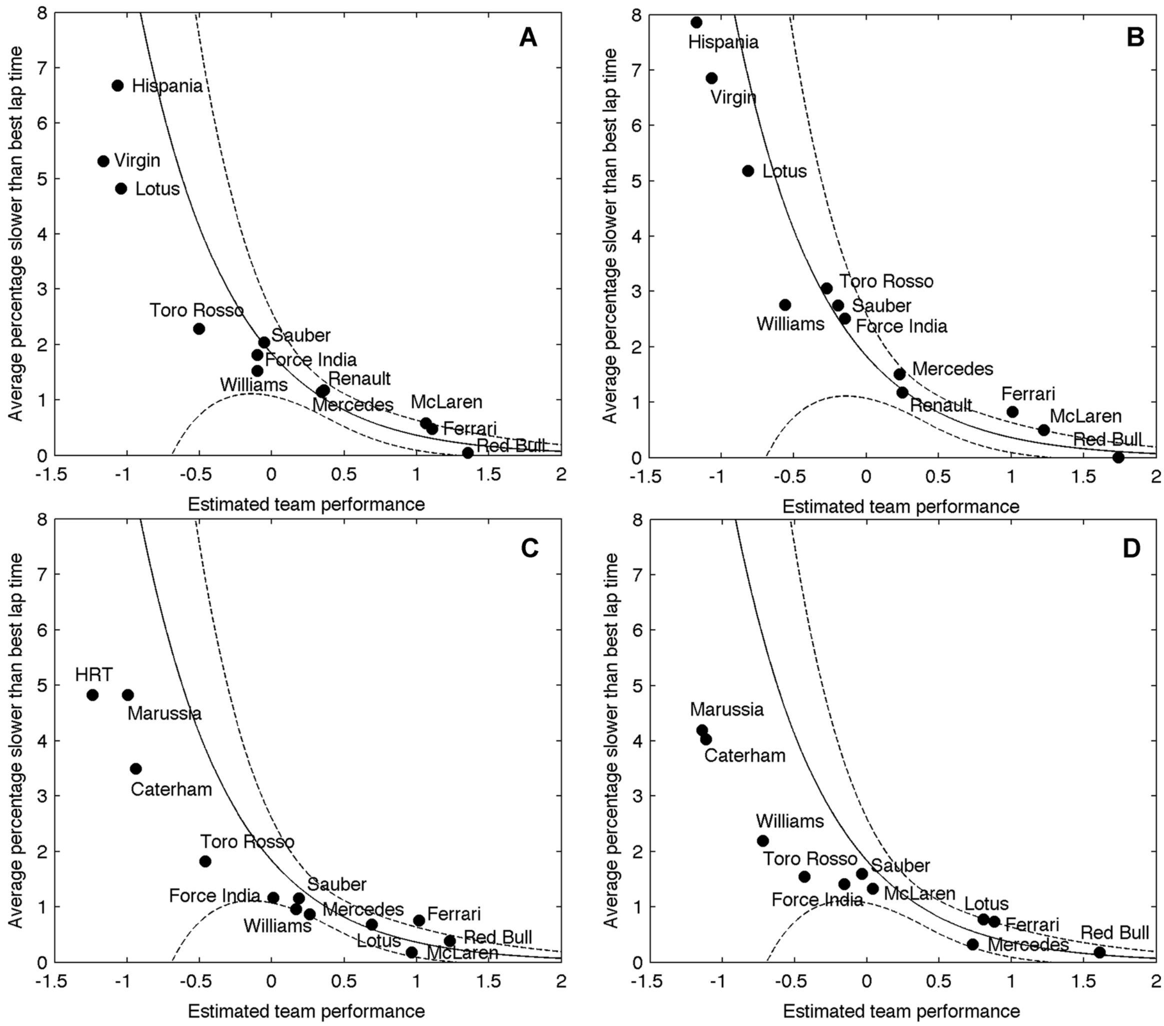

Figure 3 Lap-time data (percentage slower than best time, averaged across races) versus model estimates of each team's best team/driver performance $\hat{y}_{jk}$ (labeled data points). Data are shown for (A) 2010, (B) 2011, (C) 2012, and (D) 2013. The best exponential fit to all the data is shown (solid line; adjusted R^2=0.49), along with the 95% confidence limits (dashed lines).

highest 3-year peak performance is due to outperforming teammate Michael Schumacher across 2010–2012, and performing similarly to Lewis Hamilton in 2013. Our model does not account for changes in driver performance across the career, as discussed in detail below. It could thus be argued that Rosberg outperformed Schumacher due to Schumacher being past his peak form. If true, then Rosberg's performances from 2010 to 2012 are overrated by the model. Removing Schumacher's data from 2010 (his comeback year and the year in which he performed worst), lowers Rosberg ranking to 16th, with no increase in Schumacher's ranking. Since Rosberg is still an active driver, confidence in his ranking will improve as more data are accrued.

John Watson: Watson is generally rated outside the top 40 by experts (F1 Racing 1997, 2004; Autosport 2009), but in our analysis he has the 15th highest 3-year peak performance. This ranking is primarily due to Watson outperforming teammates Alain Prost in 1980 and Niki Lauda in 1982 and 1983. Like Nico Rosberg, this may be a case of benefiting from the model not including changes in driver performance across the career, since Prost was in his rookie year in 1980 and Lauda had just returned from a 2-year break in 1982.

Heinz-Harald Frentzen: Frentzen is usually rated outside the top 40 by experts (F1 Racing 1997, 2004; Autosport 2009), but in our analysis he has the 17th highest 3-year peak performance. With the exception of teammate Jacques Villeneuve from 1997 to 1998, Frentzen performed very well against all his teammates, dominating two highly rated drivers: Damon Hill in 1999 and Nick Heidfeld in 2003.

Jack Brabham: Brabham is rated between 12th and 18th by experts (F1 Racing 1997, 2004; Autosport 2009). Our model rates Brabham 57th for 3-year peak performance and awards him no championships. Brabham is penalized by the model for having more competitive cars and fewer non-driver failures – 2 failures to Stirling Moss's 4 out of 8 rounds in 1959 (excluding Indianapolis), and 3 failures to Surtees's 5 out of 9 rounds in 1966. He was also outperformed by Roy Salvadori in 1958, Dan Gurney from 1963 to 1965, Jochen Rindt in 1968, and Jacky Ickx in 1969.

Graham Hill: Hill is rated between 17th and 21st by experts (F1 Racing 1997, 2004; Autosport 2009). Our model rates Hill 36th for 3-year peak performance and awards him no championships. In 1962, Hill is penalized for driving a more competitive car and suffering only 1 non-driver failure to Jim Clark's 4 out of 9 rounds. In 1968, Hill is penalized for driving the dominant Lotus-Ford. Overall, Hill's rating is harmed by underwhelming performances relative to teammates. He was outperformed by teammates Innes Ireland in 1959, Jo Bonnier in 1960, Tony Brooks in 1961, Jackie Stewart in 1966, Jim Clark in 1967, Jochen Rindt in 1969, Tim Schenken in 1971, and Carlos Reutemann in 1972.

Mario Andretti: Andretti is rated between 17th and 20th by experts (F1 Racing 1997, 2004; Autosport 2009). Our model rates Andretti 48th for 3-year peak performance and awards him no championships. He is penalized for driving an extremely dominant car in 1978, being slightly outperformed by teammate Carlos Reutemann in 1979, and being heavily outperformed by teammate Elio de Angelis in 1980.

Gilles Villeneuve: Although Villeneuve never won a championship, he is rated between 10th and 13th by experts (F1 Racing 1997, 2004; Autosport 2009). Our model rates Villeneuve 50th for 3-year peak performance. Villeneuve is often praised for his spectacular and entertaining driving style. Objectively, however, our model does not support the notion that Villeneuve is one of the all-time greats.

Nelson Piquet: Piquet is rated between 9th and 13th by experts (F1 Racing 1997, 2004; Autosport 2009). Our model rates Piquet 56th for 3-year peak performance and awards him no championships. He is heavily penalized for driving very dominant cars during his championship years.

Nigel Mansell: Mansell is rated between 10th and 14th by experts (F1 Racing 1997, 2004; Autosport 2009). Our model rates Mansell 46th for 3-year peak performance and awards him no championships. Mansell is penalized for driving dominant cars, and for being outperformed by teammates Elio de Angelis and Alain Prost.

Ayrton Senna: Senna is often voted the best driver of all time in qualitative rankings (F1 Racing 1997, 2004, 2008; Autosport 2009). In our objective analysis, however, he is rated 19th for 3-year peak performance. Senna's ranking may have been slightly penalized by our use of a uniform scoring system for all years, which uses all counting races in calculating adjusted scoring rates. In 1988, Senna was outscored by his teammate Alain Prost 105 points to 94 across all races. However, due to the points scoring system in place at that time, only the 11 best finishes counted towards the championship, resulting in Senna winning with 90 points to Prost's 87. Senna may have driven differently had all the races counted, although he was known for his extremely competitive approach to driving no matter what was on the line (Hilton 1991). Nevertheless, this only accounts for a small penalty to his ranking.

Mika Hakkinen: Hakkinen is rated between 11th and 15th by experts (F1 Racing 2004; Autosport 2009). Our model rates Hakkinen 55th for 3-year peak performance and awards him no championships. Hakkinen is compared primarily to teammate David Coulthard (rated 63rd for 3-year peak performance); he performed only slightly better than Coulthard from 1996 to 2001, being outscored in 1997 and 2001. Both Hakkinen and Coulthard are penalized for driving very dominant McLarens over this period.

4.2 Previous models, model limitations, and future work

To our knowledge, there has been only one previous attempt to estimate underlying driver performances in Formula One using a statistical model (Eichenberger and Stadelmann 2009). That model used a similar mathematical framework to our own, but generated many questionable results. As noted in our Introduction, this is likely due to the use of a linear model with finishing position as the performance metric, rather than a nonlinear model with points as the performance metric. Problems may have arisen in their analysis due to the use of less data and many more parameters than in our model, potentially resulting in model overspecification. Whereas our analysis included all drivers with at least one season of three counting races (n=342), their analysis was restricted to drivers who competed in at least 40 races (n=124). This removes many drivers who provide useful comparisons with teammates, especially in the 1950s, where it excludes Alberto Ascari and Giuseppe Farina. Moreover, their model included weather and race distance effects, and assigned a unique parameter to every single chassis number. While this approach could in theory provide greater accuracy due to the greatly increased number of parameters, it undermines the ability to achieve meaningful comparisons between drivers, due to a lack of parameters in common between drivers.

Our models agree on some important points, including five of the top six drivers (Clark, Stewart, Fangio, Schumacher, and Alonso). However, they differ drastically on many others. Of the 124 drivers ranked by Eichenberger and Stadelmann (2009), 51 differ from our ranking by at least 20 positions, given in Table 6. These cases highlight the improved face validity of our model, with more reasonable rankings for highly esteemed champions of the sport (e.g., Lauda, Jones, and Mansell) and fewer undistinguished drivers with spuriously high rankings (e.g., Comas, Schell, Surer, Merzario, Boutsen, and Blundell).

As is true of any model, ours has some potential limitations. In dichotomizing non-finishes into driver and non-driver failures, we neglect the fact that some drivers may contribute to mechanical failures through rough driving or prevent mechanical failures through careful driving. We also do not attempt to attribute blame in the case of accidents, since this would require a very detailed and highly subjective analysis. For our model, we treat team reliability and speed as two independent factors. In reality, there may be some interactions, e.g., an unreliable car may need to be nursed by the driver to finish, resulting in reduced speed. Similarly, some cars may be more difficult to drive and induce more driver failures. So long as any such interactions affect all drivers for the same team to a similar extent, they will simply be absorbed into the estimated value of $\hat{\beta}_j$ and should not adversely affect predictions. However, there may also be interactions between drivers and teams, e.g., some cars may better suit certain driving styles than others. We do not attempt to model these interactions. Additionally, our model does not account for unfair treatment of drivers by the same team, including team orders. We did not include this factor because it is exceedingly difficult to objectively determine the extent to which it has affected results. Nevertheless, if one considers team orders to be distinct from driver performance (some observers may not), then drivers who received consistent preferential treatment may have slightly inflated adjusted scoring rates.

Table 6 Discrepancies between our model rankings (Phillips) and the rankings from the previous model of Eichenberger and Stadelmann (2009) (E&S09).

Driver	E&S09	Phillips	Driver	E&S09	Phillips
M Hawthorn	5	82	G Fisichella	69	33
G Hill	8	36	S Johansson	70	108
J Ickx	10	45	E Cheever	71	37
D Gurney	13	44	U Katayama	72	117
P Rodriguez	18	85	J Bonnier	73	99
P Hill	20	131	A Wurz	74	100
R Ginther	22	133	C Pace	76	35
E Comas	23	89	O Panis	79	54
M Trintignant	24	97	N Mansell	80	46
M Webber	27	59	J Verstappen	85	130
J Behra	28	119	P Streiff	86	112
F Cevert	29	65	C Klien	87	122
H Schell	30	144	J Oliver	88	142
M Surer	37	110	P Alliot	89	132
E Irvine	41	62	L Bandini	91	149
A Merzario	43	103	M Winkelhock	93	152
HH Frentzen	51	17	D Warwick	95	69
A Jones	52	21	J Scheckter	98	28
T Boutsen	54	80	S Modena	100	51
M Blundell	55	95	A Caffi	102	70
N Lauda	61	11	R Stommelen	104	136
F Massa	62	29	P Depailler	107	41
V Brambilla	63	88	G Villeneuve	112	50
JP Beltoise	64	32	JJ Lehto	117	139
J Trulli	66	30	A Nannini	123	98
M Hailwood	68	91			

The table shows all 51 cases where our rankings of the same driver differ by more than 20 places, using 3-year peak as our ranking of choice.

Perhaps the most important limitation of our current model is the omission of changes in driver performance across the career. In a few isolated cases, this leads to spurious results, such as the high rating of Tarso Marques in 2001 due to finishing in higher positions than his rookie teammate Fernando Alonso. It is reasonable to assume that drivers typically improve their performance with experience, reaching peak form near the middle of their career, and declining thereafter. This idea has been implemented in other sports performance models (Berry et al., 1999). However, analyses including such an experience effect in this model were found to achieve negligible improvement in goodness of fit, due to very large variability in the response to experience. This may be due to the fact that rookies have entered the sport with highly variable levels of experience. There have also been systematic differences between eras. For example, rookies in recent years have been allowed far less testing than they were a decade ago. Age is not a good proxy for experience either, since competitive drivers begin karting at a much earlier age today than in the past. A problem also arises in defining the experience level of drivers in 1950. Since this was the first year of the Formula One World Drivers' Championship, all drivers in 1950 were by definition in their first year. However, many of those drivers had competed for years or decades in Grand Prix races. It is unclear how to best treat this issue, and depending on how it is treated, the model can generate very strong present bias in its rankings. This is an issue that should be addressed in future models.

In summary, we have developed the first statistical model to provide sensible estimates of underlying driver performances in Formula One and predictions of lap-time data. Our model agrees with expert rankings in most cases, but also provides challenges some established expert opinions. Over time, our model can continue to be updated with new data and the rankings of drivers will consequently continue to evolve and improve in accuracy. In future, this same model could be readily applied to other motorsports with similar data structures, such as motorcycle racing, touring car racing, NASCAR, and IndyCar.

Acknowledgments: The author thanks OJ Walch, BM Whiteside, CL Murgo, the two anonymous reviewers, and the journal editor for their helpful comments on earlier versions of this paper.

References

Albert, J. 2006. "Pitching Statistics, Talent and Luck, and The Best Strikeout Seasons of All Time." *Journal of Quantitative Analysis in Sports* 2(1):Article 2.

Aitken, T. 2004. "Statistical Analysis of Top Performers in Sport with Emphasis on the Relevance of Outliers." *Sports Engineering* 7:75–88.

Autosport. 2009. "F1's Greatest Drivers." Accessed January 12, 2013 (http://f1greatestdrivers.autosport.com/).

Berry, S. M., C. S. Reese, and P. D. Larkey. 1999. "Bridging Different eras in Sports." *Journal of the American Statistical Association* 94:661–676.

Brown, M. and J. Sokol. 2010. "An Improved LRMC Method for NCAA Basketball Prediction." *Journal of Quantitative Analysis in Sports* 6(3).

Cohea, C. and M. E. Payton. 2011. "Relationships Between Player Actions and Game Outcomes in American FOOTBALL." *Sportscience* 15:19–24.

Davis, C. 2000. *The Best of the Best: A New Look at the Best Cricketers and their Changing Times*. Sydney, Australia: ABC Books.

Eichenberger, R. and D. Stadelmann. 2009. "Who is the Best Formula One Driver? An Economic Approach to Evaluating Driver Talent." *Economic Analysis & Policy* 39:389–406.

F1 Racing. 1997. 100 Greatest Drivers of all Time. June 1997 issue.

F1 Racing. 2004. 100 Greatest Drivers of all Time. June 2004 issue.

F1 Racing. 2008. 100 Greatest Drivers Ever. May 2008 issue.

F1 Racing. 2012. Man of the Year. December 2012 issue.

Henry, A. 2008. *The Top 100 F1 Drivers of All Time*. Cambridge, UK: Icon Books Ltd.

Hilton, C. 1991. *Ayrton Senna: The Hard Edge of Genius*. London, UK: Corgi Books.

Jones, B. 1995. *The Ultimate Encyclopedia of Formula One*. London, UK: Carlton Books.

Kvam, P. H. 2011. "Comparing Hall of Fame Baseball Players Using Most Valuable Player Ranks." *Journal of Quantitative Analysis in Sports* 7(3):Article 19.

Kvam, P. H. and J. S. Sokol. 2006. "A Logistic Regression/Markov Chain Model for NCAA Basketball." *Naval Research Logistics* 53:788–803.

Oberstone, J. 2011. "Uncovering Europe's Best Goalscorers from the 2009-2010 Season." *Journal of Quantitative Analysis in Sports* 7(1):Article 1.

Piette, J. and S. T. Jensen. 2012. "Estimating Fielding Ability in Baseball Players Over Time." *Journal of Quantitative Analysis in Sports* 8(3).

Rendall, I. 1991. *The Power and The Glory: A Century of Motor Racing*. London, UK: BBC Books.

Sonas, J. 2005. "Chessmetrics." Accessed January 12, 2013 (http://www.chessmetrics.com/cm/).

Smith, R. 2011. *Formula 1: All the Races: The World Championship Story Race-by-Race: 1950-2011*. Sparkford, Yeovil, Somerset, UK: Haynes Publishing.

West, B. T. 2006. “A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament.” *Journal of Quantitative Analysis in Sports* 2(3).
West, B. T. and M. Lamsal. 2008. “A New Application of Linear Modeling in the Prediction of College Football Bowl Outcomes and the Development of Team Ratings.” *Journal of Quantitative Analysis in Sports* 4(3).
White, C. and S. Berry. 2012. “Tiered Polychotomous Regression: Ranking NFL Quarterbacks.” *American Statistician* 56: 10–21.