Andrew Bell*, James Smith, Clive E. Sabel and Kelvyn Jones

Formula for success: Multilevel modelling of Formula One Driver and Constructor performance, 1950–2014

DOI 10.1515/jqas-2015-0050

Abstract: This paper uses random-coefficient models and (a) finds rankings of who are the best formula 1 (F1) drivers of all time, conditional on team performance; (b) quantifies how much teams and drivers matter; and (c) quantifies how team and driver effects vary over time and under different racing conditions. The points scored by drivers in a race (standardised across seasons and Normalised) is used as the response variable in a cross-classified multilevel model that partitions variance into team, team-year and driver levels. These effects are then allowed to vary by year, track type and weather conditions using complex variance functions. Juan Manuel Fangio is found to be the greatest driver of all time. Team effects are shown to be more important than driver effects (and increasingly so over time), although their importance may be reduced in wet weather and on street tracks. A sensitivity analysis was undertaken with various forms of the dependent variable; this did not lead to substantively different conclusions. We argue that the approach can be applied more widely across the social sciences, to examine individual and team performance under changing conditions.

Keywords: cross-classified models; formula 1; MCMC; multilevel models; performance; sport.

1 Introduction

Formula 1 (F1) is a sport of genuine global appeal. Established in 1950, F1 has also grown into a huge business enterprise, with sponsorship and commercialism drawn to the sport by the 527 million television viewers from 187 different countries (in 2010). After the Football World Cup and the Summer Olympic Games, it is the largest sporting event in terms of television audience (Judde, Booth and Brooks 2013).

Many of the F1 teams that compete employ statistical analysts to analyse race results; however, these are in general kept undisclosed so that teams are able to keep any tactical advantages these analyses offer to themselves. As such, there are only a handful of papers in the public domain that have done systematic statistical analysis of F1 race results, and these are focused on the question of who is the best driver, and do not consider the question of how much teams and drivers matter in different contexts. However, the large fan base ensures that there is a plethora of publicly available data on F1 race results online, and the potential for statistical analysis of these data is large.

This paper uses cross-classified multilevel models to produce a more complete picture of what influences performance in F1 races. As well as producing rankings of F1 drivers that control for the influence of teams, the models are able to partition variance to see the extent to which teams and drivers matter. The key methodological innovation of this paper is the use of complex variance functions, in which the variance depends on predictor variables, to see how team and driver influences have changed over time, and differ by different driving conditions, as well as to see how driver rankings vary by these conditions. Such an approach has potential application beyond F1 as the methodology is applicable in subject areas throughout the social sciences and beyond, such as when examining changing team and individual performance in firms.

2 Formula 1

The academic literature surrounding Formula 1 is relatively limited. However, that literature is cross-disciplinary, involving, for example, computational simulations of race results (Bekker and Lotz 2009; Loiacono et al. 2010), economic approaches that consider the importance of, and

***Corresponding author: Andrew Bell,** The University of Sheffield – Sheffield Methods Institute, Sheffield, United Kingdom of Great Britain and Northern Ireland, e-mail: andrew.j.d.bell@sheffield.ac.uk, http://orcid.org/0000-0002-8268-5853
James Smith, Clive E. Sabel and Kelvyn Jones: University of Bristol – School of Geographical Sciences, Bristol, United Kingdom of Great Britain and Northern Ireland.
http://orcid.org/0000-0001-8398-2190 (K. Jones)

adaptability of, F1 teams as firms (Jenkins and Floyd 2001; Jenkins 2010), knowledge transfer between teams (Jenkins and Tallman 2016) analyses of car design over time from an engineering perspective (Dominy and Dominy 1984; Dominy 1992), analyses of specific tracks (Alnaser et al. 2006) and their impacts on tourism (Henderson et al. 2010), and historical approaches to the sport (Hassan 2012). There have been a few statistical analyses of race results, although these are often limited to a few races or seasons (Bekker and Lotz 2009; Muehlbauer 2010).

2.1 Who is the best driver?

As far as we are aware, there are only two studies that have analysed data on F1 over the entirety of its history, and in both cases the aim of the studies was to find out which driver, controlling for the team that they drive for, is the greatest of all time. Whilst there are many examples of experts attempting to form all-time rankings of F1 drivers, these are almost exclusively based on subjective professional judgement and not statistical analysis. Given the differences in the cars driven by different drivers in different teams, the question of who is the best driver is a controversial one: being able to consistently win in the best car is not necessarily enough.

Eichenberger and Stadelmann (2009) consider finishing position as the dependent variable, control for team-years using dummy variables in a standard OLS-estimated single-level regression, and also control for a range of other variables relating to both drivers and the racing conditions. The results are for the most part intuitive: Juan Manuel Fangio comes out as the best driver, with other highly regarded drivers (Jim Clark, Michael Schumacher, Jackie Stewart, Alain Prost, Fernando Alonso) in the top ten. However, according to Phillips (2014), there are some unexpected results in these rankings, in particular noting Mike Hawthorn coming surprisingly high at number five.

Phillips argues that this unexpected result is caused by the use of finishing position as the dependent variable, and prefers the use of (adjusted) points scored as an appropriate measure, since he argues this is a better measure of achievement in F1; the season average of these scores is used in the analysis. Like Eichenberger and Stadelmann, Phillips controls for teams. He additionally controls for competition effects (such that drivers are penalised in the ranking for appearing in less competitive seasons) and, for driver withdrawals, separates driver faults and technical faults, to ensure drivers are not penalised for team errors. His rankings are based on drivers' 3-year (or 5-year) peak performance (rather than their whole career). For Phillips' rankings, it is Jim Clark who comes out top; Stewart, Schumacher, Fangio and Alonso make up the rest of the top 5. However, these results also throw up a few surprises; for example James Hunt is ranked at number six (Phillips argues that Hunt is indeed underrated by experts).

In sum, these two previous analyses have shown many consistencies, with drivers regarded as "greats" by experts coming out at the top. However, different decisions regarding model specification lead to different results, and both the above analyses produce some results that one might consider surprising. This is not to doubt the validity of those results – simply to state that if you ask slightly different questions, by defining rankings differently, you are likely to get slightly different results. We discuss some of these modelling decisions, in the context of our own modelling strategy, in the methods section below.

2.2 How much do teams, and drivers, matter?

Formula 1 is an unusual sport in that it is a hybrid of both a competition for individual drivers, and a competition between teams. Thus, each season there is both a drivers' and a constructors' championship, with both considered important by F1 fans. Often drivers will move between teams, and teams will change (in the technologies they use, the staff involved in designing and developing the car and race strategies, the physical components of the car, and in the drivers that make up the team) year on year. The question of whether teams or drivers are most important to formula 1 race results is of great interest to many; however, this question has been quantitatively assessed briefly only once (Phillips 2014: 267) as far as we are aware. Yet because there is a relatively large amount of movement of drivers between teams, the question can be answered with appropriate statistical techniques that can model this complexity.

Certainly, there are reasons why some teams should in general outperform others. Certain teams have more funds, are able to employ the best engineers, statisticians and tacticians, and use more advanced technology than other teams. For example, the Williams cars that were so successful in the mid-1990s included computerised driver aids; Brawn GP in 2009 used double diffuser technology, which gave them an advantage and led Jenson Button to win the championship despite having won only one race in the nine previous years of his career. As well as these specific innovations, team/car performance will depend on "factors such as aerodynamic efficiency, brakes, engines, gearbox, fuel, and more recently kinetic energy recovery systems", which change between teams and year on year (Horlock 2009: 4).

Experts generally agree that the team matters more than the driver, although the extent to which this is true is hotly debated. Driver Nico Rosberg has stated the respective contributions to be 80% teams and 20% drivers (Spurgeon 2009). Others have argued that only the best drivers are really capable of making a difference in F1 races, with Allen (2000) giving Michael Schumacher as an example of this. A key contribution of this paper is to statistically evaluate the extent to which teams and drivers matter.

2.3 Teams and drivers over time

Having said this, it seems likely that the importance of teams and the importance of drivers would change as technologies develop. On one hand, technological developments mean that highly advanced cars can really stand out; if so, the team effect becomes greater because the best funded, most prestigious teams will be able to apply those technological advancements better. On the other hand, there has been an increase in regulation that means car design is becoming increasingly homogenised (Dominy 1992). If cars are all similar, there is little that teams can do to differentiate themselves, apart from with superior tactics and employ better drivers.

2.4 Teams and drivers in different places

As well as changing over time, it is possible that the effects of teams, and drivers, could be different on different race tracks. Tracks vary between street and temporary tracks, which often contain a large number of corners, and purpose-built permanent circuits that often have fewer turns. Different tracks can suit a particular driver's style. However, the same can be said in regard to track suitability for different cars; for example, in 2011, the Red Bull car had unmatchable downforce (meaning it can carry a higher level of speed through corners) while Mercedes powered engines offered greater straight-line speed (Allen 2011). Thus, one might expect the Mercedes car to have performed better on permanent circuits, and Red Bull better on street circuits.

In basic terms, driver performance comes down to the ability to overtake one's rivals, aided in recent years by Kinetic Energy Recovery Systems and Drag Reduction Systems. Generally speaking this is most easily done on long straight stretches of track, because the overtaking car can benefit from slipstreaming behind the car in front. One might expect, therefore, that drivers might matter more on more sinuous tracks where the opportunities to overtake are reduced and so only the best drivers are able to successfully attempt overtaking manoeuvres. Indeed, many judge the Monte Carlo circuit in Monaco (a street circuit) to be the greatest test of a driver's skill in F1, with the best drivers, rather than the best cars, being rewarded (Collings and Edworthy 2004).

2.5 Teams and drivers in different weather conditions

The final factor considered here that may affect the importance of teams and drivers are the weather conditions in a particular race. F1 teams spend large amounts on weather forecasting and meteorologists, in order to predict weather minute-to-minute and thus make strategic decisions on how to deal with weather conditions. In theory, therefore, the best teams will make the best strategic decisions, which could increase the team's role in performance. However, because rain reduces the grip that cars can maintain, the ability of a driver to handle the car becomes increasingly important, and drivers are more likely to make mistakes (Spurgeon 2011). Indeed, certain drivers, such as Ayrton Senna, Michael Schumacher and Lewis Hamilton, are noted by experts for their abilities in wet weather, although in some cases this is based on a few outstanding notable performances rather than a more general trend. Overall, we might expect rain to introduce additional unpredictability into races, with even the best teams and drivers more prone to making mistakes.

3 Methodology

In sum, this paper is looking to answer three inter-related questions: who are the greatest F1 drivers of all time, how much do teams and drivers matter, and how much does the latter change over time, on different tracks, and in different weather conditions.

These questions can be answered thanks to the multilevel structure that is inherent to F1 race results data. Specifically, each observed race result can be nested within a driver, a team-year, and a team. This is not a strict hierarchical structure: since drivers move between teams over their careers, and teams contain multiple drivers, the levels cannot be hierarchically nested. Instead, it is a cross-classified structure. The variables used in the

Table 1: Variables used in this study.

Variable	Description
Fixed part variables	
Points	The dependent variable: Number of points scored, based on the scoring system used from 1991 to 2002: 1st place: 10 pts, 2nd place: 6 pts, 3rd–6th place: 4–1 points; Fractional points are awarded for lower positions, including those who fail to finish the race (i.e. the first driver to drop out will finish last, the second second from last, and so on)
Year	Year of race – from 1950 to 2014
Weather	Dummy variable coded 1 if the race is in any way affected by rain (some 15% of all observations), and 0 if not. Whilst this is a somewhat crude measure, data for a more exact measure (like proportion of race conducted in wet conditions) is not readily available
Track type	Categorical variable classifying the type of track: Permanent – a permanent track (76% of observations), Street – a race that occurs on public streets (which are temporarily closed to the public (19%), Temporary – a temporary race track that is not on public streets (5%)
Ndrivers	Number of entrants in the race. Mean = 23.3, SD = 3.1
Comp	Competitiveness of the race based on the career performance of drivers in a given race (see Section 3.2). Mean = 0.5, SD = 0.02
Random part variables	
Driver	The driver of the car (e.g. Michael Schumacher)
Team	The team name (e.g. Ferrari)
Team-year	Identifier of the team-year (e.g. Ferrari1992)

Teams are defined based on the chassis-engine-constructor combination, unless a constructor changes the chassis or engine used mid-season, in which case the team is judged to continue as that team. Whilst this is problematic where a team changes the car for one driver and not another, this problem only affects a small minority of team-years.

modelling are summarised in Table 1; these come from two online sources,[1] and incorporate all 905 F1 races (excluding Indianapolis 500 races) between 1950 and 2014.

The dependent variable $Points_i$ is the points scored by a driver in a race. Following Phillips (2014), we deploy for all the races the points scheme used between 1991 and 2002 (10 points for 1st, 6 for 2nd, and then 4 to 1 points for 3rd–6th), and use fractional points for lower positions: $(0.1^{0.2})\wedge$p-6 where p is the finishing position. Where drivers do not finish a race, they are still ranked on the basis of when they dropped out of the race (with the first to drop out being in last place). A basic, null model can be expressed algebraically as:

$$\begin{aligned} Rankit(Points)_i &= \beta_0 + u_{Driver} + v_{Team} + w_{TeamYear} + e_i \\ u_{Driver} &\sim N(0, \sigma_u^2) \\ v_{Team} &\sim N(0, \sigma_v^2) \\ w_{TeamYear} &\sim N(0, \sigma_w^2) \\ e_i &\sim N(0, \sigma_e^2) \end{aligned} \tag{1}$$

Such that σ_u^2, σ_v^2 and σ_w^2 summarise the between-driver, between-team and between-team-year[2] variance, respectively, and σ_e^2 summarises the within-race variance net of driver, team and team-year characteristics.[3] The model assumes Normality of the random effects and we have therefore used the Rankits of the points which are the expected Normal order statistics of the form that are used in a Normal probability plot (Chambers et al. 1983). The nature of the response variable is considered on more detail below.

3.1 Finding driver rankings

In the model the normalized points scored by an entrant is allowed to vary by drivers, teams, and team-years, and their respective variances (σ_u^2, σ_v^2, and σ_w^2) are estimated and assumed to be Normally distributed. The driver-level

1 www.race-database.com and www.f1-facts.com/stats. All data was available from these sites as of February 2015.

2 Alternatively, the team-year residuals can be thought of a random draw from a distribution with a mean value of the team residual.

3 One could additionally include a driver-year level in this model, to assess the extent to which drivers vary across their career. With our data, whilst this led to a modest improvement in the model according to the Deviance Information Criterion (DIC), a penalized measure of badness of fit (Spiegelhalter et al. 2002). The model took a long time for the driver-year variance parameter to converge (with a relatively low ESS score even after 500,000 iterations), and the modal estimate for that variance was zero, suggesting there is little or no variation at the driver-year level. Moreover, the driver rankings were identical to the model without the driver-year level included.

residuals, u_{Driver}, can be ranked and thus represent a rank of driver ability, controlling for team and team-year.

The advantage of this approach is that any variance is automatically partitioned into the respective levels, and there is no need to control for many variables in order to achieve appropriate rankings. Thus, and contrary to Phillips (2014), we do not need to treat driver failures and team/car failures differently – the model will automatically apportion the latter into the team or team-year levels and so they will not unfairly penalise a driver who suffers such failures. In contrast, a driver that makes a large number of mistakes (whilst his team mate does not) will be penalised (see Section 3.3.3 for more on this).

Having said this, there are a number of ways in which points scored can be affected that is not a result of teams/team years, but that should also not penalise drivers. First, where there are fewer drivers in a race, the average points scored will be higher, meaning drivers who generally compete against fewer people will tend to get more points despite not necessarily being better drivers. Second, drivers who are competing against better drivers will tend to perform worse than those competing against worse drivers. Thus we add two predictor variables to the fixed part of the model to take account of these concerns. For the former, we control for the number of drivers in a given race ($Ndrivers_i$). For the latter, we take each driver's mean finishing position (divided by the number of drivers in their races) across their career, and average these by race occasions, and use this to control for the competitiveness of the race ($Comp_i$).[4] Thus the fixed part of equation 1 can be extended to:

$$Rankit(Points)_i = \beta_0 + \beta_1 Ndrivers_i + \beta_2 Comp_i + u_{Driver} + v_{Team} + w_{TeamYear} + e_i \qquad (2)$$

Here, we would expect β_1 to be negative (given that every additional driver will decrease the average points scored), and β_2 to be positive (since racing against drivers that usually perform well should reduce the points scored by a driver). However, these values are not for the most part of substantive interest; the important thing is that they are controlled for when extracting the driver-level residuals u_{Driver}. The random part of the model remains as in equation 1.

4 Note that $Comp_i$ is a variable measured at the "race-occasion" level. We do not include race occasion as a random effect because there is not enough variation in it to make its variance significant (the only source of variation is the number of drivers, which is entirely controlled out by $Ndrivers_i$). The $Comp_i$ variable is problematic in that it uses response data to form it. However, given this variable is a control, rather than being of primary interest in itself, this method of constructing the variable seems appropriate.

3.2 Driver, team and team-year variance functions

As well as uncovering the driver level residuals, the estimated variances for the team, team-year and driver residuals allow us to compare the effects of each: that is, to see which of teams, team-years and drivers matter the most. For this we only consider data from 1979 onwards. The reason for this is that, prior to this date, the team-structure of F1 was less clearly defined: for example, wealthy drivers would enter their car in just a few races because the costs and regulations required to do so were not as prohibitive as they were/are in later years. Many teams and drivers only competed in one race. It is thus much more difficult to define the team level, since teams did not always function in the same way that they do today. Given this, and our primary interest in team and driver effects in the modern sport, only the more recent post-1979 data were used in order to delineate between team and team-year variance, where the former represents consistent team effects that persist over the years, and the latter represents within-team fluctuations where teams perform particularly well or badly in a particular year. Other than the reduction in data, the model used is exactly the same as that in equation 2.

This model is further extended to include other variables: year, weather, and track type. In order to keep the models relatively simple and to avoid convergence issues, these variables were included in the model individually in separate models. One interesting nuance of the points scored dependent variable is that it does not vary a huge amount between races (only varying by the number of drivers in a race), which means that one cannot find effects on points scored of race-level variables such as those used above. Whether it rains or not, the points scored will still be filled. However, these variables can be included in the random part of the model to see how team, team-year and driver effects vary across the variables. Furthermore, for categorical variables (that is, weather and track type), these same models can produce separate driver rankings for different values of those variables (thus showing if drivers are ranked differently in different driving conditions).

To assess the effect of year on driver, team and team-year effects, the model in equation 2 can be extend to:

$$\begin{gathered} Rankit(Points)_i = \beta_0 + \beta_1 Ndrivers_i + \beta_2 Comp_i \\ + \beta_{3i} Year_i + u_{0Driver} + v_{0Team} + w_{0TeamYear} + e_i \\ \beta_{3i} = \beta_3 + u_{1Driver} + v_{1Team} + w_{1TeamYear} \\ \begin{bmatrix} u_{0Driver} \\ u_{1Driver} \end{bmatrix} \sim N\left(0, \begin{matrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{matrix}\right) \end{gathered}$$

$$\begin{bmatrix} v_{0Team} \\ v_{1Team} \end{bmatrix} \sim N\begin{pmatrix} 0, & \sigma^2_{v0} & \\ & \sigma_{v01} & \sigma^2_{v1} \end{pmatrix}$$
$$\begin{bmatrix} w_{0TeamYear} \\ w_{1TeamYear} \end{bmatrix} \sim N\begin{pmatrix} 0, & \sigma^2_{w0} & \\ & \sigma_{w01} & \sigma^2_{w1} \end{pmatrix}$$
$$e_i \sim N(0, \sigma^2_e) \tag{3}$$

We would expect β_3, the fixed effect of Year, to be approximately zero. However, the driver, team and team-year differentials from this effect ($u_{1Driver}$, v_{1Team} and $w_{1TeamYear}$, respectively) could be non-zero, and as such, the variance at each level could vary by Year. The extent to which the variance changes are quantified with variance functions (Bullen, Jones and Duncan 1997; Goldstein 2010): the driver level, team level and team-year level variances are calculated, respectively as:

$$Variance_{Driver} = \sigma^2_{u0} + (2\sigma_{u01} * Year_i) + (\sigma^2_{u1} * Year_i^2) \tag{4}$$

$$Variance_{Team} = \sigma^2_{v0} + (2\sigma_{v01} * Year_i) + (\sigma^2_{v1} * Year_i^2) \tag{5}$$

$$Variance_{TeamYear} = \sigma^2_{w0} + (2\sigma_{w01} * Year_i) + (\sigma^2_{w1} * Year_i^2) \tag{6}$$

Note that we additionally tested for a quadratic year term (thus allowing a quartic variance function); no improvement in the model was observed based on the DIC (see below).

The model is much the same for weather (with the Year variable replaced by the dummy variable for rain). For the track type variable, there are three categories, so two dummy variables ($Street_i$ and $Temporary_i$) contrasted against the reference category Permanent_i, must be included in the model. This model is specified as:

$$Rankit(Points)_i = \beta_0 + \beta_1 Ndrivers_i + \beta_2 Comp_i + \beta_{3i} Temporary_i$$
$$+ \beta_{4i} Street_i + u_{0Driver} + v_{0Team} + w_{0TeamYear} + e_i$$
$$\beta_{3i} = \beta_3 + u_{1Driver} + v_{1Team} + w_{1TeamYear}$$
$$\beta_{4i} = \beta_3 + u_{2Driver} + v_{2Team} + w_{2TeamYear}$$
$$\begin{bmatrix} u_{0Driver} \\ u_{1Driver} \\ u_{2Driver} \end{bmatrix} \sim N\begin{pmatrix} & \sigma^2_{u0} & & \\ 0, & \sigma_{u01} & \sigma^2_{u1} & \\ & \sigma_{u02} & \sigma_{u12} & \sigma^2_{u2} \end{pmatrix}$$
$$\begin{bmatrix} v_{0Team} \\ v_{1Team} \\ v_{2Team} \end{bmatrix} \sim N\begin{pmatrix} & \sigma^2_{v0} & & \\ 0, & \sigma_{v01} & \sigma^2_{v1} & \\ & \sigma_{v02} & \sigma_{v12} & \sigma^2_{v2} \end{pmatrix}$$
$$\begin{bmatrix} v_{0TeamYear} \\ v_{1TeamYear} \\ v_{2TeamYear} \end{bmatrix} \sim N\begin{pmatrix} & \sigma^2_{w0} & & \\ 0, & \sigma_{w01} & \sigma^2_{w1} & \\ & \sigma_{w02} & \sigma_{w12} & \sigma^2_{w2} \end{pmatrix}$$
$$e_i \sim N(0, \sigma^2_e) \tag{7}$$

The variance function for the driver level becomes:

$$Variance_{Driver} = \sigma^2_{u0} + (2\sigma_{u01} * Temporary_i) + (\sigma^2_{u1} * Temporary_i^2)$$
$$+ (2\sigma_{u02} * Street_i) + (2\sigma_{u12} * Temporary_i * Street_i)$$
$$+ (\sigma^2_{u1} * Street_i^2) \tag{8}$$

Thus, the driver variance on a permanent circuit is estimated as σ^2_{u0}, for a temporary circuit it is $\sigma^2_{u0} + (2\sigma_{u01} * Temporary_i) + (\sigma^2_{u1} * Temporary_i^2)$, and for a street circuit it is $\sigma^2_{u0} + (2\sigma_{u02} * Street_i) + (\sigma^2_{u1} * Street_i^2)$. There are equivalent variance functions for the team and team-year levels.

This same model can also be used to calculate separate driver rankings for different conditions (using the entirety of the data from 1950, in order to include all drivers). Thus, in equation 7 the driver rankings for permanent circuits is given by $u_{0Driver}$, for temporary circuits they are given by $u_{0Driver} + u_{1Driver}$, and for street circuits they are given by $u_{0Driver} + u_{2Driver}$. Rankings for dry and wet weather conditions can be found in a similar way.

All models were fitted in MLwiN v2.35, (Rasbash et al. 2013), using Monte Carlo Markov Chain (MCMC) estimation (Gelfand and Smith 1990; Browne 2009).[5] Plausible but arbitrary starting values were used for an initial model, and once that model had run for 10,000 iterations, these estimates were used as starting values for the final models, which were run for 500,000 iterations (following a 500 iteration burn in). These were found to be sufficient to produce healthy-looking visual diagnostics (that is, chain trajectories that have the appearance of white noise), and effective sample sizes (ESS) of over 1000 for all parameters.

In all cases, variables were allowed to vary at each level one at a time, and model improvements were calculated using the Deviance Information Criterion (DIC, Spiegelhalter et al. 2002). Where the model shows substantial model improvement (a decrease in DIC of at least 4) when a variable's effect was allowed to vary at any level, the full model (with the variable effect allowed to vary at all three levels) was run and these results presented.

5 The default prior specifications in MLwiN were used: uniform distributions for the fixed effect estimates, inverse Gamma distributions for the variances in the null models, and inverse Wishart distributions for the variance-covariance matrices when effects were allowed to vary in other models. The models were also estimated using Uniform priors for the variances, as suggested by Gelman (2006), but the results were almost identical and the substantive conclusions did not change. For more on this see Browne (2009).

3.3 Comparison to the methods of previous F1 studies

3.3.1 Advantages of a multilevel approach

There are a number of ways in which our methods differ from those of previous attempts to rank F1 drivers (Eichenberger and Stadelmann 2009; Phillips 2014). The most notable of these is our use of multilevel models, or random effects (RE) models, rather than using fixed effects (FE) to represent drivers and teams (Bell and Jones 2015). We argue that there are a number of advantages to the RE approach. First, RE allows us to include a team as well as a team-year level in our model; with FE, only a team-year level can be included because this takes all the degrees of freedom associated with the team level. Thus, we can distinguish between enduring team "legacy" effects, that do not change year-on-year, and more transient effects, where teams perform better or worse from one year to the next. Second, the RE approach allows for the modelling of variance functions and so allows the effects of drivers or teams to vary with covariates, as well being able to find different driver rankings for discrete covariate values. Whilst this could be achieved with a FE model, it would require a large number of interaction terms in the model for which parameters would need to be estimated, and would quickly become unwieldy. Third and crucially for producing driver rankings, in RE models unreliably estimated higher-level residuals are "shrunk" to the mean – thus we are able to include, for example, drivers that have only driven in very few races, without being concerned that they might produce spurious results. Thus the residuals j at level u are in effect multiplied by reliability λ_j, calculated as:

$$\lambda_j = \frac{\sigma_u^2}{\sigma_u^2 + (\sigma_e^2 / n_j)} \tag{9}$$

where n_j is the sample size of driver-level entity j, σ_u^2 is the between-driver variance, and σ_e^2 is the level 1 variance. Thus competing in only one race and winning in a poor car is not enough in our model to do well – drivers must perform consistently well to be sure their good performances are not simply down to chance. This also means that drivers who are unlucky with random car problems, but have not been in enough races for that luck to even out, will not be unfairly disadvantaged since that unreliability is accounted for in their residual (see Section 3.3.3).

3.3.2 The choice of dependent variable

For the dependent variable, there are broadly three choices: (a) season-long points earned (as used by Phillips 2014), (b) individual race finishing position (as used by Eichenberger and Stadelmann 2009), and (c) individual race points scored. In this paper, we choose the latter. We avoid (a) because we want to utilise the full uncertainty of race results in our modelling. The shrinkage that is applied to drivers as described above would be incorrectly applied when meaned across seasons, since a driver that raced in one race in one season would be judged to have the same certainty as a driver racing in every race of that season. Phillips avoids this problem by removing drivers with very few races in a season (and removing what non-finishing results that were judged to be "non-driver failure" – see Section 3.3). An advantage of our approach is that we do not need to do this, and can model all types of high and low performance, as well as all drivers that have ever competed in any race.

The difference between individual race points scored and finishing position is actually rather small. In order to apply a Normal model to such data, the latter needs to be transformed, meaning that the resulting dependent variable is actually very similar to the finishing position (correlation of 0.99). We choose points scored because it is a more realistic measure of what is valuable to drivers, although the results found from each are rather similar.

This leaves the question of what transformation we should apply to points scored. From testing a range of different transformations, it is notable that (a) none of the transformations produce level 1 residuals that are very far from Normality, and there are no extreme outliers, and (b) there is remarkably little difference in the driver rankings found as a result of using the different dependent variables. However, because it shows the closest relation to Normality, we choose as our dependent variable the rankit transformation of Points scored.[6]

3.3.3 Treatment of driver and non-driver failure

A notable difference between our approach and that of Phillips (2014) is our treatment of driver and non-driver failures. Phillips excludes data for which drivers failed to finish a race for reasons that were not their fault – his

6 Another possibility could be to use an exploded logit model (also called a Plackett-Luce, or rank-ordered logit model) for rankings (Allison and Christakis 1994; Skrondal and Rabe-Hesketh 2003; Anderson 2014; Baker and McHale 2015; Glickman and Hennessy 2015). However, due to the complexity of our model, the large size of the dataset we use (especially after it has been "exploded"), and the already long chain lengths required to make the model converge, this was deemed unfeasible.

argument being that these results should not count against them. There are a number of problems with this approach. First, the distinction between driver and non-driver failure will always be somewhat arbitrary: for example a failure could be the fault of the car, but with more careful driving the risk of such a failure might be reduced. Second, non-driver failure that does not result in retirement from the race is not discounted in the same way, which seems somewhat inconsistent. Third, one of the advantages of our approach is our ability to use all races and drivers in history with no exclusion criteria; excluding driver-races with non-driver failures would make our analysis incomplete. And fourth, we argue that it is not necessary to make the distinction.

The latter argument rests of an assumption that there are three causes of failure: (1) failures that are the fault of the driver, (2) failures that are the fault of the team or car, and (3) random failures that are nobody's fault. We would expect worse drivers to have more driver failures, worse teams to have more team failures (and for both drivers in that team to be equally affected by such failures), and random failures to be randomly distributed across driver-races. If this is the case, then the multilevel structure will partition these failures to the appropriate levels, and only driver error will count against the driver in the driver-level residual. This approach may be problematic for drivers that only raced in a small number of races (and, due to sheer bad luck, experienced lots of random failures in those races). However, such drivers would experience a high degree of shrinkage (indicative of the unreliability of their residual – see Section 3.3.1). Over the course of a driver's career, we would expect such random failures would even out across drivers' careers, meaning any differences between drivers are most likely to be the fault of the drivers themselves.[7] In our view this approach to errors is more appropriate than any arbitrary classification into driver and non-driver faults. If a driver consistently has more failures than their teammate in the same car, this is statistical evidence that those failures were at least in part the fault of the driver, who should be "penalised" accordingly.

Other notable differences include:

– The use of a different (simpler) correction for competitiveness compared to Phillips 2014 (see Section 3.1).
– We consider drivers' entire careers, rather than their 1, 3, or 5 year peak performance (as considered by Phillips (2014). To perform well in our rankings, drivers must perform consistently well, and not just for a portion of their career.
– We control for fewer variables than Eichenberger and Stadelmann (2009), since we do not want to control out any effects that should more appropriately be included in the random effects.
– We are able to use data from all drivers and teams in history, whereas previous models have had inclusion criteria.

With each difference, we are not necessarily claiming that our models are better than previous models; rather that we are defining what a good driver is in subtly different ways that will impact the results that are produced.

4 Results

4.1 Who is the best F1 driver of all time?

Figure 1 presents the driver level residuals for what are the top 20 F1 drivers of all time according to our model[8], controlling for teams, team-years, the number of drivers in each race and the competitiveness of the race. It is Juan Manual Fangio who comes out as the top driver, followed by Alain Prost, Jim Clark, Ayrton Senna and Fernando Alonso. Of drivers currently racing, Alonso comes out top, followed by Sebastian Vettel, Lewis Hamilton, Nico Rosberg and Jenson Button.

It should be noted that there are rather wide confidence intervals around each driver. There are two reasons for this. First, the measure being estimated by the model is the (transformed) points scored in a race, not in a season, meaning the confidence bounds reflect the uncertainty that exists race-to-race, compared to smaller season-to-season uncertainty. And second, it reflects the level of uncertainty that will exist in any "best of all time" rankings, given the question of who is the best is inherently uncertain. However, because the residuals have already been shrunk to the mean, there is some protection against over-interpretation regardless.

Michael Schumacher, who holds the record for the most championships and race victories of any driver in

7 A similar logic is used, for example, to identify poorly performing hospitals on the basis of having an unusually high mortality rate, (for example see Taylor 2013). In such a case, there is no need to attempt to separate random and hospital failures – the higher mortality rate given the sample size presents enough statistical evidence.

8 Further model details, including parameter estimates, extended rankings, additional graphics, and model predictions can be found at http://eprints.whiterose.ac.uk/96995/.

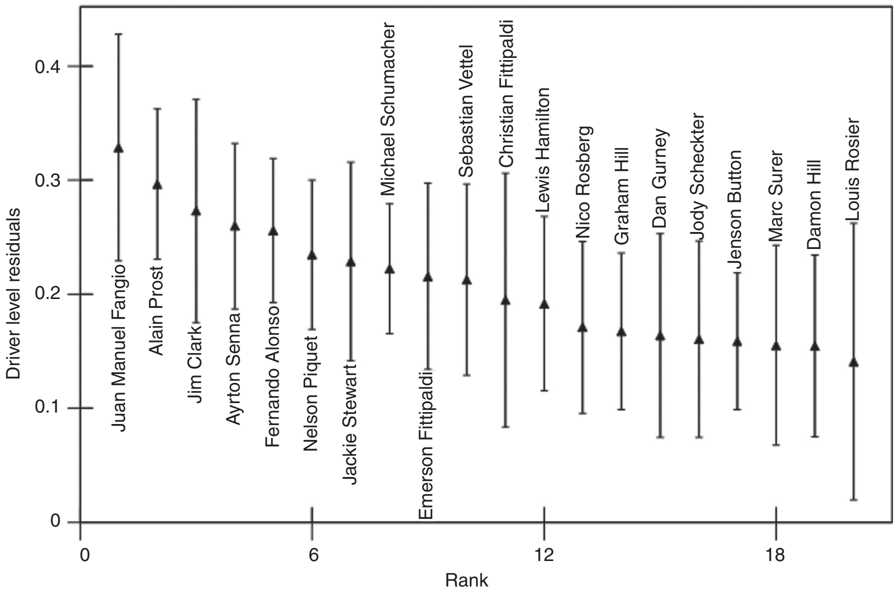

Figure 1: Plot of the top 20 driver-level residuals, with 95% Bayesian credible intervals (based on Goldstein and Healy 1995), representing the top 20 drivers of all time (1950–2014) according to our model. Number of drivers and race competitiveness are controlled. The residual value represents the difference when compared to an average driver driving for an equally good (or the same) team, with higher numbers indicating a better position. We can be confident that drivers with CIs that do not overlap the zero line are "better than average".

Formula 1, comes in a relatively modest eighth place. This is in part because those victories were won in an excellent car, but also because his ranking is dragged down by his more recent post-retirement performances (2010–2012) when he performed less well than in the main part of his career and crucially was generally outperformed by his Mercedes teammate Nico Rosberg. Thus, we re-ran the analysis with the latter section of Schumacher's career treated as a separate driver. In this formulation, pre-2006 Schumacher's ranking rises to 3rd and Nico Rosberg's ranking falls from 13th to 49th. This is because Schumacher's high standing as a driver in the model effectively deflated 2010–2012 Mercedes' team ranking in the first model, meaning Rosberg's performances appeared more impressive. When treated as separate drivers, post-retirement Schumacher performed less well, the Mercedes team effect appears greater, and so Rosberg's performances no longer stand out compared to his team.

We were additionally able to produce rankings specific to certain weather conditions and track types (not shown). In general, these showed similar results – Fangio remained top in all but one of the categories and the top drivers still populate the top positions. However, there are some interesting points to note. In particular, whilst the reputations of Ayton Senna and Michael Schumacher for being very good wet weather drivers are justified by the data (pre-2006 Schumacher is estimated to be the second best wet weather driver of all time, whilst Senna is the third best and ahead of his rival Alain Prost), the similar reputation of Lewis Hamilton is not born out statistically (his ranking does not differ between wet and dry conditions).

4.2 How much does the driver and the team matter?

The variances for the team, team-year and driver levels (for the years 1979–2014 were 0.066, 0.042 and 0.017, respectively (controlling for number of drivers and race competitiveness as in equation 2). Thus, team effects significantly outweigh driver effects, accounting for 86% of driver variation. Furthermore, the majority (about two thirds) of the team effect is constant for a team and does not change year on year (although there is substantial variation within teams, year-on-year as well). In other words, the legacy of a team outweighs any transient effects as teams change year by year.

There is also limited evidence that the importance of these levels vary by key variables. Although the confidence intervals are relatively wide, there is evidence that the importance of the team has increased over time, whilst the importance of the driver has slightly decreased (Figure 2). There is also evidence that the team-year is less influential to race results on street circuits, compared to permanent circuits (Figure 3). The confidence intervals regarding wet and dry conditions are too wide to be able

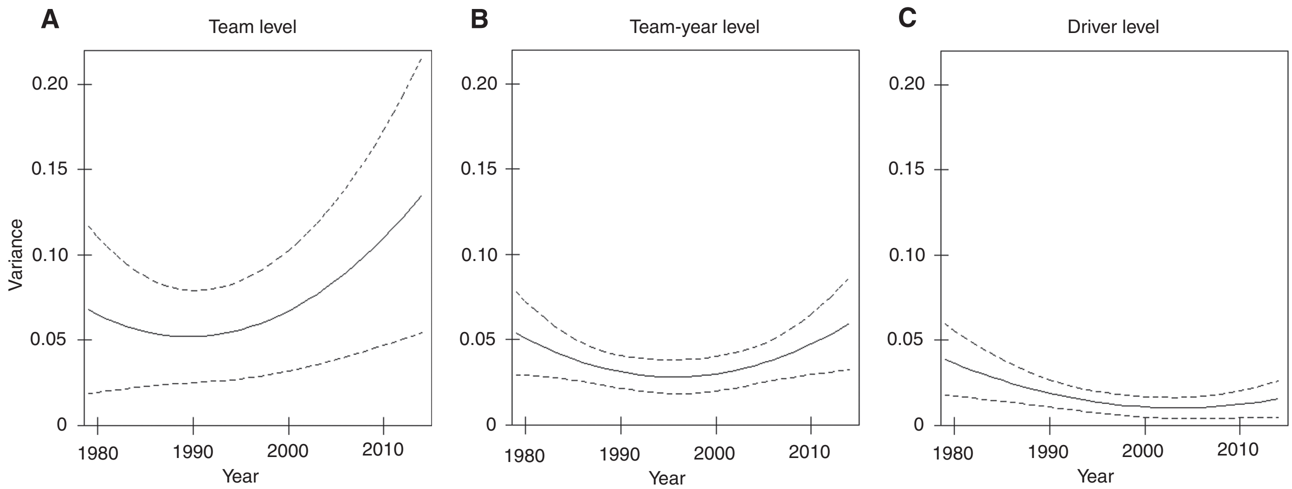

Figure 2: Variance as a function of Year (data from 1979 only). Number of drivers and competitiveness of race are controlled. 95% CIs are shown.

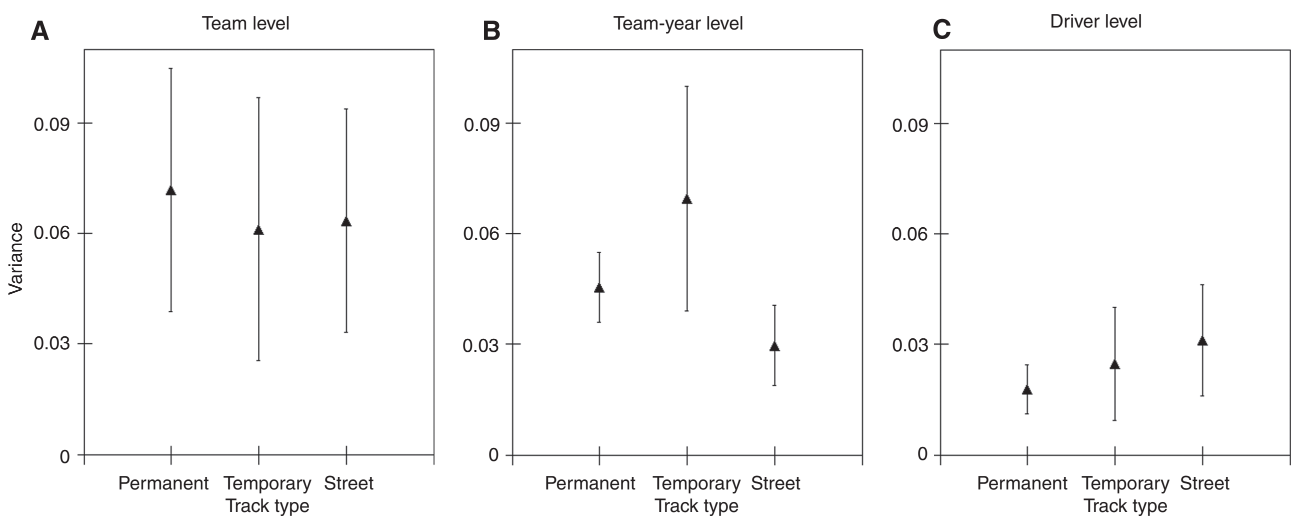

Figure 3: Variance as a function of track type. Number of drivers and competitiveness of race are controlled. 95% CIs are shown.

to make any robust conclusions (Figure 4), but the direction is in line with what one might expect: that teams are less influential in wet conditions than in dry conditions (in other words there is more uncertainty to race results in wet conditions).

4.3 Does our model predict season results?

Whilst the aim of this paper is historical ranking rather than prediction, many readers may wonder how well our model predicts the outcome of F1 seasons when the driver, team and team-year variances are all taken into account. Overall, we would expect the model to be less good at predicting season results than coming up with overall rankings, because there is more chance that a driver would be unlucky in a given season (compared to over their entire career), and, when predicting out of sample, we do not know the team-year residual attached to teams in that year.

There are two aspects to evaluating this. First, we used our model that allows the variances to vary by year and made predictions based on the varying driver, team and team-year effects, which we then averaged across each driver-year. Given we are modelling a transformation of points scored, and assuming driver ability changes over time linearly, we would expect some disagreement between our model and final championship results. However, the model does a relatively good job: the actual champion is correctly predicted in over half of

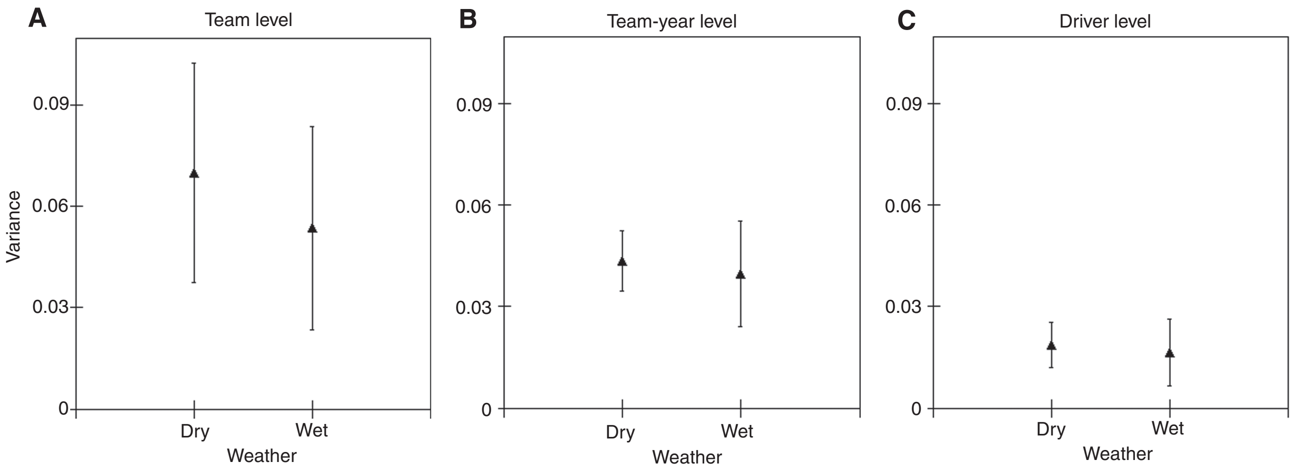

Figure 4: Variance as a function of weather (data from 1979 only). Number of drivers and competitiveness of race are controlled. 95% CIs are shown.

the seasons, and in only two instances does the predicted champion finish outside the top three. There were specific reasons for that in each of those cases: Senna's death, and Schumacher's broken leg. In sum, this is an encouraging indication that our model is performing well.

Second, we used the same model to predict ahead to the 2015 season, using team-year residuals from 2014. These predictions are not particularly accurate; there is too much variation between seasons to correctly predict results out of sample. For example, the McLaren drivers perform very well in the predictions, due to McLaren's (comparatively) decent performance in 2014 and the addition of Alonso, a very high ranking driver, to the team. The model is unable to predict McLaren's poor performance in 2015, and the fact that Alonso's quality would not have much benefit in such a poor car. Thus, whilst this model is very useful in assessing performance that has already occurred, it is less good at predicting performance into the future.

5 Discussion

5.1 The greatest driver?

As with any ranking system, our claim of who is the "best driver" should be treated with an appropriate degree of circumspection. As Figure 1 shows, there is substantial uncertainty around each of the drivers' residuals. This is not a limitation of our model; rather we are explicitly quantifying the uncertainty that inevitably exists in a ranking exercise such as this. Having said this, our results present an interesting ranking when compared to both previous statistical rankings, and subjective expert rankings.[9]

The first point to note is that in most respects, our results match those of others: nine of our top ten – Fangio, Prost, Schumacher, Alonso, Clark, Senna, Stewart, Fittipaldi and Vettel – are considered by most models and experts of be among the best drivers of all time. Our model agrees with previous statistical models (Eichenberger and Stadelmann 2009; Phillips 2014) in ranking Prost above Senna (in contrast to many subjective rankings), and in viewing drivers such as Nigel Mansell, Mario Andretti, Gilles Villeneuve and Mika Hakkinen as rather overrated by experts.

Our model differs from previous statistical attempts in not throwing up any particular surprises in the top 10, in comparison to Eichenberger and Stadelmann (who placed Mike Hawthorn in 5th) and Phillips (who placed James Hunt in 6th). Whilst they argue each has been underrated by experts, our model suggests otherwise (with Hawthorn and Hunt in 34th and 95th place, respectively). Part of the reason for our low positioning of Hunt compared to both Phillips and Eichenberger and Stadelman is his high rate of retirement, and the relatively high penalty that we place on not finishing (compared to Phillips, for example, who does not include non-driver failures in his analysis). The high performance (7th place) of Nico Rosberg in Phillips 2014 was, as Phillips suggests, a result of his partnership with an out-of-form world champion (Michael Schumacher), which artificially improved his results. In our analysis,

9 For example http://www.bbc.co.uk/sport/0/formula1/20324109 and http://f1greatestdrivers.autosport.com/.

when Schumacher is separated into two drivers, pre- and post-retirement, Rosberg's performance against the latter appears less impressive and he is placed 46th.

Perhaps the biggest surprise in our results is the high ranking of Christian Fittipaldi at number 11, despite only competing in three seasons and never making a podium finish. This ranking occurs because C. Fittipaldi consistently outperformed his team-mates, and because he never raced for a "good" team, the standard required to get a high ranking is lower. More specifically, C. Fittipaldi's teammates had relatively high rates of retirement: he gains his high ranking by being able to successfully keep a relatively poor car on the track. Of course, this model cannot say that C. Fittipaldi would have won championships had he raced for a better team, and his confidence intervals are wider than most of the other highly ranked drivers, but the results suggest that in one aspect of good race driving at least – that is, keeping a relatively unreliable car on the road – he should be highly regarded.

Other surprises are the low ranking of champion drivers such as Niki Lauda (142nd) and Alberto Ascari (76th). Lauda only performed notably well when racing for Ferrari (1974–1977) and his results dropped when racing for other, lower achieving teams. Ascari's performances can also be at least in part attributed to his team (Ferrari); he also had a high performing team mate, and his result will be shrunk back to the mean because he raced in relatively few (31) F1 races (see Section 3.3).

5.2 The team or the driver

The multilevel approach presented here has allowed consideration of how much teams and drivers matter, as well as to what extent team effects are invariant effects of a team's "legacy" and how much they change year on year. Our results show that teams matter more than drivers, and that about two thirds of the team effect is consistent over time, with one third being down to year-on-year changes. This fits with what we know about team performances: Ferrari has historically been a very high performing team, and its legacy, and the funds that come with it, ensures that it has and will remain relatively high performing (even if it has been overtaken by newer teams in recent years); there has also been non-negligible variation within teams between years – for example Red Bull's performances in 2011–2013 was exceptionally good for that team (Red Bull in 2011 was the biggest Team-Year level residual in the models), whilst Ferrari did unusually badly in 1992. Finally, although drivers undoubtedly matter, their influence is smaller than that of teams and team years.

When allowing these variances to vary by various covariates, results were produced that in general had wide confidence intervals but with intuitive directionality. It seems that teams have become more important over time, whilst drivers have become less important. There is some evidence that wet weather produces less predictable results (in that all three variances are reduced suggesting results are more down to random chance) although there is a large amount of uncertainty in this result. Finally, street tracks appear to reduce the team effect (in particular the time-varying effect) in comparison to purpose built tracks, whilst increasing the driver effect. Again there are wide confidence intervals, but this again fits with expert opinion that sees street tracks, such as Monte Carlo in Monaco, as difficult for drivers and requiring of skill that cannot be substituted by technological advances. Thus, on street tracks, top drivers are able to differentiate themselves from less good drivers in better cars.

6 Conclusions

We have presented a complex multilevel model that has allowed us to fulfil two related aims: (1) to find a ranking of F1 drivers, controlling for team effects, and (2) to assess the relative importance of team and driver effects. Whilst there is significant uncertainty in our results, our models suggest that (1) Juan Manuel Fangio is the greatest F1 driver of all time; (2) teams matter more than drivers; (3) about two-thirds of the team effect is consistent over time – a "legacy effect"; (4) team effects have increased over time but appear to be smaller on street circuits.

As with any ranking system that one could devise, this one has some flaws. First, where drivers have not changed teams over the course of their career it is very difficult to know whether their performance is the result of their car, the drivers skill, or a combination of the two (for example a driver that happens to driver a particular car well) especially if their teammate remains constant as well. Thus, the model really tells us how drivers perform against their team mates, but those team-mates are not randomly selected since good drivers will self-select into good teams. This is most clearly demonstrated by the high ranking of Christian Fittipaldi, who performed well given his low ranking team, but who has never been tested in a "good" car (that counterfactual is not in the dataset – see King and Zeng 2006). Moreover, team orders have (particularly in recent years) been known to be given to lower ranked team members, encouraging them to allow a favoured team mate to pass them – thus a "good" team driver may achieve more points simply by following team orders. However, with the observational data that we have, our models are the best we can do.

The model could be extended in a number of ways. For example, additional levels could be added to further differentiate between different attributes of the team – we could include a tyre level, an engine level, and so on, to assess what attributes of teams matter the most. It could also be interesting to see how these results differ when qualifying position, or fastest lap times, are used as the response variables.

Finally, we contend the methods used here have a potential broad appeal to researchers in social science and beyond. The cross-classified structure has potential to assess the importance of a wide range of social and economic determinants: how much do individuals, teams and companies affect worker productivity; how much do Primary Care Trusts and neighbourhoods affect health; how much do classes, schools and neighbourhoods affect educational attainment, and how much has this changed over time. All of these questions could be answered, where data is available, using models similar to those used here. The explicit analysis of variances as a function of continuous and categorical predictors allows for the assessment of performance in complex and changing circumstances reflecting the reality of the world that is being modelled.

Acknowledgments: Thanks to the anonymous reviewers and journal editors, Rick Stafford, and the Bristol University Spatial Modelling Research Group, for their helpful comments and advice.

References

Allen, J. 2000. *Michael Schumacher: Driven to Extremes*. London: Bantam Books.

Allen, J. 2011. *James Allen on F1: 2011 – Vettel Steals the Show*. London: Speed Merchants Ltd.

Allison, P. D. and N. A. Christakis. 1994. "Logit Models for Sets of Ranked Items." *Sociological Methodology* 24:199–228.

Alnaser, W. E., S. D. Probert, S. El-Masri, S. E. Al-Khalifa, R. Flanagan, and N. W. Alnaser. 2006. "Bahrain's Formula-1 Racing Circuit: Energy and Environmental Considerations." *Applied Energy* 83(4):352–370.

Anderson, A. 2014. "Maximum Likelihood Ranking in Racing Sports." *Applied Economics* 46(15):1778–1787.

Baker, R. D. and I. G. McHale. 2015. "Deterministic Evolution of Strength in Multiple Comparisons Models: Who Is the Greatest Golfer?" *Scandinavian Journal of Statistics* 42(1):180–196.

Bekker, J. and W. Lotz. 2009. "Planning Formula One Race Strategies using Discrete-Event Simulation." *Journal of the Operational Research Society* 60(7):952–961.

Bell, A. and K. Jones. 2015. "Explaining Fixed Effects: Random Effects Modelling of Time-Series-Cross-Sectional and Panel Data." *Political Science Research and Methods* 3(1):133–153.

Browne, W. J. 2009. *MCMC estimation in MLwiN, Version 2.25*. University of Bristol: Centre for Multilevel Modelling.

Bullen, N., K. Jones, and C. Duncan. 1997. "Modelling Complexity: Analysing between-Individual and between-Place Variation – A Multilevel Tutorial." *Environment and Planning A* 29(4):585–609.

Chambers, J., W. Cleveland, B. Kleiner, and P. Tukey. 1983. *Graphical Methods for Data Analysis*. Boston: Duxbury Press.

Collings, T. and S. Edworthy. 2004. "*The Formula 1 Years*. London: Carlton Books Ltd.

Dominy, J. and R. Dominy. 1984. "Aerodynamic Influences on the Performance of the Grand Prix Racing Car." *Journal of Automobile Engineering* 198:87–93.

Dominy, R. 1992. "Aerodynamics of Grand Prix Cars." *Journal of Automobile Engineering* 206:267–274.

Eichenberger, R. and D. Stadelmann. 2009. "Who is the Best Formula 1 Driver? An Economic Approach to Evaluating Talent." *Economic Analysis & Policy* 39(3):289–406.

Gelfand, A. and A. F. M. Smith. 1990. "Sample-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85(410):398–409.

Gelman, A. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis* 1(3):515–533.

Glickman, M. E. and J. Hennessy. 2015. "A Stochastic Rank Ordered Logit Model for Rating Multi-Competitor Games and Sports." *Journal of Quantitative Analysis in Sports* 11(3):131–144.

Goldstein, H. 2010. "*Multilevel Statistical Models*. 4th ed. Chichester: Wiley.

Goldstein, H. and M. Healy. 1995. "The Graphical Presentation of a Collection of Means." *Journal of the Royal Statistical Society Series A-Statistics in Society* 158(1):175–177.

Hassan, D. 2012. "*The History of Motor Sport*. New York: Routledge.

Henderson, J., K. Foo, H. Lim, and S. Yip. 2010. "Sports Events and Tourism: The Singapore Formula One Grand Prix." *International Journal of Event and Festival Management* 1:60–73.

Horlock, I. 2009. "Prediction of Formula One Race Results using Driver Characteristics. Project for Masters Degree in Engineering, University of Bristol".

Jenkins, M. 2010. "Technological Discontinuities and Competitive Advantage: A Historical Perspective on Formula 1 Motor Racing 1950–2006." *Journal of Management Studies* 47(5):884–910.

Jenkins, M. and S. Floyd. 2001. "Trajectories in the Evolution of Technology: A Multi-Level Study of Competition in Formula 1 Racing." *Organization Studies* 22(6):945–969.

Jenkins, M. and S. Tallman. 2016. "The Geography of Learning: Ferrari Gestione Sportiva 1929–2008." *Journal of Economic Geography* 16(2):447–470.

Judde, C., R. Booth, and R. Brooks. 2013. "Second Place is First of the Losers: An Analysis of Competitive Balance in Formula One." *Journal of Sports Economics* 14(4):411–439.

King, G. and L. C. Zeng. 2006. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14(2):131–159.

Loiacono, D., P. L. Lanzi, J. Togelius, E. Onieva, D. A. Pelta, M. V. Butz, T. D. Lonneker, L. Cardamone, D. Perez, Y. Saez, M. Preuss, and J. Quadflieg. 2010. "The 2009 Simulated Car Racing Championship." *IEEE Transactions on Computational Intelligence and AI in Games* 2(2):131–147.

Muehlbauer, T. 2010. "Relationship between Starting and Finishing Position in Formula One Car Races." *International Journal of Performance Analysis in Sport* 10(2):98–102.

Phillips, A. J. 2014. "Uncovering Formula One Driver Performances from 1950 to 2013 by Adjusting for Team and Competition Effects." *Journal of Quantitative Analysis in Sports* 10(2):261–278.

Rasbash, J., C. Charlton, W. J. Browne, M. Healy, and B. Cameron. 2013. "MLwiN version 2.28". Centre for Multilevel Modelling, University of Bristol.

Skrondal, A. and S. Rabe-Hesketh. 2003. "Multilevel Logistic Regression for Polytomous Data and Rankings." *Psychometrika* 68(2):267–287.

Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal Statistical Society Series B-Statistical Methodology* 64:583–616.

Spurgeon, B. 2009. "Age Old Question of Whether it is the Car or the Driver that Counts." http://formula1.about.com/od/formula1101/a/Age-Old-Question-Of-Whether-It-Is-The-Car-Or-The-Driver-That-Counts.htm (accessed 17th March 2015).

Spurgeon, B. 2011. "The Art of Racing in the Wet in Formula 1." http://formula1.about.com/od/howaraceworks/a/The-Art-Of-Racing-In-The-Wet-In-Formula-1.htm (accessed 17th March 2015).

Taylor, P. 2013. "Standardized Mortality Ratios." *International Journal of Epidemiology* 42(6):1882–1890.

Supplemental Material: The online version of this article (DOI: 10.1515/jqas-2015-0050) offers supplementary material, available to authorized users. Further materials are also available at http://eprints.whiterose.ac.uk/96995/.