Université de Fribourg / Universität Freiburg
Fakultät der Wirtschafts und Sozialwissenschaften
Seminar für Finanzwissenschaft

Who is the Best Formula 1 Driver? An Econometric Analysis

Master Thesis

under the supervision of

Prof. Dr. Reiner EICHENBERGER (1st supervisor)
and
Prof. Dr. Martin WALLMEIER (2nd supervisor)

written by

David STADELMANN
BA in Economics, stud. rer. nat.

June 2006

8th Semester
Phone: +41 (076) 542 33 48
e-mail: david.stadelmann@unifr.ch

Address:
Av. Beauregard 9
CH-1700 Fribourg

Acknowledgements

I am particularly grateful to Prof. Dr. Reiner EICHENBERGER for detailed and valuable comments and suggestions on this Master thesis. He also proposed the topic itself as well as a number of evaluation ideas and was always available for special research questions. Prof. Dr. Martin WALLMEIER was kindly prepared to be the second supervisor.

Furthermore, I am also indebted to CREMA (Center for Research in Economics, Management and the Arts) for financial support.

During the "Coffee, Biscuits and Economics Seminar" organized by the assistants in economics of the University of Fribourg a preliminary version of this paper was discussed. For mentioning possible solutions to problems and referring to special issues I am obliged to all the participants. Particularly, I would like to stress the constructive inputs received by Michael FUNK, Lorenz KÜNG, Mark SCHELKER and Florian ZAINHOFER.

I decided to write this Master thesis in English in order to exercise my writing skills. Shauna SELVARAJAH and Kevin MUSA corrected a number of spelling and grammar mistakes.

All remaining errors and shortcomings are my own responsibility.

Table of contents

List of figures

List of tables

List of equations

1 Introduction

Formula 1 drivers are fast if they have a good car, but the value of a good driver alone is still not clear. The performance of drivers and cars has not yet been independently evaluated. To achieve an effective comparison between drivers it is not sufficient to sum up their achieved points in a season or accumulated points or wins over a certain time frame. Points and wins depend firstly, on the number of races that the driver took part in, secondly, on his own personal driving capabilities and thirdly, on the car as well as the whole team.

In this Master thesis we shall try to separate the driver and the car into two different factors by applying various approaches and modern econometric procedures. This will allow us to establish different ranking lists for diverse performance measures such as race classification, point differences, time differences, starting positions, and so on. The main aim behind the whole evaluation is to estimate the impact of a Formula 1 driver independently of his car and to compare today's racers with former ones. The impact of the driver's performance can be viewed as his talent or his ability. This Master thesis can therefore also be considered as a case study for estimating capability and talent.

The first method of evaluation will be a table of performance variables corrected for the number of races a driver participated in. This crude and simple approach gives some preliminary indication of the strength of a driver. Clearly, the effects of the cars and the teams are still pertinent and have not been corrected for by using this method.

The second step of the process of finding the best driver and identifying talent will be a paired comparison methodology. As most teams employ two drivers, one approach of separating the impact of racer and car is to use systematic paired comparisons over time. When a certain driver wins in a certain car against his team partner there is an indication that the winner of this intra-team (car) match is the stronger driver. As drivers change their cars and teams additional information is generated and can be used for further comparisons. We will thereby try to bridge different eras in sports. The goal can also be seen as an attempt to construct a statistical time machine, which allows us to estimate how a driver from a certain period would perform in another. Using a maximum likelihood approach, a comparable value can be calculated for each driver indicating the likelihood of winning any paired comparison in the data panel used.

Finally, driver and car effects shall be evaluated using an econometric model. The coefficients of the resulting regression model using dummies for drivers will be interpreted as a measure of a driver's skill and ability.

A number of applications using the established database for this Master thesis are shown at the end. We shall try to answer some specific questions concerning Formula 1 races and economics and attempt to test certain hypotheses. Some of the applications shown might be of immediate interest for Formula 1 teams such as the estimation of a driver's average life cycle. Other applications with an economic background such as risk evaluations and the preliminary estimation of superstar effects in Formula 1 will also be performed. Further applications in economics and for other purposes are also proposed.

Comparing sportsmen from different eras has been discussed widely. From a statistical point of view, these comparisons are especially difficult to perform without stringent hypotheses. Different methods and aggregation procedures can lead to dissimilar results. Nevertheless, such evaluations have been performed for diverse sports before,[1] but very limited research has been done so far for Formula 1 and no articles have compared racers over different time periods by means of econometric approaches.[2] However, the data available are suited for such comparisons and today's computational power makes it possible to control for a large number of variables in the analysis. Using a number of different evaluation methods does not necessarily solve the inherent problem when evaluating achievements in sports, but it shows possible ways of measuring talent.

This Master thesis is organized as follows: In chapter 2 we present a literature survey covering several articles concerning sports economics and the economics of superstars. Next statistical evaluation methods in sports are discussed. Chapter 3 starts with a brief description of the data. Some preliminary descriptive statistics are given, followed by a simple evaluation approach. Thereafter, a model for paired comparison is used to evaluate the driver's abilities. Finally, an econometric regression model is used to look at the data again and control for a number of additional variables. Problems and limits of the applied evaluation methods are discussed in each section separately. After estimating and testing the econometric model, three possible preliminary economic applications are presented in chapter 4. In addition, a number of other interesting research questions might be answered using the model. Chapter 5 summarizes our results and concludes.

---

1 See the literature survey in chapter 2 for further details or for example JOE (1990), BERRY ET AL. (1999) and KNORR-HELD (2000).

2 A search in EconLit for the keywords "Formula One", "Formula 1", "Schumacher", "car Ferrari" and "Formula 1 racing" resulted in two articles. One of them is a case study of Ferrari's marketing and the other considers aggregation problems in rankings with an application to the 1998 Formula 1 racing season.

2 Literature survey

The following literature survey will focus first on sports economics in general and on the economics of superstars in particular. In the second part, literature on specific statistical issues of different evaluation methods with a focus on sports application will be reviewed.

2.1 Sports economics and the economics of superstars

Sports mostly received marginal attention from economists in the past. A number of articles on sports business can be found in the general management literature as well as in the literature covering questions of personal management. Nevertheless, in the past years the interest in sports has increased and in the year 2000 the "*Journal of Sports Economics*" was launched.

2.1.1 Literature overview of sports economics

In an introductory textbook TROISEN (2003) presents a broad picture of sports economics. Starting with a chapter on the issues in sports economics the author then gives an historical introduction concerning the origins of sports economics. He subsequently reviews the basics of the production of sports, the utility and costs of careers in sports and the transition from performances in sports to products for sports. Afterwards sport markets are identified and several cases of different markets are analyzed. Furthermore, TROISEN (2003) dedicates a chapter to the microeconomic basics of consumption as well as the general objectives of the sports industry. The textbook's primary focus is not necessarily on the analysis of sports form an economic perspective but rather on a number of management issues and different organizational concepts which are discussed and reviewed.

KAHN (2000) points out that professional sport offers a unique opportunity for labor research. Firstly, data from sports can be used to identify the effects of monopsony power on a worker's wage. Secondly, as personal characteristics of sportsmen are widely known as well as their performance the extent of discrimination can be estimated far more precisely than in other industries. Lastly, it is possible to estimate the impact of supervisor quality and the effect of incentives on workers' behavior as well as other interesting elements of economic theory that are generally hard to estimate because of insufficient data availability. The author shows in his paper that sports owners have some monopsony power over players in the sense the players have the option of negotiating only with one team. KAHN (2000) mentions that a major difficulty for all labor market research on discrimination is the problem of unobserved variables. Such problems are surely less severe in sports than elsewhere as detailed data sources allow to control for a variety of characteristics. Indeed, he refers to some research articles identifying racial discrimination in American baseball. The results on players' performances suggest furthermore that athletes are motivated by similar forces which concern workers in general: Pay also matters for sportsmen. The lessons to be learned from sports economics should not be discounted because of the market's high profile. Interesting and valuable insights can be gained by considering sports as a research area in order to identify different aspects of economic theory.

The design of a sporting contest bears a close relationship to the design of an auction according to SZYMANSKI (2003). The objective of his review is to discuss the contest theory literature in the context of sports. By using this approach several questions concerning the design of sporting contests can be solved such as the optimal number of contestants or how evenly balanced competing teams in different sports should be. Moreover, SZYMANSKI (2003) points out that there are many aspects of the organization of individualistic sports that could be modeled more fully with a view to establishing an optimal design. Finally the author discusses a theoretical contribution to the analysis of team sports: the so-called invariance principle. This principle states that changes in ownership rights over player services and certain types of income redistribution will have no effect on competitive balance. Most empirical papers reviewed weakly support the invariance principle.

The methodology used in sports economics can also be applied to arts research. SEAMAN (2003) states that particularly arts labor research can benefit for the insights of sports economics. According to the author, the arts and the sport sector exhibit rich similarities that suggest possible collaboration and extensive cross-referencing: earnings in both sectors tend to peak early and decline quickly, employment is often sporadic, sports careers are short and most artists also leave their careers due to limited opportunities, few become superstars but the potential for great success provides inspiration. SEAMAN'S (2003) study is mainly a survey of sports literature which tries to motivate arts economists to consider techniques from sports economics for their own research.

VON ALLMEN (2001) follows a more specific application of the economics of sports with a focus on racing. He provides a description of the payment systems in the NASCAR

Winston Cup (now known as the Nextel Cup) racing. The American NASCAR's reward structure differs significantly from other tournament based sports, which tend to have nonlinear individual-performance payouts. The compensation system in NASCAR consists of two structures: a fairly linear individual-race reward system and a nonlinear season-ending points reward system. According to the theoretical contribution of VON ALLMEN (2001) a nonlinear race-level reward for NASCAR may be inefficient, which means that the current system should not be amended. Firstly, sponsorship may provide an incentive for drivers to be consistent during the season rather than taking greater risks for an opportunity of a single race victory. Anyhow, he mentions that a driver's performance in a certain race and sponsorship revenues may be directly correlated. Secondly, the risks of racing in close proximity at high speeds could be too high for a nonlinear individual-race reward system. The increased aggressiveness could lead to significant negative externalities for other drivers. Finally, VON ALLMEN (2001) mentions that car racing depends on very high expenditures. Nonlinear race payoffs may provide sufficient revenues for some teams to fund further expenditures, thereby potentially skewing competitive balance unintentionally. From a theoretical standpoint it seems therefore that the linear reward system for races and the nonlinear season-ending points reward system may be more efficient in car tournaments.

An empirical test of VON ALLMEN'S (2001) ideas concerning the NASCAR reward system is performed by DEPKEN AND WILSON (2004). They analyze season-level data from 1949 to 2001 in order to test whether an increase in performance-points concentration leads to a disproportional increase in winnings concentration. This would correspond to an increased taking of risk by NASCAR drivers. Indeed, the authors find some initial evidence for the hypothesis. Anyhow no significant results are found for the cost hypothesis stating that a non-linear reward system would lead to a skewed competitive balance. By using Granger causality tests the authors find that the performance-points concentration does not cause winnings concentration and vice versa. Due to a lack of available data DEPKEN AND WILSON (2004) cannot test the hypothesis that sponsorship issues could play a significant role for the choice of the linear reward scheme in NASCAR racing.

In the applied management literature, tournaments in which workers are paid based on their relative performance, received little attention. ZAX AND LYNCH (2000) therefore compare incentive and sorting theories of tournament performance in road races. Their data set which includes finishing times, runner names, race names, race distances, runner ages, and runner genders was obtained from the USA Track and Field Road Running Information Center. The data generally support the hypothesis that times are faster in races offering higher prize money. This is mainly because higher prizes attract better runners. When ZAX AND LYNCH (2000) control for a driver's ability through fixed effects the incentive effects of higher prizes diminish. However, the authors do not control for car fixed effects. Their results suggest that races with high stakes only record faster times because of a sorting or selection effect of better drivers. Higher prices per se do not encourage all racers to run faster. The application to human resource management is imminent as some literature on motivation states that higher pay is not a motivator for higher performance if a single individual is concerned.

Generally it can be noted that the economics of sports tends to treat research topics that are far beyond the analysis of general economics. Mostly, researchers in sports economics focus on a relevant question in sports and use it to win insights into some economic problems such as done by KAHN (2000). In the *Journal of Sports Economics* mostly incentive schemes, sports contracts and questions of how to find a talent are answered. Our econometric analysis of Formula 1 follows this path by concentrating on the identification of ability and talent and distinguishing between technical and human influences on performance.

2.1.2 Literature overview of the economics of superstars

After this short survey concerning some literature on sports economics we would like to focus now on the economics of superstars. A foray into this branch of economic literature gives valuable insights into labor market economics and industrial organization. Success or failure is a central element of every human's labor market experience and the extent to which success translates into income is highly variable. The economics of superstars states that the distribution of rewards often appears to be such that comparatively minor talent differences can generate enormous returns, so that the distribution of income is not only a simple re-scaling of the distribution of ability. If such enormous incomes earned are the market's reward for superior talent the "superstar phenomenon" may be socially admissible. If, on the other hand, the source of the superstars high incomes is not their talent, the skewed income distributions may be perceived as inequitable by society.[3] It could therefore be argued that the main focus of a superstar analysis should not only lie on an endogenous variable of success which is measured by income but rather on a

---

3 This depends on the social norms and standards prevalent in a society and shall therefore be not the matter of further discussion in this Master thesis.

measurement of ability. [4] Our analysis in chapter 3 and 4 is concentrating on this aspect as we try to find out empirically the most talented driver. Apart from the intrinsic motivation of the research question of the thesis' title, this paper can consequently also be seen as a case study for the measurement of ability.

In his seminal paper ROSEN (1981) introduced a rigorous definition and a model of the "superstar phenomenon". He mentions that in certain kinds of economic activity there is a concentration of output among a few individuals, marked skewness in the associated distributions of income with very large rewards at the top. The superstar phenomenon is said to exist when "relatively small numbers of people earn enormous amounts of money and seem to dominate the fields in which they engage". Additionally, small differences in ability are supposed to have disproportionate effects on earnings. ROSEN (1981) mentions that both demand and supply conditions may be involved in this market result. On the demand side, lesser quality is only a very poor substitute for greater quality. Furthermore quantity and quality do not have to be substitutes either. By introducing indivisibilities in the consumption of services, consumer's attempts to minimize costs give an additional advantage to higher quality producers. On the supply side the marginal cost of output is largely assumed to be constant (or falling). The effect of scale economies makes it possible to produce large amounts of consumable services at low or falling unit costs. In addition, the service delivered by the superstar can be described as a form of joint consumption but property rights are clearly assigned to the seller as opposed to a public good. The author shows that using these ingredients leads to a model which allows explaining the differences in incomes and the convex revenue function of superstars.

In a similar vein, MACDONALD (1988) presents a dynamic version of ROSEN'S (1981) superstar model where consumers are labeled as the "audience" and producers as "performers". The implicit pricing of characteristics such as ability or talent and the associated convex returns are the distinguishing feature for all superstar models but most models do to not look closer at the distribution of such personal attributes. MACDONALD (1988) demonstrates that the predictive content of superstar models could be increased by further analyzing the distribution of talent. He shows that in the steady-state equilibrium only the young enter the occupation and earn low incomes playing to small crowds. Only the successful stay on in the industry. Those few successful performers earn high incomes playing to big crowds. The less successful ones drop out of the industry. Overall, there are

[4] Indeed, there are not many empirical articles measuring the income of superstars because of the lack of consistent data. Superstars do not only benefit from fixed salaries but also from sponsoring, marketing revenues, and so on.

few stars in the industry but as a group they serve a large fraction of the audience and earn an even larger share of the rewards.

According to HAMLEN (1991) the analysis of superstars in popular music has focused on two different views concerning the modern recording industry for popular music. The layman's view of the market for popular singers is that most consumers of this type of music have little appreciation for the quality of the voice and the singing. This is explained by ignorance about vocal qualities. The economic view, based on the ideas of ROSEN (1981) is that the industry of popular music is an astonishing example of a superstar phenomenon. Thereby, only small differences in vocal ability can be magnified into disproportionate levels of success. Both assumptions lack empirical investigation. HAMLEN (1991) tries to shed some light into this issue by empirically investigating record sales on a number of different variables including a quality measure for a singer's voice. He finds that the strongest predictor of total record sales is career longevity which comes as no surprise.[5] Gender seems to play an important role, as female singers tend to sell more records than their male counterparts. HAMLEN'S (1991) measure for voice quality is significant and positive. This indicates that consumers of popular music do indeed discern quality in their preferred singers but the size of the coefficient is significantly less than one would expect if popular music would show superstar effects. The estimated elasticity in the log-linear regression function is too low to support a superstar phenomenon for the popular music industry. Only a small amount of incomes is accessible for the period of analysis. Consequently, the effects of a singer's ability on his income cannot be analyzed as record sales do not have to translate directly into additional income.

The empirical paper of CHUNG AND COX (1994) examines the phenomenon of superstars form a different perspective. Their study employs a stochastic model as a probability mechanism underlying the consumer's choice of services. They predict by using this model that the output produced will only be delivered by a very small number of "lucky" individuals. As a measure of artistic success by performers their number of gold-records is used. The authors then apply the Yule distribution to calculate numerical fits to the data. It turns out that the used stochastic process leads an excellent prediction of the real empirical distribution explaining nearly 94 % of the observed gold-records among artists. The importance of the contribution by CHUNG AND COX (1994) is that their empirical results suggest that the superstar phenomenon could exist among individuals with equal talent.

---

[5] Career longevity could be considered as an endogenous variable. This problem is not discussed by HAMLEN (1991).

This is explained by the fact that their stochastic model does not use ability differences and views the superstar phenomenon as an implication of a probabilistic mechanism underlying the record-buying behavior of consumers. The very large incomes are therefore partly driven by chance and not by superior talent as suggested in other studies. This result is in line with a theoretical contribution by ADLER (1985) stating that large differences in earnings can exist even where there are no differences in talent at all. For ADLER (1985), stardom can be seen as a market device to economize on learning costs. Acquiring knowledge by a consumer by discussion with other consumers is easier if all share the same common knowledge. Stars can therefore exist even if their abilities are not superior to that of others.

In the late 1990s and the early 2000s concert revenues for top rock performers and ticket prices for such concerts took off. As KRUEGER (2005) mentions, the economics of superstars can help to explain the long-term trends in this industry. However, the price surge beginning in 1997 cannot by accounted for. To test different explanations for this price increases the author compiled a database of box office reports and additional data on bands and performers. The first remark to be made is that the concert revenues became more skewed in the 1980s and 1990s. To control for a star's quality KRUEGER (2005) uses the number of millimeters of print columns devoted to each artist in The Rolling Stone Encyclopedia of Rock & Roll.[6] The applied regression model yields large and increasing superstar effects. The author identifies a "technology-induced erosion of the complementarity between record sales and concert tickets" as the main explanation for the surge in ticket prices. Ticket prices have soared for the reason that recording artists have seen a decline in their income for records sales because of the possibility to download free music. Additionally it is likely that the downloading of music will put further upward pressure on concert prices and revenues.

In general the literature on the economics of superstars does not try to measure talent directly but only assumes that large income differences are due to differences in abilities. It has also been shown that differences in success can be explained by using a purely stochastic model. Measuring ability itself may therefore contribute to solving this puzzle. Possible approaches to estimating talent may also be of use for other economic branches such as the career concerns literature for example (see HOLMSTRÖM, 1999). A brief look at the literature on statistical approaches in sports may therefore be helpful for our case.

[6] It should be noted that this variable could face a possibly high problem of endogeneity.

2.2 Statistical evaluation methods for sports

There are a number of different statistical evaluation methods for sports. Some authors mention paired comparison approaches, others try to estimate aging functions and some even use simple Gini coefficients when comparing teams or athletes. It should be noted that most evaluations are done for the United States and sports teams overseas. In general, far more interest is dedicated to numerical evaluation of strengths, merits and outcomes over the Atlantic than in Europe. As shown above some theoretical and empirical economic analysis has been done for NASCAR racing but not for Formula 1 even though NASCAR is far less widely viewed than Formula 1 in most parts of the world. Even more astonishing and worthwhile mentioning is the fact that a relatively large number of applied statistical methods on sports concern American baseball, basketball or football. Yet, it is European soccer that is likely to yield higher total revenues and has also a further global reach if you "Do The Math" correctly.[7] The same is likely to be true for Formula 1. We shall focus in this section on different models and methods of sports evaluation in general.

The main rating and statistical evaluation method for sports are done on the basis of categorical outcomes of paired comparisons such as wins, draws and losses in the sport considered. The method of paired comparison has attracted the attention of people from a wide spectrum of interests: statistics, marketing research, preference measurement, and so on. Sports competition is just one possible sector of application. The main methodological and mathematical framework was laid within the years from 1960 to 1970. DAVIDSON AND FARQUHAR (1976) present a list of references on paired comparisons. As a mature research branch the *Handbook of Statistics* includes an article by BRADLEY (1984) on the method.

JOE (1990) extends the general model for paired comparisons to apply it for chess ratings. The extension of the model is used to establish a ranking for 64 top chess players since 1800. Additionally the author takes account of the peak of their career periods. To calculate the peak period a ratio of 0.25 for the peak career period to the whole career period was taken for all players. Furthermore JOE (1990) assumes that all players improve before the peak and decline after the peak at the same rate. This leads to a ranking of chess players that does not necessarily seem correct to chess experts. The top eight players are A. Alekhine, M. Botvinnik, J. Capablanca, R. Fischer, A. Karpov, G. Kasparov, E. Lasker and P. Morphy. JOE (1990) mentions that the comparisons could be improved if more data had

---

7 The *Forbes* magazine reports on its website (http://www.forbes.com/2006/01/04/soccer-football-baseball-cx_pm_0104soccer.html downloaded on the 10th of April 2006) that "soccer's superior competitiveness […] helps explain why it attracts billions of sponsorship dollars from companies like Budweiser, Coca-Cola, McDonald's, and Yahoo! […]".

been compiled in order to use covariate information. Furthermore it would be essential to extend the chain for every pair of players so that better estimates of model parameters were possible. Since not all players were active in the same decades only a limited number of pairs could be matched which each other. According to JOE (1990) this could have influenced the results and the conclusions which might be drawn in the study.

The main goal of an article by KNORR-HELD (2000) is the dynamic rating of sporting within the framework of a cumulative link model. Such a model for a comparison $a_{ij\alpha}$ of two teams *i* and *j* in event α is defined by $P(a_{ij\alpha} \leq 1) = F(\theta_r + \beta_i - \beta_j)$ where F is a distribution function, θ_r represent different control parameters and β_i is the latent ability of team *i*. The author uses an extended version of the Kalman filter algorithm for categorical data to estimate the unknown parameters. All teams are treated symmetrically by the author's approach and additional parameters are entered to represent a possible home advantage. It is important to mention that such parameters θ_r are assumed as time constant and sport's team independent and that categorical data in general does not provide much information about the temporal variation in team's abilities. KNORR-HELD (2000) applies his method to the Deutsche Bundesliga for the years 1996 to 1997 and for the same period to the American National Basketball Association and finds substantial home advantage effects. Additionally, he presents graphically the filtered and smoothed estimated abilities of selected teams. Anyhow, the abilities are not compared with each other but only the ability paths are commented.

Another highly sophisticated model is set up by GLICKMAN AND STERN (1998) for American National Football League Scores. They develop a dynamic predictive model that assumes that team strength parameters follow an autoregressive process. The distribution of the data at each point in time is specified on a number of time indexed parameters and an additional process analyses the evolution of the parameters over time. As for such state-space models analytical inferences about parameters is impossible the authors apply Markov chain Monte Carlo methods. The analysis yields the Dallas Cowboys as the best team with almost 17 point advantage over the worst, the New England Patriots for the period form 1988 to 1993 in the National American Football League.

An effort to compare the different performances of athletes from different time eras is made by BERRY ET AL. (1999). They try to use the career-overlaps to compare players whose achievements were made in different time periods. The difficulty in modeling such career-overlaps lies, according to the authors, in the effects of the aging process on performance. Therefore BERRY ET AL. (1999) develop aging functions to take account of the different abilities over time. In order to control for the age, random curves are used to represent individual aging effects. Additionally they try to separate out the difficulty of playing in different time periods and they also characterize the changing structure of the population of players. The model is applied to American National Hockey League players, professional golfers and Major League Baseball players and provides, according to the authors, reasonable estimates for all the three sports considered.

UTT AND FORT (2002) mention some pitfalls associated with measuring competitive using Gini coefficients. Usually Gini coefficients of winning percentages are calculated to measure inequality in sports league playoff outcomes within a season. The first problem is that a team cannot win all games but only all of its own games. The zero-sum nature of league play overstates the level of competitive balance. Furthermore a host of other complexities must be overcome to achieve precise estimates of competitive balance. UTT AND FORT (2002) review a number of works that did not take account of these facts. They suggest using winning percentage standard deviations as an alternative to Gini coefficients.

KLABRODA (2000) discusses problems encountered when aggregating points in sports competitions. He uses Formula 1 racing as a case study. An optimal aggregation of drivers' outcomes is impossible according to KLABRODA (2000). Such an aggregation procedure would need to have a universal definition space, that is every possibly combination of classifications should lead to a certain ranking. Furthermore, the aggregation needs to be Pareto efficient: If driver A finishes always before driver B the aggregation of points needs to lead to the result that A is better than B. Moreover, a method of giving points should be independent of irrelevant alternatives: If A wins twice against B and B once against A, driver A should be ranked ahead of driver B. However, the normal point system used in Formula 1 does not guarantee this. KLABRODA (2000) shows with simple numeric models that changing the method of aggregating points and applying, for example, ordinal rankings can lead to totally different results in the Formula 1Championships of the year 1998. The author's work should not be interpreted as a statistical approach to evaluate a driver's capability as no specific evaluations are performed. It should rather be regarded as a theoretical contribution highlighting the problems encountered when aggregating points and when establishing rankings.

For Formula 1 itself, some very basic statistical evaluation methods are used by fans' and drivers' websites on the internet. Mainly they are simple listing the number of wins, podium positions and pole positions for different drivers.

A comparatively advanced ranking is provided by "Autosport-Atlas Wildsoft F1 Ranking". Their ranking system is based on the last five year drivers' achievements. Points are counted for such achievements as wins, poles, and many others, even kilometers raced. No points are given to drivers who did not appear on the start (even if retired on the warm-up lap) and no points for drivers disqualified for any reason. All resulting points for the last 5 years are summed up. Old Grands Prix are removed and new ones are added as time is passing through. Of course this method of evaluation depends highly on the hypotheses concerning the points awarded to drivers. Changing the assumptions can lead to big and unexpected changes in the results and in the ranking. Additionally a comparison over several years is not possible and the number of races is positively correlated with the number of points. This should not be the case when evaluating the ability of a driver.

Other sites show different statistics but do not try to evaluate the impact of a driver and a car separately. They do not even account for changes in the number of points over the years. Nor do they consider that the reliability of the cars was improved over the last 50 years. This can only be done by using more refined statistical and econometric methods and by including several control variables. None of the websites tries to perform paired comparisons conditioned on experience variables or other controls and on the same cars. If comparisons between drivers are done they are just representations of the differences in points or the number of wins and therefore do not take account of the main distinguishing characteristic which is the team and the engine.

In the next chapter we will apply different statistical and econometric procedures to evaluate a driver's capability independently of the car in Formula 1. In addition, we will control for several other variables such as experience or technical problems.

3 Evaluating the drivers

This chapter focuses on different methods used to evaluate a Formula 1 driver's ability independently of the car. We start with the description of the database set up for this study. A crude evaluation approach follows aimed at identifying the main drivers whose results will be presented in this paper.[8] The drivers will then be matched with their team partners and will be compared by means of different performance measures. These matches allow us to calculate the likelihood of a specific driver winning against any other driver in a paired comparison approach using all available information. Finally an econometric model will be established to statistically separate driver and car specific effects. The capabilities of a driver and the resulting rankings of each evaluation method will then be compared for a subset of Formula 1 racers.

3.1 Data

Our main data comes from the FORIX Formula 1 Database which is provided by the magazine Autsport-Atlas via a subscription on the internet.[9] On the FORIX website tables can be obtained for all Formula 1 races starting from 1950 until the latest race.

For our calculations we downloaded information on Grand Prix tracks, drivers, cars, race and weather conditions, race classifications, race times, race points, starting grids, fastest laps, causes for dropouts and further specific information. The downloaded tables were then put together in a panel data structure for all races over the period from 1950 until the end of the racing season 2005.

3.1.1 Defining and constructing the variables

Table 1 summarizes some variables' definitions used in this paper and focuses on their characteristics. Some comments on these variables are worthwhile but no detailed comments will be made on obvious combinations or their simple modifications. Such amendments shall be explained in more detail during the ongoing analysis of the data when the variables are used in a specific context for the estimations.

---

8 As there were more than 300 drivers in Formula 1 history we have to focus on a subset of them. All results can be obtained from the author's website: http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

9 For further details see http://www.autosport.com.

Table 1
Summary of variables and definitions

Variable	*Definition*
GPRIX	Country where the Grand Prix took place
GPDISTANCE	Distance of Grand Prix circuit in kilometers
GPPERIMETER	Perimeter of Grand Prix circuit in kilometers
GPLAPS	Laps of Grand Prix circuit
WEATHERCODE	Weather conditions during Grand Prix race (-2 = "cold, rainy, overcast" ; 2 = "hot, dry and sunny")
DRIVERCODE	Driver name (used as dummy variable)
DRIVERBIRTHYEAR	Driver's year of birth (several combination are formed using this variable during the analysis)
DRIVERHOMEEFFECT	Dummy for "home field advantage" (1 if race takes place in driver's nation)
DRIVEREXPERIENCE	Number of races a driver took part in
DRIVERPEAK	Dummy for peak of career
DRIVERNBCARS	Number of different cars used during active period (modifications for tyres and lubrication are used during the analysis)
DRIVERNBWINS	Number of accumulated wins a driver has accomplished
DRIVERNBPODIUMS	Number of accumulated podium positions of a driver
CARCODE	Car identifier based on team name and engine
TYRES	Identifier for tyre produces
LUB	Identifier for lubrication used
GRIDCLASS	Start position (several combination are formed using this variable during the analysis)
GRIDTIME	Time of achieved start position (several combination are formed using this variable during the analysis)
CLASSCLASS	Classification of driver in race (several combination are formed using this variable during the analysis)
CLASSTIME	Time of achieved classification (several combination are formed using this variable during the analysis)
CLASSPOINTS	Points of achieved classification
TECHNICALOUT	Dummy variable for technical reason of dropout during race
HUMANOUT	Dummy variable for human reason of dropout during race (eg. collusion)
FASTESTLAP	Time of driver's fastest lap in race

Source: Author's definitions and designations.

For every Grand Prix in each year there is specific track information available. We mainly concentrate on the number of laps, the distance in kilometers, the weather conditions and the country in order to construct a "home field advantage" variable.

Weather conditions from FORIX were only available as plain text and had therefore to be transferred into integers to be used for an econometric analysis. We coded the conditions into five different categories starting from -2 to 2. Lower values indicate "cold, rainy and overcast" weather conditions. Higher values indicate "hot, dry and sunny" conditions. The value 0 represents the category of "cloudy/overcast, mild, partially wet" weather.

The total number of drivers taking part at least once in a Formula 1 Grand Prix is 790. As more than 50 % of these drivers did not score any points during their Grand Prix appearances we left them out of the evaluation procedure. 300 out of the 790 drivers from 1950 to the end of the racing season 2005 scored at least one or more points during their active period. These 300 drivers and the races they took part in shall represent the main database although only a subset of 50 Formula 1 runners will be explicitly represented in this paper due to a lack of space for such an enormous number of racers.

Additional variables concerning drivers are their age in years, nationality, experience and career peaks. The experience variable is an integer that counts the number of races a driver took part in. To find out the career peak of a Formula 1 racer we use the ranking of the driver in a Grand Prix. A driver's career peak is identified as the season or seasons when a driver accomplished his personal best classification in a race. As a result, every Formula 1 racer will have at least one peak season. JOE (1990) made the assumption that 25 % of the career can be considered as the peak period in his evaluation of chess players. Drivers participating over long time intervals may also have several peaks periods. If a driver raced for only one season, that is identified also as his peak year.

For the 300 drivers in points there is a total number of 750 different car models. A car model in our analysis is defined as the name of the car registered for the Grand Prix in that it is taking part. Usually the car registered is a combination of the team's name and a reference to the engine installed. Sometimes the registered car name also represents a combination of the team's name and the year of entry. It is important to mention here that as the same engines are used in different cars and the team names often do not change over time, the used identification for the car in this study captures at the same time the team's influence and the technology. This choice helps us to significantly reduce the number of independent variables and car controls without possibly affecting the results.

Furthermore, we distinguish between ten different tyre producers (including unknown producers coded as "unknown"). Information is also available on eight different types of lubrication (including the type named "unknown"). The tyres and the lubrication will enter our analysis as dummy variables where appropriate.

As far as the particular ranking variables are concerned we differentiate between a large number of performance measures. The data was again matched form the FORIX tables to the driver and to the Grand Prix to maintain the panel structure. For the classification in the race itself we have matched the points, official rankings and the times. From these variables other measures such as wins, podiums, an accumulated winning variable for each driver and similar variables are constructed.

The information on the racing times for each driver is converted into second measures. This posed a specific problem: If a driver is one or more laps behind, only the time when this driver passed the finishing line in his round is available. For example, if M. Schumacher wins with a two lap difference to R. Barrichello the second measures for M. Schumacher and his rival can be very close because R. Barrichello does not finish the race but his time is stopped as soon as he passes the finishing line after M. Schumacher. To solve this problem we use the available times of drivers that finished the race in the same round as the winner did and calculate an average time per lap. This average is then multiplied with the number of laps a driver is behind the winner and the obtained value is added to the driver's time measure. In general this procedure underestimates the time of drivers who are one or more laps behind, as the drivers who finish the race in the same round as the winner have a lower lap average than the ones that do not finish in the same lap. Nonetheless, we think that it is a reasonable proxy for the real driver's achievements. This is mainly because there is a considerable difference in finishing one or more laps behind but the time difference within one round is rather small. Additionally any adjustment of the obtained time measures by a factor would be possible but arbitrary as such a factor would have to be based on individual specific lap times which are not available. Finally, this procedure of calculating times and time differences enables us to compare different drivers that are one or more laps behind with drivers that finish in the same round as the winner.

Other performance measures are the starting grid positions and the time of the fastest lap.[10] There is also information on the reasons of dropping out of a race which is coded as a HUMANOUT or a TECHNICALOUT. FORIX notations for drop-outs such as "accident", "collision", "not classified", "suspension", "disqualified", "physical" are considered as human outs in our evaluation and are correspondingly coded. Notations like "fuel leak", "gearbox", "brakes", "electrical", "transmission", "axle", "oil leak", "engine" are taken as technical reasons for drop-outs. Particularly, the technical breakdowns might be influenced by human error. Overloading the engine, for example, is considered in our evaluation as a technical out although it is possibly a human fault that could have been prevented. Vice versa collisions might be caused by technical problems with the brakes. Although we are aware of such problems there is no possibility of correcting them. In spite of this, the overall picture of the analysis will be more accurate by adding theses variables than by not using them in further evaluations.

3.1.2 Descriptive statistics of constructed variables

The process of matching results in a panel of 750 Grand Prix, 300 driver, 750 cars and altogether 15305 lines times 40 column entries (driver, car, tyre and lubrication dummies

[10] In the beginning of the Formula 1 the information on the fastest lap of a driver is not always available.

not included). This enormous amount of data poses a problem concerning the presentation of our calculations for the drivers. Ideally, a matrix with dimensions 300 times 300 would have to be used to enter explicitly the information for every driver. As such a format is impossible the results are reported in the best applicable form in this paper. In general, we will present in detail the achievements of 50 drivers who we think are the most likely to be in a list of the best. Sometimes the focus will be put on a subset of these 50 racers. Missing and additional results can be found in the supplied dataset on the author's website.

Table 2 presents some primary descriptive statistics on Grand Prix circuit characteristics as well as those driver characteristics which are sensible to represent over the whole dataset. Note that some descriptive statistics only make sense on an individual driver's level. The number of observations *N* is therefore included to facilitate comprehension.

Table 2
Descriptive statistics of constructed variables

Variable	*N*	*Min*	*Max*	*Mean*	*Standard deviation*	*Median*
GPDISTANCE	750	52.920	804.680	328.750	78.833	307.570
GPPERIMETER	750	3145	25'579	5'666	3591.250	4601
GPLAPS	750	12	200	66.830	23.452	68
WEATHERCODE	750	-2	2	1.179	1.221	2
DRIVERAGESTART	300	19	54	28.900	6.167	28
DRIVERAGEEND	300	20	56	34.420	6.351	34
DRIVERPEAKYEARS	300	1	14	2.253	2.505	1
DRIVERNBRACES	300	1	256	51.017	54.644	30
DRIVERNBWINS	300	0	84	2.500	7.633	0
DRIVERNBPODIUMS	300	0	142	7.497	16.021	1
DRIVERNBCARSUSED	300	1	25	6.793	5.018	5
DRIVERNBTYRESUSED	300	1	5	2.063	0.971	2
DRIVERNBLUBUSED	300	1	7	2.177	1.368	2

Source: Author's calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; All reported number are rounded to three decimal places or given as integers.

Already the descriptive statistics have an interesting story to tell:

The shortest Grand Prix as far as the mileage is concerned took place in Adelaide (Australia) in 1991. Due to heavy rain the Grand Prix was stopped after 14 laps and only 24 minutes and 34.899 seconds of racing time. A. Senna the leader at this point was declared winner. The longest Grand Prix took place in Indianapolis (United States of America) from 1951 to 1960 which counted also as Formula 1 Grand Prix. The circuit's perimeter was 4'023.4 meters and had to be followed 200 times adding up to a total distance of 804.468 kilometers, exactly 500 miles.

25’579 meters was the longest perimeter of a circuit ever recorded. The race took place 1959 in Pescara (Italy) and lasted for 18 rounds adding up to a total of 460.422 kilometers. The winner of this race was S. Moss followed by J. M. Fangio. The shortest circuits were located in Monte Carlo from 1950 to 1972 with a perimeter of only 3’145 meters.

The French driver P. Etancelin started his career at the age of 54. After two years he left Formula 1 at the age of 56 having started in 12 Grand Prix. He accumulated three points during his career. Probably more astonishing is the result of L. Fagioli. He is the oldest Grand Prix winner. At the age of 53 L. Fagioli won the Grand Prix of France in 1951.

Table 2 contains some other interesting results as far as the age of the drivers is concerned. The mean and the median of the variable DRIVERAGESTART, which measures a driver’s age when he started his career, are almost identical. The same is true for the variable DRIVERAGEEND. Taken together with standard deviations of 6,167 for DRIVERAGESTART and 6,351 for DRIVERAGEEND there is some indication that it is not necessarily the age itself that has an influence on a racer’s performance.[11] It would be worthwhile at this point to consider that when F. Alonso won his first Grand Prix in Hungary in 2003 at the age of 22, he was the youngest Formula 1 racer ever to win a race. A driver’s age has of course an influence if it is considered as a proxy for his experience. Indeed, we will see in section 4.2 that a quadratic function represents well the influence of age on performance. The same is true if the constructed experience variable itself is considered.

According to the defined criteria for peak years M. Schumacher and A. Prost stand out with 14 and 13 peak years respectively. This primarily means that they won in each of these 14 and 13 years at least one Grand Prix. The career of M. Schumacher started in 1991 and is not finished yet whereas A. Prost was active from 1980 to 1993. A. Prost has therefore won at least one race in every year of his active fellowship as a driver apart from his first racing season in 1980.[12]

Concerning the distributional properties of the presented descriptive statistics, the mean and the median of variables characterizing Grand Prix are similar but the standard deviations tend to be rather high. The WEATHERCODE variable differs from this pattern in the sense that its median equals its maximum value. As we shall principally focus on bad weather conditions this does not matter but may even facilitate the analysis. The hypothesis

[11] The driver’s ageing process is therefore likely to have a minor effect. A racer’s experience is probably more important.

[12] To finish our list of interesting anecdotic results we would like to mention that most podium classifications within a year were achieved by M. Schumacher in 2002. He managed to reach 17 podiums in 17 races. It was also M. Schumacher who had the longest series of finishing races without a breakdown. He managed to pass the finishing line 24 times in a row without a technical or a human accident. As a final gimmick note that J. Watson won 1983 from the start position number 22.

is that sunny conditions do not differ significantly from each other as far as the performance of certain drivers is concerned. On the contrary, rainy, wet and cold weather significantly influences the behavior of some drivers. The coding of the weather variable allows us to estimate the influence of bad weather conditions on the number of drop-outs and to identify a rain or bad weather champion in addition to the overall champion.

The driver specific descriptive statistics are more skewed. The median for the number of peak years, the number of wins and the number of podiums differs substantially from the mean of the distribution. The same is true for the number of races: DRIVERNBRACES has a mean of 51.017 with an even higher standard deviation. Altogether this indicates a possibly sharp fall from the best drivers to the weakest ones. It is therefore possible that identifying the best drivers is easier than establishing a ranking over the whole dataset because of the skewedness of the distributions.

The mean of the number of different cars used by the racers is 6.793 with a standard deviation of 5.018. The median is exactly five in this case. Some drivers were frequently changing their cars during their active years. For the number of tyres and lubrication used the changes are minor in comparison.

3.2 Initial ranking with dependence on cars

Most comprehensive lists of Formula 1 rankings tend to publish results based purely on the number of wins, pole positions or podiums in the past. They do not consider that the conditions in Formula 1 races changed during the last 50 years. Nowadays there are typically 17 Grand Prix within a year whereas there were approximately half as many in times gone by. The simple sum of wins leads to the (incorrect) conclusion that M. Schumacher is the best Formula 1 driver ever and that he has a huge gap to his next competitors such as A. Prost. Very successful racers like A. Ascari, J. Clark, J. M. Fangio or N. Farina often do not show up on such lists. This is because the absolute number of wins or pole positions has to be put in relation to the number of races.

A simple and easy method of deriving some information on capabilities and strengths of different drivers is therefore to take relative measures of performance. Table 3 provides some primary statistics on rankings based on such measures. The performance variables are the “Number of wins”, “Wins over races”, “Podiums over races”, “Poles over races”, “Fastest laps over races” and “Races in points over races”. Needless to say, this method of evaluating a driver does not allow the distinction between a driver specific and a car specific component. Nevertheless, it is a straightforward way to obtain some relative measures of performance.

For Table 3 we took the first 50 drivers with most of the wins at the end of the driving season 2005. These racers will also be used for other representation in this paper such as for the paired comparisons and the econometric model. Our chief reflection of this choice was to use a commonly well known way of representing the drivers' achievements without loosing the focus of our study which is to evaluate a racer independently of the car. Besides, it can easily be argued that a certain number of wins is necessary in order to be considered as belonging to the best. All drivers starting from rank 48 up to 54 (included) have won three races.[13] To fill up the remaining three drivers to get 50 we took the most recently active ones. Therefore T. Boutsen, J. Herbert and H.-H. Frentzen are included in the TOP-50 list but we do not explicitly report the achievements of P. Collins, M. Hawthorn, P. Hill and D. Prioni who have also won three races.

In order to concentrate as much information as possible within our tables we use subscripts and bold letters to establish some sort of inner column ranking. As highlighted, the driver column is ordered according to the wins column. The subscript behind each number indicates the ranking within the specific column. Bold letters indicate the first ten drivers in that column. For example: A. Senna is third according to absolute "Wins", 14th according to "Races started", seventh according to "Wins over races", sixth according to "Podiums over races", and so on. The same representation will be used throughout this study and the positions of the drivers in the tables will not change. This is done to facilitate the reader's comprehension and allowing at the same time to represent much information in a very condensed way.

3.2.1 Interpretation of results

The applied method of calculating relative performance measures changes the general picture of the ranking completely. Particularly drivers that are not often referred to, such as A. Ascari or N. Farina join the ranks of well known stars. At the same time young drivers like R. Räikkönen or F. Alonso can improve their rankings substantially. Conversely, drivers like N. Lauda lose pace.

[13] These seven drivers are in alphabetical order (active years in parenthesis): T. Boutsen (1983-1993), P. Collins (1952-1958), H.-H. Frentzen (1994-2003), J. Herbert (1989-2000), M. Hawthorn (1952-1958), P. Hill (1958-1966) and D. Pironi (1978-1982).

Table 3

Descriptive driver statistics and rankings based on relative measures

Driver	*Wins*	*Races started*	*Wins over races*	*Podiums over races*	*Poles over races*	*Fastest laps over races*	*Races in points over races*
M. Schumacher *active driver*	**84_{1}**	**230_{2}**	**0.365_{3}**	**0.617_{2}**	**0.278_{5}**	**0.300_{4}**	**0.757_{3}**
A. Prost	**51_{2}**	**199_{6}**	**0.256_{6}**	**0.533_{4}**	0.166_{11}	**0.206_{6}**	**0.643_{5}**
A. Senna	**41_{3}**	161_{14}	**0.255_{7}**	**0.497_{6}**	**0.404_{4}**	0.118_{16}	**0.596_{7}**
N. Mansell	**31_{4}**	**187_{9}**	**0.166_{10}**	0.316_{15}	**0.171_{9}**	**0.160_{9}**	0.439_{24}
J. Stewart	**27_{5}**	99_{37}	**0.273_{5}**	**0.434_{8}**	**0.172_{8}**	0.152_{11}	**0.576_{8}**
J. Clark	**25_{6}**	72_{42}	**0.347_{4}**	**0.444_{7}**	**0.458_{2}**	**0.389_{2}**	0.556_{11}
N. Lauda	**25_{6}**	171_{12}	0.146_{13}	0.316_{15}	0.140_{17}	0.140_{14}	0.427_{25}
J. M. Fangio	**24_{8}**	51_{47}	**0.471_{1}**	**0.686_{1}**	**0.569_{1}**	**0.451_{1}**	**0.843_{1}**
N. Piquet	**23_{9}**	**204_{5}**	0.113_{16}	0.294_{21}	0.118_{20}	0.113_{19}	0.490_{18}
D. Hill	**22_{10}**	115_{30}	**0.191_{9}**	**0.365_{9}**	**0.174_{7}**	**0.165_{8}**	0.487_{19}
M. Häkkinen	20_{11}	161_{14}	0.124_{14}	0.317_{14}	0.161_{12}	**0.155_{10}**	0.516_{16}
S. Moss	16_{12}	66_{45}	**0.242_{8}**	**0.364_{10}**	**0.242_{6}**	**0.288_{5}**	0.530_{15}
J. Brabham	14_{13}	126_{27}	0.111_{17}	0.246_{26}	0.103_{23}	0.095_{22}	0.421_{26}
E. Fittipaldi	14_{13}	144_{24}	0.097_{22}	0.243_{27}	0.042_{34}	0.042_{40}	0.396_{27}
G. Hill	14_{13}	176_{10}	0.080_{29}	0.205_{36}	0.074_{27}	0.057_{34}	0.335_{39}
A. Ascari	13_{16}	32_{50}	**0.406_{2}**	**0.531_{5}**	**0.438_{3}**	**0.375_{3}**	**0.719_{4}**
D. Coulthard *active driver*	13_{16}	**193_{8}**	0.067_{33}	0.311_{17}	0.062_{29}	0.093_{23}	**0.570_{10}**
C. Reutemann	12_{18}	146_{21}	0.082_{27}	0.308_{18}	0.041_{35}	0.041_{41}	0.452_{22}
A. Jones	12_{18}	116_{29}	0.103_{20}	0.207_{35}	0.052_{32}	0.112_{20}	0.336_{38}
M. Andretti	12_{18}	128_{26}	0.094_{23}	0.148_{43}	0.141_{16}	0.078_{29}	0.297_{45}
J. Villeneuve *active driver*	11_{21}	151_{19}	0.073_{30}	0.152_{41}	0.086_{25}	0.060_{33}	0.325_{42}
J. Scheckter	10_{22}	111_{34}	0.090_{24}	0.297_{19}	0.027_{42}	0.045_{38}	0.477_{20}
G. Berger	10_{22}	**210_{4}**	0.048_{35}	0.229_{28}	0.057_{31}	0.100_{21}	0.448_{23}
J. Hunt	10_{22}	92_{38}	0.109_{18}	0.25_{25}	0.152_{14}	0.087_{25}	0.380_{30}
R. Peterson	10_{22}	123_{28}	0.081_{28}	0.211_{34}	0.114_{21}	0.073_{30}	0.341_{37}
K. Räikkönen *active driver*	9_{26}	85_{40}	0.106_{19}	0.353_{11}	0.094_{24}	**0.188_{7}**	0.541_{14}
R. Barrichello *active driver*	9_{26}	**214_{3}**	0.042_{39}	0.285_{22}	0.061_{30}	0.070_{31}	0.467_{21}
F. Alonso *active driver*	8_{28}	68_{43}	0.118_{15}	0.338_{12}	0.132_{18}	0.044_{39}	**0.574_{9}**
D. Hulme	8_{28}	112_{33}	0.071_{31}	0.295_{20}	0.009_{46}	0.080_{27}	0.545_{13}
J. Ickx	8_{28}	114_{31}	0.070_{32}	0.219_{30}	0.114_{21}	0.114_{17}	0.351_{34}
J. P. Montoya *active driver*	7_{31}	84_{41}	0.083_{26}	0.333_{13}	0.155_{13}	0.143_{13}	**0.619_{6}**
R. Arnoux	7_{31}	149_{20}	0.047_{36}	0.148_{43}	0.121_{19}	0.081_{26}	0.282_{47}
R. Schumacher *active driver*	6_{33}	145_{23}	0.041_{40}	0.179_{39}	0.041_{35}	0.055_{35}	0.552_{12}
T. Brooks	6_{33}	38_{48}	0.158_{11}	0.263_{24}	0.079_{26}	0.079_{28}	0.395_{28}
J. Surtees	6_{33}	111_{34}	0.054_{34}	0.216_{32}	0.072_{28}	0.090_{24}	0.36_{31}
J. Rindt	6_{33}	60_{46}	0.100_{21}	0.217_{31}	**0.167_{10}**	0.050_{37}	0.35_{35}
J. Laffite	6_{33}	**176_{10}**	0.034_{43}	0.182_{38}	0.040_{37}	0.034_{43}	0.335_{39}
G. Villeneuve	6_{33}	67_{44}	0.090_{24}	0.194_{37}	0.030_{41}	0.119_{15}	0.313_{43}
R. Patrese	6_{33}	**256_{1}**	0.023_{47}	0.145_{45}	0.031_{40}	0.051_{36}	0.285_{46}
N. Farina	5_{40}	33_{49}	0.152_{12}	**0.545_{3}**	0.152_{14}	0.152_{11}	**0.758_{2}**
C. Regazzoni	5_{40}	132_{25}	0.038_{42}	0.212_{33}	0.038_{38}	0.114_{17}	0.394_{29}
K. Rosberg	5_{40}	114_{31}	0.044_{38}	0.149_{42}	0.044_{33}	0.026_{46}	0.333_{41}
J. Watson	5_{40}	152_{18}	0.033_{44}	0.132_{46}	0.013_{43}	0.033_{44}	0.309_{44}
M. Alboreto	5_{40}	**194_{7}**	0.026_{46}	0.119_{47}	0.010_{45}	0.026_{46}	0.242_{49}
B. McLaren	4_{45}	100_{36}	0.040_{41}	0.270_{23}	0.000_{48}	0.030_{45}	0.500_{17}
D. Gurney	4_{45}	86_{39}	0.047_{36}	0.221_{29}	0.035_{39}	0.070_{31}	0.360_{31}
E. Irvine	4_{45}	146_{21}	0.027_{45}	0.178_{40}	0.000_{48}	0.007_{48}	0.342_{36}
H.-H. Frentzen	3_{48}	156_{17}	0.019_{48}	0.115_{48}	0.013_{43}	0.038_{42}	0.359_{33}
T. Boutsen	3_{48}	163_{13}	0.018_{50}	0.092_{49}	0.006_{47}	0.006_{49}	0.252_{48}
J. Herbert	3_{48}	160_{16}	0.019_{48}	0.044_{50}	0.000_{48}	0.000_{50}	0.181_{50}

Source: Author's calculations.

Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; Subscripts indicate inner column rankings; Bold letters designate the first ten ranks in a column; All reported number are rounded to three decimal places or given as integers.

The "Wins over races" variable is probably the most general measure of a Formula 1 racer's performance. As far as this variable is concerned it seems as if M. Fangio is clearly ahead of M. Schumacher. M. Schumacher occupies the third rank after M. Fangio and A. Ascari. J. Clark and J. Stewart are ahead of A. Prost and A. Senna. A similar picture emerges when "Podiums over races" are analyzed. This measure as well as the measure "Races in points over races" partly indicate the consistency and stable achievements of a driver. The first ten drivers in the ranking "Races in points over races" realized points in over 50 % of their races. Anyhow this measure does not necessarily have to represent the best results as D. Coulthard shows. In almost 60 % of the races D. Coulthard obtained some points but he is far behind if "Wins over races" are considered.[14]

Potentially good drivers in strong cars can, to a certain degree, be identified using this approach of evaluation even if they have a stronger team partner. Such Formula 1 racers would, of course, not show up in the column "Wins over races" but in one of the other columns such as "Podiums over races" or "Races in points over races". This might be useful to keep in mind when looking at the outputs of further evaluations in this section. Indeed, it is often said that R. Barrichello did not manage to win a lot of races because M. Schumacher was the number one driver in Ferrari. Taking a closer look at Table 3 shows that R. Barrichello is ranked 39th for the measure "Win over races", 22nd for the measure "Podiums over races", and 21st for the measure "Races in points over races". There might therefore be a negative bias for R. Barrichello's winning measure because he was driving against M. Schumacher. Still such a bias is limited in the sense that in every race there can only be one winner. A very strong driver winning most races in a year has a negative effect on the winning probability of all drivers (good and bad ones) in that season. As we try to identify the best Formula 1 racer a hypothetical point measure and a classification measure will be used in section 3.4 in order to evaluate overall performance.

For "Poles over races" and "Fastest laps over races" a different picture emerges once more when considering relative measure as opposed to absolute count measures.[15] These indicators point to fast drivers and training champions but not necessarily to race winners. High rankings for racers in these two columns and in the columns "Win over races" as well as "Podiums over races" designate Formula 1 racers who show good training results and fast lap times. Moreover, such drivers are able to realize their comparative advantage during the race and carry home their starting advantages as wins and podium position. The same is

[14] In fact S. Moss, D. Coulthard and C. Reutemann are the Formula 1 runners with the highest number of accumulated wins (16, 13 and 12 respectively) but never managed to win the World Championship.

[15] Note that absolute count measures are not reported here.

true when considering "Fastest lap over races" and "Wins over races" together. Drivers showing high values for these two performance measures might be considered as fast and reliable. The inverse is likely to be the case for drivers like K. Räikkönnen, for example. He seems to be a driver who is capable of achieving fast rounds but cannot realize a high relative winning score. For J. Rindt the same is can be asserted if the variable "Poles over races" is considered. It might be interesting to note that A. Senna is the winner for the absolute number of pole positions. He had 65 pole positions whereas M. Schumacher had 64 (at the end of the 2005 racing season). Still, J. M. Fangio is the relative winner for the variable "Poles over races".

It is remarkable to notice that J. M. Fangio is the winner in all relative categories. M. Schumacher is mostly occupies the positions two to five. For several measures A. Ascari and J. Clark are in front of M. Schumacher. The rankings for classification three to ten are mostly shared by N. Farina, S. Moss, A. Prost, A. Senna, and J. Stewart. Rankings from classification ten on are less obvious and depend on the performance variables considered.

3.2.2 Limits and problems of the approach

The limits of the applied method are evident. Although we take into account the number of races the analysis does not allow us to separate the effects of the driver and the car. A potentially good driver could have a very bad score because he was always in a very weak car or because he had a very strong partner using the same car. This problem can be partly eliminated by considering other measures. Moreover, such a driver can easily be compared with others also showing fewer wins because of a very strong racer in a particular season.

Besides, by simply dividing the performance variables by the number of races information on experience, age and career peaks is not considered. Certain drivers could be very strong during a limited period of the active professional time. Such drivers do not obtain a good ranking using this method of evaluation. Their good years would be divided by a huge number of races when they did not perform very well whereas drivers stopping their activity after their peak period would receive a comparatively high score. This problem is difficult to solve because it requires an exact definition of peak periods. Anyhow, we shall try to correct for this possible bias when focusing on the regression analysis in section 3.4. This difficulty can be partly tackled by using a control variable for peak years and using other performance variables such as a continuous classification measure.

Furthermore, we do not take into the team partner's characteristics by calculating simple relative measures. Drivers with a high number of races and only a small number of wins and podiums do not have to be weak if they are compared more properly, conditioning on other variables for drivers' strengths and capabilities. It is therefore necessary to evaluate the performance of a driver in comparison to his team partner taking account of a proxy for experience. Such an evaluation shall be done in the next section.

3.3 Comparisons of drivers in the same car

As has been mentioned before, simple comparisons of Formula 1 drivers based on their relative achievements is flawed because it does not take into account the different cars that were used by the drivers. In this section we will try to set up a paired comparison model in order to evaluate different drivers on the same car. Indeed, not all drivers can be directly compared to each other because they were never using the same cars. Fortunately, statistical methods enable us to calculate a likelihood of winning any specific race based on a limited number of direct comparisons (see BRADELY, 1984 for a handbook article or DAVIDSON AND FARQUHAR, 1976 for an extended bibliography). In a first step of this approach we will simply compare drivers in the same teams using their accumulated points in a certain racing season. Afterwards, the analysis will be refined and direct comparisons will be performed for every race and every driver on the same car with his partner on that car. Additionally an experience variable will be introduced and a panel will be set up allowing us to calculate the likelihood of winning any comparison. Finally, a closer look at time differences will be attempted. Note that we consider the comparison based on points in a season and the comparison of time differences as illustrations of possible ways to obtain a ranking. Our scientific and statistical focus lies on the paired comparisons in subsection 3.3.2. The other two evaluations might mainly be of interest to the general public that is less acquainted with statistical and econometric procedures.

3.3.1 Comparisons based on points in a season

The simplest way to compare two drivers is to look at their relative performance in the team and to add up the points they made during a racing season.[16] For the 750 Grand Prix from 1950 to 2005 and the 300 drivers we find a total of 2'213 combinations in the same

---

[16] The idea of using points as well as time differences (see subsection 3.3.3) in this way to compare drivers was proposed by Prof. Dr. Reiner Eichenberger, the supervisor of this Master thesis.

team and the same racing year.[17] A driver's team partners (in the future sometimes noted as rivals) are counted several times if they were in the same team over more periods. We sum up of the points of the driver and his rivals over the all the races in a season. Additionally, a variable counting the total number of points that can be achieved in a particular racing season is introduced. This number divides the sum of points achieved in the season by a driver and serves as a relative measure for performance. Such a procedure is necessary because the number of races changes over the years and therefore the number of points that can be achieved changes too. Adding up only the points for different seasons would distort the results: scoring 30 points against 20 when the total of points in a racing season equals 160 is different from scoring the same number of points when the total is 700. Moreover, not only the total number of points changes but also the method of awarding them. In the past, for example, nine points were given to the winner whereas nowadays he receives ten points. Simply adding up the points and comparing them between different Formula 1 runners leads to wrong conclusion. Using a relative measure of the points within a season solves this problem. Two output examples is of this method of evaluation is given in Table 4.

Table 4

Example for comparison based on performance measure "Points in a season"

Driver	*Rival*	*Year*	*Total points in season*	*Points of driver*	*Relative points of driver*	*Points of rival*	*Relative points of rival*
A. Prost							
	J. Watson	1980	339	5	0.015	6	0.018
	R. Arnoux	1981	375	43	0.115	11	0.029
	R. Arnoux	1982	399	34	0.085	28	0.070
	E. Cheever	1983	369	57	0.154	22	0.060
	N. Lauda	1984	384	71	0.186	72	0.188
	N. Lauda	1985	397	73	0.184	14	0.035
	K. Rosberg	1986	396	72	0.182	22	0.056
	…	…	…	…	…	…	…
K. Räikkönen							
	N. Heidfeld	2001	442	9	0.020	12	0.027
	D. Coulthard	2002	442	24	0.054	41	0.093
	D. Coulthard	2003	624	91	0.146	51	0.082
	D. Coulthard	2004	702	45	0.064	24	0.034
	J. P. Montoya	2005	738	112	0.152	60	0.081
	…	…	…	…	…	…	…

Source: Author's calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; All reported number are rounded to three decimal places or given as integers; Tables for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

[17] We have to control for the racing year in this evaluation because the team is taken to compare drivers and not a specific car. When cars are taken, technological advances are controlled for with the car itself.

As mentioned we denote as “Driver” the driver who is being evaluated and as “Rival” the racers that were his team partners in a particular year.[18] The column “Total points in season” gives the total number of points that could be accumulated in a season. The measure “Points of driver” are the points the driver accumulated and “Relative points of driver” equals “Points of driver” divided by “Total points in season”. The columns for the rival can be interpreted respectively.

In order to feature in this paper, both drivers need to have made more than ten points in a season. This is done because firstly, we do not run an experience variable and we do not perform direct comparisons within races but use yearly accumulated points. Secondly, this helps us to avoid problems of non-transitivity when a potentially good driver loses in his early years against a weaker one. Finally, we shall report here only a subset of the drivers who have already been focused at in Table 3 in the “Wins over races” column. They are the most likely to be in the list of champions and due to economy of space not all results can be reported. Again detailed numbers and comparisons for all drivers and all combinations can be found on the author’s website. For the examples of Table 4, J. Watson will not be evaluated against A. Prost because the number of points did not exceed the lower bound of ten. The same applies for N. Heidfeld when analyzed against K. Räikkönen. E. Cheever will not be reported against A. Prost because he does not show up in Table 3.

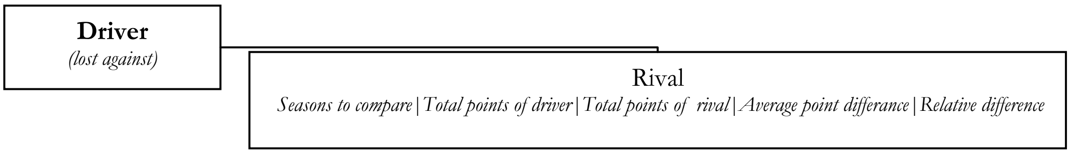
In a next step the points of the drivers are added up for each rival over the years they were driving in the same team together. The same is done for the rival’s points and the relative points of the two. By subtracting the rival’s points from the driver’s points and dividing by the number of years they competed against each other, we obtain an average number of points for which the driver is better (worse) than his rival. This variable will be denoted as “Average point difference”. Furthermore, we sum up the “Relative points of rival” minus the “Relative points of driver” to receive a “Relative difference” measure. This enables us to come up with a partial ordering.

The resulting numbers are reported in Figure 2. Unfortunately it is impossible to present all 50 drivers from Table 3. Therefore only the first 15 and the active drivers who were not matched with any of these (this is only K. Räikkönen) from Table 3, column “Wins over races” are shown here. A tree representation with these drivers would overload the page and hamper visibility. As a consequence, we adopt an approach similar to a time line where the estimated strength decreases with moving to the right. A clear disadvantage of this

[18] At the beginning of Formula 1 racing until the 1970s not all drivers were affiliated with teams. In this case the first car used in the season is the basis for grouping.

approach is that not all information available can be represented. The obvious advantage is that such a representation is easy to understand, it preserves readability and it is easy to compare drivers against each other. The boxes of Figure 2 have to be interpreted in the following way:

Figure 1
Box interpretation for comparisons based on "Points in a season"



Source: Author's representation.

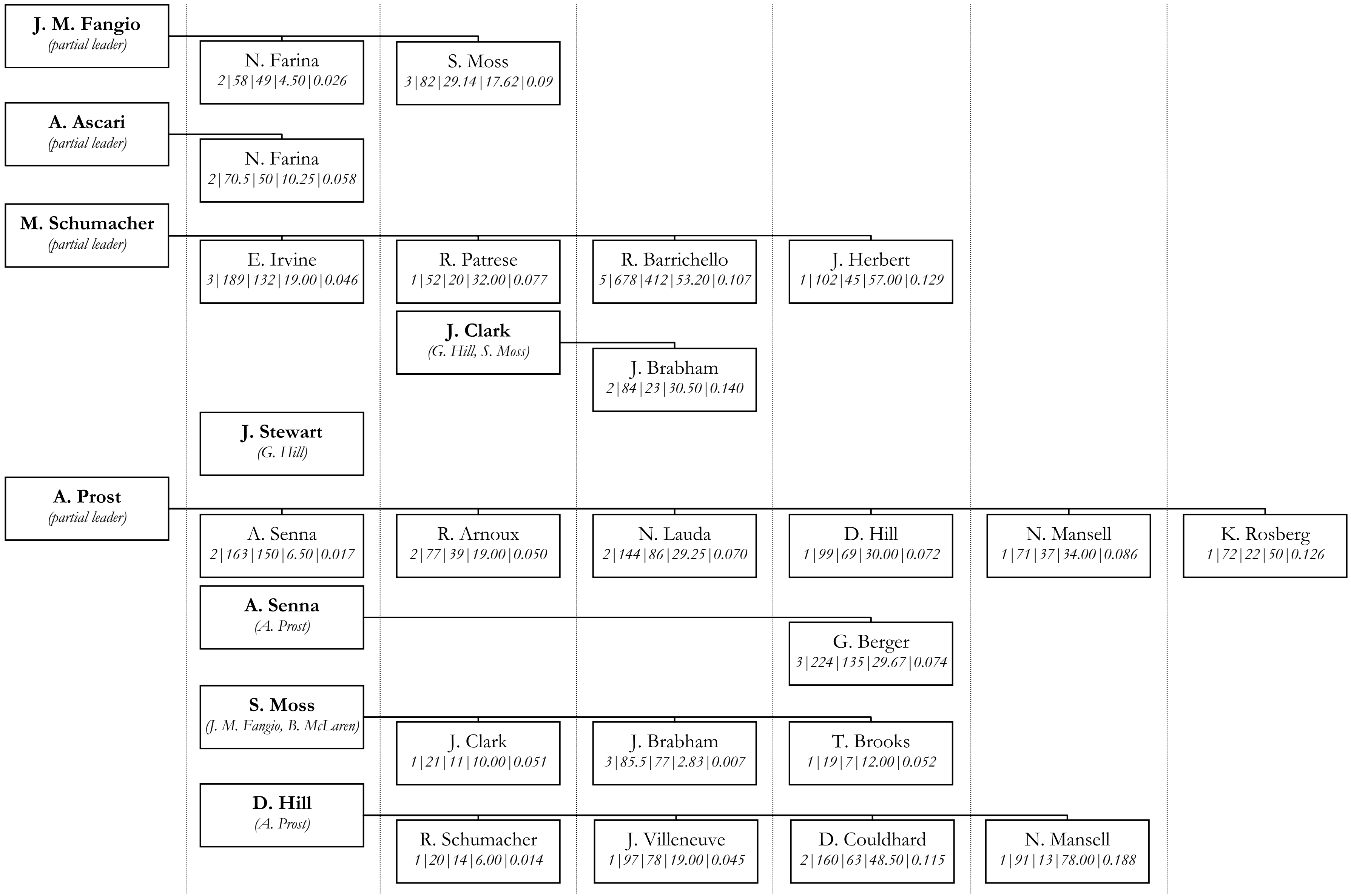
If the driver lost against no rivals he is reported as a partial leader. The reported rivals are only those that are also shown in Table 3.

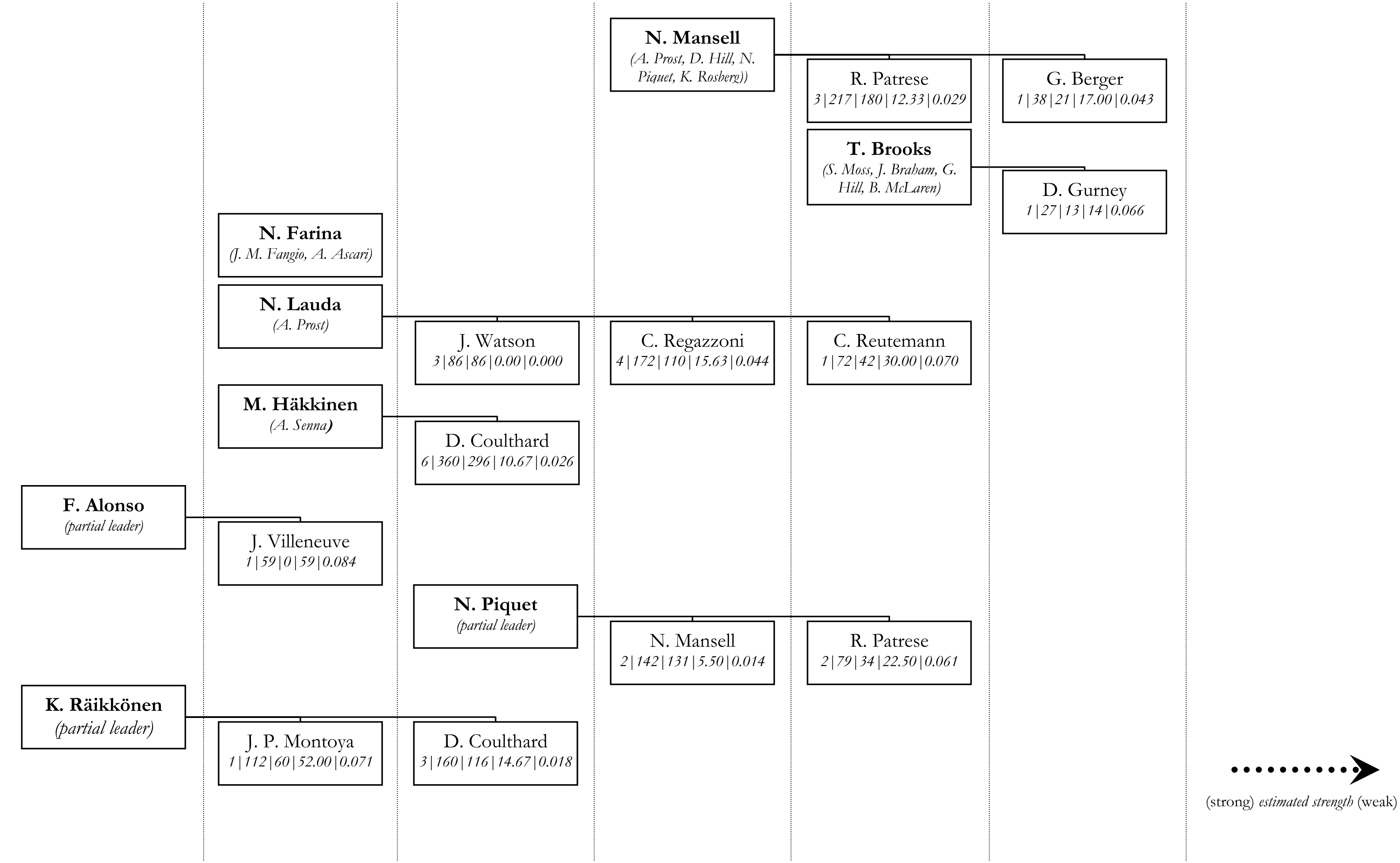
3.3.1.1 Interpretation of results of comparisons based on "Points in a season"

This method of evaluation shows some similar results as the method applied in section 3.2 but here the drivers are really compared in the same team. In later evaluations we will not focus on the team but on the car itself. As can be seen in Figure 2, J. M. Fangio, A. Ascari, M. Schumacher, A. Prost, F. Alonso and K. Räikkönen are all partial leaders. It is interesting that only these six out of the first 16 driver of Table 3 can be identified as partial leaders in this comparison. A. Senna or N. Lauda, for example, have both lost against A. Prost and can therefore be found as his rivals in this representation. Note that some drivers are fairly close to their rivals. To the extent that the "Average point difference" as well as the "Relative difference" are concerned N. Farina is very close to J. M. Fangio and A. Senna is only 6.5 points on average weaker than A. Prost. Due to the small number of observations (one to six racing seasons) the calculation of a standard deviation does not make any sense.

We would like to view this evaluation as a simple illustration of a partial order based on the number of accumulated points in a season. Nevertheless, the established figures allow some deeper insights than other available rankings to date. Clearly, using other performance variables may lead to other results.

Figure 2
Rankings based on "Points in a season"





Source: Author's representation.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosport-atlas.com; Box interpretations are explained in Figure 1; Only the first 16 and the active drivers that were not matched with any of the first 16 (this is only K. Räikkönen) from the "Wins over races" column of Table 3 are used in this representation; If a driver is not comparable to a driver according to the defined criteria, he is left out in the figure; Tables for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

3.3.1.2 Limits and problems of the approach

We have already mentioned that due to space reasons only a very limited number of drivers can be represented this way. Out of the subset of 50 drivers a choice of 16 had to be made in order to allow a readably presentation. A tree representation would lead a complete loss of visibility as the branches would not follow a classical order but would sometimes be connected under each other. The timeline approach adopted solves the problem of readability by sacrificing some information in the panel and by decreasing the number of drivers to represent. The flash indicating a racer's strength is only an estimate based on the following comparisons in this subsection. It should only be interpreted as an indication of strength.

If more drivers are included and if the limit of at least ten points in a year to enter a comparison is dropped, we face problems of transitivity of the results. This does not come as a surprise. Comparisons based on points do not need to be transitive at all as a number of other sports show. Further research and an exact listing of the intransitivities might be of interest. Practically they can be performed for every driver using the established tables in this thesis.

Additionally, the differences in average points are often minor. As no standard deviations are calculated because of the limited number of seasons to compare no hypothesis tests can be made.

Although we try to correct for a possible bias coming from the absolute number of points by dividing it with the total number of point in a season, there is still the risk left that strong rivals are wrongly taken as weak ones. This could be the case if one driver scores a lot of points in a certain period and his rival scores almost none. In the next period the pattern changes and the rival is stronger but cannot make up for the points lost in the first season. This is a general problem with any aggregation procedure for rankings as mentioned by KLABRODA (2000).

Moreover, it has to be noted that as the number of drivers increased during time competition increased too. It could be argued that it is nowadays more difficult to achieve any points because the number of rivals in a race is higher than it was in the beginning of Formula 1 racing. Still our evaluation focuses on the best drivers and therefore on those who are supposed to be able of achieving points in a race with a lot of entrants as well as in races with a limited number of entrants.

Finally, a comparison based on points in a season does not take account of what is happening in the races. A lot of new information can be obtained by evaluating every race

separately and using more refined methods. The evaluation of this subsection should only be regarded as one possible illustration of a partial ranking based on points in a season.
The next goal is to obtain a comparable number for all drivers so that a clear list of champions can be set up and not just a partial one. In the next subsection such a number shall be calculated as a likelihood of winning any comparison based on different performance variables for a driver.

3.3.2 Paired comparisons and likelihood estimates

The results of subsection 3.3.1 were established via direct comparisons of different drivers in the same team based on accumulated points and relative points. The procedure and the partial ranking of Figure 2 are not satisfying because they do not allow us to compare all drivers and they do not take account of additional information in the panel and the experience of a driver. Therefore a standard paired approach is used to obtain a ranking for all drivers.
As mentioned by STERN (1992) paired comparison data usually arise when subjects or objects are compared to elicit a strength or a certain degree of strength. Today the literature on paired comparison modeling is vast, spanning fields such as statistics, marketing, psychology and decision sciences. A common situation is to observe paired comparison data over time where the underlying value or strength of the observed unit is changing. This might occur, for example, in comparing outcomes of games played between competitors. According to JOE (1990) the approach can often appropriately describe the merits of game players. The method of paired comparisons shall be applied here in order to calculate the probability of winning any comparisons between the Formula 1 racers.

3.3.2.1 General setup of the paired comparison approach

In paired comparisons, t Formula 1 racers compete and are compared with each other in pairs on the same car. We will use the terminology of racers who win, lose or draw. The statistical methods devised are thus ranking methods and, while they are not literally nonparametric methods, they are often classified as such according to BRADLEY (1984).
Four performance variables are considered in our paired comparisons. The first one is a simple classification variable. It takes the value of 1 if a driver wins concerning race classification against his rival and 0 if the rival wins or both racers drop out of the race. The second variable is based on the podium position. It takes the value of 1 if the driver is on

the podium and his rival is not, or if the driver is better classified on the podium than his rival. The podium variable equals 0 if none of the drivers reached a podium position.

We distinguish between a classification measure and a podium measure because it could be possible in the evaluation that a bad driver wins always against an even worse one and therefore joins the ranks of the top. Using the podium measure as an additional control somewhat solves this problem.

In order to take into account the grid classification we use two performance measures. The first one evaluates the grid position in the same manner as the race classification. The second variable focuses only on the first starting row and is constructed equivalently to the podium measure.

In the pair-wise analysis we only compare two drivers if they were using the same car. Being in the same team but using different cars is not any information which is used for comparisons. The justification would be that drivers in the same team using different cars might either have a comparative advantage (if they were using a more sophisticated model) or a disadvantage (if they were using a test model). The number of cases of drivers in the same team using different cars is rather small. Albeit, in the early years of Formula 1 racing no teams were present and cars were sometimes changed by drivers within one season. Concentrating on the car instead of concentrating on the team allows an even better distinction between the capabilities of a driver and his equipment. Additionally changes in technology can be partly accounted for when comparing drivers in the same car.[19]

In order to allow for new drivers entering Formula 1 to build up experience a separate control variable is introduced. The experience variable counts the number of races a driver took part in. If a driver participated in 15 or more races he will be compared with the driver sharing the same car in a race. While a runner's accumulated experience is lower than 15 he will not be compared with other racers. 15 races were chosen because this number represents the median value of Grand Prix in a season.[20] We condition on a separate experience variable and not on age itself because we want to avoid having drivers in the panel that were only racing a few times.

Constructing the performance measures for classifications, podiums, grid positions and first row positions leads to four separate lists, each containing 16'172 symmetric matches of drivers and their cars with drivers using the same car in the same year and race. To find all these matches loops were programmed using spreadsheet software and all 300 driver were

[19] The car models changed almost every year.
[20] Though, in the beginning there were generally close to ten races in a season.

extracted from the panel of 15’305 entries. Taking account of drivers who did not race against each other driver in the same car and who were only present for less then 15 Grand Prix 171 Formula 1 racers remain. Out of the 171 total racers being matched, all 50 drivers form Table 3 are included and they have a comparatively high number of matches with other drivers.

We believe that this procedure is appropriate because firstly it allows us to correct for a time period to acquire experience. Secondly, drivers who followed Formula 1 for a few Grand Prix and received very good or very bad classifications are left out of the analysis and will not join the ranks of the champions.[21] Such drivers could possibly distort the results.

16’172 is a rather small number of matches considering that we are really comparing drivers on the same car. Such unbalanced panels are possible to evaluate. Anyway, it would be required that paired comparisons can be made via links (see BRADLEY, 1984). As most drivers where only matched against a limited number of rivals but for several times we end up with a highly unbalanced table and links can generally not be established between all different drivers.

When comparing two drivers we first add up the number of races the two drivers were competing against each other in the same cars. A correction is then performed in order to obtain the number of races that can be evaluated because of differences in a driver’s experience. After applying this correction no supplementary corrections need to be made in order to sort out any other drivers because the remaining races can be declared as comparable. As a final step, we add together the number of times a driver won against his rival, the number of times he lost and the number of times both ended in a draw.

Table 5 gives an example of this procedure. The measure of performance is race classification in this case. Again the driver under consideration is denoted as “Driver” and his partner on the same car is noted as “Rival”. Looking at A. Prost and his rival N. Lauda it can be seen that they were in 30 races on the same car. 30 of these races can also be evaluated. N. Lauda lost 17 and A. Prost eight of them and five times both drivers did not finish the race. Such a listing as presented in Table 5 is done for every driver and all the four considered performance variables. The results are transferred to a matrix with dimensions 171 times 171. This is our main panel designed for paired comparisons.

21 A classical example in this case is L. Wallard who has won one Gand Prix in 1951 and took only part in a second one in 1950 finishing sixth.

Table 5
Example for comparison based on performance measure "Classification"

Driver	*Rival*	*Races (total)*	*Races (corrected for experience)*	*Rival loses*	*Rival wins*	*Draws*
A. Prost						
	J. Watson	8	1	1	0	0
	R. Arnoux	30	26	12	7	7
	E. Cheever	14	14	10	3	1
	N. Lauda	30	30	17	8	5
	K. Rosberg	16	16	12	2	2
	…	…	…	…	…	…
K. Räikkönen						
	N. Heidfeld	17	1	0	1	0
	D. Coulthard	51	51	25	21	5
	J. P. Montoya	17	17	9	7	1
	…	…	…	…	…	…

Source: Author's calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; All reported number are rounded to three decimal places or given as integers; Tables for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

3.3.2.2 Results of the paired comparison approach

For the paired comparisons themselves a standard approach described by BRADLEY (1984) is applied. The basic paired comparison model has n_{ij} comparisons for a given number of competitors *i* and *j* where $i = 1,...,t$. For each comparison, the order is designated by $a_{ij\alpha}$ where $a_{ij\alpha} = 1$ if *i* wins against *j* in comparison number α. As a further definition, let a_i be the total number of wins for driver *i* over all comparisons with driver *j*. The general approach to analysis of paired comparisons is based on an extended model of independent Bernoulli comparisons with probabilities estimated through likelihood methods. Let $\pi_i \geq 0$ be the winning probability of driver *i* over the whole sample of drivers and p_i the maximum likelihood estimator of π_i. The resulting likelihood equations are (see BRADLEY, 1984):

$$\frac{a_i}{p_i} - \sum_{j,j\neq i} \frac{n_{ij}}{p_i + p_j} = 0, \quad i = 1,...,t \tag{1}$$

and

$$\sum_i p_i = 1 \tag{2}$$

Other and more sophisticated models are possible such as those shown by GLICKMAN (2001) who even allows for the merits of the units to adapt over time. KNORR-HELD (2000) uses advanced procedures like extended Kalman filters and smoother. Although such methods could also be applied for this paper it has to be pointed out that they usually require more data on comparisons between sportsmen over time than is available here. Furthermore, as we mainly want to achieve ordinal rankings in this subsection our results are likely to be sufficient for this purpose.

The system of equations (1) and (2) to maximize the likelihood of the searched parameters cannot be solved for analytically. Numerical iterations are necessary in order to find a result for every driver. The formula for the iteration is given as follows:

$$p_i^{(k)} = \frac{p_i^{*(k)}}{\sum_i p_i^{*(k)}} \tag{3}$$

where

$$p_i^{*(k)} = \frac{a_i}{\sum_{j,j \neq i} n_{ij} / (p_i^{(k-1)} + p_j^{(k)})}, \quad k = 1,2,\ldots \tag{4}$$

The iteration is started with an initial specification of $p_i^{(0)}$ where p_i denotes the searched parameter for driver *i*. As is the standard, we shall take $p_i^{(0)} = 1/t,\ i = 1,\ldots,t$ as the initial value and because a sufficient number of iterations is run the result will not be influenced by the choice of the initial values. We programmed equations (3) and (4) in our spreadsheet and tested the formula with an example given in BRADLEY (1984).

One problem concerning the iteration is that the probabilities of some drivers (the very weak ones) will be approaching zero. As the applied procedure uses these probabilities as divisors we face a problem of numerical stability (see SCHWARZ AND KÖCKLER, 2004). A simple way to solve this problem is to set a lower bound for the divisors and stop the iteration as soon as the probabilities of the drivers fall below this bound. Clearly, this solution implies a loss of precision. Anyway, as only a simple ranking is intended and not a direct comparison of the calculated probabilities the results are not likely to be biased. Additionally, the proposed solution avoids advanced formula manipulation in order to achieve numerical stability.

The results of the iterations for the performance measures are reported in Table 6. Note that the calculated likelihoods are multiplied with a factor of 10'000. In addition we calculate the sum of all won comparisons of a driver and divide it by the number of total matches against all rivals on the same car. This results in a separate and simpler measure of performance.

Table 6

Rankings based on paired comparisons for several performance measures

Driver	*Classifications over matches*	*Lik. of better classification*	*Podiums over matches*	*Lik. of better podium position*	*Start positions over matches*	*Lik. of better start position*	*First starting row over matches*	*Lik. of first row starting*
M. Schumacher *active driver*	**0.696**$_{2}$	**179.0**$_{3}$	**0.599**$_{2}$	**412.4**$_{5}$	**0.855**$_{6}$	**220.8**$_{5}$	**0.459**$_{7}$	**241.8**$_{7}$
A. Prost	**0.548**$_{9}$	117.5$_{11}$	**0.452**$_{6}$	**287.4**$_{7}$	0.655$_{17}$	**134.8**$_{10}$	0.262$_{13}$	110.3$_{11}$
A. Senna	0.540$_{11}$	**130.6**$_{7}$	**0.490**$_{5}$	**338.1**$_{6}$	**0.850**$_{7}$	**353.9**$_{3}$	**0.660**$_{4}$	**462.3**$_{6}$
N. Mansell	0.407$_{34}$	63.3$_{28}$	0.309$_{15}$	135.7$_{17}$	0.519$_{33}$	80.0$_{17}$	0.284$_{11}$	98.3$_{15}$
J. Stewart	0.477$_{16}$	83.7$_{16}$	**0.386**$_{9}$	204.9$_{11}$	0.636$_{19}$	63.2$_{26}$	**0.295**$_{10}$	97.4$_{16}$
J. Clark	**0.650**$_{3}$	**142.4**$_{4}$	**0.550**$_{3}$	**601.4**$_{2}$	**0.850**$_{7}$	**482.2**$_{2}$	**0.750**$_{2}$	**894.7**$_{3}$
N. Lauda	0.442$_{22}$	72.6$_{23}$	0.301$_{19}$	132.4$_{20}$	0.596$_{23}$	59.7$_{29}$	0.199$_{19}$	53.2$_{24}$
J. M. Fangio	**0.821**$_{1}$	**292.0**$_{1}$	**0.714**$_{1}$	**686.9**$_{1}$	**0.893**$_{4}$	**775.3**$_{1}$	**0.750**$_{2}$	**644.7**$_{4}$
N. Piquet	**0.542**$_{10}$	79.4$_{20}$	0.282$_{21}$	**222.2**$_{9}$	0.732$_{11}$	78.6$_{18}$	0.225$_{16}$	**165.6**$_{9}$
D. Hill	0.373$_{40}$	52.3$_{38}$	0.134$_{42}$	53.1$_{38}$	0.433$_{42}$	29.6$_{47}$	0.090$_{38}$	20.3$_{38}$
M. Häkkinen	0.486$_{15}$	91.0$_{14}$	0.308$_{16}$	133.2$_{18}$	0.719$_{12}$	90.1$_{15}$	0.247$_{14}$	108.7$_{12}$
S. Moss	0.333$_{44}$	46.6$_{44}$	0.314$_{14}$	182.3$_{12}$	**0.784**$_{9}$	**198.3**$_{8}$	**0.510**$_{6}$	**517.3**$_{5}$
J. Brabham	0.424$_{29}$	61.8$_{30}$	0.174$_{33}$	64.4$_{34}$	0.522$_{32}$	48.5$_{34}$	0.163$_{25}$	55.7$_{21}$
E. Fittipaldi	**0.567**$_{7}$	**122.4**$_{9}$	0.373$_{11}$	179.6$_{13}$	0.612$_{21}$	76.2$_{19}$	0.119$_{30}$	40.2$_{30}$
G. Hill	0.426$_{27}$	63.4$_{27}$	0.208$_{25}$	81.4$_{27}$	0.525$_{31}$	35.3$_{40}$	0.149$_{26}$	44.1$_{29}$
A. Ascari	**0.581**$_{5}$	**129.5**$_{8}$	**0.452**$_{6}$	**422.3**$_{4}$	**0.935**$_{2}$	**232.0**$_{4}$	**0.903**$_{1}$	**2099.4**$_{1}$
D. Coulthard *active driver*	0.420$_{30}$	83.5$_{17}$	0.201$_{27}$	106.6$_{25}$	0.356$_{47}$	38.8$_{38}$	0.109$_{33}$	33.4$_{32}$
C. Reutemann	0.467$_{18}$	77.7$_{21}$	0.262$_{23}$	132.6$_{19}$	0.542$_{30}$	61.3$_{27}$	0.121$_{29}$	34.1$_{31}$
A. Jones	0.426$_{27}$	59.2$_{33}$	0.262$_{23}$	113.1$_{24}$	0.639$_{18}$	60.5$_{28}$	0.180$_{22}$	50.0$_{25}$
M. Andretti	0.434$_{25}$	53.4$_{37}$	0.145$_{38}$	37.2$_{44}$	0.658$_{16}$	72.6$_{22}$	0.224$_{17}$	106.5$_{13}$
J. Villeneuve *active driver*	0.433$_{26}$	76.5$_{22}$	0.100$_{48}$	21.4$_{48}$	0.617$_{20}$	80.6$_{16}$	0.117$_{31}$	44.8$_{27}$
J. Scheckter	0.439$_{24}$	71.3$_{24}$	0.193$_{31}$	67.7$_{32}$	0.421$_{43}$	49.9$_{32}$	0.105$_{35}$	15.6$_{40}$
G. Berger	0.399$_{37}$	61.2$_{31}$	0.192$_{32}$	69.3$_{30}$	0.487$_{37}$	65.7$_{24}$	0.114$_{32}$	20.6$_{37}$
J. Hunt	0.441$_{23}$	60.2$_{32}$	0.324$_{13}$	113.9$_{23}$	**0.941**$_{1}$	**171.1**$_{9}$	**0.559**$_{5}$	**1788.4**$_{2}$
R. Peterson	0.398$_{38}$	48.0$_{43}$	0.157$_{34}$	62.9$_{35}$	0.694$_{15}$	72.7$_{21}$	0.204$_{18}$	69.2$_{17}$
K. Räikkönen *active driver*	0.493$_{13}$	113.4$_{12}$	**0.380**$_{10}$	**217.2**$_{10}$	0.592$_{24}$	64.4$_{25}$	0.197$_{20}$	58.4$_{19}$
R. Barrichello *active driver*	0.318$_{48}$	54.8$_{35}$	0.153$_{36}$	68.1$_{31}$	0.412$_{45}$	59.0$_{30}$	0.100$_{37}$	44.3$_{28}$
F. Alonso *active driver*	**0.604**$_{4}$	**183.9**$_{2}$	**0.396**$_{8}$	171.4$_{15}$	0.604$_{22}$	128.7$_{12}$	0.226$_{15}$	54.2$_{23}$
D. Hulme	0.415$_{31}$	81.6$_{18}$	0.200$_{28}$	76.8$_{29}$	0.462$_{40}$	37.3$_{39}$	0.077$_{39}$	13.6$_{42}$
J. Ickx	0.446$_{21}$	68.9$_{26}$	0.286$_{20}$	120.8$_{21}$	0.339$_{48}$	17.9$_{49}$	0.179$_{23}$	54.3$_{22}$
J. P. Montoya *active driver*	0.508$_{12}$	**140.2**$_{5}$	0.308$_{16}$	154.2$_{16}$	0.554$_{28}$	73.9$_{20}$	0.277$_{12}$	100.6$_{14}$
R. Arnoux	0.360$_{42}$	32.9$_{50}$	0.149$_{37}$	49.5$_{39}$	0.544$_{29}$	43.6$_{37}$	0.184$_{21}$	56.0$_{20}$
R. Schumacher *active driver*	0.474$_{17}$	97.6$_{13}$	0.155$_{35}$	57.1$_{37}$	0.485$_{38}$	51.9$_{31}$	0.103$_{36}$	26.9$_{34}$
T. Brooks	**0.556**$_{8}$	**120.4**$_{10}$	0.111$_{44}$	32.1$_{45}$	0.556$_{27}$	30.4$_{46}$	0.056$_{44}$	15.0$_{41}$
J. Surtees	0.400$_{35}$	49.0$_{41}$	0.333$_{12}$	**228.4**$_{8}$	**0.900**$_{3}$	**218.9**$_{6}$	**0.300**$_{9}$	**227.3**$_{8}$
J. Rindt	0.378$_{39}$	48.8$_{42}$	0.267$_{22}$	116$_{22}$	**0.889**$_{5}$	**217.8**$_{7}$	**0.356**$_{8}$	**162.3**$_{10}$
J. Laffite	0.330$_{45}$	33.7$_{48}$	0.198$_{30}$	59.4$_{36}$	0.415$_{44}$	34.1$_{42}$	0.066$_{41}$	9.6$_{44}$
G. Villeneuve	0.408$_{33}$	70.8$_{25}$	0.204$_{26}$	77.8$_{28}$	0.714$_{14}$	119.0$_{13}$	0.122$_{28}$	23.1$_{36}$
R. Patrese	0.301$_{49}$	33.0$_{49}$	0.087$_{49}$	18.0$_{49}$	0.500$_{36}$	49.8$_{33}$	0.073$_{40}$	15.9$_{39}$
N. Farina	**0.571**$_{6}$	**137.3**$_{6}$	**0.500**$_{4}$	**492.1**$_{3}$	0.571$_{26}$	32.8$_{44}$	0.036$_{45}$	66.9$_{18}$
C. Regazzoni	0.364$_{41}$	52.1$_{39}$	0.103$_{46}$	25.6$_{47}$	0.290$_{50}$	16.9$_{50}$	0.065$_{42}$	9.9$_{43}$
K. Rosberg	0.412$_{32}$	62.0$_{29}$	0.141$_{40}$	65.6$_{33}$	0.718$_{13}$	133.8$_{11}$	0.106$_{34}$	25.2$_{35}$
J. Watson	0.329$_{46}$	39.5$_{45}$	0.143$_{39}$	41.5$_{41}$	0.457$_{41}$	33.5$_{43}$	0.129$_{27}$	27.3$_{33}$
M. Alboreto	0.32$_{47}$	35.7$_{46}$	0.109$_{45}$	43.2$_{40}$	0.484$_{39}$	34.5$_{41}$	0.023$_{46}$	2.6$_{48}$
B. McLaren	0.400$_{35}$	57.4$_{34}$	0.200$_{28}$	101.5$_{26}$	0.511$_{34}$	30.6$_{45}$	0.022$_{47}$	3.7$_{47}$
D. Gurney	0.491$_{14}$	84.3$_{15}$	0.302$_{18}$	172.1$_{14}$	**0.755**$_{10}$	98.2$_{14}$	0.170$_{24}$	49.7$_{26}$
E. Irvine	0.346$_{43}$	50.7$_{40}$	0.102$_{47}$	39.7$_{43}$	0.386$_{46}$	46.9$_{35}$	0.016$_{48}$	6.6$_{46}$
H.-H. Frentzen	0.447$_{20}$	81.2$_{19}$	0.132$_{43}$	27.4$_{46}$	0.509$_{35}$	67.0$_{23}$	0.061$_{43}$	8.8$_{45}$
T. Boutsen	0.452$_{19}$	54.6$_{36}$	0.140$_{41}$	41.0$_{42}$	0.581$_{25}$	45.2$_{36}$	0.011$_{49}$	1.3$_{49}$
J. Herbert	0.289$_{50}$	35.6$_{47}$	0.044$_{50}$	9.9$_{50}$	0.333$_{49}$	28.7$_{48}$	0.000$_{50}$	0.0$_{50}$

Source: Author's calculations.

Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; Subscripts indicate inner column rankings; Bold letters designate the first ten ranks in a column; Likelihood is abbreviated as "Lik." and the reported numbers were multiplied with 10'000; "Classifications over matches" represent the number of comparisons won divided by the number of total races to evaluate for a certain driver (other measures can be interpreted respectively);.Tables for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

For hypothesis tests we apply the major test proposed by several authors, that of winning equality. The null hypothesis is:

$$H_0: \quad \pi_1 = \pi_2 = ... = \pi_t = 1/t \tag{5}$$

and the alternative hypothesis is:

$$H_1: \quad \pi_i \neq \pi_j \quad \textit{for some } i, j \tag{6}$$

The null hypothesis is tested using the likelihood ratio λ_0 and $-2\,log(\lambda_0)$ follows a χ^2 distribution with *(t-1)* degrees of freedom. For all specifications p-values of 0.000 are calculated.[22] This is a clear indication that the π_i are not equal and some of the winning probabilities differ from each other.

Supplementary datasets like for Table 5, a matrix in order to establish paired comparisons and the comparisons themselves like in Table 6 have also been established for a "Fastest laps" variable. The results are not reported here because for the early seasons of Formula 1 racing fastest laps times are only available for the winner of the race or the racer with the fastest lap time. This decreases again the number of possible matching between the drivers. Therefore the scope of paired comparisons with fastest laps is limited and also the applied ordering does not allow us to compare early Formula 1 drivers with today's racers.[23]

3.3.2.3 Interpretations of paired comparison results

In general, the relative measures (better "Classifications over matches", "Podiums over matches", and so on) lead approximately to the same ranking results as the likelihood measures do. Theoretically it would be possible that the likelihood measures resulted in a different ranking. This is probably not the case here because the panel for the comparisons is highly unbalanced and links between different drivers can hardly be made. The gained insights using paired comparisons as opposed to simple relative measures can therefore be questioned. Nevertheless, the proximity of the two different ways of measuring talent indicates that the procedure has been applied correctly and that further insights might be gained with a larger and more balanced panel.

Note that the measure for better "Classification over matches" as well as the other variables do not have to be over 0.500 for the best drivers. Firstly, there is the possibility to draw which means that especially for the indicators "Podiums over matches" and "First

[22] The hypothesis tests had to be programmed separately and were checked with an example mentioned by BRADLEY (1984).

[23] The table and the results are available on the author's website.

row over matches" a rather small value would be expected. This is clearly the case for most drivers apart from the possible champions who have extraordinary high values for all indicators. This reflects the fact that they almost never lost a comparison and that they finished the race fairly often with a podium position and started in the first row too. Secondly, concerning the values of the relative indicators, drivers of the list were competing against each other. However, as the variables can just take continuous values between 0 and 1 a high number for one driver implies necessarily a lower one for another driver in the list. A similar problem is mentioned by UTT AND FORT (2002) when using Gini coefficients to measure competitive balance.

The likelihood measures indicate the probability of winning a comparison in the whole panel for the 171 racers to evaluate. Their values should not necessarily be interpreted as distances between drivers. Clearly, higher parameters indicate stronger drivers but the limited number of possible comparisons with the whole panel of all drivers partly restricts more detailed statements on relative performance between two arbitrary drivers.

Comparing the results of Table 6 with the indicators of Table 3 leads to new conclusions.[24] The rankings presented in Table 3 are sometimes dissimilar to those of Table 6. The paired comparison approach as well as the calculated relative measures allow us to evaluate a driver independently of his car as opposed to the variables presented in Table 3. Most of the likely champions remain same but sometimes change their positions change, according to the performance measure used. J. M. Fangio, A. Ascari, J. Clark, M. Schumacher, J. Stewart and A. Senna stay under the TOP-10. A. Prost can be found on rank eleven for the likelihood measure of better classification. It is interesting to note that F. Alonso, N. Farina and K. Räikkönnen join the ranks of the top, whereas N. Mansell, D. Hill and N. Lauda loose pace. This pattern corresponds well to Figure 2 where points in a season are used to establish comparisons. It seems as if especially J. M. Fangio is ahead of M. Schumacher who occupies in general the ranks from two to seven depending to the indicator used. The rankings based on the likelihood measures are in general lower for M. Schumacher than his rankings based on the "Classifications over matches" measure or the other relative measures. This could be the case because M. Schumacher won mostly against his nine different partners but others did too. J. M. Fangio was competing against more than 20 rivals in exactly the same car and most of them were reasonably strong against their rivals.[25]

It is also worthwhile noting that different performance variables such as the "Likelihood of

---

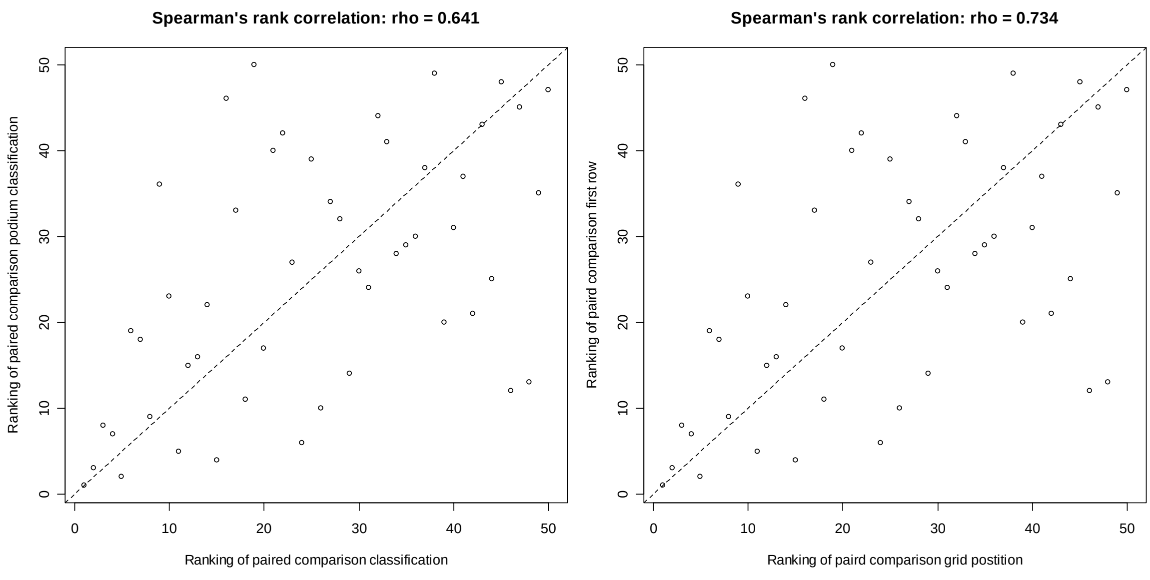
[24] No detailed comments but rather a broad overview of the results shall be given.

[25] Unfortunately some of J. M. Fangio's racing partners are not used in the evaluation because of differences in their experience.

better classifications" or the "Likelihood of better grid positions" lead to different results in the rankings. This confirms the belief that a distinction can be made between race champions and training champions. For the start positions A. Ascari seems to be a clear winner. This comes as no surprise as A. Ascari won over 90 % of the matches. A bit more astonishing is the performance of J. Hunt and J. Surtees concerning the grid measure. They are always under the first ten and J. Hunt once even achieves ranks 1 and 2. During his career J. Hunt achieved 14 pole and 24 front row positions over a total of 92 races. Mostly his partners using the same car were weaker than him concerning the start position.

Finally, we would like to show the correlation between the different rankings that were calculated by using the method of paired comparison. As rankings are compared, the usual coefficient of correlation can not be applied. Instead the Spearman's rank correlation measure is used. The results are presented in Figure 3.

Figure 3
Ranking correlations for performance measures of paired comparisons



Source: Author's calculations and representation.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; Rankings for the first 50 drivers taken from Table 6 are correlated; Left figure shows ranking correlation for "Lik. of better classification" with "Lik. of better podium position" and right figure shows ranking correlation for "Lik. of better grid position" with "Lik. of first starting row".

The diagonal line indicates when two rankings lead to exactly the same result. We focus on the comparison of the rankings resulting from the "Likelihood of better classification" with the "Likelihood of podium position" and the "Likelihood of better start position" with the

"Likelihood of first starting row". The calculated Spearman rank correlations are for a ranking comparison rather high, though it could be argued that even higher correlations would be expected because good drivers should show superior results no matter what measure is applied. We shall discuss this possible insufficiency and other problems with the results in the next subsection. Looking at the correlations between "Likelihood of better classification" and better "Classification over races" would result in far higher correlations.

3.3.2.4 Limits and problems of the approach

Overall the standard procedure of paired comparison and the results are consistent with the results of the "Wins over races" of Table 3 and Figure 2.[26] There are, however, some limits of the applied approach.

The results of paired comparisons give some indications of differences between various drivers on the same car. Here it is really the same car and not just a statistical proxy or the team but the price is a limited number of observations and therefore a limited number matches of drivers for paired comparisons. Although there is a large number of races between particular drivers and their team partners to be evaluated, in general there are not many comparisons between different drivers. Thus, most drivers stayed with their rivals for several seasons or at least several races. The matrix to calculate the maximum likelihood of the estimator for the probability to win any direct comparison is therefore highly unbalanced. As already mentioned this leads to the result that simple relative measures of better "Classifications over matches" or better "Start position over matches" are not inferior to the more sophisticated likelihood measures. Indeed, mostly the same results are achieved and the rankings are well correlated.[27] To reach a more balanced panel, additional observations could be won if the experience control was dropped and the evaluation started immediately with the first race. Anyhow this would also distort the results because drivers tend to lose comparisons against older and more experienced rivals especially in their first year. Furthermore, not a very large number of observations could be added by dropping the experience variable. Authors such as JOE (1990) faced a similar problem when he tried ranking 64 top chess players since 1800. It might be possible to achieve a better balanced panel if more data is compiled.[28]

---

[26] The same cars are used here whereas for Table 3 this is not the case. Some differences in rankings would therefore be expected but not a complete reversal.

[27] The correlations are not reported here.

[28] Remember that we "only" compiled data for the 300 drivers that were awarded at least one point during their whole career. 490 drivers did not obtain any points during their careers.

In addition, Figure 3 shows that the correlation of the two rankings, resulting from the "Likelihood of better classifications" with the "Likelihood of better podium position" and the "Likelihood of better grid position" with the "Likelihood of first row position" are not necessarily high. This is an indication that there is a certain probability that some potentially good drivers are grouped in the same team against even better ones. They therefore lose most of the comparisons. The opposite is also possible: very bad drivers could be in the same cars against rivals who perform even worse. As they win all the races the indicator for the "Likelihood of better classification" would show values that are far too high. In fact this problem is linked with the number of possible comparisons between drivers. If every driver raced in the same car against all the others such a difficulty would not exist. The application of the algorithms in order to form the pairs results in a mean of only 5.053 rivals per driver with a standard deviation of 3.198. This is a possible explanation for the somewhat weak correlations shown in Figure 3. It cannot be said here which measure is preferable. We are facing a typical problem of ranking and aggregation of results as mentioned by KLABRODA (2000). Indeed, he points out that it is fundamentally impossible to achieve a fully consistent ranking even for only one period. As we are trying to rank drivers over different periods the problem is only aggravated.

It is possible to dismiss the paired comparison results because of the assumptions on the distribution and the simple modeling process applied. For chess evaluations very sophisticated assumptions about the distribution of the strength of players were made, tested and applied as mentioned by GLICKMAN AND JONES (1999). Anyhow, STERN (1992) remarks with reference to other authors, that various models of paired comparison data lead to similar fits. His analysis of several sports datasets concerning the American football, baseball and basketball data provide adequate and comparable results. Due to reasons of space and the general problem with the unbalanced panel no tests with other distributional forms were performed. This may still be done in the future.

At the end, another problem with this approach is that the evaluation via paired comparisons is purely based on winning a comparison or loosing one.[29] However, winning for example "2 to 3" or "2 to 6" or "2 to 10" might be perceived differently. The same is true for losing. The applied method does not take account of this fact. Without using additional hypotheses to assign weights to such classification differences, these differences cannot be introduced into the analysis.

In a similar vein differences in times would be interesting to know and to evaluate. Again

[29] Indeed, this is the main idea and the strength of the approach as described by BRADLEY (1984)

winning always against a rival with a difference of one second could be interpreted differently from winning with a difference of one minute. A similar evaluation has already been done for points and teams in subsection 3.3.1. Nevertheless, comparisons shall be made in an analogous way in the next subsection for time differences on the same car.
Despite considerable room for improvement and the unbalanced data panel, we feel that our model captures the main components of Formula 1 racing outcomes. We believe that general practitioners and fans will see most of the rankings in this subsection as correct and overall appropriate.

3.3.3 Comparisons based on time differences within races

Another way to compare two drivers is to look at their performance in the same car and add up the number of seconds that differentiate them. This analysis based on time differences within races is performed in the same manner as the comparisons in subsection 3.3.1. No likelihoods are calculated because the results would be the same as for classifications: A faster driver wins a comparison against a slower one. The main difference as compared to 3.3.1, where points in a season have been used, is the following: Instead of points, time differences of seconds are added up in each race. In addition a normalized time difference is calculated. This is done because conditions can change between different races and the technological level changes every year.[30] A time difference of 30 seconds in a particular race and year does not necessarily have to be the same in another race and another year.
To construct the time differences we first look at the racing times of the driver and his rival in the same car and race and take the difference. Note that we do not compare the drivers if either one of them did not finish the race since no times are available in this instance. Therefore, only those races are used in the comparison where both drivers pass the finishing line. The construction of the time variable itself and the problems connected with the construction are described in section 3.1. There we also explain the construction of the average time for a race and what is happening if drivers are one or more laps behind.
In order to calculate a normalized time difference we take the time difference between a rival and the driver in the race and divide it by the constructed average race time. This allows us to take account of race specific characteristics such as weather conditions, long and short distances and so on. To control for drop-outs, a counter is also run for them. As

---

[30] Weather conditions can change, leading to slower races and cars become faster over time.

for the analysis in subsection 3.3.2 an experience variable is used that has exactly the same characteristics as described before and is also applied in a similar manner. The time differences and the normalized time differences are summed up for every comparable couple leading to an "Accumulated time difference" and an "Accumulated normalized time difference" measure.

Table 7 presents an output example for two drivers.

Table 7

Example for comparison based on performance measure "Time differences"

Driver	*Rival*	*Races to evaluate (without outs)*	*Driver outs*	*Rival outs*	*Accumulated time difference*	*Accumulated normalized time difference*
A. Prost						
	J. Watson	1	0	0	-88.480	-0.016
	R. Arnoux	4	15	17	-249.734	-0.047
	E. Cheever	5	3	7	-128.626	-0.025
	N. Lauda	9	9	17	104.833	0.018
	K. Rosberg	4	4	11	-223.252	-0.035
	…	…	…	…	…	…
K. Räikkönen						
	N. Heidfeld	0	1	0	0	0
	D. Coulthard	21	22	14	-448.794	-0.082
	J. P. Montoya	8	3	7	-52.185	-0.008
	…	…	…	…	…	…

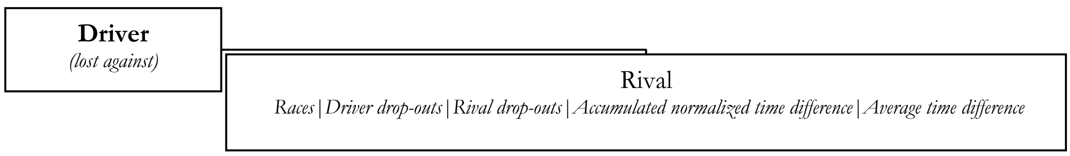
Source: Author's calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; All reported number are rounded to three decimal places or given as integers; Tables for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

Some short comments are worthwhile on Table 7. A. Prost was faster than all his rivals in the same car apart from N. Lauda taking account their experience. The "Accumulated time difference" is 104.833 seconds over nine races but it can be seen in the example of Table 7 that N. Lauda was only faster when neither N. Lauda nor A. Prost dropped out of the race. N. Lauda did not finish his races almost twice as often as A. Prost did. N. Heidfeld cannot be compared in this evaluation against K. Räikkönen because of differences in their experience.

In a next step, the "Accumulated time difference" for the rivals is divided by "Races to evaluate (without outs)" leading to an "Average time difference" measure. This gives an indication of the average time that a driver was faster (slower) than his rival when both passed the finishing line in the same car. We do not divide the "Accumulated normalized time difference" variable by the number of races as a further division is only partly meaningful and makes an interpretation of this measure even more complicated. The

correct interpretation for the "Accumulated normalized time difference" is as an accumulated difference of strength measured by time differences that are independent of Grand Prix and other technological conditions. As for the evaluation done in subsection 3.3.1 the applied procedure enables us to come up with a partial ordering of the Formula 1 racers based on their time differences. The resulting numbers are reported in Figure 5.
Again it is impossible to show results for all drivers of interest. Therefore we limit ourselves to the same drivers as used in Figure 2. The representation itself follows the same pattern as before. Strong drivers can be found on the left and the losers of the time based comparisons can be found on the right. The lines from one driver to the others indicate the ordering. Boxes in Figure 5 have to be interpreted in the following way:

Figure 4
Box interpretation for comparisons based on "Time differences"



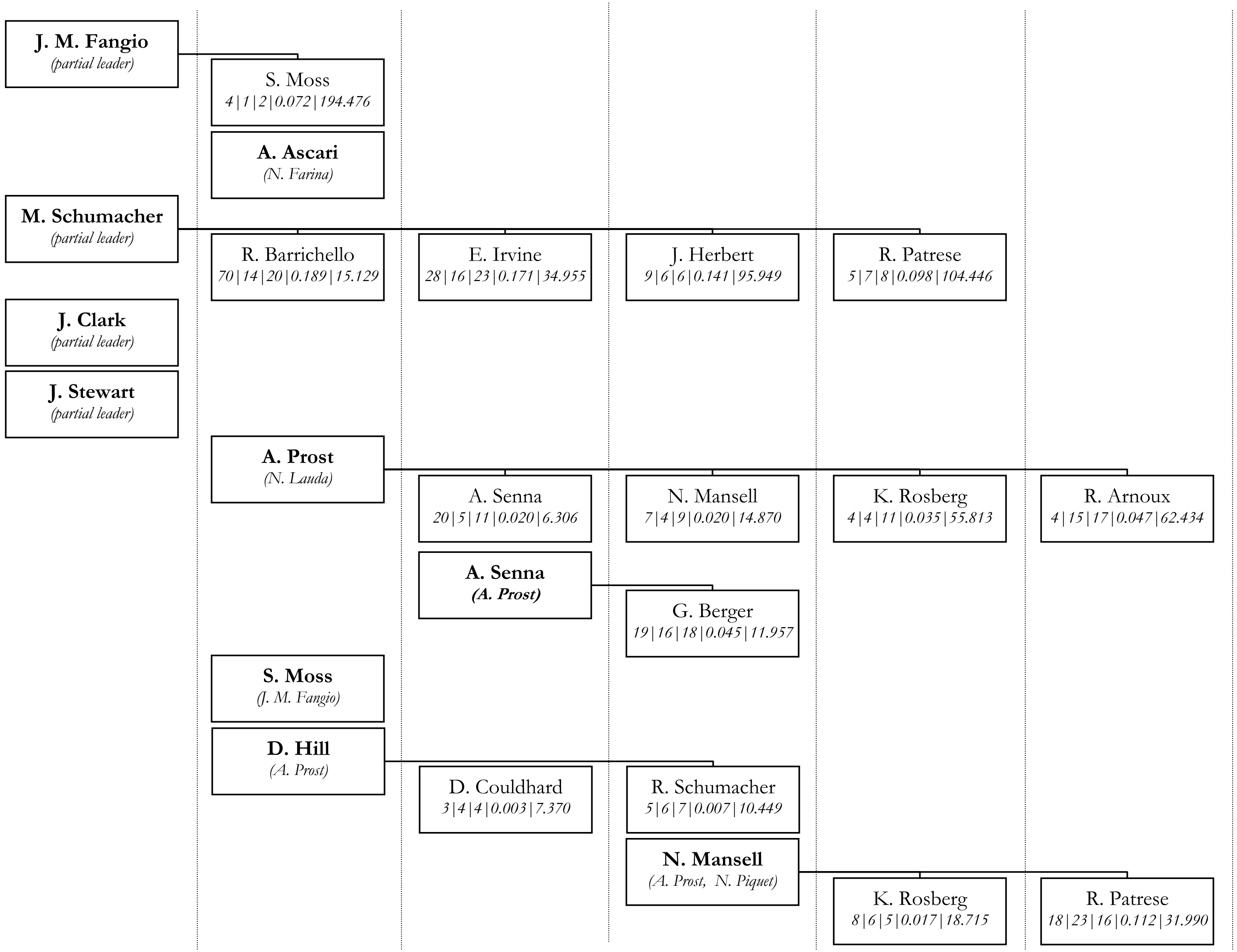
Source: Author's representation

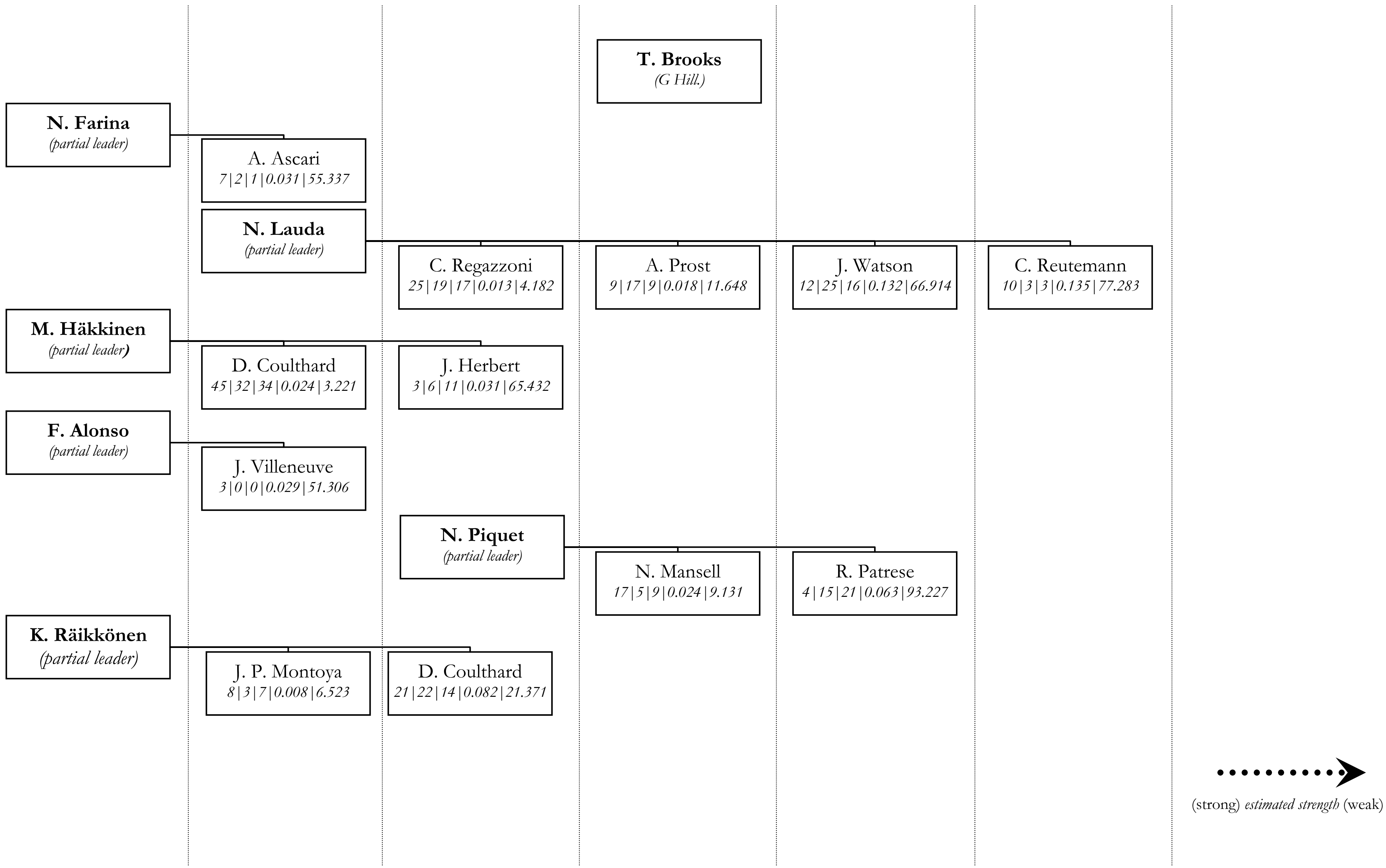
If the driver did not lose against a rival he is reported as a partial leader.

3.3.3.1 Interpretation of results of comparisons based on "Time differences"

Note that the established partial rankings in Figure 5 do not have to be the same as those in Figure 2 since point differences in a season and time differences in a race do not always have to show into the same direction. Moreover, when the "Average time difference" variable is considered some very slow races of a certain driver could bias the results when comparing the two figures. Therefore, it is necessary to consider the accumulated normalized time difference as a separate indicator too. We shall only give a very short overview over the results.
M. Schumacher, J. M. Fangio, J. Clark, J. Stewart, N. Farina, M. Häkkinen, F. Alonso, K. Räikkönen are partial leaders. A. Ascari loses in this evaluation against N. Farina and A. Prost loses against N. Lauda.

Figure 5
Rankings based on "Time differences"





Source: Author's representation
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosport-atlas.com; Box interpretations are explained in Figure 4; Only the first 16 and the active drivers that were not matched with any of the first 16 (this is only K. Räikkönen) from the "Wins over races" column of Table 3 are used in this representation; If a driver is not comparable to a driver according to the defined criteria, he is left out in the figure; Tables for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

Remember that the calculated time differences do not take into account the number of drop-outs which is reported as an additional variable. For most of the other Formula 1 racers the partial ordering remains the same as the ranking based on points in subsection 3.3.1. A. Prost is again only moderately ahead of A. Senna. So far these two drivers seem to have the same relative strength. Drivers like D. Hill and N. Mansell clearly fall behind in all the applied methods so far. For J. Clark and J. Stewart, no clear statements can be made because of a lack of valid comparisons. The "Accumulated normalized time difference" shows the same pattern as the "Average time difference". Small differences for the indicators can be interpreted as almost equal strength.[31] Still, the number of drop-outs has to be considered when evaluating two drivers against each other.

This method of evaluation also helps to solve a problem mentioned earlier. It was said that a potentially very good driver could lose a comparison against an even better one. Cleary, by applying the approach of using time differences to compare drivers the same ordering remains as in subsection 3.3.2. However, there is a possibility to measure distances in a continuous way by using "Average time differences" as well as "Accumulated normalized time differences" and not only the discrete way of a distance measure with 1 for wins, 0 losses which was used before.

Consider for example K. Räikkönen and the two reported adversaries: We notice that the "Average time difference" and the "Accumulated normalized time difference" between R. Räikkönen and J. P. Montoya is much smaller than for the former and D. Coulthard. Therefore J. P. Montoya can be put ahead of D. Coulthard in this partial analysis. Still, it has to be noted that R. Räikkönnen dropped out more often than D. Coulthard did. In a similar way the strength of M. Schumacher can be better evaluated. On average he was 15.129 seconds faster in a race than R. Barrichello, 34.955 seconds faster than E. Irvine, 95.959 seconds faster than J. Herbert and 104.446 seconds faster than R. Patrese.

Again we do not want to push the interpretations too far and would like to view this evaluation simply as an illustration of a partial order based on time differences.

3.3.3.2 Limits and problems of the approach

Yet again one clear drawback of this approach is the limited number of comparisons between drivers that are possible in the same car. Such problems with the data panel have been discussed before.

[31] No standard deviations are calculated because of the generally small number of matches.

Furthermore, there is the insufficiency that we do not directly consider drop-outs for the rankings. Indeed, the number of drop-outs is reported but it is not used to establish the ranking in combination with the time differences. If only one of the drivers dropped out of the races no comparison can be made as no times were available. This distorts the result because a driver who drops out often but is fast when he stays in the race should not necessarily receive a better ranking than one who always stays in the race. No direct correction for this problem is possible because any assumption concerning time differences if a driver does not finish a race would be arbitrary. We face again the general problem of aggregating different measures. Nonetheless, the results confirm the already perceived pattern of strength of the different drivers.

Finally, although transitivity seems to hold for the presented 16 racers it does not hold if all racers were included in the figure and the experiences variables were dropped. Further analysis of the data and the established measures and rankings has to look more closely at the problem of transitivity. Probably some interesting results and conclusions could be made by focusing on the question of non-transitivity of our measures of in subsection 3.3.1 and in this subsection. A single number as in subsection 3.3.2 for every driver avoids the permanent problem of transitivity in these methods of evaluation.[32]

In order to obtain one single and more pertinent number for every driver independent of the car, the technology and other covariates that do not directly influence a driver's capability and talent we proceed in the next section with an elaborated econometric model. This model will also be used for some applications and hypothesis tests in chapter 4.

3.4 Econometric model with driver and car effects

In this section an econometric model will be derived that includes driver and car effects.[33] At the beginning we would like to discuss some general points concerning this model and the estimation procedure.[34] Later on specific tests will follow as well as indications that the proposed specification for the derived rankings is acceptable and possibly the best that can be made in the light of a number of unknown effects.

---

[32] It could be argued though that exactly such a single number disguises the problem of transitivity and is therefore not valid.

[33] The original idea of distinguishing the car effects from the driver effects by using such a model stems from Prof. Dr. Reiner Eichenberger, the supervisor of this thesis.

[34] Consider the next three pages as a summary of the more detailed description of the model in the next subsections.

All drivers will enter the econometric model as dummy variables. The resulting coefficients will then be interpreted as a driver's ability or his talent. In order to decrease the number of dummies the cars are modeled as random effects. A similar approach for another field of research is described by LAIRD AND WARE (1982) for longitudinal data in an example at the end of their study. The data itself will be estimated by using a Generalized Least Squares fit supplied by the statistical software **R**. The code was written in a separate text file and all regressions were run after the whole programming was done. This allowed us to perform a huge number of regressions and tests in a comparatively efficient way.

For the data we shall generally use all available observations of the drivers in the regression framework. Only drivers with less than 15 races in total will be excluded. The main justification for this exclusion is that drivers not running more than 15 races could have achieved some relatively good classifications by chance during their limited careers and would therefore distort our results. The procedure of eliminating racers with less than 15 Grand Prix leaves us with a total number of 14'404 observations.

It has to be noted that we do not have a common panel structure for the data. A normal panel structure would have an indicator for the year, the race and the drivers as well as the cars. Indeed, we do have indicators for each year and the race but not all drivers were performing in all races. Once more our problem is one of a highly unbalanced panel. In order to evaluate the effect of a driver he is introduced as a dummy variable, as mentioned before. The dummy is mostly 0 for the 14'404 lines. It only takes the value of 1 if the driver it represents is in the same line in the design matrix. Therefore the dummy variable captures only the effect of being a certain driver on the dependent variable. To put it more clearly: 14'404 observations are used in the regression. A driver only participated in a number of races and in these races his dummy takes the value of 1 in the correct line where the driver really was participating.

We therefore obtain the following econometric specification formula for a dependent performance variable y_{it}:

$$y_{it} = \sum_{j=1} \alpha_j d_{j,it} + \sum_{s=1} \gamma_s d_{s,it} + X_{it}\beta + u_{it} \tag{7}$$

α_j are driver and γ_s are car specific effects. $d_{j,it}$ is a selection function that equals 1 if $i = j$ and 0 otherwise and $d_{s,it}$ equals 1 if $i = s$ and 0 otherwise. X is a design matrix including several control variables and β is the corresponding vector of coefficients. u_{it} is the error term.

Generally, such a model would be estimated applying the "Within Estimator" to estimate a fixed effects model. This is also what ZAX AND LYNCH (2000) do in their study on NASCAR racing. Similar methods like first differencing or other transformations remove the unobserved effect prior to estimation (see CAMERON AND TRIVEDI, 2005). Any other time-constant explanatory variables are removed along with α_j. It is therefore impossible to use such a method because our chief interest lies exactly in the estimation of the fixed effects for each driver. A traditional view of a fixed effects model is to assume that the unobserved effects α_j are parameters to be estimated. Thus, in equation (7) α_j is estimated along with the vector β. The way we estimate an intercept for each driver *i* is to put in a dummy variable for each cross-sectional observation along with the explanatory variables. This method is usually called a "dummy variable regression" (see WOOLDRIDGE, 2002B).

As mentioned before, we assume that the unobservable car effects are random variables that are distributed independently of the regressors. This is due to two main reasons: Firstly, we need to decrease the number of dummies in or model. Estimating a dummy for every driver and every car leads to non-interpretable results.[35] The most likely mathematical explanation for this phenomenon is that the condition number for such regressions is usually greater than 10^{18}, depending on the design matrix used. Generally speaking a very high condition number leads to problems with numerical estimations because a high number of floating point precision is lost (for more details see SCHWARZ AND KÖCKLER, 2004). Secondly, as will be shown later in this thesis the Breusch-Pagan test is highly significant and instead of estimating simple OLS we need to apply GLS and specify the covariance structure. In this case grouping by cars was used in the specification which results in an estimation of random effects for the cars. Different grouping levels are assumed to be uncorrelated in the specification.[36]

The random effects estimator exploits special features of the data. We shall estimate a feasible GLS estimator of the random effects model which is generally more efficient than a pooled OLS model. CAMERON AND TRIVEDI (2005) state that random effects estimators are inconsistent if the correct model is a fixed effects model.

---

[35] The computer produced an output but the standard errors were extraordinary high and most of the coefficients were skewed so that no interpretation was possible.

[36] We are well aware of the problems of the mentioned hypothesis. Unfortunately to date no other acceptable solution was feasible.

Usually economists tend to focus only on the sign of a coefficient and not on its value. Here explicitly the values of the coefficients have to be compared. If the values of coefficients are very close it cannot be rejected statistically that they are not equal from each other even if the standard deviation is rather low.

In the next subsection we will focus on finding a dependent variable to measure performance. A favorable measure should be a continuous one because we want to avoid running binary regressions at this stage. When taking, for example, wins as a dependent variable a huge number of zeros will be in the vector of the dependent variable.

3.4.1 Searching dependent variables

There are a number of different dependent variables that come into consideration when trying to evaluate Formula 1 drivers in an econometric model: racing times, racing speed, classifications, start positions, points, fastest lap times, and so on. We will mainly focus on the classification variable for several reasons that will be outlined in this subsection. Anyhow, regressions shall also be performed for some other variables to control for robustness.

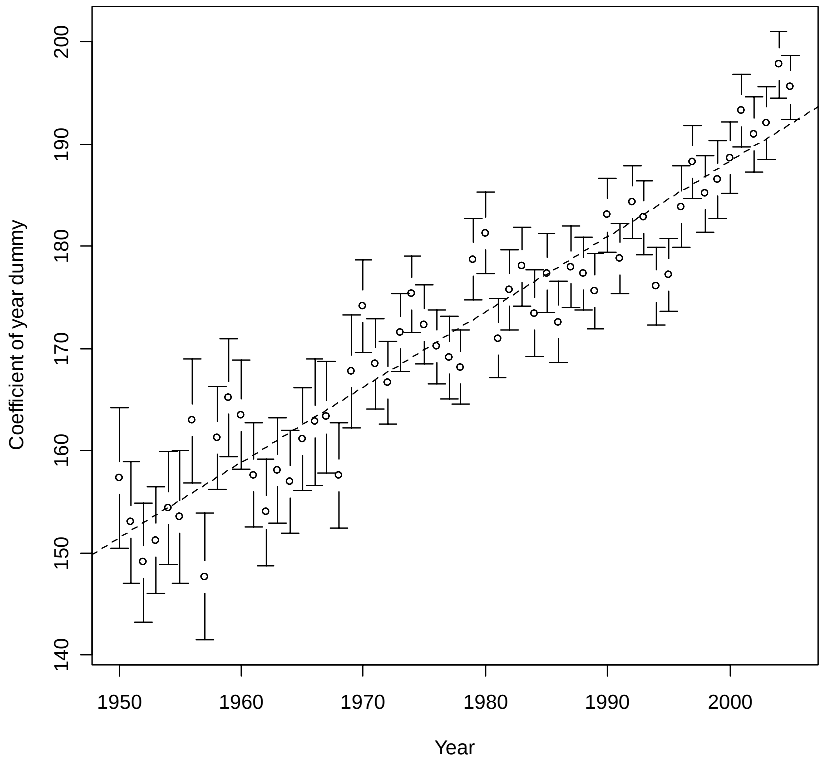
When using racing times as a dependent variable for a driver's performance there is the problem that the technological level increases continuously and cars become faster and more reliable every year. Furthermore, the weather conditions during a race have an influence on the times and speeds observed.

Estimating average speed and correcting for year dummies (proxies of technological advances) leads to the following results:

$$\frac{GPDistance}{ClassTime * 3600} = \underset{(0.239)}{3.958^{a}} \, (WeatherCode) + \beta \, (YearDummies) \qquad (8)$$
$$R^2 = 0.461$$

Note that only a subset of the 14'404 observations was used, mainly those for which the time was available. No constant was included in the regression because we used year dummies for every year from 1950 to 2005 and adding a constant would force us to drop a year dummy. The WEATHERCODE has been introduced to control for slower speeds during bad weather conditions. It can already seen that as the weather improves speeds tend to increase which does not come as a surprise at all. We do not report the coefficients of the year dummies of equation (8) explicitly. Instead they are represented in Figure 6 with their 95-%-confidence-intervals.

Figure 6
Coefficient of year dummies in speed regressions



Source: Author's calculations and representation.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; Coefficients of the year dummies stem from a linear estimation of the formula: CLASSTIME/(DISTANCEGP*3600) = WEATHERCODE + YEARDUMMIES; The dashed line represents an estimated speed trend using the coefficients.

As can be seen all year dummies are significant and positive. Additionally, a trend has been added in this figure as a dashed line.

The coefficients of the year dummy variables continuously increase over time. This makes the use of speed or time itself as a dependent variable very critical. In other words, since we do not know the driver effect, the car effect nor the technological effect on the dependent variable, we would have to estimate all those effects at the same time taking account of other problems. One ranking using speed as a dependent variable will be performed but the results can only be found on the author's website.

For fastest lap times we face the same problem plus the additional difficulty that this variable is not available for all drivers. Therefore neither speed nor fastest laps qualify as dependent variables.

Another possible candidate variable for the measurement of performance is the number of points in a race. The main problem with this variable is that the regulations changed during time and points are awarded differently nowadays than they were in times gone by. Still a hypothetical point measure can be constructed. In this case the classification of a driver is used and the points for all years are calculated for the classification according to today's regulations: first position yields ten points, second position yields eight points, third position yields six points and from the fourth position onwards the number of points decreases by one for each rank. This hypothetical measure for points will only be taken for robustness tests.

As all the measures have serious defaults, we shall mainly focus on classification as the dependent variable. The main advantage of this measure is that in every race there is a winner, a second one, and so on. The race classification itself does not depend on the technological level but mainly on a driver's ability and his car. The problem with this variable is that no classification is available for drivers if they did not finish the race. In our main specification using all observations we shall face this problem by setting the classification variable to 16 if a driver did not finish the race. When testing the robustness of our results we will perform also regressions using classifications only then, when they are really available.

3.4.2 Some preliminary specification tests

Using classification as a dependent variable some tests of the given specification in equation (7) are performed in this subsection without the use of GLS in order to justify the choice for the random effects specification. Additional explanatory variables in the design matrix are DRIVEREXPERIENCE, DRIVEREXPERIENCE^2 and TECHNICALOUT. The results of the tests for this specification can be found in Table 8. Similar specification tests have been calculated for other explanatory variables and the overall results remain the same.

For details regarding most tests, the reader is referred to DAVIDSON AND MACKINNON (1981) and standard econometric textbooks such as GREENE (1993). Other tests for the same questions are also available but were not performed here.

Autocorrelation can be rejected for the data sample because the p-value is far greater than 0.100.[37] Generally, when facing another distribution than a *Student-t* we shall report a p-value instead of different values for z, t, χ^2 or F.

The Cox-Test indicates that the inclusion of car effects increases the explanatory value of

[37] Null hypothesis are typically statements of no difference or effect. A p-value (probability value) close to zero signals that the null hypothesis is false, and typically that a difference is very likely to exist. Large p-values closer to 1 imply that there is no detectable difference for the sample size used.

the model significantly. Applying the RESET-Test on the specification indicates its correctness because the test is rejected. Furthermore, the Rainbow-Test is insignificant which means that the linearity of the true relation cannot be rejected. Only the Breusch-Pagan test is highly significant and indicating that our functional form is not completely correct but suffers from heteroscedasticity. As mentioned above, we therefore apply GLS and specify the covariance structure by grouping for cars. Finally correlating the exogenous variables with each other shows that there are no problems with multicollinearity.[38]

Table 8
Specification tests of econometric model

Test	*Brief description*	*Testvalue*	*p-value*
DW-Test	The Durbin-Watson test has the null hypothesis that the auto-correlation of the disturbances is 0.	2.041	0.760
BP-Test	The Breusch-Pagan test checks for heteroscedasticity. It fits a linear regression model to the residuals of a linear regression model (the same explanatory variables are taken as in the main regression model) and rejects the null hypothesis if too much of the variance is explained by the additional explanatory variables. Under null hypothesis the test statistic is asymptotically χ^2 distributed.	2455 (df=887)	0.000
Cox-Test	The idea of the Cox test is the following: if the first model (here a model without car effects) contains the correct set of regressors, then a fit of the regressors from the second model (here a model with car effects) to the fitted values from first model should have no further explanatory value. If it has, it can be concluded that model 1 does not contain the correct set of regressors.	NA	0.000
RESET-Test	The RESET test is a diagnostic for correctness of functional form. The basic assumption is that under an alternative model the regressors are generated by including them with taking their powers. A standard *F*-Test is then applied to determine whether these additional variables have significant influence.	1.910 (df1=4, df2=13512)	0.106
Rainbow-Test	The basic idea of the Rainbow test is that even if the true relation-ship is non-linear, a good linear fit can be achieved on a subsample in the “middle” of the data. The null hypothesis of the linear form follows an F distribution.	0.866 (df1=7202, df2=6314)	0.999

Source: Author’s calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; All reported number are rounded to thee decimal places or given as integers; The tested specification stems from equation (7) and is estimated using pooled OLS. Additional control variables in the design matrix are DRIVEREXPERIENCE, DRIVERSEXPERIENCE^2 squared and TECHNICALOUT; For the Cox-Test no car effects have been included.

3.4.3 Finding the preferred specification

In this subsection different combinations of explanatory variables are tested in order to find a preferred specification for which all driver dummy coefficients will be reported and interpreted in the next subsection. Table 9 shows seven different regressions with

[38] No specific tests for multicollinearity are reported.

classification as a dependent variable.

Table 9
Regressions for classification of drivers

Variables	*(1 OLS) Class.*	*(2 GLS) Class.*	*(3 GLS) Class.*	*(4 GLS) Class.*	*(5 GLS) Class.*	*(6 GLS) Class.*	*(7 GLS) Class.*
intercept	26.06[a]	25.138[a]	21.51[a]	10.68[a]	11.91[a]	10.94[a]	12.06[a]
	(1.712)	(2.051)	(1.921)	(0.792)	(0.759)	(0.785)	(0.754)
D_i (A. Prost)	-5.890_4[a]	-5.083_7[a]	-5.009_6[a]	-4.710_5[a]	-4.824_4[a]	-4.686_5[a]	-4.714_5[a]
	(0.762)	(0.851)	(0.812)	(0.859)	(0.817)	(0.856)	(0.817)
D_j (K. Räikkönen)	-5.618_6[a]	-5.544_4[a]	-5.304_4[a]	-4.426_6[a]	-4.619_6[a]	-4.290_6[a]	-4.516_6[a]
	(0.833)	(0.965)	(0.908)	(0.954)	(0.899)	(0.952)	(0.898)
$D_i - D_j = 0$ *(A. Prost - K. Räikkönen)*	*p-value* 0.600	*p-value* 0.481	*p-value* 0.623	*p-value* 0.619	*p-value* 0.729	*p-value* 0.540	*p-value* 0.738
DRIVERHOMEEFFECT	-0.151	-0.161	-0.194	-0.149	-0.186	-0.153	-0.188
	(0.132)	(0.128)	(0.127)	(0.128)	(0.127)	(0.512)	(0.127)
DRIVEREXPERIENCE						-0.015[a]	-9.8e-3[a]
						(2.9e-3)	(2.7e-3)
DRIVEREXPERIENCE^2						8.4e-5[a]	5.1e-5[a]
						(1.7e-5)	(1.5e-5)
DRIVERAGE	-0.952[a]	-0.902[a]	-0.601[a]				
	(0.098)	(0.117)	(0.110)				
DRIVERAGE^2	0.015[a]	0.014[a]	0.040[a]				
	(1.5e-3)	(1.8e-3)	(9.4e-3)				
DRIVERPEAK			-1.780[a]		-1.847[a]		-1.816[a]
			(0.088)		(0.088)		(0.089)
DRIVERNBCARSCONT				-0.014	5.2e-3		
				(0.033)	(0.030)		
DRIVERNBCARSCONT^2				4.7e-4	3.4e-4		
				(1.7e-2)	(1,5e-3)		
GRIDCLASS			8.0e-6[b]		7.0e-6[b]		7.0e-6[c]
			(3.7e-6)		(3.7e-6)		(3.7e-6)
WEATHERCODE			-0.081[a]		-0.083[b]		-0.083[b]
			(0.026)		(0.026)		(0.027)
TECHNICALOUT	7.227[a]	7.113[a]	7.074[a]	7.122[a]	7.081[a]	7.126[a]	7.084[a]
	(0.702)	(0.701)	(0.690)	(0.703)	(0.690)	(0.704)	(0.693)
Driver Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Car Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
N	14'404	14'404	14'404	14'404	14'404	14'404	14'404
adj. R^2	0.603	0.708	0.721	0.700	0.719	0.701	0.720

Source: Author's calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; Standard errors are reported in parenthesis; [a] Significant at 1-percent level; ; [b] Significant at 5-percent level; [c] Significant at 10-percent level; All numbers are reported with four digits; e indicates the exponential 10; Regressions (2) to (7) are performed using a feasible GLS estimator. All driver dummies are included and A. Prost and K. Räikkönen are only examples of two such dummies; Subscripts behind these drivers indicate the ranking when all coefficients are analyzed; The p-value is used for the test of the linear hypothesis whether the coefficients of the A. Prost and the K. Räikkönen dummy are equal; "Class." stands for classification; Regression results for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

A constant is always included in the regressions. Only drivers are included with a total of more than 15 races. 185 drivers remain and can be evaluated using the above mentioned procedure.

As a high number of dummy variables are used to explain the endogenous variables the general R^2 is likely to be overstated. We calculate a corrected R^2 by fitting the model and correlating the fit with the endogenous variable. In any case all our R^2s are highly significant which is partly explained by the large number of observations.

All estimations are performed for 14’404 observations. As the driver dummies are at the the correct positions they capture the influence of being a certain driver on the result. Please note that for all regressions the driver effects (driver dummies) are included and only the first specification does not include car effects.

Due to lack of space in this thesis we are not able to report the estimated coefficients for all drivers here. We therefore focus, as before, on A. Prost and K. Räikkönen as an example. Additionally, the linear hypothesis is tested for whether the driver dummy of A. Prost minus the driver dummy of K. Räikkönen equals zero. If this is the case the coefficients of the two drivers are not significantly different from each other. For this linear hypothesis test we only report a p-value. The subscript behind the coefficient of the two example drivers indicates their ranking when all the coefficients of the drivers are compared.

Consider, for example, regression (1). The coefficient of the A. Prost dummy equals -5.890 (0.762) and that of the K. Räikkönen dummy equals -5.618 (0.833). Since the dependent variable is classification, a lower coefficient indicates a stronger driver. In this specification A. Prost would be ranked fourth and K. Räikkönen sixth. The p-value is 0.600 and therefore statistically speaking the two coefficients are not significantly different from each other.

Looking closer at the coefficients and the rankings of the two example Formula 1 racers reveals that especially the ranking remains very stable over the specifications. A. Prost is mostly ranked around position five and K. Räikkönen mostly around position six. Only in specification (2) and (3) does the ranking between the two drivers changes but the p-values show again that they are not significantly different.

All regressions include a variable for a possible home advantage. This dummy is never significant but sometimes rather close to the 10-%-significance-level. A more detailed analysis of a possible home field advantage in Formula 1 racing may be of interest. Perhaps such an effect should be considered for the car and not for the driver.[39] This might be a future application for further research.

Regression (1) of Table 9 does not take account of car effects. As additional control variables the driver's age and the squared driver's age are used. As for all other specifications we also introduce a dummy for technical outs. The estimated coefficients show exactly the expected signs.

When introducing car effects in regression (2) it can be seen that they have high influence on the driver dummies. This is exactly what would be expected and it also confirms the Cox-Test of the last subsection. Both dummies decrease because some influence on the classifications can be explained via the car effects. Still both driver dummies stay significant. The introduced control variables are all significantly different from zero and show also the anticipated coefficient. Indeed the age variable and the squared age variables indicate a quadratic influence of the age itself. Being very young or very old increases the classification (indicating a lower performance). With approximately 32 years a minimum is reached (best classification) and then the classification increases again.

One method of ranking different sportsmen from the same or different periods is a comparison with the estimation of ageing function. Such comparisons are only possible, however, if careers overlap over the time as mentioned by BERRY ET AL. (1999). Another argument against focusing more formally on ageing functions for Formula 1 racers' achievements over different eras is that a number of parameters do not stay constant. The technological level of the used equipment increases, the cars themselves change, amendments are made in regulations, and so on. Additionally, BERRY ET AL. (1999) do not need to take into account a second unknown parameter influencing performance which is the car in our case. Finally, it seems to be that for Formula 1 racing the age of racers has only a minor influence on their performance[40] Controlling for age with a separate variable in a econometric model should be considered sufficient.

In regression (3) weather conditions, a career peak variable and the start position are introduced in the analysis. As anticipated weather conditions do not have an influence on classification itself. This is intuitively clear because a driver achieves a classification no matter if it is raining or not. Indeed, this was exactly the idea of using classification as a dependent variable: Only a limited number of variables have an influence on classification and most of these variables depend on the driver or the car. For that reason it is also clear

---

[39] One has just to think of the "Monza Effect" for Ferrari.

[40] Considering descriptive statistics shows that older drivers in Formula 1are not as uncommon as older athletes in other sports.

that the peak variable and starting position have a significant and negative influence on the classification. Starting from a better position leads to a better classification. When using a control for pole positions or first row starting position these controls are even more significant and have an even higher influence on the classification. Interestingly, including the DRIVERPEAK and GRIDCLASS does not seem to have a big influence on the driver dummies. However, this is only the case for the two drivers used in the representation.[41] The coefficients of the other drivers change when introducing GRIDCLASS as an additional variable. The introduction of DRIVERPEAK alone has only a minor influence which can partly be explained by the construction of this variable.

Regressions (4) and (5) do not use a driver's age as an explanatory variable. It is substituted by another proxy for experience. The number of passed cars used up to a certain race could be seen as an indicator for experience too. Anyhow, the coefficients of this variable and its squared value do not show any significant coefficients. The other coefficients remain similar to those in regressions (2) and (3).

In order to test the specification again the constructed experience measures of section 3.1 is introduced in regressions (6) and (7). This variable has the expected sign and is also highly significant. Regression (7) controls additionally for weather conditions, the career peak and the start position.

Regression (6) of Table 9 has some advantages over the rest of the specifications shown in Table 9 and also over other specifications that have been run. Firstly, the constructed experience variable is of higher interest than DRIVERAGE when teams choose their drivers. Particularly in the first 30 years of Formula 1 racing drivers were rather old as compared to nowadays. Still it is not the age itself that is important but the experience accumulated during the time. The experience variable has the advantage that it starts for every driver with the value of 0 and is therefore much easier to interpret over the large panel than the driver age variable. Besides, regression (6) is superior to (3), (5) and (7) because it does not include the starting position which is itself a performance variable nor does it include variables that are clearly not influencing the endogenous variable such as weather conditions. Furthermore, using the driver peak variable in the regressions is critical because it is a dummy that mostly equals 1 when also the driver dummy equals 1. We therefore rather not include DRIVERPEAK in our preferred specification.

Finally, the R^2 does only differ very little from all other specifications. We shall use regression (6) as our preferred specification and report a number of robustness tests on it.

---

[41] A. Prost and K. Räikkönen are both also strong in the qualification.

3.4.4 Robustness tests of econometric model

In the last subsection a preferred specification of the econometric model in equation (7) has been identified. Here we want to check the robustness of this specification when the number of observations changes. Furthermore, we will change the endogenous variable and some controls will be added and deleted. Table 10 reports a number of robustness checks. The exogenous variables are put into the columns of the table and we only look at the driver dummies in more detail as they are of interest in the further analysis.

Regression (01) represents again our preferred specification of Table 9 which has already been discussed.

Changing the endogenous variable in regression (02) from classification to the constructed point measure described in subsection 3.4.1 leads to a similar ranking as using the classification variable. The coefficients of the dummy variables change size and sign because here higher points indicate better results. It is interesting to note that in this specification the p-value for the comparison between A. Prost and K. Räikkönen indicates a significant difference between these two drivers at the 10-%-level. This can be partly explained by the construction of the endogenous variable. It takes the value of 0 when a driver reaches the position eight or worse. In fact, this means that drivers with classifications from nine onwards are treated the same way as drop-outs. This is also another reason why the constructed point variable is only used for robustness tests. It can also be noted that the R^2 is lower when the point measure is used. Still, when comparing the results at the end of this chapter it will be shown that the rankings derived from these econometric models are very well correlated.

Regression (03) includes the grid position as an additional control variable. The two presented coefficients of the driver dummies almost do not change. When looking at all coefficients some significant changes can be found. Regression (04) includes the starting position in specification (02). The results are similar.

For a test of a possible influence of the tyres on the results they are included in specifications (05) and (06). None of the tyres apart from Continental has a significant influence. Continental itself has a negative influence on the dependent variable at the 5-%-level which means that it increases performance. Maybe the significance of the tyres can also be explained by the large dataset used and the comparatively small number of tyre producers. Further research might be necessary though it is also likely that this is just a statistical coincidence.

Table 10
Robustness tests of regressions

Variables		D_i *(A. Prost)*	D_j (K. Räikkönen)	$D_i - D_j = 0$	*Controls*	*N*	R^2
(01)	Preferred specification	-4.686_5^a (0.856)	-4.290_6^a (0.952)	*p-value* 0.540	Yes	14’404	0.701
(02)	ClassToPoints in (01)	3.353_5^a (0.526)	2.629_9^a (0.590)	*p-value* 0.078	Yes	14’404	0.594
(03)	Includes grid position in (01)	-4.685_5^a (0.857)	-4.313_6^a (0.952)	*p-value* 0.566	Modified	14’404	0.701
(04)	Includes grid position in (02)	3.352_5^a (0.526)	2.640_9^a (0.590)	*p-value* 0.083	Modified	14’404	0.491
(05)	Includes tyres in (01)	-4.752_4^a (0.864)	-4.415_7^a (0.973)	*p-value* 0.611	Modified	14’404	0.703
(06)	ClassToPoints and Tyres in (01)	3.394_5^a (0.530)	2.647_9^a (0.604)	*p-value* 0.076	Modified	14’404	0.597
(07)	Whole sample	-4.703_5^a (0.856)	-4.294_7^a (0.951)	*p-value* 0.528	Yes	15’305	0.701
(08)	ClassToPoints in (07)	3.364_4^a (0.522)	2.626_9^a (0.586)	*p-value* 0.070	Yes	15’305	0.593
(09)	Includes tyres in (07)	-4.777_5^a (0.863)	-4.438_7^a (0.971)	*p-value* 0.608	Modified	15’305	0.702
(10)	Includes tyres and lubrication in (07)	-4.331_6^a (0.856)	-3.506_{11}^a (0.971)	*p-value* 0.608	Modified	15’305	0.709
(11)	Speed difference when time available	9.201_8 (7.197)	6.232_{27} (7.576)	*p-value* 0.426	Yes	7’945	0.275
(12)	Classification when time available	-5.067_{10}^a (1.068)	-3.972_{21}^a (1.158)	*p-value* 0.117	Modified	7’945	0.543
(13)	ClassToPoints when time available	4.593_9^a (0.886)	3.461_{15}^a (0.959)	*p-value* 0.049	Modified	7’945	0.567
(14)	First50 sample	6.055_6^a (0.421)	5.939_5^a (0.626)	*p-value* 0.875	Yes	6’610	0.718
(15)	ClassToPoints in (14)	4.825_5^a (0.886)	4.624_7^a (0.959)	*p-value* 0.700	Yes	6’610	0.605
(16)	Includes tyres in (14)	3.289_6^c (1.769)	3.093_5^c (1.840)	*p-value* 0.795	Modified	6’610	0.721
(17)	Start position	-7.412_{11}^a (0.818)	-5.029_{31}^a (0.964)	*p-value* 0.001	Modified	14’404	0.547

Source: Author’s calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; Standard errors are reported in parenthesis; [a] Significant at 1-percent level; ; [b] Significant at 5-percent level; [c] Significant at 10-percent level; All numbers are reported with four digits; e indicates the exponential 10;. All driver dummies are included and A. Prost and K. Räikkönen are only examples of two such dummies; Subscripts behind these drivers indicate the ranking when all coefficients are analyzed; The p-value is used for the test of the linear hypothesis whether the coefficients of the A. Prost and the K. Räikkönen dummy are equal; Results for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

When considering specification (16), for example, with a limited panel we find that Continental does not have any influence on the results. In regression (16) only Michelin has a positive significant coefficient at the 10-%-level which means that it has a negative influence on performance.

In specifications (07) to (10) the whole database of 15'305 observations is used. This means that we also include drivers that were participating in less than 15 Grand Prix. Generally speaking, the coefficients of the TOP-10 drivers and the rankings do not change when extending the database. For a number of other drivers the coefficients stay almost the same but the rankings change. Most nearest neighbors are not significantly different from each other. Again, when using CLASSTOPOINTS as a dependent variable the difference between A. Prost and K. Räikkönen's coefficient is significant at the 10-%-level. When tyres and lubrication is included in regression (10) we do not find any significant influence of the tyres any more. Lubrication does not have any influence either apart of the dummy for Mobile which is significant and negative and leading therefore to better classifications. Still it is likely that this is only a statistical artifact and not of any importance. The inclusion of tyres and lubrication has a non-negligible influence on the coefficients and a minor influence on the ranking.

In regression (11) merely the sign of the coefficients should be interpreted. In this case only observations are included if a time is available which means that the driver did finish the race. Equation (8) is used to predict the "average" speed in a year for all drivers. The difference is then calculated between the real speed in the sample and the prediction of equation (8). This difference is used as the dependent variable in regression (11). The dummy for technical outs is dropped because no outs are recorded as only observations with available times are used.[42] Before commenting on the results it should be highlighted that this specification poses a number of problems: Firstly, there are estimation uncertainties when predicting speed with the use of equation (8). Secondly, when regressions on the difference in speeds are performed the confidence intervals of the predictions are not taken into account. Thirdly, the observations used do not include technical outs. This means that a driver who usually drops out but is fast when he finishes obtains too strong results. Fourthly, speed differences between drivers in the early years of Formula 1 were bigger than nowadays and no controls are added for this fact. In the light of all these problems, regression (11) should only be taken as an illustration. Still, further research might be worthwhile on such a specification or a similar one if speed or time differences are of interest. Concerning the results, the R^2 is far lower than for all the other specifications. Both reported driver dummies are positive but not significant. The derived

[42] This admittedly strange specification was only included in the robustness tests because it was suggested as worth a try by the participants of the "Coffee, Biscuits and Economics Seminar" organized by the assistants in economics of the University of Fribourg when a preliminary version of this thesis was presented there on the 18th of May 2006.

rankings change.

In regressions (12) and (13) classification and the constructed points measure are again used as dependent variables but the observations used are the same as in specification (11). This means the 7'945 lines of the design matrix of these regressions include only observations where the drivers passed the finishing line. Not controlling for technical outs and restricting the database changes the size of the coefficients and the rankings. Additionally the R^2 falls from approximately 0.700 to 0.550. The ranking derived from these specifications shall only be used when averaging the different rankings in the next section.

In specifications (14) to (16) the number of observations is restricted again. Here only the first 50 drivers of Table 3 are considered. All of these drivers were running more than 15 races which makes it necessary to either drop a dummy or the constant. In the presented specifications the constant was dropped. Note again that smaller values represent stronger drivers. It is clear that the driver dummies change sign in regression (14) because the constant itself was positive in all other specifications using classification as a dependent variable. Interestingly, the ranking itself remains fairly consistent although some drivers change positions with their nearest neighbors or with other drivers. As the p-values show close coefficients are not significantly different from each other. Small changes in the ranking itself come as no surprise.

At the end, the grid position is used as a dependent variable in regression (17). The control variable TECHNICALOUT is not included in this specification. Consider this regression like the calculations of the "Likelihood of better grid position" in Table 6. We shall not explicitly present a ranking table using this endogenous variable here but the reader may again obtain the results of the estimation from the author's site.[43] Instead, the ranking for the start positions will be correlated with the ranking resulting from the likelihood specifications of subsection 3.3.2.

3.4.5 Rankings on preferred specification and correlation with other rankings

In this section a ranking of the different drivers will be given. The ranking will be based on our preferred specification which is regression (6) in Table 9. For all coefficients, standard deviations shall be reported as well as 95-%-confidence-intervals.

---

[43] The results show that J. Clark is the first as far as starting positions are concerned followed by J. M. Fangio, J. Hunt, S. Moss, J. Rindt, J. Stewart, A. Ascari, A. Senna, N. Farina and M. Schumacher.

Table 11
Rankings based on econometric model

Driver	*Coefficient of driver dummy*	*Standard deviation*	*p-value for comparison with neighbor*	*Confidence intervals [2.5 %;97.5%]*	*Mean of coefficients (other specifications)*	*Mean of rankings*
M. Schumacher *active driver*	**-4.997_{2}^{a}**	0.855	0.182	[-6.673;-3.321]	**-4.983_{3}**	**3.2_{2}**
A. Prost	**-4.686_{5}^{a}**	0.857	0.856	[-6.366;-3.007]	**-4.714_{5}**	**5.5_{5}**
A. Senna	**-4.283_{7}^{a}**	0.881	0.992	[-6.010;-2.557]	**-4.359_{7}**	**7.3_{7}**
N. Mansell	-2.665_{36}^{a}	0.862	0.970	[-4.353;-0.976]	-2.884_{35}	29.7_{30}
J. Stewart	**-4.157_{9}^{a}**	0.915	0.924	[-5.950;-2.363]	**-4.208_{9}**	**8.4_{9}**
J. Clark	**-4.905_{3}^{a}**	0.959	0.889	[-6.784;-3.026]	**-4.926_{4}**	**4.6_{4}**
N. Lauda	-2.751_{33}^{a}	0.862	0.979	[-4.440;-1.062]	-3.059_{27}	26.1_{26}
J. M. Fangio	**-5.947_{1}^{a}**	1.001	NA	[-7.909;-3.986]	**-5.871_{1}**	**1.4_{1}**
N. Piquet	-3.443_{17}^{a}	0.860	0.981	[-5.128;-1.757]	-3.573_{14}	15.7_{13}
D. Hill	-3.231_{18}^{a}	0.906	0.714	[-5.006;-1.455]	-3.257_{20}	21.4_{21}
M. Häkkinen	-3.507_{15}^{a}	0.883	0.950	[-5.237;-1.777]	-3.534_{16}	16.6_{15}
S. Moss	-3.206_{19}^{a}	0.945	0.972	[-5.057;-1.355]	-3.387_{18}	18.9_{17}
J. Brabham	-2.870_{27}^{a}	0.875	0.945	[-4.586;-1.154]	-3.111_{24}	24.4_{24}
E. Fittipaldi	-2.817_{28}^{a}	0.888	0.927	[-4.558;-1.076]	-3.004_{28}	26.0_{25}
G. Hill	-2.735_{34}^{a}	0.856	0.974	[-4.412;-1.058]	-2.899_{34}	30.3_{32}
A. Ascari	**-4.244_{8}^{a}**	1.102	0.964	[-6.404;-2.084]	**-4.644_{6}**	**5.7_{6}**
D. Coulthard *active driver*	-3.535_{14}^{a}	0.873	0.951	[-5.246;-1.824]	-3.467_{17}	19.2_{18}
C. Reutemann	-2.807_{29}^{a}	0.863	0.986	[-4.497;-1.116]	-2.971_{30}	29.3_{29}
A. Jones	-2.137_{44}^{b}	0.899	0.979	[-3.899;-0.375]	-2.291_{44}	42.4_{43}
M. Andretti	-1.975_{46}^{b}	0.890	0.967	[-3.718;-0.231]	-2.228_{45}	43.5_{45}
J. Villeneuve *active driver*	-2.469_{39}^{a}	0.888	0.969	[-4.210;-0.728]	-2.579_{40}	38.5_{39}
J. Scheckter	-3.113_{21}^{a}	0.924	0.988	[-4.925;-1.302]	-3.215_{23}	23.1_{22}
G. Berger	-2.971_{24}^{a}	0.858	0.988	[-4.653;-1.288]	-3.07_{26}	27.4_{28}
J. Hunt	-3.022_{22}^{a}	0.930	0.896	[-4.844;-1.200]	-3.32_{19}	19.9_{19}
R. Peterson	-2.393_{40}^{a}	0.872	0.898	[-4.103;-0.683]	-2.588_{39}	38.8_{41}
K. Räikkönen *active driver*	**-4.290_{6}^{a}**	0.952	0.540	[-6.156;-2.423]	**-4.218_{8}**	**7.9_{8}**
R. Barrichello *active driver*	-2.907_{26}^{a}	0.863	0.926	[-4.598;-1.215]	-3.079_{25}	27.3_{27}
F. Alonso *active driver*	-3.617_{11}^{a}	0.990	0.822	[-5.557;-1.678]	-3.642_{12}	**11.9_{10}**
D. Hulme	-3.572_{12}^{a}	0.900	0.952	[-5.336;-1.808]	-3.591_{13}	17.6_{16}
J. Ickx	-2.223_{42}^{b}	0.888	0.799	[-3.963;-0.483]	-2.321_{43}	42.3_{42}
J. P. Montoya *active driver*	-3.459_{16}^{a}	0.966	0.945	[-5.353;-1.565]	-3.561_{15}	16.4_{14}
R. Arnoux	-1.415_{50}	0.880	0.557	[-3.139;0.310]	-1.504_{50}	49.3_{50}
R. Schumacher *active driver*	-2.779_{30}^{a}	0.896	0.963	[-4.536;-1.022]	-2.938_{33}	29.8_{31}
T. Brooks	**-3.828_{10}^{a}**	1.057	0.702	[-5.900;-1.756]	**-3.814_{10}**	14.6_{12}
J. Surtees	-2.766_{32}^{a}	0.895	0.988	[-4.520;-1.012]	-2.967_{31}	30.8_{33}
J. Rindt	-2.980_{23}^{a}	0.959	0.956	[-4.859;-1.101]	-3.245_{22}	20.6_{20}
J. Laffite	-2.491_{38}^{a}	0.872	0.980	[-4.199;-0.783]	-2.643_{38}	37.0_{38}
G. Villeneuve	-2.508_{37}^{b}	0.977	0.821	[-4.422;-0.594]	-2.731_{37}	35.5_{37}
R. Patrese	-2.774_{31}^{a}	0.851	0.993	[-4.442;-1.106]	-2.967_{32}	31.0_{34}
N. Farina	**-4.836_{4}^{a}**	1.085	0.941	[-6.962;-2.710]	**-5.067_{2}**	**3.7_{3}**
C. Regazzoni	-1.747_{49}^{b}	0.879	0.967	[-3.471;-0.023]	-1.971_{47}	47.6_{49}
K. Rosberg	-2.379_{41}^{a}	0.897	0.982	[-4.138;-0.620]	-2.470_{41}	38.5_{40}
J. Watson	-2.153_{43}^{b}	0.865	0.901	[-3.848;-0.457]	-2.352_{42}	43.3_{44}
M. Alboreto	-1.770_{48}^{b}	0.867	0.818	[-3.469;-0.071]	-1.902_{49}	47.1_{47}
B. McLaren	-3.572_{13}^{a}	0.902	0.999	[-5.340;-1.803]	-3.701_{11}	13.9_{11}
D. Gurney	-3.124_{20}^{a}	0.939	0.911	[-4.964;-1.284]	-3.25_{21}	23.5_{23}
E. Irvine	-2.686_{35}^{a}	0.891	0.930	[-4.433;-0.939]	-2.869_{36}	34.4_{36}
H.H. Frentzen	-2.957_{25}^{a}	0.880	0.979	[-4.681;-1.232]	-2.978_{29}	31.1_{35}
T. Boutsen	-1.998_{45}^{b}	0.876	0.814	[-3.715;-0.281]	-2.108_{46}	45.5_{46}
J. Herbert	-1.891_{47}^{b}	0.872	0.884	[-3.600;-0.181]	-1.949_{48}	47.3_{48}

Source: Author's calculations.

Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; Subscripts indicate inner column rankings; Bold letters designate the first ten ranks in a column; [a] Significant at 1-percent level; ; [b] Significant at 5-percent level; [c] Significant at 10-percent level; All reported number (apart from "Mean of ranking") are rounded to three decimal places; The "Coefficients of driver dummy" stem from regression (6), Table 9 (preferred specification); The p-value is used for the test of the linear hypothesis whether coefficient i is different from coefficient j, where i and j are nearest neighbours; "Mean of coefficients" and "Mean of rankings" represent a mean of coefficients (rankings) form different regressions; Results for all drivers and additional information are available on http://homeweb4.unifr.ch/stadelmd/pub/public.data/f1.rar.

Furthermore, we always calculate a p-value for the hypothesis that the coefficients of the driver dummies are the same. To put it more clearly: We test the hypothesis if $D_i - D_j = 0$ where driver *j* is driver *i*'s nearest neighbor and *i*'s coefficient is smaller than *j*'s coefficient (driver *i* is stronger than driver *j*). Moreover, a mean of the coefficients of some robustness tests in Table 10 is calculated as well as a mean of different rankings resulting from Table 9 and Table 10. The results of the subsequent ranking are presented in Table 11.

The general picture that emerges when considering Table 11 is a similar one as in Table 6 when the column better "Classifications over matches" or "Likelihood of better classification" is concerned. Most drivers ranging under the first ten in Table 6 are also under the TOP-10 in Table 11. Their rankings tend to change either up or down. For example, J. Stewart joins the ranks of the best and N. Piquet as well as E. Fittipaldi decrease in their rankings. F. Alonso who was ranked very well under the first ten in Table 6 is ranked eleventh in this evaluation but K. Räikkönen improves from rank eleven to rank six. Regarding the other positions, ups and downs can be registered as well but the overall picture remains the same. It is worthwhile noting that J. M. Fangio is again the leader of the ranking. M. Schumacher is ranked second and J. Clark third. All coefficients, apart from one exception, represented in Table 11 are significant. This can partly be explained by the large number of observations. More comparisons between different specifications and different ranking will be done further below.

It should be noted that we always use the same 50 drivers for this representation although coefficients for all drivers have been calculated.[44] It might therefore be possible that the coefficients of some drivers not figuring under these 50 could be lower (indicating better results) than for the ones presented here. This is the case for 15 drivers that are not reported in Table 11. Those 15 drivers have mostly the following characteristics: They were usually participating in only slightly more than 15 Grand Prix, they did not win more than two Grand Prix and they achieved for their short presence in Formula 1 comparatively good classifications. Let us mention two examples of such drivers: P. Taruffi has a very low coefficient of -3.049 (1.243) but he started only 18 times, finishing once first but achieved five podium positions. F. Cevert has a coefficient of -2.973 (1.093). With 46 races he started relatively often, finishing once first and achieved 13 podium places. Without a doubt it could be said that such drivers would be of interest too because they show comparatively good rankings here. However, they are not reported since, firstly this econometric evaluation is just one of many that are done in this thesis and the mentioned

[44] This is also the case for the likelihoods of Table 6.

drivers do not always show such strong results. Secondly, we believe that a certain number of wins is necessary in order to be declared as belonging to the best and three wins, as all racers in Table 11 have, is already a very low criteria. Thirdly, the good results stem partly from the rather low number of races these drivers participated in. Fourthly, we must restrict in some way the number of drivers to be presented in this thesis because reporting numbers for all drivers would simply overstretch it.

Unfortunately looking only at the coefficients and their standard deviations alone is not sufficient for establishing a ranking. Although the coefficients are significantly different from zero they do not necessary have to be significantly different from each other. This is a serious problem when applying a ranking because statistically it would be possible that the coefficients are the same and therefore the drivers are ranked the same. We test for the linear hypothesis that a driver's coefficient is significantly different from his next neighbor's coefficient. None of the p-values comparing the nearest neighbor's is significant. For J. M. Fangio the p-value does not contain a value because he is the leader of this ranking, as he was in most other rankings before. It could now be tested if the calculated values of a driver are significantly different from his neighbors' neighbors and that so forth. Instead two other procedures shall be applied.

The p-values when comparing the first with the fifth is 0.078, for the fifth with the tenth it is 0.280, for the tenth with the 20th it is 0.418, for the 20th with the 30th it is 0.611, for the 30th with the 40th it is 0.522 and for the 40th with the 50th it is 0.090.[45] The rankings are particularly stable at the beginning and the end.

Furthermore, 95-%-confidence-intervals are reported. Note that overlapping confidence intervals do not necessarily indicate that the drivers are not significantly different from each other. The linear hypothesis test is constructed in a different manner as the confidence intervals (see WOOLDRIDGE, 2002A or any other econometric textbook). Anyhow, the confidence intervals give some idea of whether the coefficients are very close to each other or rather different.

The picture that emerges highlights again the enormous difficulties encountered when establishing a ranking, as also mentioned by KLABRODA (2000) and BERRY ET AL. (1999). Strictly statistically speaking the calculated strengths and talent measures of the drivers could be the same for most of them or at least within several groups (such as the given intervals above from one to five, five to ten and so forth).

---

[45] Only some clear conclusions for the ranking can be taken considering the p-values. Generally drivers within one of the mentioned groups could be ranked all the same.

In order to allow for a check of the stability of the given ranking two further measures were calculated. First of all, we took the coefficient of each driver of regressions (6) and (7) out of Table 9 and those of regressions (03), (05), (07), (09), (10), (12) from Table 10 and calculated their arithmetic mean. The results are reported in the last but one column. Moreover, the rankings for each driver from all regressions in Table 9 and all regressions in Table 10, apart from regression (11) and (17), were used to calculate a mean ranking. The results can be found in the last column of Table 11. In comparing the last two columns with the ranking based on the coefficients of the preferred specifications it can be concluded that the resulting rankings are almost identical, though some minor changes are possible. This can also be seen considering Table 12 where some rankings (not coefficients) from different regressions are correlated with each other. The correlations are usually very high.

Table 12
Spearman's rank correlation matrix for different specifications

	Reg-6	*Reg-7*	*Rob-03*	*Rob-07*	*Rob-09*	*Rob-10*	*Rob-12*	*Rob-14*
Reg-6	1.000							
Reg-7	1.000	1.000						
Rob-03	0.992	0.995	1.000					
Rob-07	0.979	0.983	0.995	1.000				
Rob-09	0.955	0.960	0.974	0.979	1.000			
Rob-10	0.921	0.924	0.943	0.947	0.912	1.000		
Rob-12	0.687	0.689	0.714	0.729	0.682	0.785	1.000	
Rob-14	0.914	0.920	0.943	0.955	0.951	0.918	0.766	1.000

Source: Author's calculations.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; All reported number are rounded to three decimal places or given as integers; "Reg-#" stands for the driver rankings of regression number # of Table 9 and "Rob-##" stands for the driver rankings of regression number ## of Table 10; The rankings derived from these regressions for the first 50 drivers have been correlated.

3.4.6 Limits and problems of the approach

By using the econometric model described in section 3.4 it is theoretically possible to estimate the effect of a driver independently of the car. Practical implementation shows that there remain a number of problems and limits of the approach.

From the beginning it has to been pointed out that although a high number of total observations is available the panel to estimate the drivers' effects is highly unbalanced. Usually panels over long time periods include the same units in each year. Our dataset does not contain the same drivers for every year because they simply changed over time. In a sense the high number of observations is therefore not relevant and disguises that for most drivers only 40 to 100 observations are available. As mentioned by CAMERON AND TRIVEDI (2005) it is usually assumed, when applying asymptotic theory, that the number of units tends towards infinity. Here this is not the case.

Applying several test statistics showed that the estimation procedure was highly heteroscedastic and that a pure "dummy variable regression" as proposed by WOOLDRIDGE (2002B) when trying to find fixed effects is impossible. These two facts lead us to using GLS instead of OLS and to grouping for cars. The cars themselves entered the regression as random effects. We cannot be sure if this form is completely correct. Generally the random effects estimator is more efficient than pooled OLS but if the correct model would be a fixed effects model, the random effects estimator is not consistent (see CAMERON AND TRIVEDI, 2005).

A general assumption that is implicit in all specifications is that the performance of a driver is independent of a specific car. This assumption derives directly from the use of the random effects specification. To lessen the hypothesis: cars and drivers are not correlated with each other. A good car is a good car no matter who drives it and a good driver is a good driver no matter what car he uses. The driver does not have any influence on the car. It is clear that this could be a rather strong hypothesis and does not necessarily have to be correct. Indeed, general public opinion often states that it is the driver who influences the team and tries assisting in improving the car. It is often heard, for example, that M. Schumacher had a positive influence on Ferrari. There is some evidence in the data that, for example, the so called "Schumacher effect" on the car is not generally correct. In addition, if all drivers try to improve their cars there should only be a negligible problem when estimation equation (7) with car random effects. Enhancements introduced by a driver will be captured with the dummy. In any case we neither know the driver effect nor the car effect. Estimating makes it necessary to focus on one of the two.

Sorting is another problem. It is possible that good drivers end up in good teams and use good cars. At the moment we do not see any way to capture such effects and control for them. Interacting driver and car variables is impossible because both effects are unobservable and need to be estimated: Interacting a dummy with another gives only one more dummy and increases even the possibility of singularities in the design matrix. In this instance it is worthwhile mentioning that three effects are unknown: the driver effects, the car effects and the technological changes. By estimating classifications instead of time or speed differences controlling for technological changes is not necessary but the other two unknown effects remain problematic.

Moreover, our main interest is not finding a single coefficient but a coefficient for every driver. Likewise, it is not only important when establishing a ranking that the coefficients be significantly different from zero but also that they be significantly different from each other. Most of the coefficients represented in the tables are different from zero but not all are different from each other. From a pure statistical perspective a clear ranking of the 50 mentioned drivers can therefore not be made.

In addition, the coefficients as well as the rankings are changing when the specification of the regression formula is changed. Adding other control variables, increasing or reducing the number of observations as well as other robustness tests have an influence on the dummies' coefficients. Typically the coefficients remain significant and their sign does not change but their values do. Usually economists tend to focus only on the sign and the significance of one or two coefficients and do not interpret their sizes nor if they are different from each other. To solve this problem an average over different specifications and different rankings was calculated. Still this solution is not satisfying as further changes to the specifications would result in other rankings. Unfortunately, as this is an empirical question no specification can entirely be proved as right. The only way is to justify the equations used and show that they are likely to be superior to other specifications.

Every ranking is difficult to perform. Against the given ranking in Table 11 it could be mentioned that the dependent variable "Classification" as well as the variables used for robustness tests are not appropriate. Instead, a probability for wins or for podium positions should be estimated. There is no counter-argument. The only answer that can be given here is that firstly, winning probabilities are less interesting for drivers and teams because sponsorship contracts in Formula 1 usually depend on the number of points. Secondly, when performing the comparison approaches on the same car in section 3.3 we have already calculated a "Likelihood of better podium" for each driver. Ultimately, a sheer huge number of possible dependent variables could be used in order to perform a ranking of the drivers. It could be even argued that drivers should be ranked according to the number of drop-outs or technical problems or according to the number of physical outs. Nonetheless, we have also calculated a logit specification for wins with our data. The results resulting rankings are almost the same as in Table 3, column "Wins over races". This is not unexpected as a huge number of drivers simple did not win any Grand Prix at all and the TOP-10 of Table 3 won a comparatively high number of them.

Another possible objection to the econometric model used, the specifications tested and the coefficients and rankings derived could be as follows: We do not consider drivers who never won but would have won if they had been given a better car. In other words: A possibly good driver could always be racing with a very bad car. He therefore has no chance of winning and is excluded by our design in general since the list presented in Table 6 or Table 11 is the same as in Table 3 with different indicators. We are of the opinion that this does not pose a problem for two main reasons. Firstly, the stated problem is a purely hypothetical one. Secondly, Formula 1 is a huge business and one of the main interests of any team is to have the best drivers. If a potentially good driver in a season faces a bad car but is clearly stronger than his team partner there is a higher probability that he will be offered a contract by a better team. This change in teams would increase the good driver's classification which is captured in our model.

3.5 Comparing the results of the different methods

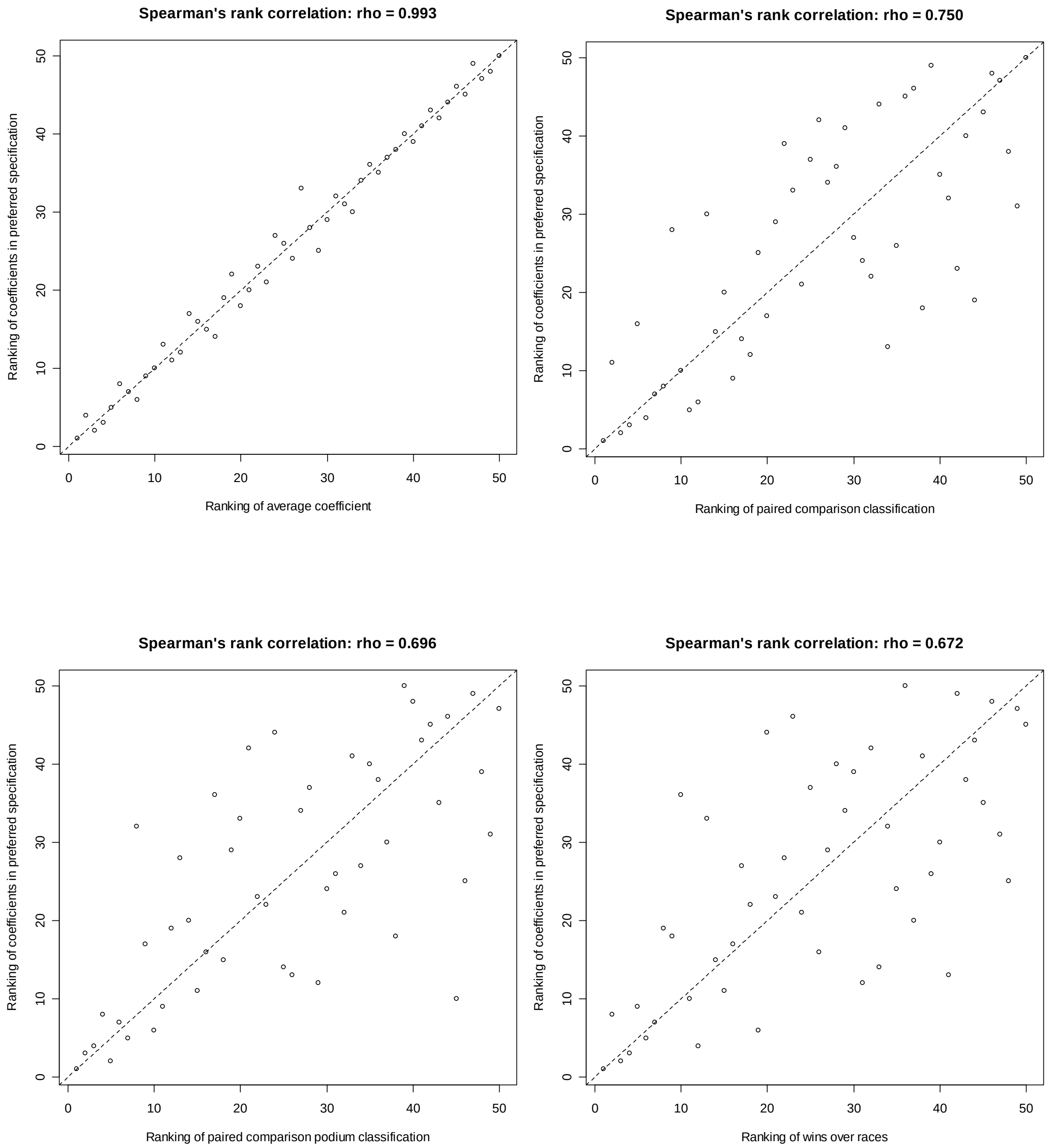
No formal attempt has been made to date to separate the effects of the car and the drivers in Formula 1 nor to evaluate different drivers from different eras. Possible ways of evaluating talent are not only lacking in sports but also in many other walks of daily life as mentioned in the literature survey.

There is no simple and clear-cut way of establishing a ranking for Formula 1 drivers. In fact, many different rankings can be established, depending on the method and the aggregation procedures applied. In this chapter an overview of descriptive statistics and some relative measures was given. The approach was then extended and different forms of comparisons on the same cars were made including the calculation of a likelihood of winning any given comparison. At the end a formal econometric model including driver dummies was set up and estimated.

In this section a closer examination of the different results will be made and the rankings shall be compared. Due to the number of different performance variables and also the number of different methods applied, only a subset of them shall be explicitly commented on and described. However, the results are similar for all evaluations done in this thesis.

Figure 7 represents some correlation plots for the different ranking procedures dealing with classification and some of its derivates such as "Likelihood of better podium" and "Wins over races". We shall mainly use the ranking resulting form the preferred specification of section 3.4 which can be found in regression (6), Table 9. Since rankings are compared the Spearman correlation measure has to be used.

Figure 7
Ranking correlations for different evaluation methods for classification



Source: Author's calculations and representation.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; "Ranking of coefficients in preferred specification" with "Ranking of average coefficient" represent the rankings from Table 11; "Ranking of paired comparison classification" and "Ranking of paired comparison podium classification" can be found in Table 6; Table 3 gives the "Ranking of wins over races"; The rankings derived from these tables for the first 50 drivers have been correlated.

As is evident from the upper left quadrant of Figure 7, the "Preferred specification ranking" of section 3.4 is highly correlated with the "Ranking of the average coefficients" from the same section. This should not come as a surprise as Table 12 gives already an indication towards this direction.

More interesting are the correlations with the rankings derived from the paired comparison model. The upper right quadrant represents the rank correlation for the "Preferred specification ranking" and the "Likelihood for better classification". The Spearman's rank correlation equals 0.750. This is a rather high value for ranking correlations especially as it could be possible that the different methods measure different aspects of a driver's talent. The first ten drivers seem to have results which are an almost identical for both approaches.

The lower left quadrant of Figure 7 gives a slightly different picture. Here the "Likelihood of better podium position" and the "Preferred specification ranking" are considered. It has to be admitted that at least the best drivers correspond quite well to the other ranking methods applied. It is even more astonishing if we consider that the p-values for the ranking based on the econometric model of section 3.4 indicate that the coefficients of the first few drivers are not significantly different form each other.

The last quadrant shows the ranking correlation with the simple performance measure "Wins over races" taken from Table 3. The estimated coefficient is the lowest but if the absolute number if wins was taken the correlation would be even smaller.

An overview of all applied methods indicates that they differ in some aspects concerning the final ranking but this could be anticipated. Despite this fact, the overall picture seems to correspond mostly to the intuition. This does not necessarily mean that the rankings are correct but it gives some hints that they are not completely far fetched. It can be shown that a "correct" ranking cannot exist since it always depends on the variables chosen and the aggregation procedures applied. The evaluations made here use comparatively few hypotheses to establish rankings.[46]

We shall also consider ranking correlations for the grid positions. As no explicit results for the starting position were reported in section 3.4 this will just be done briefly.[47] Figure 8 represents two ranking correlations using the results that can be derived from the regression (17) of Table 10.[48] It can be observed that the ranking correlations from the econometric model and the paired comparison approach tend to diverge more than was the case in Figure 7. A number of drivers still achieve the same rankings in both methods.

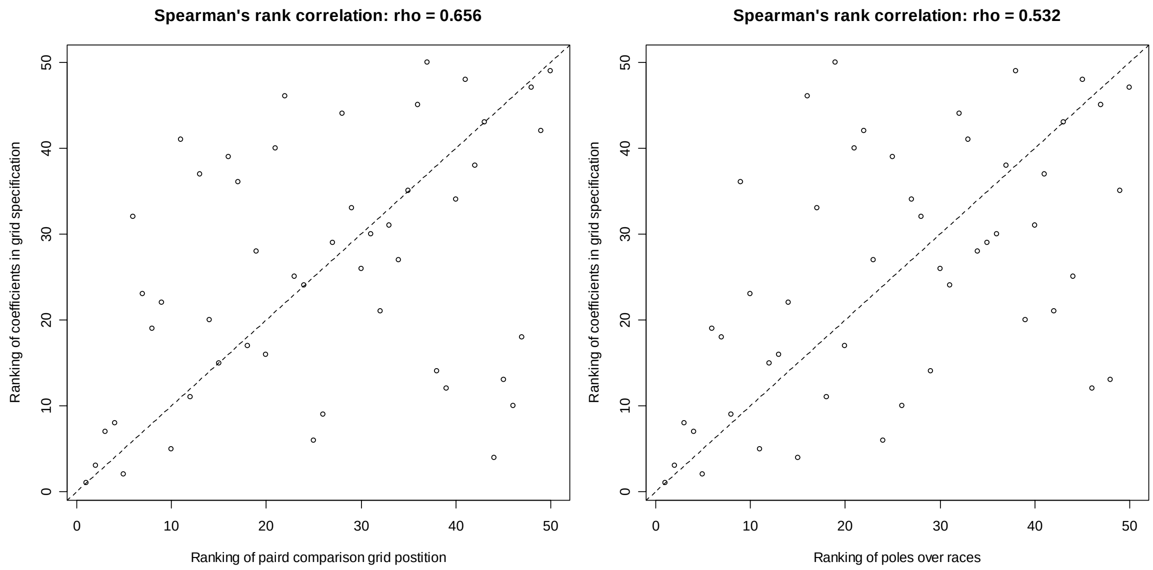
---

[46] In the literature survey Autosport-Atlas Wildsoft F1 Ranking was mentioned which depends on an enormous amount of different hypotheses.

[47] The results of the econometric model for starting positions can be obtained from the author's website.

[48] All insufficiencies and problems discussed in subsection 3.4.5 also apply to this specification.

Figure 8
Ranking correlations for different evaluation methods for grid position



Source: Author's calculations and representation
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; "Ranking of coefficients in grid specification" stem from regression (17), Table 10; "Ranking of paired comparison grid position" can be found in Table 6; Table 3 gives the "Ranking of poles over races"; The rankings derived from these tables for the first 50 drivers have been correlated.

Focusing on the second representation in Figure 8 we note that the rankings diverge even more, though this does not catch us unaware. The relative measures of Table 3 do not seem to be well correlated with the more sophisticated rankings resulting from the other methods.

With these comparisons, the end of the evaluation of the drivers' abilities and talent is reached. The next chapter will present some applications that are possible using the econometric model of section 3.4. In addition, the calculated driver coefficients will be used for a first identification of superstar effects in Formula 1 and we will also try to evaluate risk behavior under rainy weather conditions.

4 Applications

Extensive time and work has been invested in the database used for the evaluations done in chapter 3. However, only a minor fraction of possible uses and application of the database itself have been looked at. We consider this thesis mostly as a case study for the evaluation of talent. In this chapter some additional applications for the database, the calculated rankings and the econometric model derived in section 3.4 will be studied. As mentioned by KAHN (2000) sports can be used to gain some insights into specific economic problems. Only a small number of questions concerning Formula 1 and economics shall be answered since all possible applications would certainly overload this thesis. Starting with an evaluation of the weather's influence on ranking and risk behavior, a calculation of an average life cycle of a driver will follow. Moreover, the some preliminary superstar effects will be measured. The chapter concludes with a brief discussion of other possible future applications and a short "fun" application. Some applications in this chapter are performed in a rather cursory manner but without sacrificing scientific exactness and correctness. Therefore this chapter should serve purely as an illustration of the hidden possibilities of using the established database, the rankings and the applied methods and as an inspiration for more research.

4.1 The weather's influence

Although there is already an extensive theoretical, empirical and experimental literature on risk behavior and risk aversion it might be interesting to take a look at sports.[49] Formula 1 is especially interesting for this question as the risks of fatal injury in this sport are high. It will be shown that while there are more drop-outs registered when the weather is bad, the increase in drop-outs is not significant. Further distinguishing between human and technical drop-out can partly explain this fact. Higher risk due to bad weather is only partly compensated for by more careful driving.

The risks that drivers take seem to be independent of the weather. When the weather is good more drop-outs caused by technical problems are registered whereas when the weather is bad human drop-outs increase. Altogether there seems to be no influence of the

[49] For a classical article see KAHNEMAN AND TVERSKY, (1979); For experimental literature see HOLT AND LAURY (2002).

weather on the number of drop-outs and drivers always face the same risk independent of the weather. We shall also use the instance when working with weather conditions to identify the best driver in rainy weather.

4.1.1 Compensating the risk of a drop-out

Let us first consider equation (8) again which was used to estimate the year dummies for technological advances represented in Figure 6. This time we are not interested in the year dummies but in the coefficient of the weather variable.

$$\frac{GPDistance}{ClassTime * 3600} = \underset{(0.239)}{3.958^{a}} \ (WeatherCode) + \beta\,(YearDummies) \qquad R^2 = 0.461 \tag{8}$$

As expected, this variable is highly significant and positive indicating that better weather conditions lead to an increase in speed. However, this alone does not prove that drivers are adapting their risk-taking by driving slowly when the weather is bad. Another dependent variable has to be considered.

The question that shall be addressed now is whether the weather condition also has an influence on the number of drop-outs. The variable for drop-outs is a binary measure with values 0 and 1. Therefore it cannot be estimated with OLS but by using a Logit model. The following is a simple specification that only includes WEATHERCODE as an explanatory variable as well as year dummies:[50]

$$Dropout = \underset{(0.014)}{0.020} \ (WeatherCode) + \beta\,(YearDummies) \qquad pseudo\ R^2 = 0.017 \tag{9}$$

In equation (9) the same number of observation is used as in our preferred specification of section 3.4. Note that regression (9) does not contain a constant because all year dummies are included.

The general coefficient of determination R^2 which is generally interpreted as a goodness-of-fit statistic cannot be used when modeling binary data. We therefore calculate a *pseudo* R^2 with the log-likelihood value of the model. Our *pseudo* R^2 is commonly known in the literature as McFadden's R^2 which compares the maximum value of the log-likelihood-function of the fitted model with the maximum log-likelihood-function of a model containing only a constant (for other coefficients of determination for logistic regressions see also MENARD, 2000).

[50] The number of dropouts was higher in the past because the cars were less reliable then. This makes it necessary to include dummies for each year.

It is interesting to note that the weather itself does not seem to have any significant influence on the probability of a drop-out. This result holds even when the specification is changed.[51] If anything, the coefficient is positive indicating that more drop-outs happen when the weather is good.

A closer look at the data reveals the story behind this result. When estimating the probability of a technical drop-out like in equation (10) we find a positive and significant coefficient:

$$\begin{gathered} TechnicalDropout = \underset{(0.015)}{0.061^{a}}\ (WeatherCode) + \beta\,(YearDummies) \\ pseudo\ \mathrm{R}^{2} = 0.077 \end{gathered} \tag{10}$$

On the other hand, when estimating the probability of a human drop-out in equation (11) we find a negative and highly significant coefficient:

$$\begin{gathered} HumanDropout = \underset{(0.0201)}{-\,0.075^{a}}\ (WeatherCode) + \beta\,(YearDummies) \\ pseudo\ \mathrm{R}^{2} = 0.179 \end{gathered} \tag{11}$$

It seems therefore that drivers do not change their risk-takings significantly due to weather conditions. The weather coefficient of equation (11) indicates that the probability of human outs due to accidents and collisions decreases significantly when the weather gets better. Equation (10) shows that probably cars are more demanded when the weather is good leading to a higher technical drop-out rate then.

Consequently, most Formula 1 drivers seem to accept a certain risk associated with their profession. They augment their speed when the weather turns good but this increase is associated with more technical problems. On the other side better weather conditions lead to significantly less human drop-outs like collisions or accidents. Taken together the risk of a drop-out is independent of the weather and drivers always seem to face the same overall risk of failure.[52]

4.1.2 Problems and limits of the approach

A number of shortcomings of the last section can easily be identified.

Firstly, the drop-out variable does not necessarily need to measure the risk taken by a

---

[51] For example when the year dummies are dropped.

[52] We used this instance when working on weather conditions to estimate the best driver in the rain. For this we estimated the specification of regression (6) of Table 9 with a restricted dataset containing only weather conditions with WEATHERCODE smaller or equal to 0. The procedure is exactly the same as the one applied in section 3.4. The number of observations is 2'992 and the R^2 equals 0.703. The resulting ranking for the first 10 is as follows: N. Farina, J. M. Fangio, A. Ascari, A. Senna, G. Villeneuve, K. Räikkönen, M. Schumacher, J. Ickx, S. Moss, J. Rindt.

driver. Dropping out because of an engine problem is not necessarily correlated with a risk of an incident that could harm the driver himself. Looking at death cases in Formula 1 would be a possibility for further research. Although, the death toll is at 23 drivers (excluding the Formula 1 Grands Prix Indianapolis 500) very high, only 15 were killed in Grand Prix, the rest were killed during qualification. A detailed analysis of deaths may therefore fail due to a limited number of observations. Nonetheless, it might be enlightening to look closer at accidents themselves and their influence on the driver's behavior.

Moreover, a number of other specifications for the regressions in this section could be imagined. In some tests we have included other control variables and dropped the time dummies. The overall results stay the same. As mentioned in the introduction of this chapter, only a brief overview of possible applications and their first interesting results for economics and Formula 1 management shall be made. This explains why no further details were given here.

Likewise, the drop-out variables employed are not exactly independent of each other. Information on the reason of drop-outs was compiled from FORIX Formula 1 Database and coded correspondingly into a human and a technical drop-out variable. Anyhow, it is possible that a technical problem with the brakes leads to a collision. Our coding method would result in a human out although it was a technical problem. Particularly when collisions are analyzed we do not have any information on who caused them. It could therefore be the case that some very risk-taking drivers are responsible for a number of collisions involving rather risk-adverse drivers.

It would be possible and interesting to include and evaluate the influence of other driver specific factors such as being married or having children on the risk attitude. Unfortunately our database does not yet contain such indicators for all drivers.

It could also be argued that drivers increase their risk taking at the end of a season or when they know that they are monitored because new contracts will be written. We shall try to look at such factors in the future.

Although this very short evaluation of the weather's influence was rather superficial, it managed to shed some light on the question of how people change their behavior when risks increase or decrease. For this evaluation it seems that Formula 1 drivers are prepared to live with a certain risk of drop-out. They adapt their speeds according to the weather but the overall risk of drop-out does not depend on the weather.

4.2 Preliminary estimation of a racer's average life cycle

In this section a brief overview of different statistics of a Formula 1 driver's average life cycle will be given. Such statistics might be interesting for teams and sponsors when deciding to whom new contracts will be given.

In Figure 9 four different histograms are presented dealing with various aspects of the drivers' life cycles from our database. The first histogram in the upper left quadrant shows that most Formula 1 drivers started their career between the age of 25 and 30. Considering career length almost 140 drivers of the 300 used for this representation ended their career within one or two year after the start. Only very few drivers can look back at careers in Formula 1 that were longer than 10 years. The third histogram in the lower left quadrant shows that most drivers reached their personal peak year one year before their career end. Only a small percentage of drivers have long peak periods. The construction of the peak variable allows for peak periods that extend over several years. [53] This is also shown in the last quadrant. For almost 70 % of the drivers the peak period was only one single year as most racers were only present in Formula 1 for one season. This pattern corresponds to the theoretical model for superstars of MACDONALD (1988). Similar conclusions are also reached by SEAMAN (2003) when he compares sports with the arts. It can be observed that most careers are short and decline quickly. Only very few become superstars.

4.2.1 Estimating the impact of age and experience

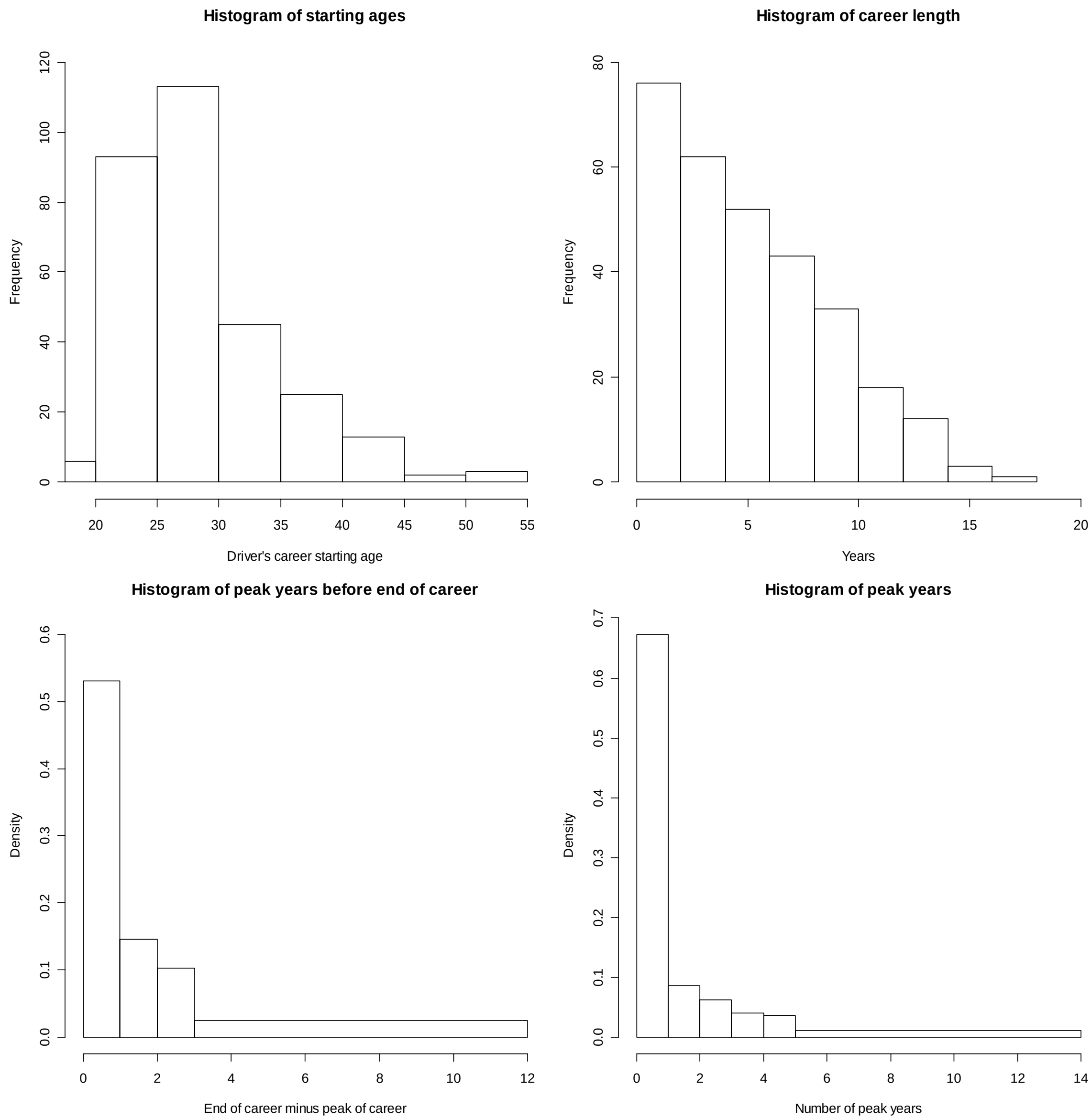
Here the impact of age and experience on the classification variable also used in section 3.4 shall be estimated. The focus will be on Formula 1 Grand Prix from 1950 to 2005 and therefore over the whole period. The estimation procedure used is exactly the same as for regression (2) and regression (6) of Table 9. The focus is on the quadratic term of these two regressions. The results for the regression including age as a control variable are as follows:

$$Classification = \underset{2.051}{25.14^{a}} - \underset{(0.014)}{0.902^{a}}\,(DriverAge) + \underset{(1.8e\text{-}3)}{0.117^{a}}\,(DriverAge)^2 + \beta X \qquad (12)$$
$$R^2 = 0.708$$

The results for the regression including experience as a control variable are:

[53] For more information on the construction of the peak variable see section 3.1.

Figure 9
Histograms of life cycle variables

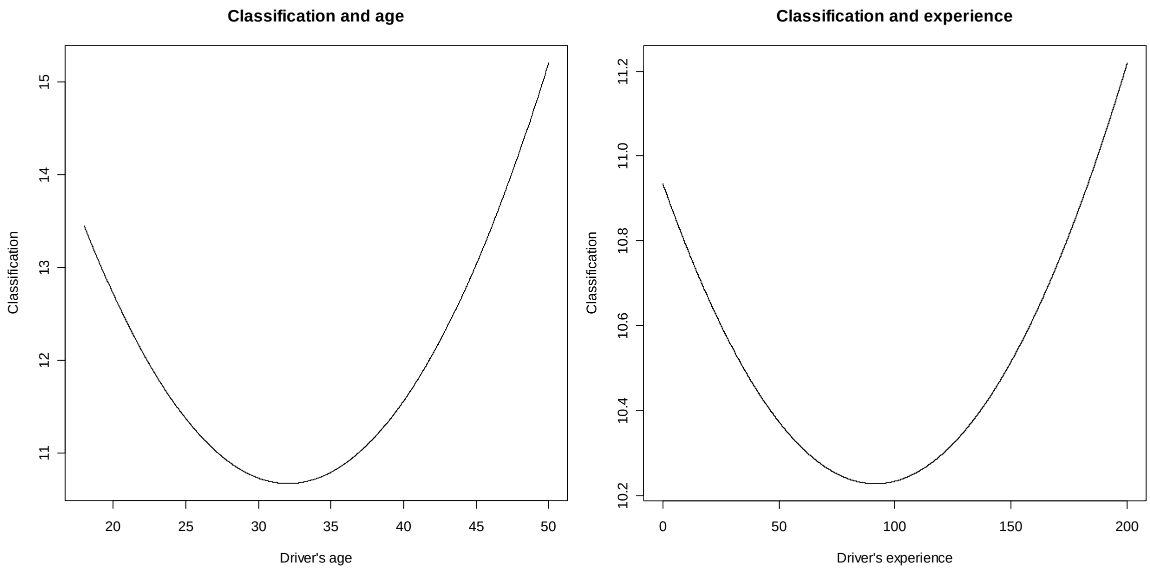


Source: Author's calculations and representation.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; Lower histograms use densities on the y-axis.

$$Classification = \underset{0.784}{10.93^{a}} - \underset{(2.9e\text{-}3)}{0.015^{a}}(Experience) + \underset{(1.7e\text{-}5)}{8.4e\text{-}5^{a}}(Experience)^{2} + \beta X \quad (13)$$
$$R^{2} = 0.701$$

The specifications were already described in section 3.4 and an explicit look at the quadratic terms is of importance here. Equation (12) and (13) show the expected signs for both coefficients. The intercept as well as the coefficients are highly significant. Figure 10 gives a representation of the quadratic forms for the two specifications.

Figure 10
Representation of quadratic influence of age and experience on classification



Source: Author's calculations and representation.
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; Quadratic forms stem from equations (12) and (13).

No confidence intervals are reported in this representation.

As can be seen when the whole period form 1950 to 2005 is analyzed, best classifications are on average attained at the age of around 32 and after an experience of approximately 95 races. The greater influence of age on classification than experience can partly be explained by the construction of the experience variable. It starts with a value of 0 and goes for some drivers up to values of over 200 whereas the age variable starts at around 20 follows a significantly less skewed distribution. Furthermore, as many drivers quit Formula 1 racing after a year or two the experience variable usually does not reach high values for such runners. It could therefore be necessary when further research is done on this question to look at a selected number of racers.

4.2.2 Problems and limits of the approach

Generally the same problems can be mentioned here as in subsection 3.4.6. However, there are also a number of other insufficiencies that stem mainly from the fact that our aim in this chapter is to provide some ideas of using Formula 1, the established database and methods for more research.

The use of the whole dataset over the 55 years could be criticized. Without a doubt, drivers in the early years of Formula 1 were generally older than today's racers. It could also be more revealing for Formula 1 managers to look at the last few years of racing. In a separate test we used data for the last 20 years. The overall pattern of the picture does not change which means that the coefficients have the same signs and remain significant. However, the minima of Figure 10 change to 29 years and approximately 90 races for the experience variable.

Furthermore, the figures of this section always represent an average driver. The life cycles of different Formula 1 racers could be very different from each other. Estimated confidence intervals are rather large for both representations.

Likewise, as we include a relative large number of drivers who remained for a rather short time in the Formula 1 business our results for the effect of age and experience on race classification might be partly driven by them. Changing the specification leads to trivially different results that shall not be represented here because the exact and more sophisticated estimation of a mean life cycle alone could fill one more paper.

4.3 First notions of superstar effects

An extensive part of the literature survey has been dedicated to the economics of superstars where the main contribution was the theoretical article by ROSEN (1981). Unfortunately no income data for Formula 1 racing was available. Although some contracts sums for today's drivers can be obtained the more important income source mainly sponsorship contracts and marketing revenues are not accessible. For former Formula 1 drivers not even contract sums could be found.

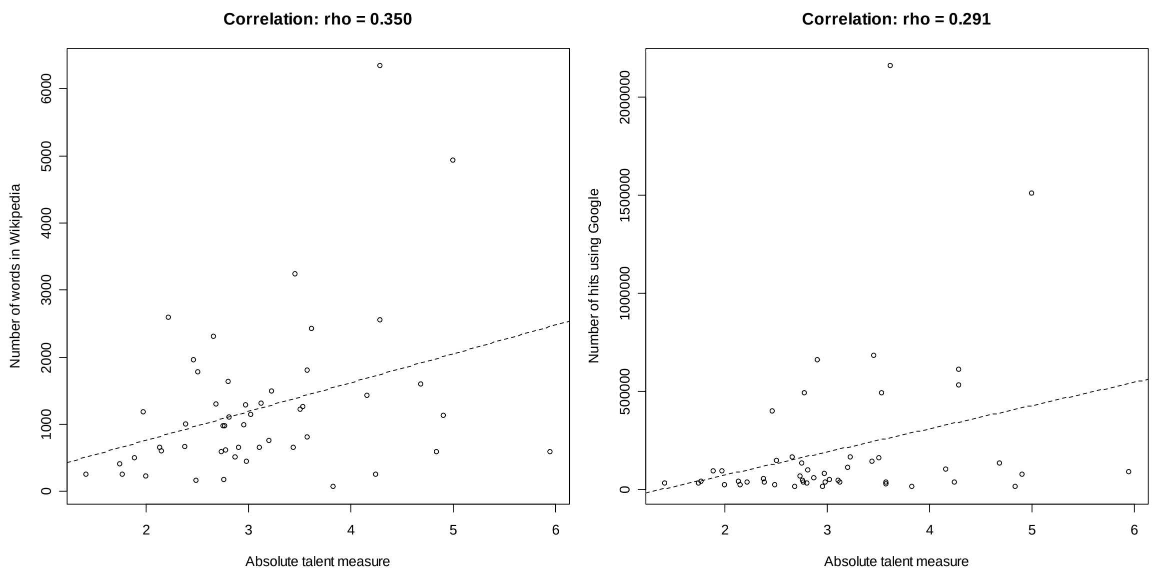
4.3.1 Preliminary estimation of superstar effects

Although the data situation for the income of the drivers is not the best, some first notions of superstar effects can still be estimated using a very crude way of measuring high profile. The general publicity shall be measured as the number of Google hits and the number of words for the first 50 drivers out of Table 3 in the English Wikipedia encyclopedia which is a similar measure as the one used by KRUEGER (2005). Both variables have a number of drawbacks but we shall try to correct for some of them. Once more, the results are only preliminary ones. As the number of Wikipedia words as well as the number of Google hits change over time it is important to mention that measures of the variables were all obtained within two days, the 17th and the 18th of May 2006. In order to obtain the number of Wikipedia words we looked at the English version of the site and counted the words for all 50 drivers out of Table 3. For the Google hits we used the following search string: "Formula 1" "first and second name of driver".[54]

Like HAMLEN (1991) has done for popular music we shall now try to observe the effects of the estimated talent on the dependent variables crudely measuring high profile. The absolute coefficients for the drivers of Table 11 will be used as a measure for the driver's talent. Figure 11 gives some primary correlations.

Figure 11

Correlation of a driver's talent and general publicity



Source: Author's calculations and representation
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on autosportatlas.com; Data for Wikipedia words and Google hits was compiled by the author; As talent measures the absolute coefficient of Table 11 are used.

The dashed line represents the linear regression where the dependent variables are Wikipedia words and Google hits respectively and the independent variables are a constant and the talent measure. Note that this time no Spearman correlations are calculated but standard correlation measures and that the indicated "rho" is not the coefficient of the talent variable but the correlation.

Although the fit is far from perfect, the correlation between the two variables is

54 Specific problems with both variables will be discussed in the next subsection.

astonishingly high given the raw nature of the endogenous variables and the problems associated with estimating the driver's ability variable.

In Table 13 we try to control for other exogenous variables in order see if the shown correlations are not purely a statistical artifact. Note again that the absolute coefficients are used to measure talent.

Table 13
Preliminary estimation of superstar effects

variables	*(1) Wiki*	*(2) Wiki*	*(3) Wiki*	*(4) Google*	*(5) Google*	*(6) Google*
Intercept	-32.91	-829.58	-1709	-1.2e5	-1.9e5	-2.7e5
	(514.1)	(862.02)	(1020)	(1.2e5)	(2.1e5)	(2.6e5)
absolute coefficients	358.4[b]	370.8[b]	421.6[b]	6.5e5	6.5e4	7.2e4[c]
(talent measure)	(160.74)	(166.9)	(165.7)	(3.9e5)	(4.1e4)	(4.2e4)
ACTIVEDRIVER	979.61[b]	1199[b]	592.7	7.6e5[a]	7.7e5[a]	7.1e5[a]
	(411.16)	(426.4)	(537.7)	(9.8e4)	(1.0e5)	(1.3e5)
CAREERLENGTH		57.65	65.75		5176	6572
		(49.21)	(48.61)		(1.2e4)	(1.1e4)
DEADDURINGCAREER		659.1	799.9[c]		5.9e4	6.9e4
		(445.0)	(451.9)		(1.0e5)	(1.1e5)
CAREERPERIOD			418.4			3.9e4
			(269.2)			(6.8e4)
ENGLISHSPEAKING			-175.3			-4.9e4
			(338.4)			(8.6e4)
N	50	50	50	50	50	50
adj. R^2	0.184	0.266	0.326	0.576	0.569	0.603

Source: author's calculations
Data from 1950 until 2005 (included) racing season; FORIX Formula 1 Database on http://www.autosport.com; Standard errors are reported in parenthesis; [a] Significant at 1-percent level; ; [b] Significant at 5-percent level; [c] Significant at 10-percent level; All reported number are represented with four digits; e indicates the exponential 10;. "Wiki." Stands for Wikipedia; Data for Wikipedia words and Google hits was compiled by the author; Driver coefficients (talent measures) stem Table 11.

All Wikipedia regressions show adjusted R^2 or around 0.250 which is surely an acceptable value for the used variables. The Google hit regressions have even higher adjusted R^2 at values of around 0.600 which can be easily explained by the huge influence of the control variable ACTIVEDRIVER.

As can be seen in the first three regressions of Table 13 the driver coefficients measuring talent is always positive and significant. For the last three specifications it is positive. The Google hits are not significant but close to the 10-%-level.

As can be seen in Figure 11 there are two drivers in the data set that have very high values for the dependent variables. Concerning Wikipedia words these two drivers are M. Schumacher and A. Senna. For Google hits F. Alonso and M. Schumacher could be considered as outliers. When excluding these drivers from the analysis the fits get even better and the talent measure has a higher influence and is more significant.

The control variable for the activity of a driver is a dummy taking the value 1 when the driver was active in the year 2005. ACTIVEDRIVER is positive and significant in all specifications apart from regression (3). This variable drives the results in the Google hit regressions.

In order to control for a possible influence of myths a dummy variable for a death during the career is introduced in regressions (2) and (5). The variable is positive but not significant. The same is true for a control of career length. The coefficient of CAREERLENGTH is positive but not significant in all specifications. If we do not control for talent the CAREERLENGTH variable gets significant.

Moreover, a measure for the career period and the fact of speaking English is included in regressions (3) and (6). The CAREERPERIOD variable is introduced with the same idea as the ACTIVEDRIVER variable. It could be that drivers from the early ages of Formula 1 racing receive less attention in today's modern world. The CAREERPERIOD variable takes a value of 1 for career ends during 1950 to 1969, a value of 2 for career ends during 1970 to 1989 and a value of 3 for all other career ends. When the control for active drivers is included, the coefficient of the variable CAREERPERIOD is positive but not significant. Including CAREERPERIOD also leads to an insignificant coefficient for the ACTIVEDRIVER variable in regression (3). A test of "ACTIVEDRIVER and CAREERPERIOD = 0" can be rejected with a p-value of 0.040.

The variable ENGLISHSPEAKING should control for the possibility that drivers coming from an English-speaking country obtain more publicity in the English Wikipedia or generally from the English websites dominated internet. This does not seem to be the case.

Overall the performed evaluations indicate that talent has a positive influence on general publicity.

4.3.2 Problems and limits of the approach

Although the regressions show a positive influence of talent on general publicity it is far too early to explicitly mention superstar effects in the instance.

Firstly, the evaluation is only performed for the first 50 drivers and indeed at least the first ten of them may already be considered as superstars by the general public. Otherwise they would not be found so easily on the internet and they would not have entered Wikipedia with several hundred words. The analysis should therefore be extended to all drivers for whom an explicit talent measure is available. However, other authors such as HAMLEN (1991) when estimating the influence of voice on record sales in popular music, uses also a sample of persons that could already be considered as stars.

Another serious problem concerns the dependent variables itself. The number of Google hits changes almost daily and the spirit of Wikipedia is that everyone can edit it. It could be the case that fans of rather weak drivers entered extensive information on their preferred racers. The measures used are therefore very superficial and volatile.

Moreover, it has already been mentioned that the talent variables themselves could be viewed as incorrect. Admittedly and as discussed in the last chapter every ranking has its drawbacks. Furthermore, other measures of talent used in the literature when empirically testing the effects of superstars are mostly constructed in a qualitative manner as for HAMLEN (1991) or are not used at all as for CHUNG AND COX (1994). This thesis used a quantitative measure which was derived under difficult circumstances from the data.

Finally, the economics of superstars as introduced by ROSEN (1981) tries to explain income differentials. Small changes in talent can imply large differences in income. However, in this paper a simple linear model has been estimated and not a convex function and no income data was used. In addition, no elasticities have been considered in this preliminary evaluation. Estimations for all variables with their logarithms have been run like by HAMLEN (1991). The results changed only partially but a simple logarithmic form of the equation cannot compensate for more sophisticated estimation techniques that might be necessary.

4.4 Outlook and further research questions

This chapter has presented three economic applications of the talent measures derived in this thesis. All these applications should only be seen as preliminary efforts, as each of them could be treated as a follow up case study for sports economics, Formula 1 management questions or as a test of economic theories. None of the estimations was highly sophisticated and all could be improved. The aim was to give a brief overview over some applications of the model and the questions that could be answered. An enumeration of some other possible questions that could be treated in the same framework now follows:

- Does faster driving translate into higher publicity or income?
- Is it more difficult nowadays to be a good Formula 1 driver than in the past

because of more competitors?

- Do past wins have an effect on future wins? Do past wins increase the confidence of a driver?
- How exactly can the peak of the career be found and estimated?
- Do teams try to maximize one of their driver's points or do they maximize the team's points?
- Do fast drivers have more ponsors or are other considerations such as reliability, attractiveness, and so on more important?
- Do drivers systematically perform better in advance of new contracts?
- Are drivers "lame ducks" at the end of their contracts?
- What is the influence of special driver characteristics on the risks taken?
- How should the point system in Formula 1 be organized in order to minimize accidents?
- Do new entrants in Formula 1 drive out older or less competitive racers?
- Who has the monopoly power in Formula 1: drivers or teams?
- Are Formula 1 teams discriminating against certain drivers?

This is just a short list of economic questions that can potentially be analyzed using the results from this thesis. However, the general public may be even interested in other questions and so called "fun" applications:

- Which is the best car independent of the driver?
- How can future results be predicted?
- What would have happened if A. Senna had not died?
- Who is the best driver on a certain circuit?
- What would have happened if J. Siffert had not had any technical problems with his car?

There is almost an infinite amount of questions that could be posed and evaluated.

For the sake of amusement let us very briefly consider the last "fun" application. J. Siffert, Swiss Formula 1 driver from Fribourg, was actively participating in Formula 1 from 1962 to 1971. During these years he started in 96 Grand Prix and dropped out 39 times because of technical reasons and 13 times because of accidents, collisions and so on (human outs). Therefore, over 40.62 % of his Grand Prix ended with a technical out but this was fairly common during that time. Out of the remaining races J. Siffert finished six with a podium position, two of them were even wins. In the preferred specification of section 3.4 he is classified on position 70. Strictly statistically speaking, J. Siffert could even be ranked within ranks 40 to 50 because comparing drivers in these ranks does not lead to significant differences. As argued above he did not qualify for the list because he only won two races.

By slightly modifying regression (02) of Table 9 and changing the specification of the dependent variable to the accurate representation of the regulations of points during J. Siffert's time we can predict his points under the assumption that he had not had any technical outs. Under the mentioned assumption and the hypothesis that the points of all other drivers remained the same in the considered years J. Siffert would have been two times under the first three in the World Championship. In 1969 he would have ranked third instead of fifth, and in 1971 also third instead of fourth.

We do not want to discuss all shortcomings and problems of this very hypothetic prediction nor do we want to mention the prediction confidence intervals for such an evaluation.[55] The aim of this "fun" application was mainly to show that a number of questions can be answered and that these questions do not have to be of a dry and uninteresting nature.

55 Indeed, considering the confidence intervals J. Siffert could have even ranked last in these two Championships.

5 Conclusions

Formula 1 drivers are fast and good cars make them even faster. In this Master thesis the main focus was on evaluating the talent of a driver independently of the car.

The literature survey in chapter 2 gave a general insight into the wide field of sports economics, the economics of superstars and statistical evaluation methods used in sports. Nowadays sports economists do not only focus on pure economics. Instead, economic tools are applied to sports and the field of sports serves as base for different hypothesis tests concerning economic questions. Furthermore, the economics of superstars focuses on the explanation of large income differentials in the light of small differentials in talent and ability. Particularly, theoretical papers assume that the difference in talent is small when superstars are compared with the general public but no attempt is made to verify this hypothesis.

Our evaluation of Formula 1 drivers might be criticized as pure empiricism, but when considering the enormous number of Formula 1 fans, there is an intrinsic value of trying to establish a ranking of racers. Furthermore, Formula 1 and sports itself can be seen as a laboratory for several hypothesis tests in economics: the measurement of talent is just one example.

We focused on the three main methods of evaluating a driver independently of his car in chapter 3. The first method of evaluation was a table of performance variables corrected for the number of races in which a driver participated. This very crude and simple approach shed more light onto the matter of Formula 1 than most fan-sites do. The second step of the process of finding the best driver was a paired comparison methodology. Three different methods were applied and for the paired comparison approach all drivers were ranked. Finally, driver and car effects were evaluated using an econometric model. The coefficients of the resulting regression model were interpreted as the strengths of the drivers. In all evaluations we faced problems with the data and the methodology. We discussed these problems and hinted to different possible solutions. Some of the solutions were applied in later chapters but a number of problems remain unresolved to date. Comparing different outcomes in sports is still a difficult statistical research area and simple methods are not available. It can be shown that a unique and "correct" ranking in sports cannot exist because every aggregation procedure depends on a number of hypotheses and assumptions.

Our approach was to use several different methods for the evaluation of a driver's talent.

Indeed, it can be said that the resulting rankings are relatively well correlated and that the results are not completely counter-intuitive. They are certainly superior to the usual rankings done by diverse sports magazines. Particularly, J. M. Fangio is better than M. Schumacher. J. Clark, N. Farina, A. Prost, K. Räikkönen, A. Senna, A. Ascari, J. Stewart and F. Alonso can usually be found under the TOP-10 in all evaluations.

Chapter 4 used the extensive database that was constructed for this thesis as well as the results of the evaluations of chapter 3 to show that interesting economic questions can be answered. Firstly, we found that drivers seem to accept a certain amount of risk. When the weather is good they drive faster and thereby increase the probability of technical drop-outs. When it is raining they slow down but the number of human drop-outs increases all the same. Overall, no clear impact of the weather on risk-taking can be identified and it seems that the probability of a drop-out is independent of the weather. Furthermore, an average life cycle of a driver was estimated and it was shown that only a handful of drivers can look back to long careers. Most drivers already drop out after only one or two seasons. Finally, we considered possible effects of superstars and found that increasing talent had a positive influence on general publicity.

There are a many questions that can be analyzed using a similar framework as the one used here. Those problems extend not only to economic questions, but also to questions that are in the general interest of the public. Some initial hints have been made in section 4.4 and a short evaluation of J. Siffert's Championship positions under certain hypothesis has been made.

In general we believe that the field of sports is rich and that our case study for Formula 1 rankings and economic applications in Formula 1 is just one small element that was worth a closer look. In the future we hope to use the gained insights in econometrics and statistics not only for Formula 1 but also for other branches.

Bibliography

ADLER M. (1985): "Stardom and Talent", *The American Economic Review*, Vol. 75, No. 1, pp. 208-212.

VON ALLMEN P. (2001): "Is the Reward System in NASCAR Efficient?", *The Journal of Sports Economics*, Vol. 2, No. 1, pp. 62-79.

BALTAGI B. H. (2001): Econometric Analysis of Panel Data, 2nd edition, John Wiley & Sons Ltd., Chichester.

BERRY S. M., REESE S. C., LARKEY P. D. (1999): "Bridging Different Eras in Sports", *The Journal of the American Statistical Association*, Vol. 94, No. 447, pp. 661-676.

BRADLEY R. A. (1984): "Paired Comparisons: Some Basic Procedures and Examples", In: P. R. Krishnaiah and P. K. Sen (Eds.), *Handbook of Statistics*, Vol. 4, pp. 299-326, Elsevier Science Publishers.

CAMERON A. C., TRIVEDI P. K. (2005): Microeconometrics – Methods and Applications, Cambridge University Press.

CHUNG K. H., COX R. A. K. (1994): "A Stochastic Model of Superstardom: An Application of the Yule Distribution", *The Review of Economics and Statistics*, Vol. 76, No. 4, pp. 771-775.

DAVIDSON R., FARQUHAR P. H. (1976): "A Bibliography on the Method of Paired Comparisons", *Biometrics*, Vol. 32, No. 2, pp. 241-252.

DEPKEN C. A., WILSON D. P. (2004): "The Efficiency of the NASCAR Reward System: Initial Empirical Evidence", *The Journal of Sports Economics*, Vol. 5, No. 4, pp. 371-386.

DAVIDSON R., MACKINNON J (1981): "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, Vol. 49, No. 3, pp. 781-793.

FARAWAY J. J. (2005): Linear Models with R, Chapman and Hall/CRC, Boca Raton Florida.

FORT R., MAXCY J. (2003): "Competitive Balance in Sports Leagues: An Introduction", *The Journal of Sports Economics*, Vol. 4, No. 2, pp. 154-160.

GLICKMAN M. E. (2001): "Dynamic paired comparison models with stochastic variances", *The Journal of Applied Statistics*, Vol. 28, No. 6, pp. 673-689.

GLICKMAN M. E., JONES A. C. (1999): "Rating the chess rating system", *Chance*, Vol. 12, No. 2, pp. 21-28.

GLICKMAN M. E., STERN H. S. (1998): „A state-space model for National Football League scores", *The Journal of the American Statistical Association*, Vol. 93, No. 441, pp. 25-35.

GREENE W. H. (1993): "Econometric Analysis", 2nd edition, Macmillan Publishing Company, New York.

HAMILTON J. D. (1994): Time Series Analysis, Princeton University Press, Princeton.

HAMLEN W. A. (1991): "Superstardom in Popular Music: Empirical Evidence", *The Review of Economics and Statistics*, Vol. 73, No. 4, pp. 729-733.

HOLLAND P. W. (1986): "Statistics and Causal Inference", *The Journal of the American Statistical Association*, Vol. 81, No. 396, pp. 945-960.

HOLMSTRÖM B. (1999): "Managerial Incentive Problems. A Dynamic Perspective", *The Review of Economic Studies*, Vol. 66, No. 1, pp. 169-182.

HOLT C. A., LAURY S. K. (2002): "Risk Aversion and Incentive Effects," *The American Economic Review*, Vol. 92, No. 5, pp 1644–1655.

HÜBNER G. (2003): Stochastik für Informatiker und Mathematiker, 4. Auflage, Vieweg & Sohn Verlagsgesellschaft, Braunschweig.

JOE H. (1990): "Extended Use of Paired Comparison Models, with Application to Chess Rankings", *Applied Statistics*, Vol. 39, No. 1, pp. 85-93.

KAHN L. M. (2000): "The Sports Business as a Labour Market Laboratory", *The Journal of Economic Perspectives*, Vol. 14, No. 3, pp. 75-94.

KAHNEMAN D., TVERSKY A. (1979): "Prospect theory: An Analysis of Decision under Risk", *Econometrica*, Vol. 47, No. 2, pp. 263-292.

KLADROBA A. (2000): „Das Aggregationsproblem bei der Erstellung von Rankings. Einige Anmerkungen am Beispiel der Formel 1 Weltmeisterschaft 1998“, *Jahrbücher für Nationalökonomie und Statistik*, Vol. 220, No. 3, pp. 302-314.

KNORR-HELD L. (2000): “Dynamic Rating of Sports Teams”, *The Statistician*, Vol. 49, No. 2, pp. 261-276.

KOENKER R. (1981): “A Note on Studentizing a Test for Heteroscedasticity”, *The Journal of Econometrics*, Vol. 17, No. 1, pp. 107-112.

KRUEGER A. B. (2005): “The Economics of Real Superstars: The Market for Rock Concerts in the Material World”, *The Journal of Labor Economics*, Vol. 23, No. 1, pp. 1-30.

LAIRD N. M., WARE J. H (1982): „Random-Effects Models for Longitudinal Data“, *Biometrics*, Vol. 38, No. 4, pp. 963-974.

LALONDE R. J. (1986): “Evaluating the Econometric Evaluations of Training Programs with Experimental Data”, *The American Economic Review*, Vol. 76, No. 4, pp. 604-620.

LIAO J. G., MCGEE D. (2003): “Adjusted coefficients of determination for logistic regression”, *The American Statistician*, Vol. 57, No. 3, pp. 161-165.

LYNCH J. G., ZAX J. S. (2000): “The Rewards to Running: Prize Structure and Performance in Professional Road Racing”, *The Journal of Sports Economics*, Vol. 1, No. 4, pp. 323-340.

MENARD S. (2000): “Coefficients of determination for multiple logistic regression analysis”, *The American Statistician*, Vol. 54, No. 1, pp. 17-24.

MACDONALD G. M. (1988): “The Economics of Rising Stars”, *The American Economic Review*, Vol. 78, No. 1, pp. 155-166.

ROSEN S. (1981): “The Economics of Superstarts”, *The American Economic Review*, Vol. 71, No. 5, pp. 845-858.

SEAMAN B. A. (2003): “Cultural and Sport Economics: Conceptual Twins?”, *The Journal of Cultural Economics*, Vol. 27, pp. 81-126.

SIEGFRIED J., ZIMBALIST A. (2000): “The Economics of Sports Facilities and Their Communities”, *The Journal of Economic Perspectives*, Vol. 14, No. 3, pp. 95-114.

SCHWARZ H.-R., KÖCKLER N. (2004): Numerische Mathematik, 5. Auflage, Teubner Verlag, Wiesbaden.

STERN H. (1992): “Are all linear paired comparison models empirically equivalent?”, *Mathematical Social Sciences*, Vol. 23, pp. 103-117.

SZYMANSKI S. (2003): “The Economic Design of Sporting Contests,” *The Journal of Economic Literature*, Vol. 41, No. 4, pp. 1137-1187.

TROISEN G. (2003): Sportökonomie – Ein Lehrbuch in 15 Lektionen, Meyer & Meyer Verlag, Aachen.

UTT J., FORT R. (2002): “Pitfalls to Measuring Competitive Balance With Gini Coefficients”, *The Journal of Sports Economics*, Vol. 3, No. 4, pp. 367-373.

WOOLDRIDGE J. M. (2002A): Introductory Econometrics – A Modern Approach, 2nd edition, South Western College Publishing, Manson.

WOOLDRIDGE J. M. (2002B): Econometric Analysis of Cross Section and Panel Data, The MIT Press, Cambridge Massachusetts.