Westfälische Wilhelms-Universität Münster

The Uncertainty-of-Outcome Hypothesis and Competitive Balance in Sports

Martin Langen

The Uncertainty-of-Outcome Hypothesis and Competitive Balance in Sports

Inauguraldissertation
zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften durch die
Wirtschaftswissenschaftliche Fakultät der
Westfälischen Wilhelms-Universität Münster

vorgelegt von
Martin Langen
aus Osnabrück

-Münster, 2013-

Wirtschaftswissenschaftliche Fakultät
Der Westfälischen Wilhelms-Universität

Universitätsstraße 14-16, 48143 Münster

Erstberichterstatter:	Prof. Dr. Aloys Prinz
Zweitberichterstatter:	Prof. Dr. Alexander Dilger
Dekan:	Prof. Dr. Christoph Watrin
Tag der mündlichen Prüfung:	21. Januar 2014

Martin Langen

The Uncertainty-of-Outcome Hypothesis and Competitive Balance in Sports

MV WISSENSCHAFT

Wissenschaftliche Schriften der WWU Münster

Reihe IV

Band 8

Universitäts- und
Landesbibliothek Münster

Martin Langen

The Uncertainty-of-Outcome Hypothesis and Competitive Balance in Sports

MV WISSENSCHAFT

Wissenschaftliche Schriften der WWU Münster
herausgegeben von der Universitäts- und Landesbibliothek Münster
http://www.ulb.uni-muenster.de

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Dieses Buch steht gleichzeitig in einer elektronischen Version über den Publikations- und
Archivierungsserver der WWU Münster zur Verfügung.
http://www.ulb.uni-muenster.de/wissenschaftliche-schriften

Martin Langen
„The Uncertainty-of-Outcome Hypothesis and Competitive Balance in Sports"
Wissenschaftliche Schriften der WWU Münster, Reihe IV, Band 8

Die Reihe „Wissenschaftliche Schriften der WWU Münster" erscheint im Verlagshaus Monsenstein und
Vannerdat OHG Münster
www.mv-wissenschaft.com

ISBN 978-3-8405-0095-4 (Druckausgabe)
URN urn:nbn:de:hbz:6-04309403598 (elektronische Version)

direkt zur Online-Version:

Satz: Martin Langen
Umschlag: MV-Verlag
Druck und Bindung: MV-Verlag

Table of Contents

List of Tables

Table of Figures

1 Introduction

"The nature of the [sports] industry is such that competitors must be of approximately equal "size" if any are to be successful; this seems to be a unique attribute of professional competitive sports."

- Simon Rottenberg, 1956, p.242

The demand for professional sport is closely linked to the competition's uncertainty of outcome. Without this uncertainty, spectators would lose interest in watching sport events, which would be predictable and dull. For Simon Rottenberg, the founder of sport economic research, this hypothesis is one of the central characteristics of the professional sports industry. In contrast to other industries, sports competitors create a good conjointly. There can be no competition without an opponent. Furthermore, it is not sufficient merely to have an opponent, it also has to be similar in terms of sporting quality. Otherwise, the competition would be too one-sided and therefore predictable for the spectators. Thus, in professional sports, individual decisions lead to the creation of a good, the quality of which depends on the difference in quality of the various opponents. The uncertainty-of-outcome hypothesis seems to be so intuitive and quintessential for sport competition, that the difficulties involved in describing and interpreting it economically is initially surprising. Of peculiar importance is the topic of competitive balance in sports, due to the many regulatory interventions frequently justified by the need to improve this competitive balance. These interventions constrain free competition, restrict the freedom of contract and often conflict with the general law. However, without a comprehensive understanding of competitive balance in sports, these market interventions are hardly justifiable. Hence, providing new insight into this issue is the aim of the present thesis. Accordingly, reasons for competitive imbalances within sport leagues, beyond the well-established ones, are considered. Furthermore, different measures of outcome uncertainty, as well as its numerous dimensions, are presented. Moreover, the interdependence between competitive balance and the demand for a professional sport event is discussed. In order to gain more in-

sight into the phenomenon of competitive balance in sports, existing approaches are extended and new methods introduced.
In Chapter 2, Chapter 3 and Chapter 4, on the basis of a general model of sport team investment behavior, model extensions are presented, which provide a more sophisticated explanation of talent allocation in a sport league. Several sport teams compete in the factor market over talent, and on the field with their talent. The theoretical equilibrium level of competitive balance in sports leagues are often derived from and analyzed by means of models based on the first mathematical formulization of El-Hodiri and Quirk (1971). In these models, sport league teams invest in talent which is an input factor in the production of winning probability of a team. It is this probability that generates a team's revenue. The specification of the revenue, cost and production function as well as the team's objective have been discussed comprehensively in the economic literature, and the assumptions regarding the function specifications exert a major influence on the competitive balance predicted by the models and the ability to alter it. However, the only consensus in the literature is on the basic properties of the revenue function. According to the uncertainty-of-outcome hypothesis, the revenue function should be concave with a maximum value at a winning percentage below one.
As a slight variation of the famous Sigmund Freud quotation, a sports economist could say: "The great question that has never been answered is: What does a sports team want? “ The discussion on the true objective of sport teams goes back to Sloane (1971). Before that research article, sport teams were regarded as behaving like regular firms, so their assumed objective was that of maximizing profits. Since Sloane (1971), there has been broad debate on whether teams may in fact attempt to maximize their sporting success instead, and thus their employed talent, while being restricted to a budget constraint often specified as break-even. When confronted with the large losses, that sports teams sometimes incur, some economists have stated that teams are not even restricted by this zero profit condition and teams therefore overinvest in talent. However, the relationship between investment and profit is not unique to the sports indus-

try. Teams invest in talent, which produces a winning probability, which generates revenue, which is then used to invest in more talent. In the sport economic literature, one-period models of investment are often used, which are not sufficient to depict this cycle of investment and profits. To shed some new light on the question of whether the objective of teams really is to maximize profits or rather sporting success, in Chapter 2 an alternative objective is presented. Players are a tradable good with an inherent value and which do generate additional value for a team. With this approach, teams not try to maximize sporting success or profits, rather than maximizing their value, which means that they attempt to maximize success and profits. This is a more accurate description of the abovementioned investment circle. Furthermore, it can be shown that low profits or even losses are unproblematic for sports teams, as long as they are in accordance with value maximization, and therefore do not reflect overinvestment.

Beside the objective of win maximization, the conclusion that sport teams tend to overinvest in talent can be derived by applying alternative equilibrium concepts. The well-established approach in the sport economic literature, given that sport is assumed to be a non-cooperative game, is to find the Nash equilibrium for a sport league. Alternatively, it is possible to determine an evolutionarily stable equilibrium in a sport league, which is a refinement of the Nash equilibrium. Grossman (2013) shows that teams invest more in the evolutionarily stable equilibrium than in the Nash equilibrium. The difference between the two equilibria results from the limited number of teams in a sport league. For an infinite population, every evolutionarily stable equilibrium is also a Nash equilibrium, but not the other way round. In Chapter 3, the approach of Grossman (2013) is extended by applying the concept of an optimal aggregate-taking strategy from Alós-Ferrer and Ania (2005), which is a sufficient condition for a globally stable evolutionarily stable equilibrium. By restricting the parameters of the model to reasonable values, it can be shown that neither a Nash nor an

evolutionarily stable equilibrium concepts do in fact lead to overinvestment, in contrast to what Grossman (2013) stated.
A basic result of the analysis of sports leagues is that the different revenue potentials of teams are one of the main reasons for an unequal distribution of talent in a league. Teams, for which sporting success has a higher value, accumulate more talent and are therefore more successful. That talent will end up where it has the highest value is one of the central results of Rottenberg (1956). Existing models assume that teams differ in terms of talent valuation, because they are located in different markets. A team's market determines the revenue potential a team is able to realize with their sporting success and comprises, among other things, the number of attendees, their willingness to pay and sponsor potential. Thus, differences in sporting success are attributable to different market sizes. Greater differences in market size should therefore lead to a greater competitive imbalance or dominance by teams from larger markets. By comparing this theoretical prediction with real league outcomes, it can be shown that this causality does not reveal the whole story. In reality, large market dominance is not reflected in this explicitness. According to Kuper and Szymanski (2009), the seven largest metropolitan areas in Europe, namely Istanbul, Paris, Moscow, London, St. Petersburg, Berlin, and Athens, were never able to win the main trophy in European Club football (European Champion Clubs' Cup, Champions League). Aside from their unusual constitution from the largest metropolitan region (Randstad, Madrid, Rhine-Ruhr, Rhine-Main, Manchester-Liverpool are as large or even larger than the abovementioned regions) and the fact that Chelsea F.C. from London won the Champions League in 2012, it is nevertheless evident that the link between size and success are not as simple as the theory suggests. That a large market does not necessarily lead to success is also clear from looking at the most successful European teams. Hardly ever is the all-time leader of a country also from the largest market. In Germany, the most successful team is from Munich, in England from Liverpool, or in recent years Manchester, in France from Marseille and in Italy from Turin. Thus, in the biggest European football leagues, with the exception of

Spain and the most successful team being from Madrid, the all-time leaders are not from the largest markets. In Chapter 4 a model is presented in which the markets in which the teams are located are not given exogenously, but determined endogenously. With endogenous markets, teams from large markets may be challenged by teams in their own markets. This leads to regional competition and lower revenue potential for contested market teams, as well as to smaller-market-team dominance. The implications of regional competition on the effectiveness of revenue-sharing agreements are presented, as well as evidence on important European leagues, which shows the extent of regional competition.
In Chapter 5 the competitive balance of a league is considered from a different perspective. Besides determining the reasons for competitive imbalance within a sports league, another relevant topic in sports economics is that of evaluating the degree of imbalance in an actual league. However, analyzing a league outcome and determining its level of competitive balance is challenging, due to the many dimensions of outcome uncertainty. Single matches can be uncertain, as well as a season's outcome or a number of consecutive seasons. The most common means of measuring competitive balance in sport economics is to analyze the distribution of wins per season. However, these measures do not consider the level of competitive balance due to volatility in team rankings from consecutive seasons. The degree of uneven point distributions is one dimension of the competitive balance of a league, although whether the same teams end up recurrently in the same positions is another. The similarity between consecutive seasons is analyzed by means of a geometrical approach. Seasonal outcomes in terms of point distribution and team ranking can be interpreted as vectors, and to determine the similarity between two vectors, the angle between them can be calculated. This concept is developed for a general sport league and applied to data for the German Bundesliga. It can be shown that, although point distributions become more even over time, seasons have become more similar in recent years. The former suggests an improvement and the latter a deterioration of competitive balance.

There are not only economic reasons for a certain level of competitive balance, but also sport policy. The rules of a particular sporting competition are an important factor determining the competitive outcome. For many sporting competitions, in the sense of a Formula One season or a football league, one element of these rules is the mechanism which aggregates numerous events into one single ranking. In one of the world's major sports, the Formula One Championship, this is done with a scoring vector. This scoring vector has had several changes over its history, with the last one in 2010. In Chapter 6, these rule changes are discussed from the perspective of social choice and it will be shown that a scoring vector is the only appropriate aggregation mechanism for Formula One. This is followed by a discussion about the many dimensions of competitive balance in this sporting event. The uncertainty of outcome in Formula One comprises, for example, the length of time a season is undecided, as well as a maximum equal point distribution or the number of different champions. All these elements add up to uncertainty of outcome and no single scoring vector achieves a high competitive balance in all these areas.

With the uncertainty of outcome being a fundamental component of a sport competition, it also has an influence on the spectator demand for this competition. The question remains of whether there might be an ideal competitive balance maximizing the demand for a sport competition. With a maximum imbalance, a sporting event would lack suspense, which is fundamental to demand. If all competitors are of equal strength, the result will be that of pure and equal chance. Whether the demand favors one of the two extremes, or a competitive balance in the middle, is discussed in Chapter 7 with respect to Formula One racing. One argument supporting the existence of a more imbalanced competition is that the audience might favor some protagonists more than others. An important feature of racing sport, like motorsport, cycling or athletics, especially compared to team-sports, is that all competitors compete against each other simultaneously. This emphasizes the uncertainty of outcome resulting from duels at the top. Even though the competition as a whole might be extremely unbalanced, the results at the top are uncertain, due to a small group that are similar in level to each other but superior to the rest of the competitors. Furthermore, the audience might enjoy watching their favorites being challenged by outsiders, more than interactions between opponents of equal strength. In such cases, a higher competitive imbalance would lead to a greater attractiveness of the sport. Using data relating to German TV viewers of Formula One races, it is evident that a too balanced competition is unattractive for TV Viewers, whereas fiercely contested duels at the top catch their attention. Chapter 8 concludes.

2 Players as Assets in Professional Sport Teams

Martin Langen, Aloys Prinz

Abstract: According to the sport economics literature, professional team sports, such as European football, are prone to overinvestment. According to this view, teams overinvest in players to maximize their winning probabilities, ignoring the fact that their competitors are doing likewise. Therefore, the relevant Nash equilibrium is a variant of the prisoners' dilemma. In this paper, it is shown that players constitute assets for teams and, hence, the stock of talented players has generic value for them. As a consequence, to understand investment in players from an economic perspective, a portfolio approach is required. In a two-period model, investments in talented players are studied. Taking account of the asset value of players, it is evident that no 'overinvestment' occurs after all. Instead, observable team decisions on the stock of employed talented players can be considered rational.

Keywords: team sport, overinvestment, players' asset value, talent portfolio

JEL classification: C72, L83, L2, D43, G32

1 Introduction

A U.S. American baseball team was recently sold for over two billion dollars. Moreover, as can easily be seen in the Internet, all teams of the German football Bundesliga, for example, are valued according their market value. This valuation entails a kind of 'goodwill', as already carried out for all major firms (for instance, according to the accounting rules of IFRS).[1] However, in sport economics, teams and their players have not so far been analyzed as assets. Instead, investments in players are regarded as particularly too high when teams maximize their winning probabilities rather than their profit (Dietl, Franck and Roy, 2003, Dietl, Franck and Lang, 2008; Grossmann and Dietl, 2009; on overinvestment in talent through profit maximization, see Whitney, 1993; Dietl and Franck, 2000; Lang, Grossmann and Theiler, 2011). The reason is that almost all team sport models consider one period only and they do not account for the value of their stock of players. From a portfolio analysis point of view, this is hardly justifiable.

In this paper, to the best of our knowledge, the value of teams (its stock of talented players) is taken into account for the first time in sport economics. In a two-period model of a league, we show how teams do invest rationally in players. In order to do so, the value of the stock of players is adjusted periodically according to the winning probabilities. In sport leagues, the profits of a team depend partly on past success, so that players can increase or decrease future profit opportunities and thus the value of a team. Furthermore, players often have contracts that are extended by one period only, so that teams are able to trade them and they thus constitute tradable assets for the teams. In this model players are treated as assets that provide a generic value for the team. Hence, buying and selling talent should not only be analyzed from the perspective of periodic revenue, but also as an investment in valuable assets.

[1] Baetge, Klönne and Weber (2013) discuss the potential for and limits of an objective economic valuation of football players with monetary values.

The paper is structured as follows. Section two contains a general two-period model of a league, accounting for team investments in talented players. In section three, the model is specified to both calculate and compare profits between teams. A discussion of the model and its results is presented in section four with section five concluding.

2 Players as assets: General model

2.1 Player allocation in the first and second period

Define the profit π_i of a team i, i ε I, I = {1, …, n}, in a two-period model as:

$$\pi_i = R_{i,1}(w_{i,1}) - c_1 t_{i,1} + r\{R_{i,2}(w_{i,2}) + \theta_i(w_{i,1}) - c_2(t_{i,2})\} \tag{1}$$

with:

$R_{i,j}(\ldots)$: revenue of team i in period j, j = {1, 2}, $R'_{i,j}(\ldots)>0$, $R''_{i,j}(\ldots)<0$,
$w_{i,j}$: winning probability of team i in period j,
c_j: per-unit cost of talent in period j (constant),
$t_{i,j}$: employed stock of talent of team i in period j,
r := 1/(1+ρ), ρ: discount factor,
θ_i: increased value of team i in Period 2 as a function of winning probability (success) in Period 1.

With respect to the profit function, note that a conventional team profit function, extended over two periods, reads as follows:

$$\pi_i = R_{i,1}(w_{i,1}) - c_1 t_{i,1} + r\{R_{i,2}(w_{i,2}) - c_2(t_{i,2})\}. \tag{2}$$

Comparing (2) with (1), the team valuation term θ_i is missing in (2). Equation (2) defines profit as revenue, depending on the team's winning probabilities in both periods, minus the costs of talent employed. However, note that (2) contains flow values only, i.e., stock values are not accounted for. The valuation function θ_i determines the 'goodwill' a team acquires during the first period, it represents the team's value change in the second period due to success in period 1. Inserting equation θ_i into equation (2) yields equation (1).

In addition to the notation and the definitions presented so far, several further assumptions are required to analyze profit maximization given by (1). First of all, it is assumed that the talent available for a league is fixed at $\sum_i t_{i,j} := T$ for $j = 1,2$. Moreover, as is usual in sport economics, the winning probabilities of the periods are determined by a logit contest success function (see Tullock, 1980; Hirschleifer, 1989; Skaperdas, 1996; for a discussion of the contest success function in sports, and on talent supply, see Fort and Winfree, 2009):

$$w_{i,j} = \frac{t_{i,j}}{T}. \tag{3}$$

Notice that with a fixed supply of talent $\frac{\partial w_{i,1}}{\partial t_{i,1}} = \frac{1}{T}$.

Inserting (3) into (1) gives:

$$\pi_i = R_{i,1}\left(\frac{t_{i,1}}{T}\right) - c_1 t_{i,1} + r\left\{R_{i,2}\left(\frac{t_{i,2}}{T}\right) + \theta_i\left(\frac{t_{i,1}}{T}\right) - c_2 t_{i,2}\right\}. \tag{4}$$

Hence, the optimal employment of talent is determined for the first period by:

$$\frac{\partial R_{i,1}}{\partial w_{i,1}} + r\,\frac{\partial \theta_i}{\partial w_{i,1}} = c_1\, T, \tag{5}$$

and for the second period by:

$$\frac{\partial R_{i,2}}{\partial w_{i,2}} = c_2\, T. \tag{6}$$

The left side of (5) and (6) encompasses the marginal total revenues of employing talent and the right side, total marginal costs. Note that for $\theta_i = 0$, as well as $\frac{\partial \theta_i}{\partial w_{i,1}} = 0$, the usual optimal employment condition for talent emerges.

Hence, increases (or decreases) in marginal team value are decisive for the player asset value model presented here. For $\theta_i > 0$, the (discounted) marginal value of an additional unit of talent is given by: $r\,\frac{\partial \theta_i}{\partial w_{i,1}}$.

Proposition 1: In a two-period team sport league with *n* teams and a fixed number of *T* talent units, the demand for talent by a team *i* is larger (smaller) in a league where players' asset values and value changes are accounted for, than in a league where players' asset values are not considered.

Proof: Proposition 1 follows directly by a comparison of equation (5) and equation (6) for $\theta_i = 0$ and $\theta_i > (<) 0$, respectively.//

A first implication of Proposition 1 is that in completely balanced leagues with

$$t_{i,1} = \frac{T}{n}, \forall i \in I, \quad (7)$$

the allocation of players to teams is exactly the same, irrespective of whether or not players' asset values are taken into account. Furthermore, the talent allocation in leagues with and without consideration of the players' asset value the will differ, as long as the marginal asset value is different between the teams.

Secondly, with the available number of talent units fixed, the reallocation of talent among teams means that some will increase their winning probabilities, whereas the remaining teams will have lower winning probabilities. In this sense, the reallocation of players is a zero-sum game. Especially with a fixed talent pool, a higher demand for talent leads to higher marginal costs. However, it is crucial to note that investment decisions according to (4) are rational and they cannot be considered 'overinvestments', because they are in fact profit maximizing. As a consequence, more investment in talent cannot be attributed generally to 'winning maximization' in contrast to 'profit maximization'. Observing apparent 'overinvestment' may not be valid, if the asset value of players is considered.

2.2 Player reallocation in the second period

In the previous section, teams decided at the start of each period on the talent allocation between them. But in a more periodic model, they could also give players long-term contracts and trade talent in the second period. In this section, teams may buy or sell players to maximize profits. Talent is bought and sold at a transfer price p per unit of talent. The transfer price is price p which clears the player transfer market (transfer market equilibrium):

$$p\colon \sum_{b \epsilon I} t_b(p) = \sum_{s \epsilon I} t_s(p), \quad (8)$$

with:

b ε I: teams that want to buy some units of talent at the end of Period 1, and

s ε I: teams that want to sell some units of talent at the end of Period 1.

Teams chose talent at the beginning of the first period, which also determines the marginal costs per unit of talent. In the second period, each team *i* decides on the level of talent for Period 2 and the transfer of talent $\Delta t_i = t_{i,2} - t_{i,1}$, which is the difference in the levels of talent employment between periods 1 and 2. The transfer decisions of all teams determine the transfer price p.

In general, for each team, the new profit function is determined as follows:

$$\pi_i = R_{i,1}\left(\frac{t_{i,1}}{T}\right) - c\, t_{i,1} + r\left\{R_{i,2}\left(\frac{t_{i,2}}{T}\right) + \theta_i\left(\frac{t_{i,1}}{T}\right) - p\left(t_{i,2} - t_{i,1}\right) - ct_{i,2}\right\}. \quad (9)$$

Comparing (4) with (9), the determination of profit for Period 1 is the same. However, for the second, several changes are to be considered. Firstly, the value of the existing player stock at the end of Period 1, θ_i, depends on the talent employed (which determined the winning probability in period one). Secondly, the net investment in players, i.e., $t_{i,2} - t_{i,1}$, is accounted for. Note that transfer payments are paid (received) if Δt_i > (<) 0. As a consequence, transfer payments are either larger or smaller than zero, respectively.

Maximizing (9) with respect to the demand for talent in periods 1 and 2, $t_{i,1}$ and $t_{i,2}$, yields the following first-order conditions for profit maximization:

$$t_{i,1}:\ \frac{\partial R_{i,1}}{\partial w_{i,1}} + r\frac{\partial \theta_i}{\partial w_{i,1}} = (c - r\,p)T \quad (10)$$

$$t_{i,2}:\ \frac{\partial R_{i,2}}{\partial w_{i,2}} = (c + p)\ T. \quad (11)$$

The first-order condition for the employment of talent in Period 1 entails the discounted marginal value of the first-period stock of talent at the end of Period 1, *rp*, as well the discounted marginal team value $r\frac{\partial \theta_i}{\partial w_{i,2}}$; both values are unambiguously larger than zero. For the optimal talent demand in the second period, the marginal transfer payment for an additional unit of talent, p > 0, enters the first-order condition.

To shorten the notation, the following abbreviations will be used:
$\frac{\partial R_{i,j}}{\partial w_{i,j}} := R'_{i,j}$, $\frac{\partial \theta_i}{\partial w_{i,1}} := \theta'_i$.
The intertemporal optimum of talent employment results from combining equations (10) and (11):
$R'_{i,1} + r\theta'_i - (c - rp)T = R'_{i,2} - (c + p)T,$
$$R'_{i,2} - R'_{i,1} = r\theta'_i + pT(1 + r). \tag{12}$$
Solving (12) for the transfer price of a talent unit, p, yields:
$$p = \frac{R'_{i,2} - R'_{i,1} - r\theta'_i}{T(1+r)}. \tag{13}$$
From (13), it follows that a transfer price larger than zero requires:
$$p > 0 \Leftrightarrow R'_{i,2} > R'_{i,1} + r\theta'_i. \tag{14}$$
Proposition 2: (a) Intertemporally optimal decisions on the units of talent (players) in sports teams requires taking into account both the marginal value of additional talent units, as well as the asset value of the existing stock of talent (incorporated in players). (b) Transfer payments smooth the intertemporal talent allocation among teams, as they increase first-period and decrease second-period talent demand.
Proof: (a) It is worthy emphasizing that for conventional profit maximization, an intertemporal optimum of talent employment requires:
$$R'_{i,1} = R'_{i,2}. \tag{15}$$
Comparing equation (15) with equation (12) reveals that both equations are equal only if, at the same time, the marginal stock value change in player assets, θ_i, as well as the transfer price for a unit of talent, p, are zero. Hence, equations (15) and (12) will not necessarily be equal. According to (14), the transfer price for a talent unit may be larger than zero. Moreover, inserting equation (13) into the first-period optimality condition of talent employment, equation (10), and solving the equation for θ_i, results in:
$$\theta'_i = \frac{1}{r}\left[c(1 + r)T - R'_{i,1} - rR'_{i,2}\right]. \tag{16}$$
Hence, the marginal stock value change of players is also not necessarily zero. Consequently, the intertemporal optimality conditions in (12) and (15) may differ.

(b) The transfer value of a talent unit, p, enters the optimality conditions for periods 1 and 2 on the respective right sides of equations (10) and (11). However, the sign of this transfer price is negative in equation (10) and positive in equation (11). This means that the transfer price increases talent employment in the first period and decreases it in the second. In this way, a team's talent employment is smoothed intertemporally. //

2.3 Revenue sharing and transfer payment sharing

In order to improve competitive balance, revenue sharing is advisable. Such a policy seems effective at first glance, as the reason for an unbalanced league is unequal potential team revenue. There is a lengthy debate in the sport economics literature on the effectiveness of revenue sharing agreements whose results depend to a large extent on specific assumptions. Stated first by Rottenberg (1956) and formalized by El-Hodiri and Quirk (1971), revenue sharing does not affect talent allocation within a league. When sharing does not affect the equalization of marginal product of talent investments across teams, it does not change talent allocation in a league (Winfree and Fort, 2012). This holds true in the basic model with profit maximizing teams and a fixed talent supply (e.g. Szymanski, 2003; Vrooman, 1995). If the talent supply is perfectly elastic, Rottenberg's (1956) invariance principle does not hold (Szymanski and Kesenne, 2004; Kesenne, 2005). The principle does not hold either if the team's objective is win-percentage maximization (Kesenne, 2006; Vrooman, 2008). Moreover, a non-linear tax-subsidy scheme renders the invariance principle invalid (Marburger, 1997), as does a progressive payroll tax (van der Burg and Prinz, 2005).

To check whether revenue sharing has an effect in a model of flow and stock values of talent, it is assumed that in the respective flow model, the invariance result for revenue sharing applies.

Revenue sharing without long-term contracts

There are several possible revenue sharing schemes presented in the literature (see, e.g., Vrooman, 2007); in this paper so-called pool revenue sharing is applied. Each team retains a share α of its own revenue. The remainder of its own revenues, $1-\alpha$, is used to fund a revenue pool that is divided equally among all teams. Without long-term contracts, the profit of a team *i* is given by:

$$\pi_i = \alpha\left\{R_{i,1}\left(\frac{t_{i,1}}{T}\right) + r\,R_{i,2}\left(\frac{t_{i,2}}{T}\right)\right\} + \frac{1-\alpha}{2}\left\{R_{i,1}\left(\frac{t_{i,1}}{T}\right) + R_{j,1}\left(\frac{t_{j,1}}{T}\right) + r\left[R_{i,2}\left(\frac{t_{i,2}}{T}\right) + R_{j,2}\left(\frac{t_{j,2}}{T}\right)\right]\right\} - c_1 t_{i,1} + r\left\{\theta_i\left(\frac{t_{i,1}}{T}\right) - c_2 t_{i,2}\right\}.$$

Note that in a two-team league $\frac{\partial w_j}{\partial t_i} = -\frac{\partial w_i}{\partial t_i}$; thus, with a fixed supply of talent,

$$\frac{\partial w_j}{\partial t_i} = -\frac{1}{T}.$$

Hence, the optimal employment of talent is determined in the first period by:

$$\alpha\,R'_{i,1} + \frac{1-\alpha}{2}\left\{R'_{i,1} - R'_{j,1}\right\} + r\,\theta'_i = c_1 T, \tag{17}$$

and in the second period by:

$$\alpha\,r\,R'_{i,2} + \frac{1-\alpha}{2} r\left\{R'_{i,2} - R'_{j,2}\right\} = c_2\,T. \tag{18}$$

The equilibrium conditions implied by (17) and (18) are:

$$R'_{i,1} + r\,\theta'_i = R'_{j,1} + r\,\theta'_j \tag{19}$$

for the first period and:

$$R'_{i,2} = R'_{j,2}. \tag{20}$$

for the second period. These conditions do not differ from those implied by (5) and (6), so that revenue sharing does not affect the league's equilibrium talent allocation.

Revenue sharing with long-term contracts

With long-term contracts, talent becomes a marketable commodity in the second period. Thus, with a pooled revenue sharing agreement, the revenues of a team *i* are:

$$\pi_i = \alpha\left\{R_{i,1}\left(\frac{t_{i,1}}{T}\right) + r\,R_{i,2}\left(\frac{t_{i,2}}{T}\right)\right\} + \frac{1-\alpha}{2}\left\{R_{i,1}\left(\frac{t_{i,1}}{T}\right) + R_{j,1}\left(\frac{t_{j,1}}{T}\right) + r\left[R_{i,2}\left(\frac{t_{i,2}}{T}\right) + R_{j,2}\left(\frac{t_{j,2}}{T}\right)\right]\right\} - c t_{i,1} + r\left\{\theta_i\left(\frac{t_{i,1}}{T}\right) - c\,t_{i,2} - p(t_{i,2} - t_{i,1})\right\}. \tag{21}$$

The respective first-order conditions for talent employment are:

$$t_{i,1}: \ \alpha R'_{i,1} + \frac{1-\alpha}{2}\{R'_{i,1} - R'_{j,1}\} + r\,\theta'_i = T(c - r\,p), \tag{22}$$

$$t_{i,2}: \ \alpha\, r\, R'_{i,2} + \frac{1-\alpha}{2} r\{R'_{i,2} - R'_{j,2}\} = (c + p)\,T. \tag{23}$$

Both first-order conditions also imply a league equilibrium independent of α.

Hence, treating players as assets for teams does not alter the ineffectiveness of revenue sharing in a two team league with a fixed talent supply and profit maximizing teams.

Transfer payment sharing with long-term contracts

However, the result may be changed if revenue sharing is replaced by *transfer payment sharing* only. Suppose that a team *s* selling talent obtains a share of α of the transfer payment, share (1-α) of the payment flows into the sharing pool and is distributed equally among the teams. Team *s* sells $(t_{i,2} - t_{i,1})$ units of talent at price p in a two-team league and earns the following payment after the redistribution of transfer payments:

$$TP_s = p\left[\alpha(t_{s,1} - t_{s,2}) + \frac{1-\alpha}{2}(t_{s,1} - t_{s,2})\right] = p\frac{1+\alpha}{2}(t_{s,1} - t_{s,2}).$$

For a team *b* buying the respective units of talent, the net transfer payment with transfer revenue sharing is:

$$TP_b = -p\left[(t_{b,2} - t_{b,1}) - \frac{1-\alpha}{2}(t_{s,1} - t_{s,2})\right] = -p\frac{1+\alpha}{2}(t_{b,2} - t_{b,1}).$$

In the model applied here, the respective equations (22) and (23) change for both teams to:

$$t_{i,1}: \ R'_{i,1} + r\,\theta'_i = \left(c - \frac{1+\alpha}{2} r\,p\right)T, \tag{24}$$

$$t_{i,2}: \ R'_{i,2} = \left(c + \frac{1+\alpha}{2} p\right)T. \tag{25}$$

The intertemporally optimal decision rule is then:

$$R'_{i,2} - R'_{i,1} = r\theta'_i + \frac{1+\alpha}{2} pT(1 + r). \tag{26}$$

Proposition 3: Linear transfer payment sharing in a pool system changes the intertemporal demand for talent within teams for $0 < \alpha < 1$. With concave revenue functions, more talent is demanded in the second period, relative to the talent demand in the first period.

Proof: Since the right side of equation (26) decreases due to transfer payment sharing (because of (1+α)/2 < 1), talent demand is reallocated from period 1 to period 2. //

Proposition 4: The allocation of talent among teams is not changed by linear transfer payment sharing.

Proof: A two-team league equilibrium with transfer payment sharing derived from (24) and (25) requires: $R'_{i,1} + r\,\theta'_i = R'_{j,1} + r\,\theta'_j$ for the first period and $R'_{i,2} = R'_{j,2}$ for the second.

Both equilibrium conditions are independent of the sharing parameter α. Hence, transfer payment sharing has no effect on talent allocation among teams in a league. //

3 Players as assets: Specified model

In the following analysis, a specified version of the general model is examined for a two-period two-team league.

Profits of each team i = 1,2 are given by:

$$\pi_i = \pi_{i,1} + r\,\pi_{i,2}. \tag{28}$$

As before, revenues are $R_{i,j}$, j = 1,2 as periods and costs are denoted by $C_{i,j}$. With parameter $\theta_i \neq 0$, teams are able to earn more (less) profit in period 2 if they were more (less) successful in period 1. Future revenues are discounted as before by the factor r = 1/(1+ρ) (with ρ as the discount rate).

The valuation function f_i is given by:

$$f_i(w_{i,1}) = \theta_i \cdot \left(w_{i,1} - 0.5\right). \tag{29}$$

According to equation (29), a team's value depends on a valuation parameter θ_i, as well as on the difference in the team's own winning probability $w_{i,1}$, realized in the first period, minus the average winning probabilities over all teams, given by: $\frac{1}{n}\sum_i w_{i,1} = \frac{1}{n}$. For two teams, the average winning probability is obviously 0.5.

The valuation parameter θ_i determines the 'goodwill' a team acquires during the first period. Notice that the valuation parameter, as well as its revenue potential, is team-specific. From equation (29), it follows that the value of a team may increase or decrease, depending on the difference between the team's own winning probability and the average winning probability of all teams. The team's stock value increases (decreases) if its winning probability is higher (lower) than that of all teams together. Although a great variety of valuation functions is conceivable, the valuation function defined by equation (29) seems to be sufficient for this paper.

The revenue earned by team *i* in period *j* is given by the following specification (see, for instance, Vrooman, 2007):

$$R_{i,j} = m_i w_{i,j} - \frac{\beta}{2} w_{i,j}^2. \tag{30}$$

with *m* as the market size, *w* as the winning probability, and β as a fan preference parameter for competitive balance. The cost function $C_{i,j}$ is assumed to be linear with constant marginal costs c_j for each unit of talent, the usual assumption in sports economics.

3.1 Optimizing talent employment in both periods

The (discounted) total profit of team *i* is given by:

$$\pi_i = R_{i,1} - C_{i,1} + r\,(R_{i,2} - C_{i,2} + f_i). \tag{31}$$

Inserting the specifications defined by equations (29) and (30), the profit maximization program is as follows:

$$\max_{t_{i,1},t_{i,2}} \pi_i = m_i w_{i,1} - \frac{\beta}{2} w_{i,1}^2 - C_{i,1} + r\left[m_i w_{i,2} - \frac{\beta}{2} w_{i,2}^2 - C_{i,2} + \theta_i \cdot \left(w_{i,1} - 0.5\right)\right]. \tag{32}$$

Together with the contest success function (3), the two first-order conditions are:

$$t_{i,1}: m_i - \beta\, w_{i,1} + r\,\theta_i = c_1\, T \tag{33}$$

$$t_{i,2}: m_i - \beta\, w_{i,2} = c_2\, T. \tag{34}$$

Together with the adding-up constraints, the talent employment optimum and, thus the winning percentages are given by:

$$w_{1,1} = \frac{\mathrm{m}_1 - \mathrm{m}_2 + \beta + r\,(\theta_1 - \theta_2)}{2\beta} \tag{35}$$

and

$$w_{2,1} = \frac{\mathrm{m}_2 - \mathrm{m}_1 + \beta + r\,(\theta_2 - \theta_1)}{2\beta} \tag{36}$$

for the first period and

$$w_{1,2} = \frac{m_1 - m_2 + \beta}{2\beta} \tag{37}$$

and

$$w_{2,2} = \frac{m_2 - m_1 + \beta}{2\beta} \tag{38}$$

for the second period.

Let the 'goodwill' parameters be equal for both teams, $\theta_1 = \theta_2$. It then holds that:

$$w_{1,1} = \frac{m_1 - m_2 + \beta}{2\beta} \tag{39}$$

and

$$w_{2,1} = \frac{m_2 - m_1 + \beta}{2\beta} \tag{40}$$

in the first period.

Hence, in this case, the valuation of earlier success has no influence on the allocation of talent. Differing 'goodwill' parameters imply that equal success may have different effects on the valuation of the stock of players. If such a difference occurs, talent allocation will adapt accordingly and talent allocation is changed by 'goodwill'.

In addition, the valuation of a team's talent stock will alter marginal costs. In a closed league, marginal talent costs are endogenous. In the model specified here, marginal talent costs are given by:

$$c_1 = \frac{m_1 + m_2 - \beta + r\,(\theta_1 + \theta_2)}{2T} \tag{41}$$

and

$$c_2 = \frac{m_1 + m_2 - \beta}{2T}. \tag{42}$$

If both stock valuation parameters are equal, i.e., $\theta_1 = \theta_2$, marginal costs are the same for both periods. Nevertheless, even if the parameters are the same, there is an economic effect, because the valuation parameters enter the winning probabilities as differences, whereas marginal costs are incorporated as a sum. Therefore, a higher marginal valuation of talent in the second period implies lower profits in the first period. The 'goodwill' value of the teams' stock of players increases the payments to players in Period 1, as a result of the more intense competition for talent. Hence, low profits in one period do not necessarily imply ‘overinvestment’ or ‘non-profit-maximizing behavior’, as suggested in the sports economics literature (Whitney, 1993; Dietl and Franck, 2000; Dietl, Franck and Roy, 2003; Dietl, Franck and Lang, 2008; Grossmann and Dietl, 2009; Lang, Grossmann and Theiler, 2011).

3.2 Talent reallocation in the second period

If teams buy talent for both periods and trade them before the second period, the maximization program is given by:

$$\max_{t_{i,1},t_{i,2}} \pi_i = R_{i,1} - C_i + r\left[R_{i,2} - C_i + \theta_i\left(w_{i,1} - 0.5\right) + p\left(t_{i,2} - t_{i,1}\right)\right]. \quad (43)$$

In combination with the contest success function in (3), the two first-order conditions are:

$$t_{i,1}: m_i - \beta\, w_{i,1} + r\,\theta_i = (c - r\,p)\,T \quad (44)$$

and

$$t_{i,2}: m_i - \beta\, w_{i,2} = (c + p)\,T. \quad (45)$$

The respective winning percentages are given by:

$$w_{1,1} = \frac{\mathrm{m}_1 - \mathrm{m}_2 + \beta + r\,(\theta_1 - \theta_2)}{2\beta} \quad (46)$$

and

$$w_{2,1} = \frac{\mathrm{m}_2 - \mathrm{m}_1 + \beta + r\,(\theta_2 - \theta_1)}{2\beta} \quad (47)$$

for the first period and

$$w_{1,2} = \frac{\mathrm{m}_1 - \mathrm{m}_2 + \beta}{2\beta} \quad (48)$$

and

$$w_{2,2} = \frac{\mathrm{m}_2 - \mathrm{m}_1 + \beta}{2\beta} \quad (49)$$

for the second period.

Consequently, the allocation of talent does not differ from that of the previous section 3.1. If the valuation parameters are the same for both teams, the allocation of talent will be the same for both periods and thus no trading of talent will occur at the start of Period 2. The value change due to former success changes the allocation of talent, whereas trading talent for money does not. However, trading talent changes the distribution of profits, as the marginal talent costs c are given by:

$$c = \frac{m_1 + m_2 - \beta(1+r) + r(m_1 + m_2 + \theta_1 + \theta_2)}{2(1+r)T}. \qquad (50)$$

The price p for a unit of transferred talent results in:

$$p = \frac{r(\theta_1 + \theta_2)}{2(1+r)T}. \qquad (51)$$

Finally, the profits of teams with, as well as without talent transfers between teams (sections 3.1 and 3.2, respectively), are compared (Table 1). Since the talent allocations are the same with and without transfer payments, the profit differences are determined by the talent cost differences.

Table 1: Cost (profit) differences between the model with and without players transfer payments

	Costs model in 3.1	Costs model in 3.2	Cost difference (model 3.1 – model 3.2)
Period 1	$c_1 t_{i,1}$	$c\ t_{i,1}$	$(c_1 - c\ )t_{i,1} = \frac{r^2(\theta_1+\theta_2)}{2(1+r)T} t_{i,1} > 0$
Period 2	$c_2 t_{i,2}$	$c\ t_{i,2} + p\left(t_{i,2} - t_{i,1}\right)$	$-\frac{r(\theta_1+\theta_2)}{2(1+r)T} t_{i,2} - \frac{r(\theta_1+\theta_2)}{2(1+r)T}\left(t_{i,2} - t_{i,1}\right) < 0$
Total (Period 1 + Period 2)	$c_1 t_{i,1} + r\ c_2 t_{i,2}$	$c\ t_{i,1} + r\ c\ t_{i,2} + r\ p\left(t_{i,2} - t_{i,1}\right)$	$t_{i,1}(c + rp - c_1) - t_{i,2}(rp - rc + rc_2) = 0$

Although a talent transfer system changes neither the allocation of talent nor total profits in a two-period model, it nevertheless changes the period in which profits are earned. With a transfer system, profits are higher in the first period.

4 Discussion

In this paper, the asset value of the stock of players is incorporated into a standard model of sport economics. This approach requires an intertemporal model instead of the usual one-period model. In contrast to the results from the standard approach (Dietl, Franck and Roy, 2003; Dietl, Franck and Lang, 2008; for overinvestments in talent with profit maximization, see Whitney, 1993; Dietl and Franck, 2000), in our approach no 'overinvestment' could be found. Especially the usual result that winning maximization - in contrast to profit maximization - is prone to overinvestment in talent does not to be valid. From an observational viewpoint, large investments in talent – even with losses being incurred – cannot be attributed to an objective maximizing wins. Hence, winning and profit maximization are not easily distinguishable from one another empirically.

In contrast to overinvestment, player hoarding may occur in sport teams. It should be emphasized, however, that overinvestment and player hoarding are different team strategies. Player hoarding is a strategy which is similar to a well-known industry strategy of 'raising rivals' costs'. In a sport league, dominant teams may earn so much money that they can buy talent that is not even employed in matches. To prevent such a strategy, closed U.S. leagues restrict the number of players a team is allowed to employ. In contrast, European professional soccer leagues have not imposed such restrictions so far.

Investing in players as assets seems to be a preferred strategy especially in imperfect capital markets. In perfect capital markets, the risks of playing in a league might be covered in various different ways that do not require players as assets. Moreover, financing sport teams with loans or issuing shares would be perfect substitutes. In imperfect capital markets, treating players as assets might be necessary to finance a team; in this case, players could be seen as collaterals for risky credits. Since credits and equity are no longer perfect substitutes, credits will be preferred. Material collaterals that can be sold are necessary in order to secure credits. Therefore, players as assets may emerge over time because of capital market imperfections.

5 Conclusions

In this paper, players of sport teams are modeled as assets. In addition to the usual approach in sport economics, an intertemporal model of sport teams is employed to analyze the consequences of players as assets in team sports.

The most important result is that investing in talented players should not generally be dubbed 'overinvestment', since player assets can be sold, if necessary. Although such approaches are quite common in industrial economics, they are new in sport economics. Revaluations of assets do indeed have consequences for the employment of talented players. In such a situation, winning teams may, for instance, achieve an increase in their team value and thus invest optimally through buying more talent, as indicated by Tobin's q-theory of investment (Tobin, 1969; Hayashi, 1982; Abel et al., 1996; Adda and Cooper, 2003, pp. 187 ff.).

References

Abel, A. B., A. K. Dixit, J. C. Eberly, R. S. Pindyck (1996), Options, the value of capital, and investment, *Quarterly Journal of Economics*, Vol. 111, No. 3, pp. 753-777.

Adda, J., R. Cooper (2003), Dynamic Economics [Investment, pp. 187-213], The MIT Press: Cambridge, Massachusetts, London, England.

Baetge, J., H. Klönne, C.-P. Weber (2013), Möglichkeiten und Grenzen einer objektivierten Spielerbewertung im Profifußball, *Discussion Paper*, WWU Münster (submitted to KoR IFRS).

Dietl, H. M., E. Franck (2000), Effizienzprobleme in Sportligen mit gewinnmaximierenden Kapitalgesellschaften – eine modelltheoretische Untersuchung, *Zeitschrift für Betriebswirtschaft*, Vol. 70, pp. 1157–1175.

Dietl, H. M., E. Franck, M. Lang (2008), Overinvestment in team sports leagues: A contest theory model, *Scottish Journal of Political Economy*, Vol. 55, No. 3, pp. 353-368.

Dietl, H. M., E. Franck, P. Roy (2003), Überinvestitionsprobleme in einer Sportliga, *Betriebswirtschaftliche Forschung und Praxis*, Vol. 55, No. 5, pp. 528–540.

Fort, R., J. Winfree (2009), Sports really are different: The contest success function and the supply of talent, *Review of Industrial Organization*, Vol. 34, No. 1, pp. 69-80.

Grossmann, M. H. Dietl (2009), Investment behavior in a two period contest model, *Journal of Institutional and Theoretical Economics*, Vol. 165, No. 3, pp. 401-417.

Hayashi, F. (1982), Tobin's marginal q and average q: A neoclassical interpretation, *Econometrica*, Vol. 50, No. 1, pp. 213-224.

Hirshleifer, J. (1989), Conflict and rent-seeking success functions: Ratio vs. difference models of relative success, *Public Choice*, Vol. 63, No. 2, pp. 101–112.

Lang, M., M. Grossmann, P. Theiler (2011), The sugar daddy's game: How wealthy investors change competition in professional team sports, *Journal of Institutional and Theoretical Economics*, Vol. 167, No. 4, pp. 557–577.

Marburger, D. R. (1997), Gate revenue sharing and luxury taxes in professional sports, *Contemporary Economic Policy*, Vol. 15, No. 2, pp. 114-123.

Skarperdas, S. (1996). Contest success functions, *Economic Theory*, Vol. 7, No. 2, pp. 283–290.

Tobin, J. (1969), A general equilibrium approach to monetary theory, *Journal of Money, Credit and Banking*, Vol. 1, No. 1, pp. 15–29.

Tullock, G. (1980), Efficient rent-seeking. In: J. M. Buchanan, R. D. Tollison, G. Tullock (Eds.), *Toward a Theory of the Rent-Seeking Society*, Texas A& M University Press, College Station, TX.

Van der Burg, T., A. Prinz (2005), Progressive taxation as a means for improving competitive balance, *Scottish Journal of Political Economy*, Vol. 52, No. 1, pp. 65-74.

Vrooman, J. (2007), Theory of the beautiful game: The unification of European football, *Scottish Journal of Political Economy*, Vol. 54, No. 3, pp. 314-354.

Whitney, J. D. (1993), Bidding till bankrupt: Destructive competition in professional team sports, *Economic Inquiry*, Vol. 31, No. 1, pp. 100-115.

3 Evolutionarily Stable Strategies in Sports Contests: An Alternative Approach

Martin Langen, Aloys Prinz

Abstract: Grossmann (2013) presented a model of a sports league with an evolutionary stable strategy (ESS) equilibrium that differs from the Nash solution (NS). In the ESS equilibrium teams invest more in talent and gain lower profits than in the respective Nash equilibrium. In this note the seemingly contradiction between NS and ESS is clarified by applying the so-called concept of an "optimal aggregate-taking strategy (ATS)" introduced by Alós-Ferrer and Ania (2005). Furthermore, the general result of Grossmann (2013) is qualified.

Keywords: sports leagues, Nash equilibrium, evolutionary stable strategies, optimal aggregate-taking strategies, rent overdissipation

JEL classification: L83, C72, C73

1 Introduction

In Grossmann (2013) the existence of an evolutionary stable strategy (ESS) for sport teams is proposed that represents an equilibrium that is different from the Nash solution (NS) for a sport league. In this ESS teams invest more in talent and gain lower profits. This result is somewhat astounding since an ESS is commonly interpreted as a Nash refinement, i.e., an ESS should be also a Nash equilibrium. In this note it is intended to clarify the seemingly contradiction by applying the so-called “optimal aggregate-taking strategy (ATS)” introduced by Alós-Ferrer and Ania (2005). Moreover, the general result of Grossmann (2013) is qualified.

The rest of the note is structured as follows: First, the Grossmann (2013) model is briefly presented. In the third section, the ATS approach is applied on the teams’ investment decisions in a sport league. The fourth section concludes.

2 Grossmann’s approach

In the Grossmann model $n \geq 2$ symmetric teams compete to win a contest. Profits for team i are the difference between revenues R_i and costs C_i:

$$\pi_i(x_1, \dots, x_n) = R_i(m_i, x_1, \dots, x_n) - C_i(x_i). \quad (1)$$

Teams invest x_i in order to increase their winning probability p_i that are given by a Tullock contest success function (CSF):

$$p_i(x_1, \dots, x_n) = \frac{x_i^r}{\sum_{j=1}^{n} x_j^r}. \quad (2)$$

Revenues are specified in a usual way by:

$$R_i(m_i, x_1, \dots, x_n) = m_i p_i - \frac{b}{2} p_i^2.$$

Costs are given by a convex cost function with constant marginal costs, c. The parameter b represents the effect of competitive balance on the demand for the sport event and m_i represents the market size of a team i. Since all teams are identical market sizes are equal: $m_i = m \; \forall i = 1, \dots, n$.

The objective of team owners is to maximize their utility, which consists of a weighted sum of profits and winning probability. The weight for profits is α and for winning percentages $1-\alpha$. A team owner's utility is given by:

$$u_i = \alpha\,\pi_i + (1-\alpha)p_i = \alpha(R_i - c\,x_i) + (1-\alpha)p_i. \qquad (3)$$

Nash equilibrium

Clubs maximize (3) with respect to x_i; this gives *n* first-order conditions (FOCs):

$$\frac{\partial u_i}{\partial x_i} = \alpha\left(m\,\frac{\partial p_i}{\partial x_i} - b\,p_i\frac{\partial p_i}{\partial x_i} - c\right) + (1-\alpha)\frac{\partial p_i}{\partial x_i} = 0. \qquad (4)$$

In a symmetric solution, $x_i = x_j = x$ for $j = 1,..,n \neq i$, $p_i = \frac{1}{n}$ and $\frac{\partial p_i}{\partial x_i} = \frac{r(1-n)}{n^2 x}$ hold. Therefore (4) can be reduced to the symmetrical Nash equilibrium investment per team as:

$$x_{NS} = \frac{r(n-1)}{n^2\alpha\,c}\left(\alpha\left(m-\frac{b}{n}\right) + (1-\alpha)\right). \qquad (5)$$

In NS, all clubs invest x_{NS} in talent and no club has an incentive to deviate from (5) as it is the best response to itself.

ESS

To derive the evolutionary stable strategy ESS, Grossmann (2013) considers team 1 as a representative team. Under ESS, this team tries to maximize its *relative* utility compared to their opponent $j \neq 1$. Suppose that all other clubs choose x_{ESS}. Team 1's optimal choice of x_1 must be equal to x_{ESS} in the following maximization program for x_{ESS} being an ESS:

$$\max_x U_1(x, x_{ESS}) \equiv u_1(x, x_{ESS}, \dots, x_{ESS}) - u_j(x, x_{ESS}, \dots, x_{ESS}) \text{ with } j \neq 1. \qquad (6)$$

The FOC for this maximization program reads:

$$\frac{\partial U_1}{\partial x} = \alpha\left(m\frac{\partial p_1}{\partial x} - b\,p_1\frac{\partial p_1}{\partial x} - c\right) + (1-\alpha)\frac{\partial p_1}{\partial x} - \left(\alpha\left(m\,\frac{\partial p_{ESS}}{\partial x} - b\,p_{ESS}\frac{\partial p_{ESS}}{\partial x}\right) + (1-\alpha)\frac{\partial p_{ESS}}{\partial x}\right) = 0. \qquad (7)$$

In a symmetric solution $p_1 = p_{ESS} = \frac{1}{n}$ holds true, thus (7) can be simplified to:

$$\left(\frac{\partial p_1}{\partial x} - \frac{\partial p_{ESS}}{\partial x}\right)\left(\alpha\left(m-\frac{b}{n}\right) + (1-\alpha)\right) - \alpha\,c = 0.$$

Likewise, due to symmetry, $\frac{\partial p_1}{\partial x} = \frac{r(n-1)}{n^2 x}$ and $\frac{\partial p_{ESS}}{\partial x} = -\frac{r}{n^2 x}$; this implies the ESS solution:

$$x_{ESS} = \frac{r}{n\alpha c}\left(\alpha\left(m - \frac{b}{n}\right) + (1-\alpha)\right) = \frac{r}{cn^2}\left(n\left(m - 1 + \frac{1}{\alpha}\right) - b\right). \qquad (8)$$

The first expression on the right-hand side of equation (8) shows that x_{ESS} and x_{NS} are similar, but not identical (whereby the second expression on the right-hand side of equation (8) is Grossmann's (2013) solution). As a consequence, x_{ESS} and x_{NS} are different, implying that the evolutionary stable strategy equilibrium is not a Nash equilibrium.

In order to see the relation between NS and ESS as well as their divergence for finite populations, the following simplification is applied.

The maximization of relative utility (6) can be reformulated to:

$$\max_x U_1(x, x_{ESS}) \equiv u_1(x, x_{ESS}, \dots, x_{ESS}) - \frac{1}{n}\sum_{j=1}^{n} u_j(x, x_{ESS}, \dots, x_{ESS}). \qquad (9)$$

With $\sum_{j=1}^{n} p_j = 1$, $\sum_{j=1}^{n} \frac{\partial p_j}{\partial x_i} = 0$ and the symmetrical equilibrium condition $w_j = \frac{1}{n}$ for $j = 1, .., n$, the FOC for the maximization program in (9) reads:

$$\frac{\partial U_1}{\partial x_1} = \frac{\partial u_1}{\partial x_1} + \frac{\alpha\, c}{n} = 0. \qquad (10)$$

According to equation (10), the FOC for an evolutionary stable strategy is given by the sum of the FOC for a Nash strategy plus a constant term. Inserting equation (4) into equation (10) and rearranging terms gives again the ESS solution:

$$x_{ESS} = \frac{r(n-1)}{n^2 \alpha\, c\left(1-\frac{1}{n}\right)}\left(\alpha\left(m - \frac{b}{n}\right) + (1-\alpha)\right) = \frac{r}{n\alpha\, c}\left(\alpha\left(m - \frac{b}{n}\right) + (1-\alpha)\right). \qquad (11)$$

The expression 1/n in equation (11) results from the constant term. As is easy to see, for an infinite population ESS and NS coincide.

3 An alternative approach

Although Grossmann's solution is new in sports economics, it is not new in game theory (see the literature quoted by Grossmann). Moreover, more recent game theoretic literature provides a tool to present Grossmann's results by an alternative approach that is more rigorous, the so-called "optimal aggregate-taking strategy" (ATS), introduced by Alós-Ferrer and

Ania (2005). ATS is generally applicable in games where the payoff of one player depends on its own strategy choice as well as on an aggregate of all players' choices. The main reasons to apply this concept are first that an ATS is in certain cases more comprehensive than ESS, i.e., "a strict ATS is a sufficient condition for globally stable ESS" (Alós-Ferrer and Ania, 2005, p. 507) and second that an ATS is much easier intuitively to understand than ESS. Although Grossmann's approach (based on Schaffer, 1988) is mathematically sound, it seems not so at first glance. The reason is that the impression is generated as if another utility function is maximized to get the ESS strategy equilibrium than a Nash strategy one. As Ania (2008) pointed out, the difficulty is that a pure strategy Nash equilibrium is not ESS with a finite population playing the respective game, but that also the ESS equilibrium is not a Nash equilibrium (and, hence, might be itself beaten). Although this is a well-known result of game theory (see, e.g., Gintis, 2009, p. 240) it can be confusing nevertheless.

The above mentioned ATS may help to understand better the difference between Nash and ESS equilibria in games with finite populations as in a sport league. The formalization of Grossmann's investment in talent of sport leagues teams with an ATS reads as follows (see Alós-Ferrer & Ania, 2005, p. 507):

$$\max_{\mathrm{x_i}} u_i\big(x_i, g(x^*, \ldots, x^*)\big) = \alpha(R_i - cx_i) + (1-\alpha)p_i$$

$$= \alpha \cdot \left[m \cdot \left(\frac{x_i^r}{\sum_{j=1}^n {x^*}_j^r}\right) - \frac{b}{2}\left(\frac{x_i^r}{\sum_{j=1}^n {x^*}_j^r}\right)^2\right] + (1-\alpha)\left(\frac{x_i^r}{\sum_{j=1}^n {x^*}_j^r}\right) - \alpha c x_i. \qquad (12)$$

Equation (12) transforms the strategic choice of team i into a game where the team chooses its talent investment by supposing that all other teams will not change their choices due to the choice of its talent investment. This means that team i is playing against all other teams at once instead of team-by-team basis concerning talent investment. This can most easily be recognized by the winning probability p_i (which is given by a Tullock contest success function as in Grossmann, 2013):

$p_i = \left(\frac{x_i^r}{\sum_{j=1}^{n} {x^*}_j^r} \right).$

The numerator represents team i's choice of talent investment whereas the denominator represents the sum of all teams' optimal talent investments, in other words, the aggregate of the game. Note that the latter is *not* influenced by the talent investment of team *i*. The first-order condition for the maximization program in equation (9) reads:

$$\frac{\partial u_i(\ldots)}{\partial x_i}_{x_i = x^*} = \alpha \left[\frac{r}{n x_i} \left(m - \frac{b}{n} \right) \right] + (1 - \alpha) \frac{r}{n x_i} - \alpha c = 0 \, .$$

Therefore:

$$x^* = x_i = \frac{r}{\alpha c n} \left[\alpha \left(m - \frac{b}{n} \right) + (1 - \alpha) \right], \quad n \cdot x^* = r \left[\frac{\alpha \left(m - \frac{b}{n} \right) + (1 - \alpha)}{\alpha c} \right]. \quad (13)$$

These are the same values as in Grossmann (2013), but derived with a shortcut. However, note that this value of x^* is a globally stable ESS for *strict* ATS only. ATS is strict, if the $u_i(\ldots)$ function in equation (12) is strictly concave in x_i (Alós-Ferrer and Ania, 2005, p. 507), i.e. if:

$$\frac{\partial^2 u_i\left(x_i, g(x^*, \ldots, x^*)\right)}{\partial x_i^2} < 0.$$

Since

$$\frac{\partial^2 u_i\left(x_i, g(x^*, \ldots, x^*)\right)}{\partial x_i^2} = \frac{r x_i^{r-2}}{n (x^*)^r} \left\{ (r-1)[\alpha(m-1) + 1] - \frac{(2r-1)\alpha b x_i^r}{n (x^*)^r} \right\}, \quad (14)$$

the sign of the second derivative depends on the term in the $\{\ldots\}$ brackets:

$$sign \, \frac{\partial^2 u_i\left(x_i, g(x^*, \ldots, x^*)\right)}{\partial x_i^2} = sign \left\{ (r-1)[\alpha(m-1) + 1] - \frac{(2r-1)\alpha b x_i^r}{n (x^*)^r} \right\}. \quad (15)$$

Hence, the sign depends on:

$$\Delta = \frac{n (x^*)^r [\alpha(m-1) + 1] - 2\alpha b x_i^r}{n (x^*)^r}, \quad (16)$$

and

$$\Gamma = \frac{n (x^*)^r [\alpha(m-1) + 1] - \alpha b x_i^r}{n (x^*)^r}. \quad (17)$$

Inserting Δ from equation (16) and Γ from equation (17) into equation (15) yields:

$$sign\ \frac{\partial^2 u_i(x_i, g(x^*, \dots, x^*))}{\partial x_i^2} = sign\ [r\,\Delta - \Gamma]. \tag{18}$$

Note the Γ is larger than zero for $x_i^* > 0$ in any equilibrium, which follows from (13), whereas the sign of Δ may be larger or smaller than zero. Hence the sign of the second derivative of the utility function in the game context is defined as follows:

$$\frac{\partial^2 u_i(x_i, g(x^*, \dots, x^*))}{\partial x_i^2} < 0 \tag{19}$$

if:

$(a)\ \Delta \leq 0: r > 0,$

$(b)\ \Delta > 0: 0 < r < \frac{\Gamma}{\Delta}.$

Consequently, the general result of Grossmann (2013) holds true only in case (a). In case (b), however, it cannot be said that x_i^* is ESS for *all* $r > 0$. This is an important restriction of the generality of Grossmann's result even in the case of a specified revenue function.

These results are summarized in the following proposition:

Proposition 1: The *optimal aggregate-taking strategy* ATS given by $x^* = x_i = \frac{r}{\alpha cn}\left[\alpha\left(m - \frac{b}{n}\right) + (1-\alpha)\right]$ is a globally stable ESS for $\Delta \leq 0$ and $r > 0$ or for $\Delta > 0$ and $r < \Gamma/\Delta$, respectively. For $\Delta > 0$ and $r \geq \Gamma/\Delta$, it cannot be said generally whether x^* is an ESS.

Proof: See above.//

Remark: Given $\Delta > 0$, it follows $\Gamma/\Delta > 1$ by comparing equations (16) and (17). Hence, $1 < r < \Gamma/\Delta$. This means that x^* is an ESS even if the Tullock contest success function exhibits increasing returns, up to a certain extent.

Rent overdissipation

Grossman defined the rent dissipation D as the ratio of the sum of investments $\sum_{i=1}^{n} x_i$ in equilibrium and the total sum of expected rents in the league $\sum_{i=1}^{n} p_i m = m$. For two reasons this definition of total expected rents in the league is a bit surprising. First, the competitive balance parameter b does not enter the equation. As a consequence, the total sum of expected rents in the league (which Grossman, 2013, uses as an equivalent to the feasible revenues) is higher than the maximally feasible revenue of the leagues, except for $b = 0$. Second, the maximum of the expected rents is also independent of α. For $\alpha < 1$ teams generate utility from revenues as well as from winning percentages. Nevertheless, the total sum of talent investments is compared with feasible revenues only for the analysis of rent dissipation. Furthermore, another crucial aspect is the cost function for talent. In the Grossmann (2013) model, teams invest an amount x_i in order to increase their winning percentage; hence their costs are exactly x_i. The price for investing an additional unit of x_i is then exactly 1; moreover, there is no justification in the paper for unit costs of $c \neq 1$ for talent. In this setup the possibility of overdissipation does not seem surprising. To analyze rent dissipation carefully, we stick in the following to Grossmann's definition of rent-dissipation, D, but consequently we set $\alpha = 1$, $c = 1$ and $b = 0$.

Nash equilibrium

Grossman stated that even for teams attempting to maximize individual profits, overdissipation is possible. This is not correct, as already shown in, e.g., Hehenkamp et al. (2004).

For $\alpha = 1$, $b = 0$ and $c = 1$, the optimal Nash investment strategy of a team is given by:

$$x_{NS} = \frac{r(n-1)}{n^2} m. \tag{20}$$

The profits of team I read:

$$\pi_i = m\, p_i - x_i. \tag{21}$$

In the symmetrical league equilibrium, each team invests the same amount and has the same winning probability. For x_{NS} to be a Nash equilibrium, profits must not be negative because otherwise either one team in a general model or each team in a symmetrical model is better off when investing nothing. It follows that:

$$\pi_i \geq 0 \Leftrightarrow \frac{m}{n} \geq \frac{r(n-1)}{n^2} m \Leftrightarrow r \leq \frac{n}{n-1}. \tag{22}$$

Hence the necessary condition for the existence of a unique Nash equilibrium in pure strategies is $r \leq \frac{n}{n-1}$.

Rent overdissipation requires (see also Grossmann, 2013):

$$D = \frac{\sum_{i=1}^{n} x_i}{m} > 1 \Leftrightarrow n\, x_{NS} > m \Leftrightarrow r > \frac{n}{n-1}. \quad (23)$$

Comparing (23) and (22) reveals that rent overdissipation is not compatible with a unique Nash equilibrium in pure strategies if teams are profit maximizers.

ESS

For $\alpha = 1$, $b = 0$ and $c = 1$, it follows form equation (16) $\Delta = m$ and from equation (17) $\Gamma = m$ in an ESS of talent investment. However, from (18) $r < 1$ is required for a globally stable ESS under these circumstances. This condition results also from the non-negativity of profits:

$$\frac{m}{n} > x_{ess} \Leftrightarrow r < 1. \quad (24)$$

Rent overdissipation requires (see also Grossmann, 2013):

$$n\, x_{ess} > m \Leftrightarrow n\frac{r}{n}m > m \Leftrightarrow r > 1. \quad (25)$$

As a consequence, rent overdissipation is not feasible in a globally stable talent investment ESS.

Proposition 2: For $\alpha = 1$, $b = 0$ and $c = 1$, rent overdissipation is neither feasible in a unique Nash equilibrium of talent investment nor in a globally stable ESS.

Proof: Compare (23) and (25) with (22) and (24), respectively. //

As a consequence, there will be no rent overdissipation, no matter whether the league teams play Nash or evolutionary stable strategies. However, as is the case in Grossman (2013), there might be rent overdissipation when teams maximize utility and ignore the zero-profit restriction.

4 Conclusion

Although it is well-known that finite populations may render evolutionary stable strategies (ESS) non-Nash, these strategies themselves are not always unbeatable. To shed more light on the problem of how to interpret Nash and ESS strategies, Alós-Ferrer and Ania (2005) proposed the concept of an *optimal aggregate-taking strategy* (ATS). An ESS is determined by maximizing *relative* performance with respect to opponents which may create the impression as if a different objective function is maximized in contrast to Nash strategies. To overcome this impression, an ATS is defined as the maximization of the own objective function under the assumption that one's own decision does not change the aggregate decisions of all players. In this way the difficulty implied by the ESS approach is overcome. In addition to that, a strict ATS is sufficient for a globally stable ESS as proven by Alós-Ferrer and Ania (2005).

In this note, the ATS approach is employed to check the validity of Grossmann's (2013) results for team talent investments in a sports league context. Although the main results of Grossmann (2013) could be verified with the ATS approach, a restriction of the generality of the results is also shown. As a consequence, it might be not only easier to calculate ATS in a league competition context than ESS, but also clearer to interpret as well as to determine the global stability of the implied ESS.

Moreover, for profit-only motivated teams and for specifications of the revenue function ($b = 0$) and the cost function ($c = 1$) it is shown that a unique Nash equilibrium in pure strategies as well as a globally stable ESS is not compatible with rent overdissipation.

References

Alós-Ferrer, C., & A. B. Ania (2005), The evolutionary stability of perfectly competitive behavior, *Economic Theory,* Vol. 26, No. 3, pp. 497-516.

Ania, A. B. (2008), Evolutionary stability and Nash equilibrium in finite populations, with an application to price competition, *Journal of Economic Behavior & Organization,* Vol. 65, No. 3-4, pp. 472-488.

Gintis, H. (2009), *Game theory evolving*, 2nd ed., Princeton, New Jersey, Princeton University Press.

Grossmann, M. (2013), Evolutionarily stable strategies in sports contests, *Journal of Sports Economics (forthcoming),* DOI: 10.1177/ 1527002512470957.

Hehenkamp, B., W. Leininger, A. Possajennikov (2004), Evolutionary equilibrium in Tullock contests: Spite and overdissipation, *European Journal of Political* Economy, Vol. 20, No. 4, pp. 1045-1057.

Schaffer, M. (1988), Evolutionarily stable strategies for a finite population and a variable contest size, *Journal of Theoretical* Biology, Vol. 132, No. 4, pp. 469–478.

4 Regional Competition and Competitive Balance in Sports Leagues

Martin Langen

Abstract: A well-known finding in of sport economics is that differences in the winning probabilities of sport teams are determined largely by unequal market sizes and, therefore, revenue potential. Larger markets yield higher marginal revenue and thus, more units of talent being employed by the respective teams. In contrast to the existing literature, this paper presents a model with endogenous market sizes. Teams are able to choose their location at which they are based, implying a direct effect on their own, as well as on their opponent's potential market size. This regional competition in sport leagues is analyzed in terms of overall payoff, competitive balance, and effectiveness of revenue sharing agreements for profit-, as well as for win-maximizing team behavior. It can be shown that the usually assumption of a strictly positive correlation between market size and success does not hold in general. Furthermore, evidence relating to major European football leagues is presented, indicating that a larger market does not necessarily imply more successful teams.

JEL: L83, C72, D43

Keywords: Regional competition, competitive balance, European football

1 Introduction

Sports teams compete in leagues in which some teams are distinctly more successful than others. The fact that sport teams are based in separate regional markets is considered to be the most important reason. These markets offer different revenue potential, which leads to different marginal revenue per unit of talent employed by the respective teams. Thus, they hire different amounts of talent which results in differences of sporting success. But taking the markets for granted seems counter-intuitive, as teams or at least the player talent, is highly mobile. In contrast to existing models of sports leagues, the approach of this paper is to allow for endogenous markets. Teams are able to choose their location, so that a larger market not only increases the marginal revenue per unit of talent, but also the likelihood of competitors entering this market. Under these circumstances, the strict relationship between market size and success no longer holds.

Rottenberg (1956) was the first to analyze the competitive outcome of a league. He showed that the allocation of talent in a league is independent from the distribution of profits. The strong ties between economic and sporting competition in leagues were analyzed by Neale (1964) and formalized by El-Hodiri and Quirk (1971). There is now an extensive literature dealing with league outcomes and the resulting competitive balance. The objective of the teams remains a point of discussion, with some arguing that teams maximize profits (e.g. Fort and Quirk, 1995; Vrooman, 2007), others that teams maximize their win-percentage (e.g. Sloane, 1971; Kesenne, 2007) and some for utility maximizing behavior (Dietl et. al., 2011), as a weighted sum of the former two. For a comparison of the former two, see Fort and Quirk (2004). All these league types have been analyzed and there is broad consensus in sport economics literature that, independent of the exact outcome, due to the owner's specific objective, the size of the market in which a team plays is the essential factor explaining team quality differences in terms of winning success (Vrooman, 1995, 2007). Teams from larger markets have a greater incentive to employ more talent, and are capable of doing so. Therefore, they are more successful.

From this premise, it remains unclear why teams from the largest regions are often not the most successful ones in a league. In this paper, the phenomenon of underperforming large-market teams (or overachieving smaller-market teams), is explained in terms regional competition. Besides competing for talent and with talent, teams are also able to compete over markets. In existing models of sports leagues, the markets of the teams are given exogenously, whereas in this paper, a model is presented with endogenous markets for teams. Teams are able to choose the distance to nearby opponents; this location choice affects their own as well as their opponents' revenue potential. In short, with endogenous markets, teams are able to choose between different locations in order to maximize their revenue potential.

Sharing a market is quite common in European as well as in American sport leagues. In Europe there are, for example, the football teams of Real Madrid and Atletico Madrid in Spain or AC Milan and Inter Milan in Italy. In American sport leagues, there are the baseball teams of the Yankees and the Mets in New York or the basketball teams of the Clippers and the Lakers in Los Angeles. A prominent difference between American and European sport leagues is the league entrance system. In Europe, there is a system of promotion and relegation, which offers at least the possibility for any team to compete in the highest league. In America the league decides about the entrance of new franchises. It is quite common for the franchises to have exclusive territorial rights guaranteed by the league. If these rights can no longer be granted, compensation payments are common (e.g., the Anaheim Ducks had to pay $25 million to the Los Angeles Kings (NHL) in 1992). Thus, in both systems, regional competition occurs and mainly in the largest markets of the league.

Various kinds of team movements are feasible. In all American sports leagues, franchises may relocate for example the former Seattle Supersonics, became the Oklahoma City Thunders in 2008 (NBA), the former

Houston Oilers became the Tennessee Oilers in 1997 (NFL), and the former Quebec Nordiques became the Colorado Avalanche in 1995 (NHL). In other words, the players, as well as the team's owner, moved geographically to form a new team. In European sports leagues, such movements are uncommon. However, with a system of promotion and relegation, new teams enter a league every year, accompanied by a reallocation of talent. One difference between America and Europe with respect to these movements is whether the talent moves with or without the team. A detailed analysis of promotion and relegation systems is provided in Szymanski and Valletti (2005). A third possibility is that teams reallocate within their region merely by reallocating their stadium, examples include the move of Arsenal London in 2006 (English Premier League) or the New Jersey Nets becoming the Brooklyn Nets in 2012. In the model presented below, all such movements are feasible, as the decisive parameter is the effect these talent movements have on the markets in which the teams operate.

There is an extensive general economic literature dealing with endogenous market sizes (e.g. Mankiw and Whinston, 1986), as well as literature focusing on sports economics. However, in sports economics the literature focuses on the subsidies teams are able to obtain from cities. Friedman and Mason (2004) develop a stakeholder approach explaining public subsidies for sport facilities. For sport leagues, Owen (2003) used a model of a closed league, with more potential markets than teams in the league. In this environment, teams can threaten to leave the cities, in order to gain subsidies. However, teams are able to move into regions which already have a team and, furthermore, regional competition is not only about gaining subsidies. The interdependencies associated with a reallocation of talent with respect to nearby teams, as well as with regard to the whole league, cannot be neglected. The model presented in this paper focuses on the interactions of team movements and regional competition. Aside from the lack of a theoretical approach in the literature, there are, however, empirical studies demonstrating the effects of a relatively small distance between sport teams on stadium attendance. Buraimo et. al. (2007) as well as Buraimo and Simmons (2009) demonstrated the negative impact on the

English Premier League. Additionally, Winfree et. al. (2004) presented a similar study on major league baseball. In contrast, Breuer and Römmelt (2009) reported that a second team in the same market had no negative impact on the ticket sales of clubs in the German soccer Bundesliga.
The remainder of the paper is structured as follows. In Section 2, a model of regional competition is presented. In a two-stage game, teams choose their level of talent and preferred degree of regional competition. The basic model is introduced in Section 2.1 and the second stage equilibrium for a profit-maximizing league is presented in Section 2.2. The equilibrium for a win-percentage-maximization league is derived in Section 2.3. The effects on the league outcome are shown in Section 2.4, and the influence of revenue sharing on regional competition in Section 2.5. In Section 3, some empirical findings are presented, showing that regional competition indeed has an influence on league outcome. Section 4 concludes.

2 Regional competition in sports leagues

2.1 Basic model

The profits π_i of each team i in a n-team league are given by revenue R minus cost C:

$\pi_i = R_i[m_i(d_i), w_i(t_i, t_{-i})] - C_i(t_i)$; for $i=1,..,n$ (1)

with m_i as the market size of team i and w_i as its winning probability. The winning probability is given by a logit contest success function (CSF), which defines a team's winning probability by the ratio of employed talent t_i to total league talent: $w_i = t_i / \sum_{j=1}^{n} t_j$. This CSF is widely used in the sport economic literature. As in Szymanski (2003) or Dietl et al. (2011), revenue is specified by: $m_i w_i - \frac{\beta}{2} w_i^2$, which is a concave function of w with a global maximum at $m_i/\beta = w_i$. The parameter $\beta > 0$ reflects fans' preferences for competitive balance and assumed to be the same for all teams. The costs are a function of t_i with constant marginal costs c, which are the same for each team.

In contrast to the existing literature, the market size m_i in this model is not given purely exogenously, but depends on d_i, the distance to the nearest competitor. The market of a team is given by the product of a game's visitors s_i and the price p_i they are willing to pay for one unit of winning percentage: $m_i = s_i(d_i)\, p_i(d_i)$. The uniformly distributed inhabitants of a region decide to attend a league game on the basis of utility maximization. The inhabitants face different travelling costs per stadium visit. As there are different travelling costs, some attendees might prefer watching another team if it is close enough to their own location. The total number of attendees is given by: $s_i = \delta 1_i\, d_i^{\gamma 1_i}$, with $\delta 1_i$ being a region-specific parameter capturing a region's population density and regions size; $\gamma 1_i$ determines the effect of regional competition on the number of attendees. With a regional competitor, some of the inhabitants lower costs by attending the competitor's games, so that regional competition may reduce the attendance of a given team's matches. The greater the distance to a competitor, the more people attend a local game, thus $\gamma 1_i > 0$. The ticket price attendees are willing to pay is given by $p_i = \delta 2_i\, d_i^{\gamma 2_i}$. The parameter $\delta 2_i$ captures a regional purchasing power and $\gamma 2_i$ shows how regional competition affects the attendees' utility.

The ticket price is a result of individual utility maximization. Matches against a nearby opponent, i.e., local derbies, are often fierce and intense and it is assumed that these matches provide a higher utility level than a match of more distant teams. The willingness to pay the price of watching a match is assumed to decline in d, thus $\gamma 2_i < 0$. By defining $\delta 1_i + \delta 2_i = \delta_i$ and $\gamma 1_i + \gamma 2_i = \gamma_i$, the market of a team is given by: $m_i = \delta_i d_i^{\gamma_i}$. The parameter δ_i indicates the specifics of a region with regard to the level of revenue that may be generated within this region; γ_i is intended to represent the volatility of this revenue for the case that there is another sports team nearby. Whether a team's market increases or decreases in d depends on the sign of γ_i. With γ_i being positive ($\gamma_i > 0$), the effect on the number of stadium attendees is larger than the effect on the price. In this case, the market of a team decreases with more regional competition. If the price effect dominates the effect on attendees, a team's market increases with more regional competition, i.e., $\gamma_i < 0$.

The kind of revenue function, as well as contest success function and cost function is standard in sport economics literature.[1] However, in the model presented here, teams are also able to decide on the degree of regional competition, d_i, that affects their own market size, as well as that of regional competitors. The degree of regional competition affects profits and the marginal products of talent directly via the market size m, as well as indirectly via the winning probability and the cost function.

The decision about the degree of regional competition is modeled in a two-stage game. In the first stage, teams choose the degree of regional competition, and in the second stage, they determine the amount of talent. Since teams are assumed to make rational decisions under perfect information, the game can be solved by backward induction. Employing the concept of a subgame perfect Nash equilibrium, the teams decide on the optimal degree of regional competition, or the distance to the markets of other teams in the league, contingent on their choices of talent. The shorter the distance d between two adjacent teams, the greater the degree of regional competition and vice versa, with no regional competition at all as the lower bound.

2.2 Profit maximization

A well-established approach since Rottenberg (1956) in the sports economics literature is to model profit maximization as the objective of teams in a sports league. For all choices of regional competition, the Nash equilibrium for the subgame at the second stage is given by profit maximization. For the solution, the assumption about the kind of talent supply emerges as crucial. There are closed and open league models, with the former assuming a fixed talent supply and with the latter assuming indefinitely available talent. A fixed supply of talent is the traditional approach used since El-Hodiri and Quirk (1971) as well as, for instance, Fort and

[1] For detailed information about the CSF, see Tullock (1980) and Skaperdas (1996).

Quirk (1995). Starting with Szymanski (2004), Szymanski and Kesenne (2004) as well as Kesenne (2005), this approach has been criticized. The latter authors have stated that it is necessary to take into account non-cooperative behavior of sport teams by assuming an infinite availability of talent. An overview of this discussion can be found in Eckard (2006), as well as a game theoretic extension in Madden (2011). However, Winfree and Fort (2012) demonstrated that by differentiating between the direct decision on talent and the more indirect investment in talent, non-cooperative behavior may be modeled with a fixed supply of talent. In this paper, it is assumed that the supply of highly skilled sports talent is fixed. Therefore, one team's talent gain is a talent loss for the rest of the league.

Proposition 1: The equilibrium winning probabilities of each team in a profit maximization league with fixed talent supply are given by:

$$w_i = \frac{n\, m_i - \sum_{j=1}^{n} m_j + \beta}{n\, \beta} \tag{2}$$

with an total league payroll of:

$$c \sum_{j=1}^{n} t_j = \frac{-\beta + \sum_{j=1}^{n} m_i}{n}. \tag{3}$$

Proof: The objective of each team is profit maximization, $\max_{t_i} \pi_i$; with the profit definition in (1), this yields:

$$\frac{\partial R_i}{\partial w_i}\frac{dw_i}{dt_i} - \frac{\partial C_i}{\partial t_i} = 0 = \frac{1m_i - \beta\, w_i}{\sum_{j=1}^{n} t_j} - c, \forall i \in I. \tag{4}$$

With the adding-up constraint $\sum_{i=1}^{n} w_i = 1$ and the *n* first order conditions (FOCs) in (4), a set of *(n+1)* linear equations is obtained for determining *n* winning probabilities and a total league payoff, $c \sum_{j=1}^{n} t_j$. The solution of this system of equations is straightforward, with the resulting equations (2) and (3) above. //

With marginal revenue equaling marginal cost, the results of Proposition 1 are common in the sports economics literature; more recent examples include Vrooman (2007) and Szymanski (2003). Note that the result of Proposition 1 relies heavily on the assumption of a fixed supply of talent. With this fixed supply, one team's talent gain is another team's loss such that: $\partial \sum_{j=1}^{n} t_j / \partial t_i = 0$, and therefore: $\partial w_i / \partial t_i = 1 / \sum_{j=1}^{n} t_j$. In an open league, the talent choice of one team does not affect the choice of another team directly: $\partial t_k/\partial t_i = 0$ with $k \neq i$, and therefore: $\partial \sum_{j=1}^{n} t_j /\partial t_i = 1$. Thus, $\partial w_i/\partial t_i$= $(1 - w_i)/ \sum_{j=1}^{n} t_j$ holds in an open league. This results in a set of n quadratic first-order conditions and the adding-up constraint. Given the assumptions about the revenue function in this paper, this system of equations is not explicitly solvable for *n>2*.

2.3 Win maximization

Going back to Sloane (1971), the maximization of sporting success, defined as winning percentages, is considered to be a legitimate objective of sports teams. With $\partial w_i/\partial t_i > 0$ and win maximization as the objective of teams, a budget restriction is required to depict teams' investment behavior. In a win-maximization league, teams will invest as much as possible in talent to increase their winning percentages. A common assumption in the literature (see Kesenne, 2007) is to assume a zero-profit condition for all teams. Although Fort and Quirk (2004) question this approach, in this present paper, the zero-profit condition is applied, as it seems to be an obvious necessity, at least in the long run.

Proposition 2: The winning probabilities of each team in a win-maximization league are given by:

$$w_i = \frac{2\, n\, m_i - 2\, \sum_{j=1}^{n} m_j + \beta}{n\, \beta}, \tag{5}$$

with a total league payroll of:

$$c\, \sum_{j=1}^{n} t_j = \frac{-\, \beta + 2\, \sum_{j=1}^{n} m_i}{2\, n}. \tag{6}$$

Proof: With a zero-profit constraint, teams will invest up to the point where profit equal cost, which is equivalent to average profit equaling average cost. In league equilibrium, the average profits of all teams are equalized.

$$\pi_i = 0 = \frac{R_i - C_i}{w_i} = m_i - \frac{\beta}{2}\, w_i - c\ \ \sum_{j=1}^{n} t_j\ \ , \forall i\ \in I. \tag{7}$$

These n reaction functions, together with the adding-up constraint, represent a set of *(n+1)* linear equations whose results determine n winning probabilities and a total league payoff $c\ \sum_{j=1}^{n} t_j$. //

Comparing equations (5) and (2) shows that the winning percentages in a win-maximization league are always higher than in a profit maximizing league for teams having market sizes above the league average market size $\sum_{j=1}^{n} m_j/n$. The winning percentages of the teams with market sizes below the league average are higher in the profit maximization league. Comparing equations (6) and (3) demonstrates that a win-maximization league also has a higher overall league payroll. The higher competitive imbalance, as well as the higher league payroll in the win-maximization league, is in accordance with the sports economics literature; for a two team league, see Vrooman (2007).

2.4 Basic model with regional competition

The standard approach in the literature is to assume that there is no regional competition at all, so that the league equilibrium is determined by equation (2) or (5), respectively. If the teams know that they can influence their own market size, as well as that of others, they become aware of the results of individual profit or win maximization in the second stage of the game defined here. Therefore, they will anticipate this outcome when they choose the degree of regional competition in the first stage of the game.

Proposition 3: Regardless of the objective of the teams, the decision about the degree of regional competition depends only on whether or not regional competition increases the team's winning percentage. Independent of the teams' objective, a single team i gains from increased regional competition as long as

$$\frac{\partial m_i}{\partial d_i} < \frac{1}{n}\sum_{j=1}^{n} \frac{\partial m_j}{\partial d}. \tag{8}$$

Proof: For a profit maximizing team, the subgame perfect Nash equilibrium is derived from the first-order conditions: $m_i - \beta w_i = c\ \sum_{j=1}^{n} t_j$. Inserting the first-order conditions into equation (1) and taking account of $c\ t_i = w_i c \sum_{j=1}^{n} t_j$, equation (1) is reduced to:

$$\pi_i = \frac{\beta}{2} w_i^2. \tag{9}$$

Inserting the equilibrium winning percentage of equation (2) into equation (9) gives the objective function of a profit-maximizing team in the first stage of the game. A profit-maximizing team profits from regional competition as long as $\frac{\partial \pi_i}{\partial d_i} < 0$. The derivation of equation (9) with respect to d, taking account of $w_i > 0$, gives (8). A win-percentage-maximizing team is profiting from a decreasing distance if $\frac{\partial w_i}{\partial d_i} < 0$, derivation of (5) with respect to *d* gives (8). //

Regional competition has an impact on winning probabilities and the total league payoff $c\ \sum_{j=1}^{n} t_j$. Independent of the teams' objectives, in a two stage game, the crucial question is whether the teams' winning probabilities increase or decrease with the distance *d*. If the change of its own market is smaller than the average league market change, a team is better off with regional competition.

To show the basic mechanism of regional competition, an *n*-team league, with $n > 2$ and two markets, *e* and *f,* are considered, that are sufficiently close to each other to allow for regional competition. In this setting, $d_e = d_f = d$. The markets of the other teams are not affected by this decisions of team e and f, so that $\partial m_k / \partial d = 0$ for $k \neq e, f$. By defining the market elasticity of regional competition $\varepsilon_i = \frac{\partial m_i}{\partial d} \frac{d}{m_i}$, and by multiplying both sides of (8) by $d/(m_i\, m_j)$, it follows from (8) that team *i* benefits from regional competition if:

$$\varepsilon_i < \frac{m_j}{m_i} \frac{1}{(n-1)} \varepsilon_j. \qquad (10)$$

The determinants of regional competition are defined by the relative market elasticities of regional competition, the relative market sizes and the league size. Crucial for the interpretation of the inequality (10) is the derivative $\frac{\partial m_i}{\partial d} = \gamma_i \delta_i d^{\gamma_i - 1} = m_i{}'$, so that (10) can be written as:

$$\gamma_i < \frac{1}{n-1} \frac{\delta_j}{\delta_i} d^{\gamma_j - \gamma_i}\, \gamma_j. \qquad (11)$$

As equation (8) shows, the positive effect of regional competition on revenue can be attributed to the positive effect it could have on the winning probabilities of the teams. This of course includes a negative effect on the winning percentages of a regional competitor. The more teams in a league, the weaker this negative effect will be, which explains why the likelihood that inequality (11) holds decreases in n. The relationship δ_j/δ_i on the of inequality (11) shows that the greater the inequality between two regions with respect to their revenue potential, the more likely a team i gains from regional competition, as long as it is the team with the lower revenue potential. Furthermore, not only the absolute effect of regional competition on the teams has an impact on the decision as to whether such a competition occurs or not, but also the difference between these effects. Yet, both effects in inequality (11), the absolute as well as the relative, point in the same direction. Hence, the chances that inequality (11) holds increases in γ_j and decreases in γ_i.

If only two teams in a league have the opportunity to enter regional competition, and the other team is unable to avoid it, there are only three possible combinations of market derivations:

- the markets of both teams increase with regional competition,
- the markets of both teams decrease with regional competition,
- only one team's market increases, the other's decreases.

Table 1 provides an overview of cases where there is regional competition in league equilibrium, given that the partial derivatives of market sizes with respect to d are constant for all i and unequal to zero for e and f. Under these circumstances, the degree of competition is reduced to a bang-bang solution: either both teams avoid competition or the highest possible degree of competition is realized. Interior solutions are feasible if the assumption of a constant derivative $m_i{}'$ is abandoned. For sake of simplicity, this is not considered here.

The specifics of a region with regard to regional competition can be defined by $\theta = \frac{m_e{}'}{m_f{}'} = \frac{\delta_e}{\delta_f}\frac{\gamma_e}{\gamma_f}d^{\gamma_e-\gamma_f}$. The parameter θ consists of the relative revenue potential and the relative effects of regional competition, as well as the distance between the teams. To find the team which gains from regional competition, three different characteristics of θ have to be considered; θ can be inside the open interval $(\frac{1}{n-1}, n-1)$ or outside it on both sides. Notice that the lower bound decreases in n whereas the upper bound increases in n; therefore, the interval increases in n. The more teams a league hosts, the better the chances that θ is within this interval. Under the assumption that $m_e{}'$ and $m_f{}'$ are unequal to zero, all possible constellations are shown in Table 1.

Table 1: Best-responses of the teams regarding regional competition

Assumption about m_e', m_f'	Possible relation of θ	Best-response function of the teams	Regional competition
$m_e' > 0$, $m_f' > 0$	$n-1 < \theta$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes
	$\frac{1}{n-1} < \theta < n-1$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} > 0$	
	$\theta < \frac{1}{n-1}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes
$m_e' < 0$, $m_f' < 0$	$n-1 < \theta$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes
	$\frac{1}{n-1} < \theta < n-1$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} < 0$	yes
	$\theta < \frac{1}{n-1}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes
$m_e' > 0$ $m_f' < 0$	$\theta < \frac{1}{n-1}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes
$m_e' < 0$ $m_f' > 0$	$\theta < \frac{1}{n-1}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes

As mentioned above, crucial for the sign of the $m_i{}'$ is the sign of γ_i, which captures the effect of regional competition. The market for a team decreases with higher regional competition, if the effect of a higher willingness to pay is smaller than the effect on attendance. If the former is greater than the latter, then the opposite is true. The higher the absolute value of γ_i, the more sensitively the team's market reacts to regional competition.
With two teams in a n-team league deciding about regional competition and both teams' markets decreasing with regional competition, one team

nevertheless benefits as long as θ is not in the abovementioned interval defined by the league size. Only if it is within the interval both teams are better off avoiding regional competition. If θ is outside the interval, the team with the less sensitive market gains from regional competition; the team with the more sensitive, and in this case decreasing market, has reduced profits. Thus, even if the effect on its own market is negative, but comparatively small, a team could enter regional competition if this has a strong negative effect on an opponent. The gaining team loses attendance, which is not fully offset by an increasing price, although this is in the end offset by higher profits due to an increased winning probability. Even though a team reduces its own market, if it is able to reduce the market of another team even more, it could still gain from regional competition. If the markets of both teams decrease in *d,* there will be regional competition at the league equilibrium. However, only if θ is inside the interval $(\frac{1}{n-1}, n-1)$, do both teams benefit from this competition. If θ is outside the interval, the team with the more sensitive market will be the beneficiary of regional competition and the other team will be worse off. The latter team's market increases with regional competition, but the former market increases so much more that it ends up with lower winning percentage and lower profits. Even a team's market which increases with regional competition, does not necessarily imply increasing profits, as regional competition has an effect on talent allocation and competitive balance. Hence, whenever there is a business expansion effect for one team, and a business-stealing effect for another team, regional competition will occur in a sport league. If γ_e and γ_f have opposing signs, only one team's market increases, while the other teams market decreases in regional competition. These cases are straightforward, as the team with the increasing market always gains from regional competition.

2.5 Competitive balance, league outcome and regional competition

As shown above, regional competition influences the talent allocation in a league. As it then affects the winning probability of each team, the analysis of competitive balance requires an appropriate measure of talent inequality. To measure the degree of competitive balance in a league, one option is to use concentration ratios of talent allocation. Several different measures of concentration have been applied in sports economics. One well known measure is the Gini coefficient used by Schmidt and Berri (2001; 2002) or Utt and Fort (2002); another one is the relative entropy of Horrowitz (1997). Without further assumptions about the market sizes *m* and their ascending or descending ordering, the Herfindahl-Hirschman Index (HHI) is a useful and appropriate measure of talent concentration in a league (see, e.g., Depken, 1999; Owen et al., 2007). In a sport league context, the HHI is defined as the sum of the squares of the talent shares of teams:

$$HHI = \sum_{i=1}^{n}\left(\frac{t_i}{\sum_{j=1}^{n} t_j}\right)^2 = \sum_{i=1}^{n} w_i^2. \tag{12}$$

The higher the HHI, the greater the competitive imbalance in a league, with a maximum of 1; the lower the HHI, the more even the distribution of talent, with a completely balanced talent allocation at $HHI = 1/n$. The influence of regional competition on competitive balance is analyzed by focusing again on the regional competition decision of teams *e* and *f*.

Proposition 4: Independent of the assumptions about team objectives and for a fixed as well as a flexible supply of talent, regional competition improves competitive balance if:

$$\frac{\partial m_e}{\partial d}(n\, w_e - 1) + \frac{\partial m_f}{\partial d}\left(n\, w_f - 1\right) > 0. \tag{13}$$

Proof: Regional competition improves competitive balance as long as $\partial HHI/\partial d > 0$ holds. Derive (12) with respect to *d,* considering the adding up constraint and $n, \beta > 0$ gives (13). //

The terms $(n\, w_e - 1), \left(n\, w_f - 1\right)$ are positive or negative whenever the talent levels of *e* and *f,* respectively, are above or below the average level of talent in the league. Whether regional competition improves or worsens the competitive balance of a league depends to some extent on the exact values of the league parameters.

If both teams *e* and *f* have talent levels above the league average, competitive balance will improve if their markets decline with regional competition and it will deteriorate if their markets increase with regional competition. With *e* (*f*) being the team whose market decreases (increases) in *d*, competitive balance improves as long as $\frac{\delta_e}{\delta_f}\frac{\gamma_e}{\gamma_f}d^{\gamma_e-\gamma_f} > -\frac{n\,w_f-1}{n\,w_e-1}$; otherwise competitive balance would deteriorate.

If both teams have talent levels below the league average, competitive balance will improve if their markets increase with regional competition or it will deteriorate if they decrease with regional competition. With *e* (*f*) being the team whose market decreases (increases) in *d*, competitive balance improves as long as $\frac{\delta_e}{\delta_f}\frac{\gamma_e}{\gamma_f}d^{\gamma_e-\gamma_f} < -\frac{n\,w_f-1}{n\,w_e-1}$; otherwise regional competition would worsen the competitive balance of the league.

If one team has a talent level above and the other team a talent level below the league average, the competitive balance will improve if the original market declines in size and the latter market increases in terms of regional competition; if it is the other way around, the opposite holds. If team *e* (*f*) has a talent level above (below) the league average and both markets are increasing in *d,* regional competition improves the competitive balance as long as $\frac{\delta_e}{\delta_f}\frac{\gamma_e}{\gamma_f}d^{\gamma_e-\gamma_f} > -\frac{n\,w_f-1}{n\,w_e-1}$. If both markets decrease with regional competition, the league has a higher competitive balance if $\frac{\delta_e}{\delta_f}\frac{\gamma_e}{\gamma_f}d^{\gamma_e-\gamma_f} < -\frac{n\,w_f-1}{n\,w_e-1}$.

Apart from total league competitive balance, an interesting effect of regional competition is the possibility to change the orderings of teams with respect to their relative talent. A standard finding presented in the literature is that teams with a larger market earn higher marginal revenue and therefore employ a larger amount of talent. Furthermore as equation (2) shows, a larger market indeed means a higher winning probability. But in contrast, equation (8) shows the negative impact a larger market could have. With a larger market, it also becomes more likely that a competitor might enter the region. Given the parameters of a league and relative market sizes, a league is not only composed of teams from regions where the teams are natural monopolies, but also of larger regions where the teams compete for markets. With d^* as the distance without regional competition and d^{**} as the distance with regional competition, a team i from a region will end up with less employed talent after the entrance of a competitor j, as long as $m_i(d*) - m_i(d**) > \frac{1}{n-1}(m_j(d*) - m_j(d**))$. Hence, the strict relationship between market size and success in a sport league does not apply to regions that are large enough to allow for regional competition.

2.6 Revenue sharing and regional competition

In sport leagues, teams sometimes agree (or are forced) to participate in some form of revenue sharing. The common justification for this is to improve the competitive balance. There is an extensive debate in the literature about the effectiveness of revenue sharing agreements on talent allocation. The so-called invariance principle, stated by Rottenberg (1956) and first formalized and proven by El-Hodiri and Quirk (1971), implies that revenue sharing does not have an impact on talent allocation among teams. Winfree and Fort (2012) stated that revenue sharing agreements are ineffective in the basic model, if equilibrium in the talent investment market equates the marginal product of talent investment across owners (see Runkel, 2011, for a comprehensive analysis of the influence of the respective CSF and talent supply on the effects of revenue sharing). However, this result depends to a large extent on the assumptions about team objectives and talent supply, as the authors note themselves. Assuming profit maximizing teams with a fixed supply of talent, several papers (e.g. Szymanski, 2003; Vrooman, 1995) showed that revenue sharing does not affect talent allocation. Fort and Quirk (1995) demonstrated that if teams earn additional revenue that is not shared, the invariance principle does not hold. Furthermore, it does not hold if the supply of talent is completely elastic (Szymanski and Kesenne, 2004; Kesenne, 2005) or if absolute rather than relative talent levels matter for team revenue (Kesanne, 2000).

Moreover, when teams behave as win-percentage maximizers, it does not hold either, as shown in Kesenne (2006) and Vrooman (2008).
Whether or not the invariance principle holds, depends on modeling assumptions and specifications (revenue function, talent supply, contest success function and owners' objectives). As shown above, regional competition operates through its effect on the winning probabilities of teams. Therefore, whenever revenue sharing has an influence on talent allocation, it also effects the regional competition of teams. The model specifications of this paper allow only for the specified analysis of an *n=2* profit maximization league with a fixed supply of talent.[2] Fort and Quirk (2007) proved, for a more general model (but with a different CSF), the existence and uniqueness of revenue sharing equilibria for *n>2*. Nevertheless, in this paper, *n=2* is sufficient to show the effect of revenue sharing, in combination with regional competition, even in a closed talent market with profit-maximizing teams. The usual expectation in this case is that revenue sharing is not effective concerning talent allocation.
There are several forms of revenue sharing (see, e.g., Vrooman, 2007); in this paper, a pool revenue sharing system is applied. Each team keeps a share α of its generated revenues, with the remaining share being used to fund a pool of money which is divided equally among all teams.
The profit of a team *i* is given by:

$$\pi_i = \alpha R_i + \frac{1-\alpha}{n}\left(\sum_{j=1}^{n} R_j\right) - c\, t_i. \tag{14}$$

Proposition 5: The relative allocation of talent in a $n = 2$ profit maximization league with a closed talent market in the second stage of the game is independent of the revenue-sharing parameter α, because the winning percentages are given by:

$$w_i = \frac{m_i - m_j + \beta}{2\beta}. \tag{15}$$

In contrast, the overall league payroll, as well the teams' profits depend on α:

$$c\left(t_i + t_j\right) = \frac{\alpha(m_i + m_j - \beta)}{2}, \tag{16}$$

[2] See Appendix for detailed information.

$$\pi_i = \frac{(2\alpha-1)\beta^2+(m_i-m_j)^2+2\beta(m_i+m_j(1-2\alpha))}{8\beta}. \tag{17}$$

Proof: Both teams maximize their profit function (14) in the second stage of the league game by choosing their respective talent level. At equilibrium, marginal revenues are equal:

$$\alpha \frac{\partial R_i}{\partial w_i}\frac{\partial w_i}{\partial t_i} + \frac{1-\alpha}{2}\left(\frac{\partial R_i}{\partial w_i}\frac{\partial w_i}{\partial t_i} + \frac{\partial R_j}{\partial w_j}\frac{\partial w_j}{\partial t_i}\right) = \alpha \frac{\partial R_j}{\partial w_j}\frac{\partial w_j}{\partial t_j} + \frac{1-\alpha}{2}\left(\frac{\partial R_j}{\partial w_j}\frac{\partial w_j}{\partial t_j} + \frac{\partial R_i}{\partial w_i}\frac{\partial w_i}{\partial t_j}\right). \tag{18}$$

With $\partial w_i/\partial t_j = -\,\partial w_j/\partial t_j = -1/T$, equation (16) is reduced to: $\partial R_i/\partial w_i = \partial R_j/\partial w_j$. In combination with the adding-up constraint, this yields equation (15). Inserting equation (15) into equation (14), taking account of $c\ t_i = w_i c(t_i + t_j)$, yields equations (16) and (17). //

In this model, revenue sharing does not directly affect the talent allocation within the league in the second stage of the league game. For $n = 2$, equation (15) equals equation (2), which is a standard result in the literature (see, e.g., Vrooman, 2007). However, revenue sharing has an effect on the profits of teams. This effect in turn impacts on the degree of regional competition chosen in the first stage of the game; regional competition then has an influence on the competitive balance of the league.

Proposition 6: In a two-team league, team i is profiting from regional competition with revenue sharing if:

$$\varepsilon_i < \frac{m_j}{m_i}\left(\frac{m_i-m_j-\beta(1-2\alpha)}{m_i-m_j+\beta}\right)\varepsilon_j. \tag{19}$$

Proof: Differentiating equation (17) with respect to d and expanding the result with $d/(m_i m_j)$ yields (19). Note that $m_i - m_j + b$ must be larger than zero, since this expression is the numerator of the winning probability according to equation (15). //

By defining $\theta_i(\alpha) = \frac{m_i-m_j-\beta(1-2\alpha)}{m_i-m_j+\beta}$, inequality (19) can be written as:

$$\varepsilon_i < \frac{m_j}{m_i}\theta_i\varepsilon_j. \tag{20}$$

The revenue sharing parameter α has an influence on the degree of competitive balance, as $\theta_i \leq 1$ depends on α and $\partial\theta_i/\partial\alpha = 2\beta$. Hence θ_i decreases with an increasing level of revenue sharing. With no revenue sharing, i.e., $\alpha = 1$, equation (19) is equal to equation (10) for $n = 2$. As $m_i - m_j + b$ is the numerator of the winning probability in equation (3), it is larger than zero. Given the parameters $\gamma_i, \gamma_j, \delta_i, \delta_j$, the probability that inequality (12) holds, decreases for both teams with a higher level of revenue sharing. Hence, regional competition is less likely to occur in a league with revenue sharing than in one without sharing. Table 2 provides an overview of cases with regional competition under revenue sharing.

Without revenue sharing, there is only one parameter combination for which no regional competition occurs (see Table 1). With revenue sharing, it becomes more likely that all teams are better off by avoiding regional competition. Revenue sharing makes this one case not only more likely, in addition, with different signs of market elasticity, revenue sharing effectively creates the possibility that no team prefers regional competition. If both teams can increase their market through regional competition, revenue sharing also increases the possibility that both teams can increase their profits.

Although revenue sharing does not affect competitive balance in a profit-maximizing closed league, in the second stage of the game, it nonetheless changes the teams' profits. Thus revenue sharing has an effect on the location choice of the teams, which therefore has an effect on the talent allocation. Teams gain from regional competition to some extent, because of the business-stealing effect. However, if their opponent's revenues is part of their own profits, business-stealing may reduce these profits, such that thus revenue sharing may decrease the level of regional competition. As a consequence, even in a league where revenue sharing is not directly effective, it can nevertheless have an indirect effect on talent allocation, if the teams can choose their own location. If revenue sharing has an impact on the location choice, talent allocation may differ between a league with revenue sharing and one without it.

Table 2 Best-responses of the teams regarding regional competition in cases of revenue sharing

Assumption about m_e', m_f'		Possible relation of θ_e, θ_f	Best-response function of the teams	Regional competition	Regional competition more likely with revenue sharing?
$m_e' > 0$; $m_f' > 0$		$\frac{1}{\theta_f} < \frac{m_e'}{m_f'}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	
		$\theta_e < \frac{m_e'}{m_f'} < \frac{1}{\theta_f}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} > 0$		yes
		$\frac{m_e'}{m_f'} < \theta_e$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes	
$m_e' < 0$; $m_f' < 0$		$\frac{1}{\theta_f} < \frac{m_e'}{m_f'}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes	
	$\theta_f < 0$	$\theta_e < \frac{m_e'}{m_f'}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	yes
	$0 < \theta_f < 1$	$\theta_e < \frac{m_e'}{m_f'} < \frac{1}{\theta_f}$			
		$\frac{m_e'}{m_f'} < \theta_e$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	
$m_e' > 0$; $m_f' < 0$	$\theta_f < 0$	$\theta_e < \frac{m_e'}{m_f'}$; $\frac{1}{\theta_f} < \frac{m_e'}{m_f'}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	yes

Table 2 Best-responses of the teams regarding regional competition in cases of revenue sharing

$m'_e > 0;$ $m'_f < 0$	$0 < \theta_f < 1$	$\theta_e < \frac{m_e{}'}{m_f{}'}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	yes
		$\frac{m_e{}'}{m_f{}'} < \theta_e;$ $\frac{m_e{}'}{m_f{}'} < \frac{1}{\theta_f}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} > 0$		yes
		$\theta_e < \frac{m_e{}'}{m_f{}'} < \frac{1}{\theta_f}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes	
	$\theta_f < 0$	$\frac{1}{\theta_f} < \frac{m_e{}'}{m_f{}'} < \theta_e$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	
	$0 < \theta_f < 1$	$\frac{m_e{}'}{m_f{}'} < \theta_e$			
$m'_e < 0;$ $m'_f > 0$		$\frac{m_e{}'}{m_f{}'} < \theta_e;$ $\frac{m_e{}'}{m_f{}'} < \frac{1}{\theta_f}$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	yes
	$\theta_f < 0$	$\theta_e < \frac{m_e{}'}{m_f{}'}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} > 0$		yes
	$0 < \theta_f < 1$	$\theta_e < \frac{m_e{}'}{m_f{}'};$ $\frac{1}{\theta_f} < \frac{m_e{}'}{m_f{}'}$			
	$\theta_f < 0$	$\frac{1}{\theta_f} < \frac{m_e{}'}{m_f{}'} < \theta_e$	$\frac{\partial \pi_e}{\partial d} < 0; \frac{\partial \pi_f}{\partial d} > 0$	yes	
	$0 < \theta_f < 1$	$\frac{m_e{}'}{m_f{}'} < \theta_e$			
		$\theta_e < \frac{m_e{}'}{m_f{}'} < \frac{1}{\theta_f}$	$\frac{\partial \pi_e}{\partial d} > 0; \frac{\partial \pi_f}{\partial d} < 0$	yes	yes

3 Regional origin of most successful European football teams

Determining the right current or potential market sizes of sports teams has been discussed extensively in the empirical literature. In this paper, the approach of Schmidt and Berri (2001) is adopted to approximate the market size of sports teams using the size of the local population. However, teams from less crowded areas could nevertheless be successful, as long as they have strong support from firms, such as VfL Wolfsburg (Volkswagen) in Germany and FC Sochoux (Peugeot) in France. Moreover, defining the right catchment area provides an additional challenge. With the population of a city or region as a proxy for market size, this might be overestimated or underestimated. Buraimo et al. (2007) proposed the population in different areas near a stadium, as well as the demographics and years of league membership, as proxies for market size. Additionally the preferences of residents for sport, as well as income levels might differ between regions. In this paper, local population is used in this manner to provide an approximate empirical analysis.

According to sport economics, it is probable that the largest European cities host the most successful European football teams, because they own the largest markets. To verify this, in Table 3, the all-time league records of the best teams of six major European football leagues are depicted. The ten most successful teams in terms of aggregated wins in the respective leagues are presented, along with the cities (ranked by population size), where the clubs are located. If more than one club was hosted by the same city, the arithmetic mean of the corresponding ranks is attributed to the respective clubs. The reported value of Kendall's tau presents the rank-correlation between success in the respective league and the size of the teams' host city, i.e., the size of their home market.

With a tau value of 0.705, the Scottish Premier League has by far the highest correlation between city size and team success. On the other hand, in France, with a tau value of 0.111, there is almost no correlation at all between team success and host city size.

Table 3: All-time most successful teams in six European football leagues

Primera Division		*Rank of the City*	*Seria A*		*Rank of the City*
1	Real Madrid C.F.	1.5	1	Juventus F.C.	6.5
2	FC Barcelona	3.5	2	F.C. Internationale Milano	3.5
3	Athletic Bilbao	9	3	A.C. Milan	3.5
4	Atletico Madrid	1.5	4	A.S. Roma	1.5
5	Valencia C.F.	4	5	ACF Fiorentina	10
6	RCD Espanyol	3.5	6	Torino F.C.	6.5
7	Sevilla F.C.	5.5	7	Lazio Roma	1.5
8	Real Sociedad	10	8	Bologna F.C. 1909	9
9	Real Zaragoza	8	9	S.S.C. Napoli	5
10	Real Betis	5.5	10	U.C. Sampdoria	8
$\tau = 0.469$			$\tau = 0.138$		
Fußball Bundesliga		*Rank of the City*	*Scottish Premier League*		*Rank of the City*
1	FC Bayern Munich	2	1	Celtic F.C.	1.5
2	SV Werder Bremen	7	2	Rangers F.C.	1.5
3	Hamburger SV	1	3	Heart of Midlothian F.C.	3.5
4	VfB Stuttgart	5	4	Aberdeen F.C.	5
5	Borussia Dortmund	6	5	Hibernian F.C.	3.5
6	Borussia Mönchengladbach	9	6	Kilmarnock F.C.	10
7	1. FC Köln	3	7	Dundee United F.C.	6
8	1. FC Kaiserslautern	10	8	Motherwell F.C.	8
9	FC Schalke 04	8	9	Inverness Caledonian Thistle F.C.	7
10	Eintracht Frankfurt	4	10	Dunfermline Athletic F.C.	9
$\tau = 0.2889$			$\tau = 0.705$		

Table 3: All-time most successful teams in six European football leagues

Premier League		*Rank of the City*	*Lique 1*		*Rank of the City*
1	Liverpool F.C.	5.5	1	Olympique de Marseille	1
2	Everton F.C.	5.5	2	FC Girondins de Bordeaux	5
3	Arsenal F.C.	2	3	AS Saint-Étienne	6
4	Manchester United F.C.	7.5	4	AS Monaco	9
5	Aston Villa F.C.	4	5	FC Sochaux-Montbéliard	10
6	Manchester City F.C.	7.5	6	Olympique Lyonnais	2
7	Newcastle United F.C.	9	7	RC Lens	8
8	Tottenham Hotspur F.C.	2	8	FC Nantes	3
9	Chelsea F.C.	2	9	FC Metz	7
10	Sunderland A.F.C.	10	10	RC Strasbourg	4
$\tau = 0.141$			$\tau = 0.111$		

Source: fussballdaten.de (2011), cities are ranked according their population size.

With exception of France and Germany, the largest regions in terms of inhabitants of all countries considered here are represented in the all-time-best list of football teams. Furthermore, in four of the six leagues, again with the exception of France and Germany, some of these regions have more than one club among the all-time best teams. Evidently, market size matters, at least to some extent. Nevertheless, only in half of the leagues, do the most successful clubs come from the largest region. In Italy, the most successful club is hosted by the city of Torino (Turin), which is relatively small compared to Rome. The most successful teams in England are not from London, just as the best French teams are not from Paris, although London and Paris are by far the largest cities of their countries with respect to population size. However, three of the ten most successful teams in England come from London.

Table 4: Regional competition in the German Bundesliga

Metropolitan Region	Million inhabitants 2008	Different clubs playing in the Bundesliga	Actual clubs in the Bundesliga	Titles in the Bundesliga	Titles per 1 million inhabitants
Rhine-Ruhr	11.69	14	5	11	0.942
Central Germany	6.90	2	0	0	0
Berlin/Brandenburg	5.95	5	0	0	0
Rhine-Main	5.52	4	2	0	0
Stuttgart	5.29	2	1	3	0.567
Munich	5.60	3	1	22	3.93
Hamburg	4.29	2	2	3	0.700
Hanover-Braunschweig-Göttingen-Wolfsburg	3.88	3	2	2	0.515
Nuremberg	3.60	1	1	1	0.278
Bremen-Oldenburg	2.73	1	1	4	1.470
Rhine-Neckar	2.36	3	1		0
Rest		10	2	2	
Share of the Metropolitan Regions		0.80	0.89	0.96	

Source: IKM (2010), bundesliga.de (2011).

The previous findings are emphasized by results for the first three German football league divisions in the season 2009/2010. The cities are ranked by population size, whereby all German cities were considered, not only those with a professional football team. The correlation of ranks of the top three division teams with the population size ranks of cities, yields a Kendall tau of 0.195.

Table 4 shows the competition across regions in the German football (soccer) Bundesliga. Of all the clubs that ever played in the Bundesliga, 80 percent came from one of the 11 German *metropolitan areas*, which are the largest population agglomerations in Germany. In the season 2009/2010, only two clubs were not located in these regions. All champions came from one of these regions, except one, namely 1. FC Kaiserslautern, a two-time winner of the national football championship.

These empirical findings support the hypothesis that a large market does not necessarily imply also having one of the best performing teams. By far the most crowded area in Germany is the Rhine-Ruhr area, with nearly 12 million inhabitants, and indeed, the majority of clubs in the Bundesliga are located in this region. But in comparison with the region of Munich in Bavaria, the Ruhr area is not that successful. Although the Ruhr area has a population size more than twice that of Munich, teams from the Ruhr area have won only half as many titles. Some large areas like Berlin or Frankfurt did not produce a single national champion, whereas the much smaller Bremen area supported a team that won the national championship four times. To some extent, this is the consequence of regional competition. Larger regions seem to trigger stronger local competition, whereas in smaller regions, the focus of support is on just one club.

Table 5 presents the results of an OLS-regression, providing evidence of the negative effect of local competition on the number of years a team is able to stay in the highest division of the German football league. The dependent variable is the number of years a club has been in the Bundesliga. The independent variables are the sum of years when one or more additional teams within a radius of 50 km also played in the Bundesliga, and the population size of the catchment area. To approximate the catchment area of a team, the inhabitants per *metropolitan area*, or of an area of comparable size, are used. Teams remained longer in the German Bundesliga, the larger their catchment area and the fewer the number of nearby opponents. All estimated coefficients are statistically significant at the 5% level. These results provide provisional empirical evidence of the effects of regional competition on success.

Table 5: Regional competition and time in the German Bundesliga

Dependent variable	years in Bundesliga
Independent variable	**Coefficient**
C	8.764**
Sum of years of competitors in BL.	-0.107**
Size of catchment area (in million inhabitants)	2.650**
F-Statistic	3.418**
Adjusted R-squared	0.090

*** 1% level of significance, ** 5% level, * 10% level.
Source: Own estimation based on data from bundesliga.de (2011).

4 Conclusion

A top sports league is a highly competitive environment and, since potential revenue is different in different locations, talent allocation will also differ between teams. However, teams do not necessarily remain in one place. Either teams or talent can change their location. This has an effect on the revenue potential of other clubs through a business-stealing or a business-expansion effect. A regional reallocation of teams or talent is thus accompanied by a change in talent allocation *between* the teams.
In this paper, a model of regional competition is presented. In a two stage game, it is shown that, independent of the objectives of teams, regional competition will occur if it increases the winning probability of at least one team. Teams might choose to compete with one another, due to an increased own winning probability or to lower the winning probability of another team. Whether regional competition enhances the league's competitive balance cannot be stated with certainty, but it is evident that a large-market team might end up with less talent than it would employ

without regional competition. A strict relationship between market size and success in sports leagues does therefore not hold. The reason is that large regions attract more teams than smaller ones. Therefore small-market teams might be more successful, as they do not face local competition over their market.

Furthermore, it can be shown that revenue sharing, although not directly effective in the reallocation of talent, nevertheless has an effect on talent allocation, as it prevents the occurrence of regional competition, at least to some extent. Revenue sharing, may evidently internalize some of the negative externalities of regional competition. For major European football (soccer) leagues, as well as for the German Bundesliga, empirical evidence is presented that regional competition exists and exerts a negative impact on the winning probabilities of teams within the respective region.

Further empirical research is required to expand this initial empirical analysis by providing an appropriate definition and measurement of market size for sport teams. The number of inhabitants, as well as regional income, regional economic growth and regional preferences for football in general might be appropriate measures. Also, the role of team-specific preferences of inhabitants of a region could be taken into account more specifically. Another topic for future research entails the effects of supranational tournaments and their influence on the regional competition of teams.

References

Breuer, M., B. Römmelt (2009), Ticket Demand Depending on the Number of Local Clubs, *International Advances in Economic Research,* Vol. 15, No. 2, pp. 265-266.

Bundesliga (2011), URL: http://www.bundesliga.de [Accessed 14.05.2011].

Buraimo, B., D. Forrest, R. Simmons (2007), Freedom of Entry, Market Size, and Competitive Outcome: Evidence from English Soccer, *Southern Economic Journal,* Vol. 74, No. 1, pp. 204-213.

Buraimo, B., R. Simmons (2009), Market Size and Attendance in English Premier League Football, *International Journal of Sport Management and Marketing,* Vol. 6, No. 2, pp. 200-214.

Depken, C. A. II (1999), Free-agency and the Competitiveness of Major League Baseball, *Review of Industrial Organization,* Vol. 14, No. 3, pp. 205-217.

Dietl H. M., M. Grossmann, M. Lang (2011), Competitive Balance and Revenue Sharing in Sports Leagues With Utility-Maximizing Teams, *Journal of Sports Economics,* Vol. 12, No. 3, pp. 284-308.

Eckard, E. W. (2006), Comment: "Professional Team Sports Are Only a Game: The Walrasian Fixed-Supply Conjecture Model, Contest-Nash Equilibrium, and the Invariance Principle", *Journal of Sports Economics,* Vol. 7, No. 2, pp. 234-239.

El-Hodiri, M., J. Quirk (1971), An Economic Model of a Professional Sports League, *Journal of Political Economy,* Vol. 79, No. 6, pp. 1302-1319.

Fort, R., J. Quirk (1995), Cross-Subsidization, Incentives and Outcomes in Professional Team Sports Leagues, *Journal of Economic Literature*, Vol. 33, No. 3, pp. 1265-1299.

Fort, R., J. Quirk (2004), Owner Objectives and Competitive Balance, *Journal of Sports Economics,* Vol. 5, No. 1, pp. 20-32.

Fort, R., J. Quirk (2007), Rational Expectations and Pro Sports Leagues, *Scottish Journal of Political Economy,* Vol. 54, No. 3, pp. 374-387.

Friedman, M. T., D. S. Mason (2004), A Stakeholder Approach to Understanding Economic Development Decision Making: Public Subsidies for Professional Sport Facilities, *Economic Development Quarterly,* Vol. 18, No. 3, pp. 236-254.

Fussballdaten (2011), URL: http://www.fussballdaten.de/, [Accessed 10.06.2011].

Horowitz, I. (1997), The Increasing Competitive Balance in Major League Baseball, *Review of Industrial Organization*, Vol. 12, No. 3, pp. 378-387.

IKM (2010), Regionales Monitoring 2010, Bonn, Mannheim, URL: http://www.deutsche-metropolregionen.org/fileadmin/ikm/01_monitoring/IKM_Monitoring2010.pdf, [Accessed 02.06.2011].

Kesenne, S. (2000), Revenue Sharing and Competitive Balance, *Journal of Sports Economics*, Vol. 1, No. 1, pp. 56-65.

Kesenne, S. (2005), Revenue Sharing and Competitive Balance: Does the Invariance Proposition Hold?, *Journal of Sports Economics,* Vol. 6, No. 1, pp. 98-106.

Kesenne, S. (2006), The Win Maximization Model Reconsidered: Flexible Talent Supply and Efficiency Wages, *Journal of Sports Economics,* Vol. 7, No. 4, pp. 416-427.

Kesenne, S. (2007), The Peculiar International Economics of Professional Football in Europe, *Scottish Journal of Political Economy*, Vol. 54, No. 3, pp. 388-399.

Madden, P. (2011), Game Theoretic Analysis of Basic Team Sports Leagues, *Journal of Sports Economics,* Vol. 12, No. 4, pp. 407-431.

Mankiw N. G., M. D. Whinston (1986), Free Entry and Social Inefficiency, *Rand Journal of Economics,* Vol. 17, No. 1, 1986, pp. 48-58.

Neale, W. C. (1964), The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition, *Quarterly Journal of Economics*, Vol. 78, No. 1, pp. 1-14.

Owen, J. G. (2003), The Stadium Game: Cities Versus Teams, *Journal of Sports Economics,* Vol. 4, No. 3, 183-202.

Owen, P. D., M. Ryan, C. R. Weatherston (2007), Measuring Competitive Balance in Professional Team Sports using the Herfindahl-Hirschman Index, *Review of Industrial Organization,* Vol. 31, No. 4, pp. 289-302.

Rottenberg, S. (1956), The Baseball Players' Labor Market, *Journal of Political Economy*, Vol. 64, No. 3, pp. 242-258.

Runkel, M. (2011), Revenue Sharing, Competitive Balance and the Contest Success Function, *German Economic Review*, Vol. 12, Issue 3, pp. 256-273.

Schmidt, M. B., D. J. Berri (2001), Competitive Balance and Attendance: the Case of Major League Baseball, *Journal of Sports Economics*, Vol. 2, No. 2, pp. 145-167.

Schmidt, M. B., D. J. Berri (2002), Competitive Balance and Market Size in Major League Baseball: A Response to Baseball's Blue Ribbon Panel, *Review of Industrial Organization*, Vol. 21, No. 1, pp. 41-54.

Skaperdas, S. (1996), Contest Success Functions, *Economic Theory,* Vol. 7, No. 2, pp. 283-290.

Sloane, P. (1971), The Economics of Professional Football: The Football Club as a Utility Maximizer, *Scottish Journal of Political Economy*, Vol. 17, No. 2, pp. 121-146.

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

Szymanski, S., S. Kesenne (2004), Competitive Balance and Gate Revenue Sharing in Team Sports, *The Journal of Industrial Economics*, Vol. 52, No. 1, pp. 165-177.

Szymanski, S. (2004), Professional Team Sports Are Only a Game: The Walrasian Fixed-Supply Conjecture Model, Contest-Nash Equilibrium, and the Invariance Principle, *Journal of Sports Economics,* Vol. 5, No. 2, pp. 111-126.

Szymanski, S., T. M. Valletti (2005), Promotion and Relegation in Sporting Contests, *Rivista di Politica Economica,* Vol. 95, No. 3, pp. 3-39.

Tullock, G. (1980), Efficient Rent-Seeking, In: J. M. Buchanan, R. D. Tollison, G. Tullock (Eds.), *Toward a theory of the rent seeking society,* pp. 97-112, College Station: Texas A&M University Press, 1980.

Utt, J., R. Fort (2002) Pitfalls to Measuring Competitive Balance with Gini Coefficients, *Journal of Sports Economics*, Vol. 3, No. 4, pp. 367-373.

Vrooman, J. (1995), A General Theory of Professional Sports Leagues, *Southern Economic Journal,* Vol. 61, No. 4, pp. 971-990.

Vrooman, J. (2007), Theory of the Beautiful Game: The Unification of European Football, *Scottish Journal of Political Economy*, Vol. 54, No. 3, pp. 314-354.

Vrooman, J. (2008), Theory of the Perfect Game: Competitive Balance in Monopoly Sports Leagues, *Review of Industrial Organization,* Vol. 34, No. 1, pp. 5-44.

Winfree, J. A., J. J. McCluskey, R. C. Mittelhammer, R. Fort (2004), Location and Attendance in Major League Baseball, *Applied Economics*, Vol. 36, No. 19, pp. 2117-2124.

Winfree, J., R. Fort (2012), Nash Conjectures and Talent Supply in Sports League Modeling: A Comment on Current Modeling Disagreements, *Journal of Sports Economics,* Vol. 13, No. 3, pp. 306-313.

Appendix

Profit maximization, Revenue sharing, fixed talent supply with n>2

The allocation of talent is a result of individual profit maximization. With marginal revenues equaling marginal costs of n teams we obtain a set of n first order conditions. The sum of all winning probabilities is unity $\sum_{i=1}^{n} w_i = 1$ and in a closed league n equations stating that one teams gain of talent is everybody else loss of talent $(\sum_{j=1}^{n} \partial t_j)/\partial t_i = 0$. With revenue sharing the variables needed to be determined by this system are n winning probabilities, the marginal costs c of the league and, with n teams and $\partial t_i/\partial t_i = 1$ given, $n*(n-1)$ partial derivations $\partial t_j/\partial t_i$ for $j = 1, \dots, n \neq i$. Thus a set of $2n+1$ equations should determine $n + n(n-1) + 1 = n^2 + 1$ variables, which is only possible for $n = 2$. This follows from the fact, that with every additional new team 5 new variables have to be determined but only two new equations are obtained. Therefore for n>2 the equation system is underdetermined. One possibility to solve this set of equations is to make assumptions about the partial derivations $\partial t_j/\partial t_i$; e.g. Kesenne (2000) solved this problem by assuming that every team takes the equal share of talent loss of one teams talent gain, but each other assumption seems to be equally reasonable.

Profit maximization, Revenue sharing, flexible talent supply with n>2 – win maximization

In an open league the partial derivations are all determined with $\partial t_i/\partial t_i = 1$and $\partial t_j/\partial t_i = 0$ for all $j \neq i$ but with $n > 2$ the system of equations becomes quadratic and cannot be solved explicitly. With the average revenues being the best-response function of the teams, we also obtain a set of equations which explicit solution cannot be given.

5 Vector Similarity as a New Measure of Dynamic Competitive Balance in Sports Leagues

Martin Langen

Abstract: A new measure of competitive balance in sport leagues is proposed, based on a vector approach to measuring similarity. The standard measures, such as, the win percentage standard deviation ratio, consider the inequality per season only. Analyzing the inequality of a competition over time requires a different approach. The measure proposed here is well suited to determining the dynamic competitive balance of a league. As the application on the German Bundesliga indicates, the new measure is a useful complement to the existing measures of competitive balance and should enhance the analysis of outcome uncertainty in sports leagues.

Keywords: Sport economics, Vector similarity, Dynamic competitive balance, German Bundesliga

JEL: L83

1 Introduction

The demand for sporting events is closely associated with the uncertainty of outcome. The underlying assumption is that predetermined sport contests are uninteresting for viewers. However, analyzing the uncertainty of outcome hypothesis in a sport league is challenging, due to the many dimensions of uncertainty. In a sport league, one aspect of uncertainty is the outcome of a single match, yet another is the outcome of the whole season. Over time, some teams are more successful than others, so that the degree of unbalance of an entire season is an incomplete measure of competitive balance and outcome uncertainty.

As there is considerable variety in the interpretation of the uncertainty of outcome hypothesis, there are a number of different measures for capturing the various effects. The methods employed so far to measure the inequality of season outcomes do not consider changes in relative standings of the teams, or are unable to display the composition changes in a league over time, due to a promotion and relegation system. Therefore the contribution of this paper to the existing literature is to propose an indicator that determines the similarity of consecutive seasons, by measuring a season's outcome through a vector of team ranking, where the similarity of consecutive vectors is determined mathematically by the angle between them. A higher level of vector similarity implies less volatility in relative team standings, with a higher degree of predetermination of a league's seasonal outcome and, therefore, a lower degree of competitive balance.

The basic idea of outcome uncertainty in sport economics is so essential for every sporting contest, that the efforts devoted in the empirical literature to finding evidence of a clear connection between outcome uncertainty and the demand for sport events is hardly surprising. That a sport competition needs at least some degree of balance in order to be interesting for spectators has been well established in the economics literature since Rottenberg (1956) and Neale (1964). More recently, Szymanski (2003) distinguished between three dimensions of outcome uncertainty: (1) uncertainty of outcome concerning a single match, (2) uncertainty of seasonal outcomes and (3) uncertainty of consecutive seasons. In sharp contrast to the effort devoted to measuring outcome uncertainty, Borland and Macdonald (2003) found, in their review of the empirical literature, only mixed evidence of the relevance of outcome uncertainty. They conclude that if there is any relationship at all between outcome uncertainty and demand for sport, only seasonal and championship uncertainty matters, whereas there does not seem to be an influence of match uncertainty on overall demand.

The rest of this paper is structured as follows. A brief overview of the most commonly used measures of seasonal and championship uncertainty is presented in Section 2. In Section 3, a new method for measuring the similarity of consecutive seasons is presented. This method is applied to the point distribution, as well as to the rankings of seasons, so as to capture outcome uncertainty resulting from changes in relative team standings in a league. In Section 4, the measurement method is applied to determine the level of outcome uncertainty in the German Bundesliga. Additionally, the new index is compared with the well-known indices presented in the second section. Section 5 concludes.

2 Measures of outcome uncertainty

In order to measure *seasonal* outcome uncertainty, the standard deviation of winning percentages in a particular season is frequently used. The standard deviation may be set in relation to that of a totally unbalanced league (Gossens, 2006). However, comparison with an ideally balanced league is more common and was first suggested by Noll (1988) and applied by Scully (1989). They compared the actual outcome of a league with an idealized standard error of a league with a maximum degree of competitive balance. Quirk and Fort (1992) argued that a totally balanced league has n equally strong teams and each team has a fifty-fifty probability of winning each match they play. With g matches, the standard deviation of an ideally competitively balanced league would be $0.5/\sqrt{g}$. The major leagues in European football are double round-robin tournaments; thus, the standard deviation of a perfectly balanced league is $0.5/\sqrt{2(n-1)}$. The more teams in a league, the lower this idealized standard deviation of the league.

The spread around the average of wins in a league is an intuitive method of measuring the competitive balance of one season, but it is an inappropriate measure of the changes in competitive balance over time or of outcome uncertainty from changes in relative team rankings, because it treats teams anonymously. With n teams in a league, where a win is awarded w points and a tie t points and the teams play against each other twice, the maximum number of points per season is reached when a team is a winner in each game, i.e., the whole league has a sum of $2n(n-1)w$ points. If each game ends in a tie, the minimum number of points per season for a league is given by $4n(n-1)t$. As long as a win is rewarded with twice as many points as a tie, the absolute number of points is fixed for the league. However, with the introduction of the three-point system in European football, the absolute number of points awarded to the teams per season is no longer fixed. Therefore, the standard deviation cannot be used to compare the competitive balance over different seasons within the same league.

To analyze the dissimilarity of distributions, several measures are commonly used in economics. One of the best-known tools is the Gini coefficient, which was applied, for instance by Schmidt and Berri (2001, 2002) to determine the competitive balance in baseball. Since it is impossible for one team to win all games played in the league, Utt and Fort (2002) adjusted the Gini coefficient for an ideal unbalanced league. They showed that the unadjusted Gini coefficient overestimates competitive balance. Further criticism of the Gini coefficient by Utt and Fort (2002), like an unbalanced schedule, are not relevant for the major European football leagues, as each team plays against any other team in the league twice. Other measures of inequality or concentration are, for instance, the Herfindahl-Hirschman Index (HHI), applied by Depken (1999) and Owen et al. (2007), and relative entropy (Horowitz 1997).

The standard deviation approach, as well as the other inequality measures mentioned so far, determines the inequality of the point distribution of one season. Considering just the rank of the teams, with *n* teams in a league, there are *n!* possible team rankings in a league with the same inequality of point distribution. These inequality measures do not consider that changes in the relative team rankings matter for competitive balance. Even if a league is unbalanced according to the abovementioned measures, it may nevertheless be competitively balanced between seasons if relative team rankings change from season to season.

Sport viewers may not only be interested in whether a season was balanced, but also in whether the same teams always win. Even if their supported team failed in one season, viewers may still be interested in the team, as long as they hope that it might succeed in the future. Two seasons can have the same unequal outcome in terms of the standard deviation of winning percentages, and at the same time differ drastically in terms of relative team standings. One of the first researchers to address this issue was Humphreys (2002). He introduced the competitive balance ratio (CBR) as an extension to the standard deviation approach. He also used the standard deviation winning percentages of one season, but in addition the standard deviation over several seasons per team. The CBR is appropriate for displaying the abovementioned outcome uncertainty due to rank volatility. Nonetheless, the application to European football leagues remains problematic. Because of the widely used system of promotion and relegation, it is not possible to calculate the standard deviation for every team, as some are replaced over the seasons. Furthermore, as indicated above, the results are biased by the three-point-rule and the moving point average of leagues. With games ending in ties, the standard deviation measures may change without affecting the league rankings. Moreover, to calculate team-specific win variations and the winning percentages ratio, it is necessary to select a number of seasons to construct subsamples and the CBR. Without further studies of the determinants of viewer demand, the choice of length of these subsamples will remain arbitrary. For a league which uses a two-point system (2 points for a win, 1 point for a tie), or a league system in which only wins are counted (without the possibility of a tie), the CBR is an appropriate measure of dynamic competitive balance, as long as the league is closed. These limitations show that the CBR may be applied to North American sport leagues, although some problems, such as unbalanced game schedules and play-offs, still persist.

Lenten (2009) constructed a measure to capture the mobility of teams over time, based on a winning percentage average and the gap to an ideal competitive balanced league, but again, the application to European football remains problematic, due to the moving average of overall league points resulting from the three-point-system. Rank correlation coefficients like Kendall's Tau (Grott, 2008), or Spearman's rho (e.g. Maxcy and Mondello, 2006, or Andreff and Raballand, 2009), are not appropriate for measuring the overall similarity of seasonal rankings, as they are not applicable to point distributions. Moreover, the promotion and relegation system again complicates the application to European football leagues.

Championship uncertainty is another dimension of outcome uncertainty. The more balanced a league, the more teams are potential winners of the championship. The distribution of championships is then analyzed by simply counting championships or using inequality measures for championships as in Quirk and Fort (1992). This approach might be useful for American sport leagues, where a championship is the only title that can be won. The competition in European football leagues is such that it is not only about national championships. Teams also compete to gain the right to participate in a European competition in the next season. Buzzacchi et al. (2003) compared the number of different teams in the top positions of leagues with outcomes suspected for a totally balanced league. Compared to the short-term outcome uncertainty proposed in this paper, this is a dynamic measure showing the long-term outcome uncertainty of a league. Moreover, in addition to the suspense resulting from the race to the top positions, the usual relegation system creates additional spectator interest in the bottom of the league. For teams that know they will not achieve top positions, the mere avoidance of relegation might be considered a success. In European football leagues, there are hardly any positions in the ranking where spectators and fans find the competition boring. Even teams in the midfield can fight for a surprising slight shift towards the top and the right to play a supranational tournament next season, just as they should be prepared to struggle against relegation. These multi-dimensional aspects of outcome uncertainty in European football leagues have not yet been reflected in the sport economics literature. Accordingly in this paper, a new measure of league volatility is proposed that can capture more of these aspects. Manasis et al. (2013) describe European football as a three-level tournament (Championship title, European tournament, Relegation) and developed a concentration ratio for all three tournaments within a season. The new measure in this paper contains all positions in a league's ranking system. As long as there is sufficient variability in the relative standing of teams from season to season, a sport league might be perceived as balanced, although the abovementioned static methods would yield different conclusions. In contrast to other measures of dynamic competitive balance, the proposed season similarity index does not treat teams anonymously. Furthermore, the measure is applicable for every point system regime, as well as for leagues with promotion and relegation systems.

3 Season-to-season similarity

The competitive balance of a league does not depend exclusively on the inequality of outcome in one season; the degree of change from season to season exerts an additional affect. More changes from season to season indicate a higher degree of unpredictability of seasonal outcome or a higher level of uncertainty, whereas a higher degree of similarity of seasons signals a lower level of competitive balance. In the following analysis, a measure of changes between seasons (i.e., the similarity of consecutive seasons) is proposed, based on a vector angle concept, sometimes referred to as the cosine measure, e.g. Jones and Furnas (1987); Busch

(1998).[1] In terms of this concept, the outcome of each season is interpreted as a vector, whose dimension is given by the number n of teams $i=1,..,n$ in a league. Let the vector $a=(a_1,\ldots,a_n)$ represent the seasonal outcome at time t and the vector $b=(b_1,\ldots,b_n)$ the outcome at t-1. The *similarity between two vectors* can be determined by the cosine between them. With α as the angle between these vectors, the cosine is given by the Euclidean dot product, which is the dot product of the two vectors, divided by both vector lengths:

$$\cos\alpha = \frac{a \cdot b}{\|a\|\|b\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}. \qquad (1)$$

The cosine of each angle is within the range -1 to 1. If the two vectors, and therefore the seasons they represent, are identical, the angle α between them will be zero with a cosine value of 1. The application of the cosine measure to seasonal outcomes of sport leagues limits the feasible realization of angles. In the range employed here, the cosine is a strictly monotone decreasing function of α. Consequently, smaller values of the cosine display a higher degree of dissimilarity between vectors. In the following analysis, the similarity between two seasons is compared by their point distributions (referred to by the index p) as well as the relative changes in ranking positions (referred to by the index r). With the maximum feasible angle between two vectors, the season-to-season similarity indices σ_p and σ_r are normalized to values between 0 and 1, with 1 indicating that they are identical and 0 meaning the greatest possible change between seasons.

Point-distribution similarity

The distribution of points gained in a league is the foundation of the abovementioned static measure of outcome uncertainty within one season. By comparing the teams anonymously, these measures are not appropriate for capturing outcome uncertainty, due to relative changes in seasonal

---

[1] In chemistry, as well as in physics, vector angles are also used to measure similarities and dissimilarities; see, for instance, Klein (1995) and Ruch, Schranner and Seligmann (1978).

rankings. The point distribution of two consecutive seasons will always be an ordering from the most points to the fewest points. Thus, the vectors a, b will always point in roughly the same direction. Therefore, the angles between the two seasons' vectors will be small. The greatest difference between seasons would be achieved by comparing a maximum unbalanced with a totally even league. The latter is given by the identical number of points for each team. A totally unbalanced league ranking is given by one team winning all its games and the second team winning all but those against the first team and so on.

Proposition 1: The similarity between two seasons with respect to their point distribution is given by:

$$\sigma_p = \frac{\cos \underline{\alpha_p} - \cos \alpha_p}{\cos \underline{\alpha_p} - 1}, \tag{2}$$

with $0 \leq \sigma_p \leq 1$ yielding the maximum feasible angle: $\cos \underline{\alpha_p} = \sqrt{\frac{(n-1)}{4(2n-1)}}$.

Proof: The maximum feasible angle between two vectors representing seasonal point distributions is defined as the angle between a totally balanced and a totally unbalanced league. Without loss of generality, let a^p be the vector of a completely unequal outcome of a double round-robin tournament with n teams and let b^p be the vector of a completely equal outcome of the same tournament with w points for a win and t points for a tie. Then, a^p and b^p are given by:

$$a^p = (2(n-1)w, 2(n-2)w, \dots, 2w, 0w),$$
$$b^p = (2(n-1)t, \dots, 2(n-1)t). \tag{3}$$

With $\sum_{i=0}^{n-1} i = \frac{1}{2}(n-1)n$ the maximum angle is given by:

$$\cos \underline{\alpha_p} = \frac{a^p\, b^p}{\|a^p\|\|b^p\|} = \frac{2(n-1)^2 n\, t\, w}{4\, n\, t\, w\sqrt{(n-1)^3(2n-1)}} = \sqrt{\frac{(n-1)}{4(2n-1)}}. \tag{4}$$

//

For an explanation of Proposition 1, note that the lowest degree of similarity between two seasons, given by (4), is determined solely by the number of teams playing in the league and is independent of the number of points rewarded for a win or a tie. The index σ_p is therefore applicable to all kind of point systems.[2] If two seasons have the same point distribution, it follows that $\sigma_p = 1$. If there is a change from a totally balanced to a totally unbalanced league, or vice versa, it follows that $\sigma_p = 0$. Higher index values indicate a higher degree of similarity of the point distribution.

Rank similarity

Changes in relative team standings are more interesting than changes in the point distribution. As mentioned, the vectors representing seasons according to their point distribution will always point roughly in the same direction, as they are orderings from the most to the fewest points. However, since the vector a_r denotes the order of the ranks $a_r=(1,2,\ldots n\text{-}1,n)$, each team is represented by its rank for the respective season *t*. The next season's vector b_r is the ordering of the previous season *t-1,* but with each team being assigned the rank of Season *t*, for instance $b_r=(2,\ldots,n\text{-}1,n,1)$. In this example, the last team of Season *t-1* was the first team in Season *t,* whereas all other relative positions remained unchanged. This approach is easily applicable in a closed-league system. In an open league, with a system of promotion and relegation, as in all major European soccer leagues, it is necessary to decide how to represent promotion and relegation. In this paper, the rank numbers of the relegated teams of Season *t-1* are assigned to the promoted teams of Season *t*. With this approach, both vectors will have the same length.

Proposition 2: The similarity between two seasons, with respect to relative team standings, is given by:

$$\sigma_r = \frac{\cos \alpha_r (2n+1) - (n+2)}{n-1}, \qquad (5)$$

with $0 \leq \sigma_r \leq 1$ and the maximum feasible angle given by: $\cos \underline{\alpha_r} = \frac{2+n}{1+2n}$.

[2] Furthermore, it is also applicable to vectors with different Euclidean lengths. An overview of how the cosine reacts to different Euclidean lengths can be found in Egghe and Leydesdorff (2009).

Proof: Both vectors are permutations of each other with n rank numbers. The product of both lengths is the sum of the n first square numbers:

$$\| a_r \| \| b_r \| = \sqrt{\sum_{k=1}^{n} k^2}\sqrt{\sum_{m=1}^{n} m^2} = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}. \qquad (6)$$

The maximum angle is determined by the largest number of possible changes in relative rankings. Because the denominator of (1) is fixed by the number of teams *n,* the maximum angle and therefore, the lowest feasible cosine, is given by the lowest dot product. The highest number of changes in relative standings is achieved by a completely reversed ordering. The lowest feasible dot product is attained when the highest value of one vector is multiplied by the lowest value of the other vector, the second highest by the second lowest, and so on. This is a second degree arithmetic progression equal to: $\sum_{i=1}^{n}[(n+1)i - i^2]$. The maximum angle is therefore:

$$\cos \underline{\alpha_r} = \frac{6\sum_{i=1}^{n}[(n+1)i-i^2]}{n(n+1)(2n+1)} = \frac{2+n}{1+2n} \qquad (7)$$

and the minimum angle is zero. With these values for α_r, the index σ_r can be normalized to values between zero and one. //

Note that for two seasons which are equal in relative rankings, it follows that $\sigma_r = 1$. If the largest feasible number of changes occurs, then the index value is given by $\sigma_r = 0$. Higher index values indicate a greater similarity of the relative ranking.

Proposition 3: The average value of σ_r in a totally balanced league is 0.5. Values higher (lower) than 0.5 indicate that the relative ranking volatility is lower (higher) than in a balanced league.

Proof: The average value of σ_r for all *n!* possible changes from season to season, which occur with the same probability, is given by:

$$\frac{\sigma_r}{n!} = -\frac{n+2}{n-1} + \frac{1}{n!}\frac{6(2n+1)}{n(n+1)(n-1)}\sum_{k=1}^{n!} a * b^k, \qquad (8)$$

with $\sum_{k=1}^{n!} a * b^k$ being the sum of all possible rank changes. Thus:

$$\frac{\sigma_r}{n!} = -\frac{n+2}{n-1} + \frac{1}{n!}\frac{6(2n+1)}{n(n+1)(n-1)}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} i\, j\right)\frac{n!}{n} = \frac{6n^2(n+1)^2}{4n^2(n+1)(n-1)} - \frac{n+2}{n-1} = 0.5. \qquad (9)$$

Proposition 4: The index σ_r places no weight on the absolute position in a rank change. As long as the changing teams have the same distance between each other in the ranking, the index will have the same value.
Proof: Let b^1 and b^2 be two vectors representing the rankings of seasons in which only two teams, both times having the same distance *j* in the ranking, change positions, compared with a season *a,* then the difference between the indexes σ_{r1} and σ_{r2} will be zero, independent of the absolute positions of the teams.

$$\sigma_{r1} - \sigma_{r2} = \frac{6\ a\ b_1}{n(n+1)(n-1)} - \frac{n+2}{n-1} - \frac{6\ a\ b_2}{n(n+1)(n-1)} + \frac{n+2}{n-1}$$

$$\Leftrightarrow \sigma_{r1} - \sigma_{r2} = \frac{6}{n(n+1)(n-1)}(a\ b_1 - a\ b_2)$$

$$\Leftrightarrow \sigma_{r1} - \sigma_{r2} = \frac{6\left(-\left(i^2+(i+j)^2\right)+2i(i+j)+(i+a)^2+\left((i+a)+j\right)^2-2(i+a)(i+a+j)\right)}{n(n+1)(n-1)}$$

$$\Leftrightarrow \sigma_{r1} - \sigma_{r2} = \frac{6}{n(n+1)(n-1)}(0) = 0 \qquad (10)$$

//
Hence, the index measures relative positional changes only and captures the rank volatility and rank similarity of a league.

4 Competitive balance in the German Bundesliga

4.1 Season-to-season similarity

With these similarity measures, the dynamic competitive balance of the German Bundesliga can be analyzed. Figure 1 shows the development of seasonal similarity measures for the German Bundesliga and Table 1 presents the respective descriptive statistics. It is clear that seasons differ only marginally in their point distributions.
Even given that season outcomes in points are an ordering and that their vectors, therefore always point roughly in the same direction, the similarity is nearly always above 0.9. The changes from season to season are relatively small and the point distributions of a season are also similar each year. By contrast, the changes in relative team standings are much more pronounced. As noted above, in a totally balanced league, each ranking and therefore each value between one and zero for the index, has the same

probability. On average, a competitively balanced league would have a rank volatility of 0.5. In the German Bundesliga, the average rank volatility is 0.765, with a minimum rank volatility of 0.551. The latter represents the largest season-to-season change in relative team standings and occurred in the 1969/1970 season. Overall, the team rankings in the German Bundesliga are largely determined by their previous relative positions in the league.

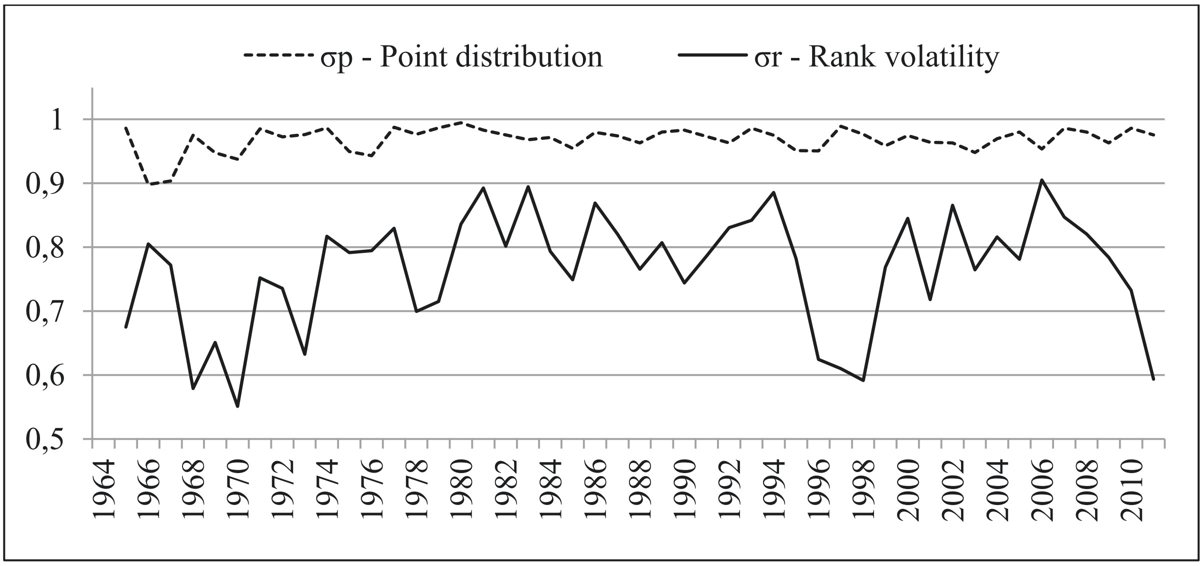

Figure 1: Season-to-season changes in the German Bundesliga from 1964 to 2011 (based on data from bundesliga.de, 2011)

Table 1: Descriptive statistics of season-to-season changes in the German Bundesliga from 1964 to 2011

	σ_p - Point distribution	σ_r - Rank volatillity
Mean	0.968	0.765
Median	0.975	0.784
Maximum	0.995	0.905
Minimum	0.898	0.551
Std. Dev.	0.020	0.090

Source: Own estimation based on data from bundesliga.de (2011).

4.2 Season-to-season similarity and static competitive balance

The season-to-season similarity measure introduced in this paper is a useful supplement to the established static measures of competitive balance. Figure 2 shows the development of various measures of competitive balance of the German Bundesliga.

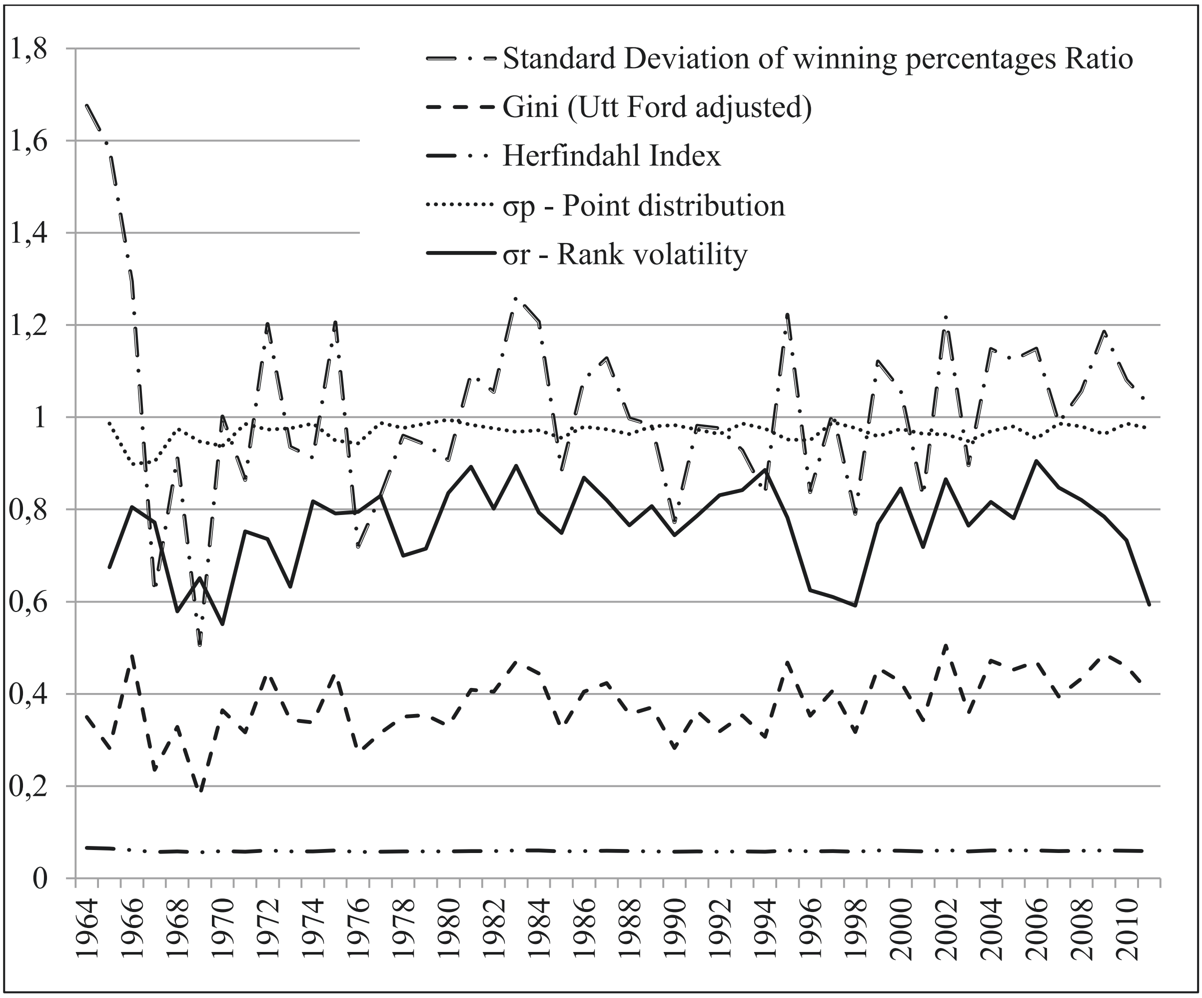

Figure 2: Measures of the competitive balance in the German Bundesliga 1964 to 2011[3] (based on data from bundesliga.de, 2011)

The standard deviation of winning percentage ratio indicates that seasons in the Bundesliga are sometimes more unbalanced, and sometimes more balanced, than would be expected in an ideally balanced league. This

3 Note the different dimensions of the measures.

could be interpreted as if the German Bundesliga was more or less competitively balanced and that there was a high degree of outcome uncertainty. However this impression may be misleading as the rank volatility index proposed in this paper indicates that the relative standing of teams does not vary that much. As mentioned, in a totally balanced league, any outcome has the same probability. Thus, all values of rank volatility between zero and one should have the same probability and, consequently, the average value would be 0.5. This is not the case for the German Bundesliga, as the average value is 0.765. Hence, the changes in relative rankings are considerably smaller than in a balanced league. Even if the outcome of individual seasons might be considered balanced, the league as a whole seems to be unbalanced. The correlations between the various measures for competitive balance considered here for the German Bundesliga are shown in Table 2. First of all, the high correlation between the standard deviation of winning percentages and the Herfindahl Index is noteworthy. Furthermore, the Gini coefficient is highly significantly correlated with both measures, so that these three indices seem to evaluate the same kind of outcome uncertainty. Even using different approaches to measure the inequality in seasonal outcomes, they nevertheless evaluate the competitive balance of leagues in a similar manner.
The standard deviation of the winning percentage ratio, as well as the Herfindahl Index, is not significantly correlated with the two season-to-season similarity indices. Consequently, the competitive balance of one season seems to be independent of the outcome uncertainty which originates from relative positional changes from season to season. The two abovementioned measures of season similarity are not correlated with each other. The similarity in point distribution of succeeding seasons is independent of the similarity in relative rank positions. More surprising is the significant, but nevertheless relatively low correlation of the Gini coefficient and the rank volatility measurement, as both differ largely with respect to their approach to measuring competitive balance.

Table 2: Correlation of measures of competitive balance for the German Bundesliga 1964 to 2011

	Standard Deviation of winning percentages ratio	Gini (Utt Fort adjusted)	Herfindahl Index	σ_p - Point distribution	σ_r - Rank volatillity
Standard Deviation of winning percentages ratio	1	0.634***	0.952***	0.084	0.240
Gini (Utt Fort adjusted)		1	0.519***	0.020	0.339**
Herfindahl Index			1	0.062	0.159
Season-to-season changes					
σ_p - Point distribution				1	0.060
σ_r - Rank volatility					1

*** 1% level of significance, ** 5 % level of significance, * 10% level of significance.
Source: Own estimation based on data from bundesliga.de (2011).

Table 3 shows the average values for the measures used here over several decades of the German Bundesliga. In the last ten years of the sample, the standard deviations, as well as the Herfindahl Indices, yield values slightly above their averages, indicating a decreasing competitive balance in the German Bundesliga. On the other hand, the standard deviation of winning percentage ratio was lower in the new millennium than at the beginning of the Bundesliga. This indicates an increasing competitive balance. Over the entire period under consideration, the average value for the standard deviation ratio, as well as for the Herfindahl Index, oscillates moderately around their total average values. This indicates that the point inequality does not vary very much over time and it is actually close to the expected value in a totally balanced league. The Gini coefficient increases marginally over time, implying a decreasing competitive balance. However, these indices measure the outcome of seasons only, ignoring relative team standings. Treating teams anonymously biases the analysis of competitive balance in a league. The index for season-to-season similarity in team ranks increases over time for the German Bundesliga. This is a clear sign of a decreasing competitive balance. The position of the teams seems more fixed now than they were in the first decade of the German Bundesliga.

Table 3: Average values for measures of competitive balance for the German Bundesliga 1964 to 2011

	60's	70's	80's	90's	00's	Over all
Standard Deviation of winning percentages ratio	1.098	0.956	1.060	0.948	1.065	1.021
Gini (Utt Fort adjusted)	0.310	0.355	0.394	0.362	0.434	0.379
Herfindahl Index	0.060	0.058	0.059	0.059	0.060	0.059
Season-to-season changes:						
σ_p - point distribution	0.992	0.996	0.997	0.996	0.996	0.996
σ_r - rank volatility	0.696	0.732	0.823	0.747	0.815	0.765

Source: Own estimation based on data from bundesliga.de (2011).

5 Conclusion

A number of different measures can be used to determine the competitive balance of a sport league. However, outcome uncertainty has many dimensions and the majority of commonly employed measures focus on the imbalance within one season. They treat teams anonymously and concentrate on the point distributions or the win averages. As an additional dimension of competitive balance, the similarity between consecutive seasons is considered in this paper. The inequality of point distribution is an important part of the competitive balance, but the seasonal similarity does affect the competitive balance over time as well. Existing measures of dynamic competitive balance are barely applicable to an open league with a three-point system. In order to measure the similarity between consecutive seasons in leagues such as those of European football, a season-to-season similarity index is proposed in this paper. This is well suited to measuring the similarity between relative team positions in a league's ranking, as well as the similarity between point distributions. Both the team ranking and point distribution of one season can be represented by a vector. The similarity between consecutive seasons is then given by the angle between these vectors.

The similarity indices introduced in this paper indicate a development in the competitive balance of the German Bundesliga, which contradicts the well-established indices, e.g. the standard deviation. In the last few years, the spread around the average in win percentages reveals no clear development in the competitive balance, but at the same time, the volatility of rankings decreased, which clearly shows a lower degree of competitive balance. At present, the similarity between relative team standings is at an all-time high in the German Bundesliga. This implies a high degree of predictability of league outcomes.

As topics for further research, the dynamic competitive balance for all major European sport leagues can be recommended, with the aim of comparing the results with the static measurements. Furthermore, the application of the season-to-season similarity indices to leagues with unbalanced league schedules or playoff systems should be considered. Finally, the effect of the dynamic competitive balance on the demand for league sports should also be investigated.

References

Andreff, W., G. Raballand (2009), Is European football's future to become a boring game?, in: W. Andreff, ed., *Contemporary Issues in Sport Economics: Participation and Professional Team Sports*, Cheltenham: Edward Elgar, pp. 131-167.

Borland, J., R. MacDonald (2003), Demand for sport, *Oxford Review of Economic Policy*, Vol. 19, No. 4, pp. 478-502.

Bundesliga (2011), URL: http://www.bundesliga.de [Accessed 19.07.2011].

Busch, P. (1998), Orthogonality and Disjointness in Spaces of Measures, *Letters in Mathematical Physics,* Vol. 44, No. 3, pp. 215–224.

Buzzacchi L., S. Szymanski, T. M. Valletti (2003), Equality of Opportunity and Equality of Outcome: Open League, Closed Leagues and Competitive Balance, *Journal of Industry, Competition and Trade,* Vol. 3, No. 3, pp. 167-186.

Depken, C. A. II (1999), Free-agency and the competitiveness of Major League Baseball, *Review of Industrial Organization*, Vol. 14, No. 3, pp. 205-217.

Egghe, L., L. Leydesdorff (2009), The relation between Pearson's correlation coefficient r and Salton's cosine measure, *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 5, pp. 1027-1036.

Gossens, K. (2006), Competitive balance in European football: comparison by adapting measures: national measures of seasonal imbalance and top3, *Rivista di Diritto ed Economia dello Sport*, Vol. 2, No. 2, pp. 77-122.

Grott, L. (2008*), Economics, Uncertainty and European Football: Trends in Competitive Balance*, Cheltenham: Edward Elgar.

Horowitz, I. (1997), The increasing competitive balance in Major League Baseball, *Review of Industrial Organization*, Vol. 12, No. 3, pp. 378-387.

Humphreys, B. R. (2002), Alternative measures of competitive balance in sports leagues, *Journal of Sports Economics*, Vol. 3, No. 2, pp. 133-148.

Jones, W. P., G. W. Furnas (1987), Pictures of relevance: a geometric analysis of similarity measures, *Journal of the American Society for Information Science*, Vol. 38, No. 6, pp. 420-442.

Klein, D. J. (1995), Similarity and dissimilarity in posets, *Journal of Mathematical Chemistry*, Vol. 18, No. 2, pp. 321-348.

Lenten, L. J. A. (2009), Towards a new dynamic measure of competitive balance: a study applied to Australia's two major professional 'football' leagues, *Economic Analysis and Policy*, Vol. 39, No. 3, pp. 407-428.

Manasis, V., V. Avgerinou,, I. Ntzoufras, J. J. Reade (2013), Quantification of competitive balance in European football: development of specially designed indices, *IMA Journal of Management Mathematics*, Vol. 24, No. 3, pp. 363-375.

Maxcy, J., M. Mondello (2006), The Impact of Free Agency on Competitive Balance in North American Professional Team Sports League, *Journal of Sport Management*, Vol. 20, No. 3, pp. 345-365.

Neale, W. C. (1964), The peculiar economics of professional sports: a contribution to the theory of the firm in sporting competition and in market competition, *Quarterly Journal of Economics*, Vol. 78, No. 1, pp. 1-14.

Noll, R. G. (1988), Professional Basketball, *Stanford University Studies in Industrial Economics,* Working Paper No. 144.

Owen, P. D., M. Ryan, C. R. Weatherston (2007), Measuring competitive balance in professional team sports using the Herfindahl-Hirschman index, *Review of Industrial Organization*, Vol. 31, No. 4, pp. 289-302.

Quirk, J., R. D. Fort (1992), *Pay Dirt: the business of professional team sports*, Princeton, NJ: Princeton University Press.

Rottenberg, S. (1956), The baseball players' labor market, *Journal of Political Economy*, Vol. 64, No. 3, pp. 242-258.

Ruch, E., R. Schranner, T. H. Seligmann (1978), The mixing distance, *Journal of Chemical Physics*, Vol. 69, No. 1, pp. 386-392.

Schmidt, M. B., D. J. Berri (2001), Competitive balance and attendance: the case of Major League Baseball, *Journal of Sports Economics*, Vol. 2, No. 2, pp. 145-167.

Schmidt, M. B., D. J. Berri (2002), Competitive balance and market size in Major League Baseball: a response to Baseball's Blue Ribbon Panel, *Review of Industrial Organization*, Vol. 21, No. 1, pp. 41-54.

Scully, G. (1989), *The business of Major League Baseball*, Chicago, IL: University of Chicago Press.

Szymanski, S. (2003), The economic design of sporting contests, *Journal of Economic Literature*, Vol. 41, No. 4, pp. 1137-1187.

Utt, J., R. Fort (2002) Pitfalls to measuring competitive balance with Gini coefficients, *Journal of Sports Economics*, Vol. 3, No. 4, pp. 367-373.

6 The Election of a World Champion

Martin Langen, Thomas Krauskopf

Abstract. This paper examines the mechanisms by which a World Champion is chosen in the Formula One Championship. Furthermore it is analyzed whether there is a best method to do this. For this purpose we will discuss the methods used by the Fédération Internationale de l'Automobile (FIA) since the founding of the World Championship in 1950. We show how the election of a method affects the Formula One contest. We then give insight whether there is a best method to select a World Champion or not. We therefore discuss Arrow's Impossible Theorem with respect to this sports contest. Moreover we simulated several seasons and compared different scoring vectors with respect to indicators that might be important for viewer's demand.

Keywords: Social Choice, Aggregation rules, Ranking, Sport, Formula One

JEL Classification: D71 (Social Choice), L83 (Sports), C63 (Simulation Modeling)

1 Introduction

The Formula One World Championship is a car race series with a long tradition. It started in 1950 and became one of the biggest sports events worldwide. During the last sixty years there were several rule changes regarding the organization of the competition between drivers and teams and with respect to the determination of the world champion. In the beginning of 2009 the Fédération Internationale de l´Automobile (FIA), which organizes the Championship, intended to change, among other things, the rules for choosing the Formula One World Champion.[1] This was not the first time the FIA modified the rules. Since 1950, there were actually four major changes. However the negative reaction to the suggested rule change was quite intense. The intention of the FIA was to modify the selection of the world champion via a scoring function and to install a rule where only the numbers of victories are decisive. The announcement of this concept started a public debate after some newspapers noted the fact that several of the past world championships would have had different outcomes if the new rules would had been employed.[2] The following protests were successful and the FIA first postponed the introduction of the new rules to 2010 and then skipped the plan altogether. Instead the FIA introduced a modification of the traditional scoring vectors.

However there are several questions that need to be answered: Why had the protest been so intense? How much would the rule change have mattered? The central question, though, is: Is there a best way to determine a world champion and if there is how does it look like?

There are several studies that did research on the topic Formula One. Kipker (2002) and Krauskopf, Langen and Bünger (2010) did research in television viewer's demand and Mastromarco and Runkel (2004) examined the relationship of rule changes and competitive balance. Furthermore Stadelmann and Eichenberger (2008) wanted to find the best driver of all

---

[1] Further changes dealt with budget limits and technical limitations.

[2] Of course it is too simple to review a rule change by comparing what this rule would have done in the past without anticipating a behavioral change of the actors. This is common economic knowledge since Robert Lucas (1976).

time. But as far as we know there are no studies examining different ways of the world champion determination.

To analyze this question we examine in section 2 different scoring vectors and within a simple model we show in how many cases different scoring vectors would nevertheless produce the same world champion. To analyze whether there is a best way to determine the best driver we examine in section 3 the Arrow Theorem and apply it to the Formula One competition. We show that with some simple assumptions the only possible method in this sport contest is to select the world champion via a scoring function. In section 4 we analyze some aspects besides deciding the world championship that might have relevance for choosing an appropriate scoring vector. We analyze the rule change of 2010 and give some insight on how a best scoring vector could look like. The fifth section concludes.

2 Scoring vectors of the Formula One

A scoring function assigns for every alternative, depending on the place in a single ranking, a specific number of points and is aggregating several rankings to one final ranking by ordering via the total points of each alternative. This basic scheme can be modified in many ways. In the history of the Formula One World Championship six different scoring functions or more precisely scoring vectors were used.[3]

Table 1 shows all the vectors that were used, labeled with the year they were established. With the proposed but now rejected rules everything would have stayed the same except for the determination of the first place in the final ranking. For this purpose only the first places in each race would have been taken into account. This can be interpreted in terms of social choice as a simple plurality vote (Gaertner 2006) and the vector

[3] Until 1991 it was usual that not all results were counted. Until 1959 the driver with the fastest lap in a race earned an addition point. As we are interested only in the difference the vectors make we are subtracting from these facts.

would look like a2010fia.[4] It is important to notice that the suggested vector should only determine the world champion. All the other places in the final ranking would have been determined by a2003. One reason for this could have been that a simple plurality vote in some cases does not result in a complete transitive order. This paradox was first described by Condorcet (1785) and occurred in the Formula One in 2002 (Soares et al., 2005).

Table 1: Scoring vectors in the Formula One

Scoring vector	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**
a1950	8	6	4	3	2	0	0	0	0	0
a1960	8	6	4	3	2	1	0	0	0	0
a1961	9	6	4	3	2	1	0	0	0	0
a1991	10	6	4	3	2	1	0	0	0	0
a2003	10	8	6	5	4	3	2	1	0	0
a2010	25	18	15	12	10	8	6	4	2	1
a2010fia	1	0	0	0	0	0	0	0	0	0
a2010fota	12	9	7	5	4	3	2	1	0	0

It is interesting, especially when changing from one scoring vector to another, in how many cases different scoring vectors deliver different aggregations. It is trivial that different scoring vectors can lead to different rankings, except if they are monotone transformations of each other. Saari (1984) has shown how many different rankings can be generated out of the same profile by using different scoring functions. Some further remarks on this topic in the Formula One came from Kladroba (2000). In the following we are examining only in how many cases different scoring vectors are leading to the same world champions applied to the same season outcome.

[4] The Formula One Teams Association (FOTA) suggested the a2010fota vector as an alternative for the a2010fia in the discussion of 2009.

For a small number of different outcomes this is quite easy and the simpler problems can be solved by using pen and paper. But in the case of the Formula One the numbers are far too big. With m drivers and n races in the season there are $(m!)^n$ possible different outcomes. By only examining the *p* places where points can be earned the number slightly reduces to $\frac{m!}{((m-p)!)^n}$. Taking the values of the 2009 season with *m=20, n=17* and *p=8* we get approximately $9{,}96 * 10^{164}$ different possible outcomes and this is far too great to calculate.
To measure the differences between two scoring vectors we therefore tested them with random samples. First we created a random outcome of one race and by doing this *n* times, we got a random season. Then we used the two scoring vectors we wanted to compare to determine the world champion for that simulated season. We simulated 1000 seasons determining the fraction of concordant world champions and repeated this 1000 times. We implicitly assume that every season outcome has the same possibility, i.e. all drivers are equally strong. We did this to give insight in all theoretical possible outcomes, not to give a realistic model.

Table 2: Descriptive Statistics of the comparison of a2003 and a2010fia

Mean	**Median**	**Max**	**Min**	**Std. Dev.**
0.529	0.529	0.584	0.476	0.016
Skewness	**Kurtosis**	**Jarque-Bera**	**Observations**	
-0.047	3.052	0.483	1000	

According to Table 2, which shows the descriptive statistics of the comparison between the vectors a2003 and a2010fia, the results are normally distributed. This is also confirmed by the significant low value of the Jacque-Bera test. The central limit theorem states that the mean of a sufficiently large number of independent random variables each with finite mean and variance will be approximately normally distributed. So our sample is sufficiently big enough and in Table 3 we are only reporting the means of our observations.

Table 3: Comparison of the vectors used in the Formula 1

	a1960	a1961	a1991	a2003	a2010	a2010fia
a1950	0.926*	0.894	0.894	0.766	0.741	0.542
a1960	1	0.928*	0.872	0.809	0.755	0.555
a1961		1	0.941	0.772	0.759	0.616
a1991			1	0.7301*	0.731	0.661
a2003				1	0.908*	0.529
a2010					1	0.367
a2010fia						1

Comparing all the vectors ever used in the Formula One there are several interesting properties. First of all, the values marked with an asterisk in Table 3 are the ones where a switch in the Formula One rules actually occurred. It is obvious that the rules suggested for 2010 would mark the biggest change in the history of the Formula One. Only in 52.9% the vector a2003 and a2010fia deliver the same world champion. The second biggest change in the history was in 2003 with 73% accordance to the previous year's vector. The vector actually introduced in the beginning of 2010 has accordance with the previous vector of 90.8%. The changes before 2003 were not only small in the degree of the vector change but also in the effects these changes have onto the world champion decision.

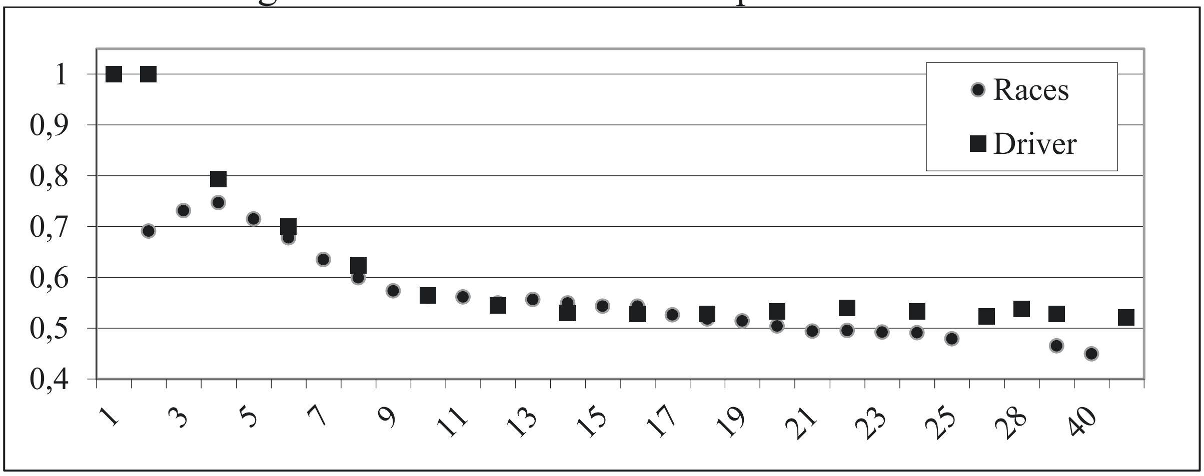

Figure 1: Comparison of a2003 and 2010 with depending on the number of drivers and races

This is not as radical as making only victories count but raises the question of how many additional points for the first place in the a2003 are necessary to obtain the same result as the a2010fia. We compared the a2010fia and a2003 and gave additional points for the first place in the a2003.

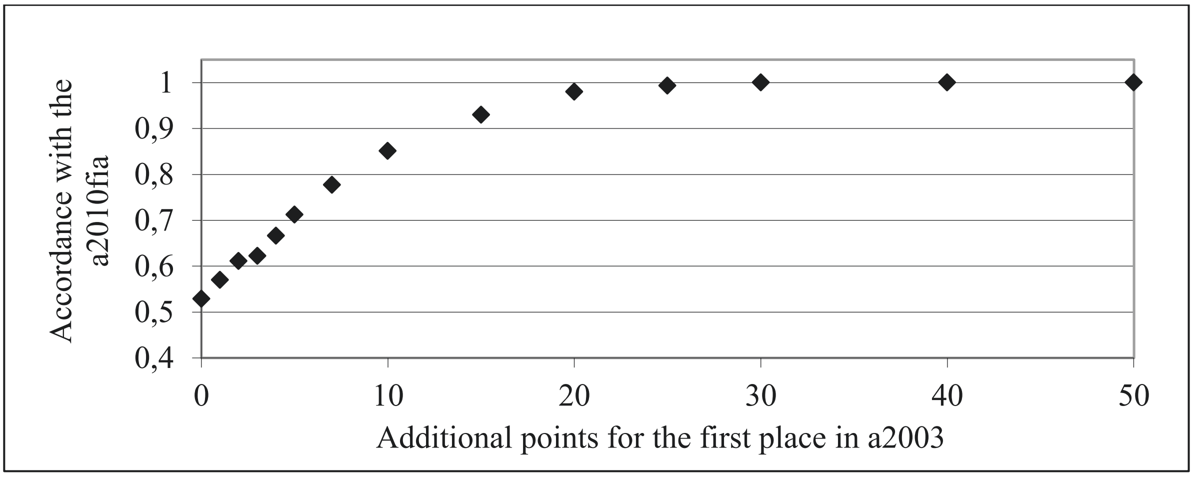

Figure 2: Accordance of a2010fia and modified versions of a2003

In Figure 2 we can of course see an increase which converts to 1, which means total conformity. It reaches total conformity with 30 additional points for the first place, hence 40 points in total. However also with 10 points, only twice as much as in a2003 amount, we get conformity of around 85%. The discussion of 2009 seems less dramatic from this perspective. As the numbers of drivers or races per season differ from time to time we tested the differences of the a2003 and a2010fia for alternative numbers of drivers and races. We can see in Figure 1 that for two drivers the vectors do not differ at all, which is not surprising considering that ties are decided with the respective other vector. After this the compliance steadily declines to a value of around 0.5 and with more than 11 drivers does not seem to alter anymore.

The picture is slightly different for alternative numbers of races. For one race the value is of course 1, has a second peak at 4 and after this is constantly going down. Taking a Formula One season with more than four races for granted we can say that more races mean more different results with these two vectors. Looking at the discussion of introducing a new point system at the beginning of the 2009 season it was common sense to strengthen the importance of the first place. The FOTA suggested a vector with a raise of 2 points for the first place.

3 Some Background from Social Choice Theory

After showing the differences the vectors of the Formula One have made in the election of a world champion we now try to identify a best method. Since there are several parallels between the determination of a winner in a Formula One season and the aggregation of preferences, social choice theory may provide some insights.[5] Aggregating the outcomes of all races into the determination of a world champion is similar to the aggregation of different preferences into the choice of a collective action.

Arrow (1951) mentioned four preferable conditions an aggregation function should comply with and he furthermore showed that no aggregation rule could simultaneously meet all four requirements. The four conditions are the unrestricted domain condition, the weak Pareto principle, the independence of irrelevant alternatives and non-dictatorship. The unrestricted domain condition requests that no individual preference ordering should be excluded *a priori*. If all individuals prefer one alternative with respect to another, the weak Pareto principle implies that this preference should also hold for the aggregated preference. The independence of irrelevant alternatives declares that the comparison of two alternatives should only depend on these two alternatives. The non-dictatorship condition demands that no individual should be able to determine the result of the preference ordering no matter how the other individual's preferences look like.

These conditions are to be applied to a Formula One season. To have a "fair" championship there should not be the possibility of excluding any theoretical possible outcome of the championship *a priori*. This is certain-

[5] For this we are subtracting from the fact that there might be changes throughout the season like strategy, driver´s form etc..

ly true for the championship. If the non-dictatorship condition was violated there would be one, and only one, race that determines the season's ranking making all other races irrelevant. This cannot be in the interest of the FIA so this condition should not be violated. In fact an even stronger version of non-dictatorship holds, called anonymity (Gaertner 2006). Anonymity means, that all races in the season are equally important for the final ranking. It is easy to see that also the weak Pareto principle should be satisfied. As there are only strict preferences in each race with the weak Pareto principle, a driver who is always better than another driver should also be better in the final ranking.[6] In combination with the other conditions the weak Pareto principle leads to neutrality. This means that by changing the positions of any two drivers in every race the same change will occur in the championship ranking. At first sight it should be clear that if these conditions are not met an aggregation rule would neglect fundamental intuitions about fairness in a sports contest. As shown by Arrow no rule can satisfy all conditions mentioned above. Therefore, all feasible aggregation rules employed by the Formula One will violate the independence of irrelevant alternatives. The comparison of two drivers in the final ranking will depend on the performance of the other drivers.

Apart from the conditions mentioned by Arrow, an aggregation rule in sport competition should also satisfy the condition of consistency. That means that drivers who are in the best set of every subset of the season, should also be in the best set of the whole season. Considering these fundamental rules, Young (1975) stated the characteristics of the aggregation rule: "A social choice function is anonymous, neutral and consistent if and only if it is a scoring function". If it is not possible to satisfy all preferable conditions, a more or less pronounced balance between these conditions should be found. By declaring some conditions to be more important than others, the only stable solution to aggregate the results of the races is a

---

6 This might not apply for those ranks in the races where zero points are earned. This can be disregarded when only the top positions in the championship are considered.

scoring function. And as we have seen, since the founding of the Formula One championship, this has been exactly the way the FIA aggregated the results.

4 Is there a best scoring vector?

The Arrow Theorem shows that the only adequate method is to aggregate the results via a scoring function. This leaves the question for the appropriate scoring vector. There are endless different possibilities how a scoring vector can be constructed. Some remarks for scoring vectors in sports contests came from Petigk (1990). In the Formula One six different scoring vectors were used. To answer the question whether there is a best scoring vector we first need an aim the scoring vector should be best in. The first but not the only goal is of course the selection of the season's best driver. But even the definition of the best driver depends highly on normative considerations. Is a constant driver better than a driver with a higher variance in places but more top results? Furthermore the selection of the best driver might be the most important but not the only objective the scoring vector should accomplish. The FIA, as the organizer of the world championship, could have a lot of alternative aims. We assume that the FIA is an organization with the goal of profit maximization. In the end the revenues the FIA can generate are highly depending on spectators' interest. Spectators′ interest is generating direct revenues and indirect via commercial revenues. So we think the objective goal of the FIA is the maximization of spectators' interest. A lot of the variables which are used to describe spectators' interest are based on the concepts of fairness and suspense or uncertainty of outcome (e.g. Simmons, 2009).
In the following we again simulate Formula One seasons and adopt different vectors to generate rankings. We then compare the results of these aggregations and show how the scoring vectors affect variables that are important for the viewer's demand. In contrast to the simulation in section 2 we now used different winning probabilities for the drivers. We assign for the drivers, who can also be interpreted as combinations of drivers and

cars, units of talent. The talent units of one driver divided by the sum of all drivers' talent units is his winning probability. The complete assignment of talent units is a talent distribution. We used six different talent distributions to compare the vectors and the particular talent units per driver can be seen, along with the Gini coefficient of these distributions, in Table 4.

Table 4: Talent distributions

Driver Distribution	1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	1	1	1	1	
2	4	3	2	1	1	1	1	1	1	1	
3	20	19	18	17	16	15	14	13	12	11	
4	10	6	3	1	1	1	1	1	1	1	
5	10	9	9	8	8	7	14	13	12	11	
6	1024	512	256	128	64	32	16	8	4	2	
Driver Distribution	11	12	13	14	15	16	17	18	19	20	**Gini**
1	1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1	1	1	0.204
3	10	9	8	7	6	5	4	3	2	1	0.317
4	1	1	1	1	1	1	1	1	1	1	0.397
5	10	9	8	7	6	5	4	3	2	1	0.499
6	1	1	1	1	1	1	1	1	1	1	0.844

In the first distribution every driver has one unit of talent which means that the possibility of every place is the same for every driver. In the second distribution driver 1 has four units of talent, driver 2 three units and so on. In the third distribution talent is constantly decreasing. The forth distribution is a more extreme version of the second distribution. In the fifth distribution we modeled three teams that are stronger than the rest and with a slight inequality in drivers' talent. In the sixth distribution, beginning with the tenth best driver, the drivers have the double amount of talent units compared to the next best driver.

A race simulation, for example with the second distribution, is calculated in the following way. We begin with the total talent population.[7] From this population a driver is chosen randomly, represented by one number. This is the winner of the first race. After this we removed all the other talent units of this driver in case there are any left. Then we are repeating this for all places until we have a whole race outcome. By repeating this *n* times we get a season existing of *n* races. For every value we present in the following, 1000 seasons were simulated

Next we have to choose variables that are proxies for viewer's attention. We assume that the longer a world championship is undecided the more suspense it has. Another cause for suspense might be the total amount of different champions over a number of seasons. For some viewers it could be important that the best driver becomes champion. The competitive balance or uncertainty of outcome is frequently used in the sport economics to describe viewer's attention (e.g. Quirk and Ford (1992)). Usually competitive balance is measured with the Gini coefficient. We compared different scoring vectors according to how they transform the unequal distributed talent into a point distribution at the end of the season. Our last variable is the number of cases, where the vector is not able to decide the world championship because of ties.

In Figure 3 we can see the average number of races after which the season is decided. For the most talent distributions (except the 6th distribution) the vectors which were actually used in the Formula One do not differ to a great extent. The a2010fia compared to the actually used vector delivers for some distributions longer and for some other shorter seasons. It is noticeable that the Borda[8] vector is generating the highest number of undecided races for all distributions.

---

[7] The vector in this case would be: (1,1,1,1,2,2,2,3,3,4,5,6,7,8,9,10,11,12,13,14,15, 16,17,18,19,20).

[8] The vector is named after the French mathematician Jean-Charles de Borda. The vector assigns one point for the last place and for every better place always one point more.

Figure 4 shows the number of different world champions the scoring vectors generate. For most of the cases the choice of a scoring vector does not matter as the number of different world champions does not change. Exceptions are the talent distributions 2 and 3 and the vector a2010fia. With both distributions the a2010 delivers more different world champions than the a2003 but less than the a2010fia.

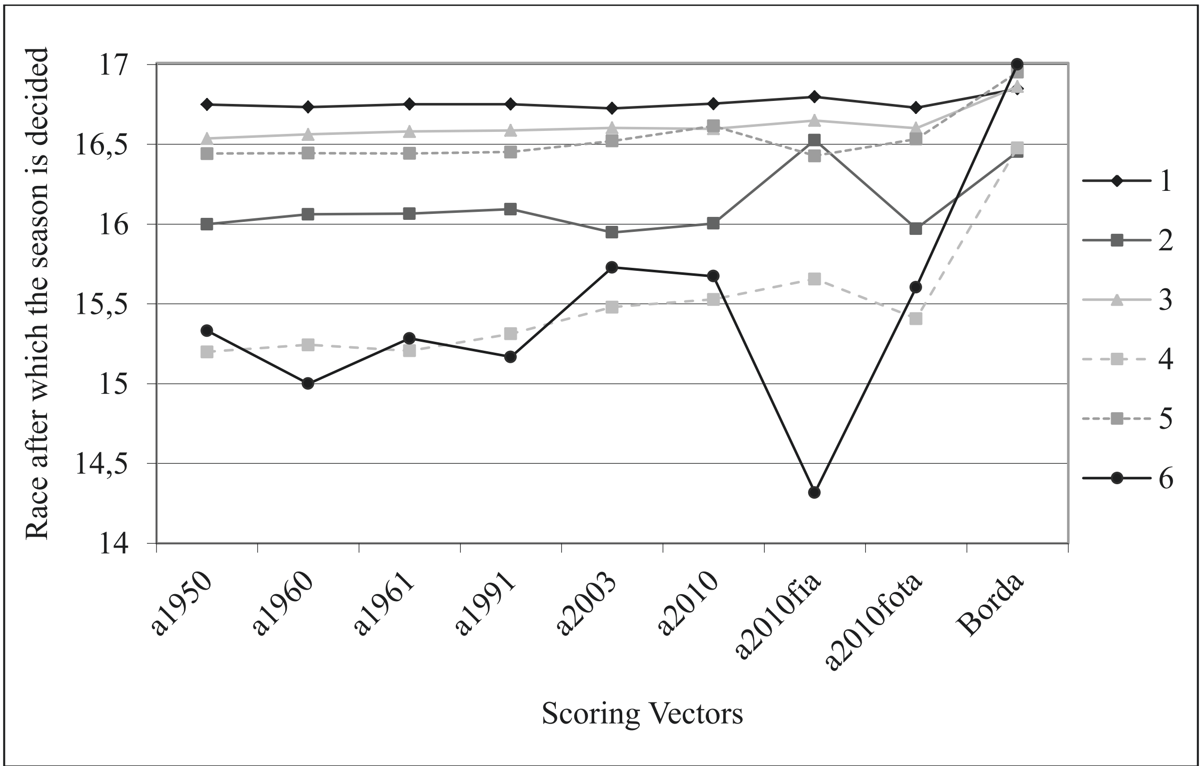

Figure 3: Average number of races after which a season is decided

As mentioned above the definition of the best driver highly depends on normative considerations. In our simulations we therefore tested the possibility of the most talented driver becoming world champion. As we can see in Figure 5 nearly all scoring vectors deliver similar results. So in most cases the vectors are not determining the probability for the best driver becoming world champion. The one exception is the vector a2010fia. For all cases where there actually is a most talented driver the appliance of the vector a2010fia leads to a lower probability for the best driver becoming world champion.

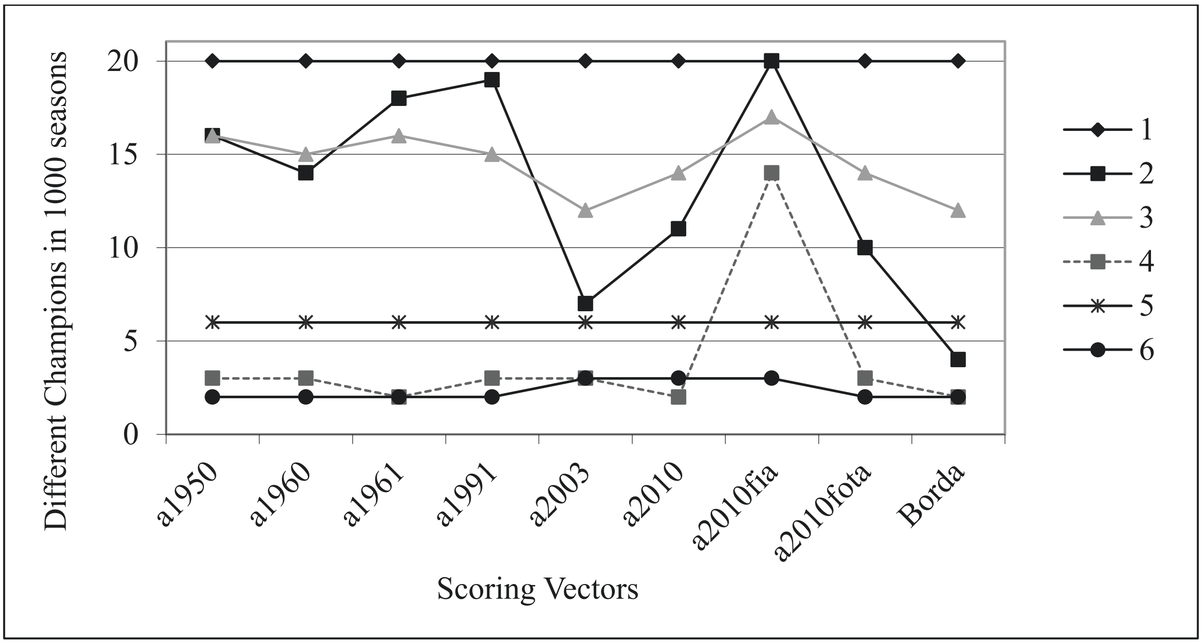

Figure 4: Different Champions in 1000 seasons

In figure 6 the Gini coefficients of the different talent distributions were plotted against the Gini coefficients of the point distribution in the final ranking. The graphs show how the vectors transform the inequality of talent into an unequal point distribution. On the 45° line the inequality of these two distributions is the same. We see that the highest inequality of scored points for every talent distribution, even with equally distributed talent, was produced by the a2010fia. The Borda scoring vectors has for every talent distribution the lowest corresponding Gini coefficient in point distribution. The second lowest coefficient for all talent coefficients has got vector a2010 and the third lowest a2003.

Every vector produces an unequal point distribution and the level of inequality depends, among other things, on the number of ranks the scoring vector is assigning point to. For this reason vector a2010fia, with only one point rank, always produces the highest inequality. The vectors which were used between 1960 and 2003 do not differ in the number of point ranks and produce nearly the same relation between the two Ginis. With the a2003, a2010 and the Borda vector the equality in the point distribution is rising with the number of point ranks.

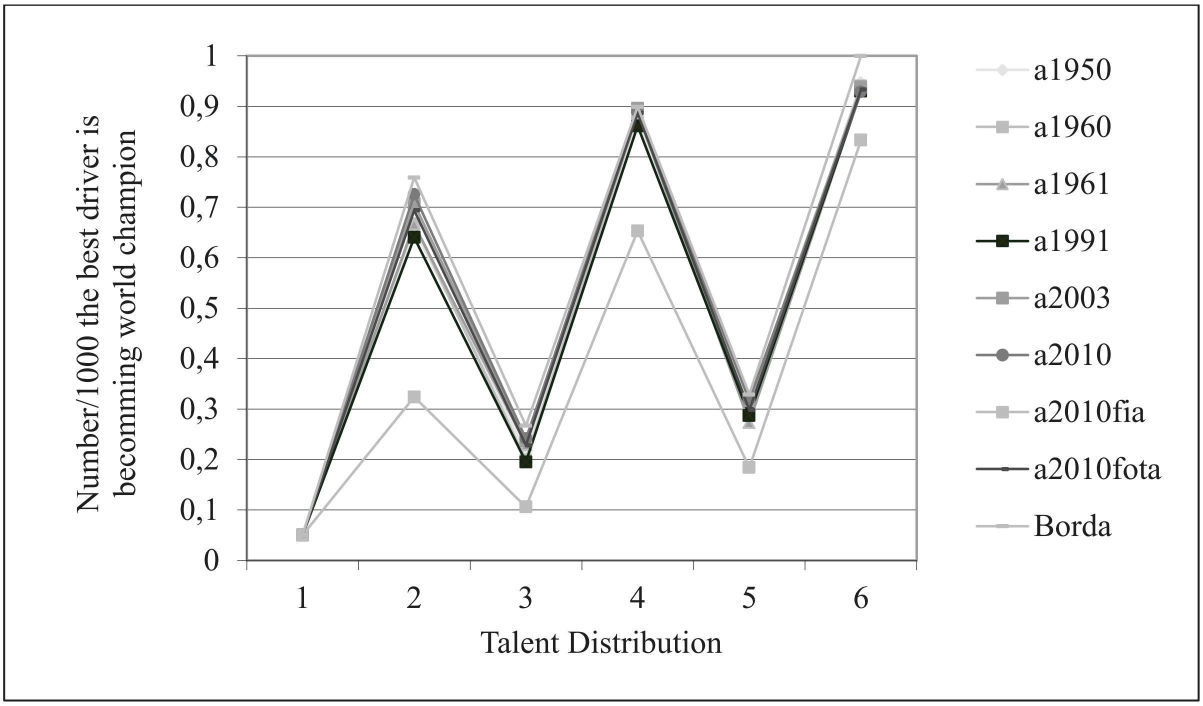

Figure 5: How often is the best driver becoming world champion

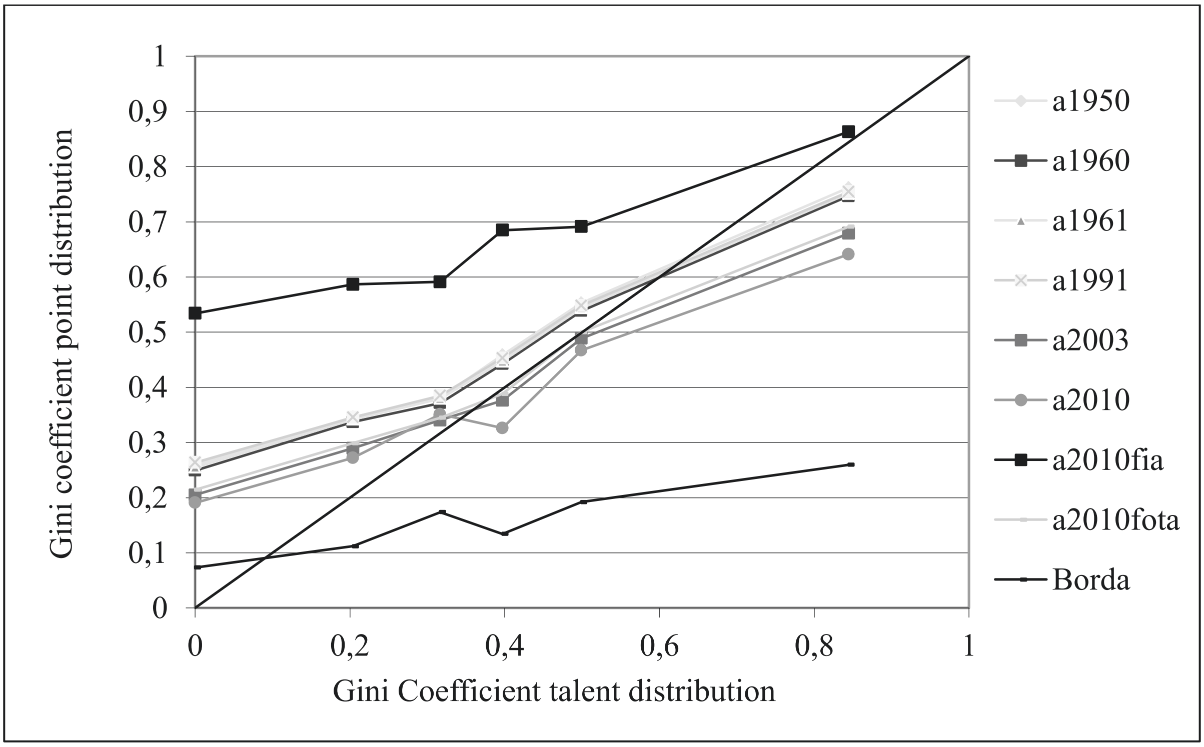

Figure 6: The effect of scoring vectors on competitive balance

Table 5 provides the fraction of 1000 simulated seasons a vector was not able to determine a world champion because ties occurred. The a2010fia has always got the highest fraction whereas the Borda has got the lowest in most of the cases. Comparing the a2003 and a2010 the latter is better in deciding world championships.

Table 5: In how many cases is the vector unable to decide the world championship?

	a1950	a1960	a1961	a1991	a2003	a2010	a2010fia	a2010fota	Borda
1	0.082	0.071	0.085	0.074	0.061	0.028	0.485	0.056	0.034
2	0.029	0.017	0.028	0.015	0.021	0.007	0.306	0.017	0.012
3	0.055	0.054	0.043	0.037	0.038	0.021	0.363	0.032	0.023
4	0.014	0.016	0.022	0.016	0.008	0.006	0.161	0.009	0.012
5	0.032	0.046	0.047	0.036	0.044	0.019	0.304	0.029	0.049
6	0.007	0.010	0.013	0.003	0.010	0.005	0.062	0.007	0.012

5 Conclusions

In this paper we have shown that the introduction of a majority vote in the election of a Formula One world championship would indeed have been the biggest change in the history of different scoring systems in the Formula One. We have shown with the Arrow Theorem that the only possible method to select a world champion in the Formula One is by aggregating the races via a scoring function. The questions whether there is a best scoring vector or not cannot be answered in general because it highly depends on normative considerations. We assumed that the FIA is a profit maximizer and therefore wants to maximize the viewer's demand. Thus we compared different scoring vectors for different talent distributions according to variables that are considered to be important for viewer's demand.

The vector the FIA introduced in 2010 does not affect the average real season; the one suggested at first would have had a bigger impact for this variable. The new vector also does not change the probability of the most talented driver becoming world champion. But for most distributions the

vector actually used in 2010 raises the number of different world champions. It also produces for all talent distributions a lower inequality in the point distribution. Altogether the introduction of the new vector seems to be a slight improvement in the variables we choose to analyze.
In general we can say that vectors with a higher number of point ranks generate longer undecided seasons. This leaves less interesting races at the end of the season. These vectors also generate lower inequality in the point distributions and thus a higher competitive balance. On the other hand fewer point ranks make the success of an underdog more realistic. Thus such vectors generate a higher number of different champions. To establish a best scoring vector more research on the specifics of viewer's demand is needed. Future studies should also examine the suspense in the races or in other words the fight for better positions in a race and this effect on viewer's demand. Our study shows that the decision for a longer or shorter point vector has positive as well as negative effects on viewer's demand. Therefore empirical studies of the importance of different factors for the viewer's demand may provide the solution for a best scoring vector.

References

Arrow, K. J. (1951, 1963), Social Choice and Individual Values, 2nd ed., New York: John Wiley & Sons.

Condorcet, Marquis de (1785), Essai sur l`application de l`analyse à la probabilité des decisions rendues à la pluralité des voix, Paris.

Gaertner, W. (2006), A Primer in Social Choice Theory, 1st ed., Oxford: Oxford University Press.

Kipker, I. (2002), Determinanten der kurzfristigen TV-Nachfrage in der Formel 1 – Superstar- vs. Spannungseffekt und Implikationen für die Wettbewerbspolitik, in: H. Dietl (eds.), *Globalisierung des wirtschaftlichen Wettbewerbs im Sport*, Schorndorf, Hofmann-Verlag.

Kladroba, A. (2000), The Problem of Aggregation Arising in the Process of Building Rankings, *Jahrbücher für Nationalökonomie und Statistik*, Vol. 220, pp. 302-314.

Krauskopf, T., M. Langen, B. Bünger (2010), The search for optimal competitive balance (forth-coming), Münster.

Lucas, R. E. (1976), Econometric Policy Evaluation: A Critique, *Carnegie-Rochester Conference Series on Public Policy*, Vol. 1, pp. 19-46.

Mastromarco, C., M. Runkel (2004), Rule Changes and Competitive Balance in Formula One Motor Racing, *Discussion paper 2004-16*, Department of Economics, University of Munich.

Petigk, J. (1990), Power Punkte Pleiten, 1st ed., Leipzig: Urania.

Quirk, J., R. D. Fort, (1992), Pay Dirt, 1st ed., Princton: Princeton University Press.

Saari, D. G. (1984), The Ultimate of Chaos Resulting from Weighted Voting Systems, *Advances in Applied Mathematics*, Vol. 5, No. 3, pp. 286-308.

Simmons, R. (2009), The demand for spectator sports, in: W. Andreff, S. Szymanski (eds.), *Hand-book on the Economics of Sport*, Cheltenham, UK, Northampton, MA, USA: Edward Elgar.

Soares de Mello, J., L. Gomes, E. Gomes, M. Soares de Mello (2005), Use of ordinal multi-criteria methods in the analysis of the Formula 1 World Championship, *Cadernos Ebape*, Vol. 3, No. 2, pp. 1-8.

Stadelmann, D., R. Eichenberger (2008), Wer ist der beste Formel 1 Fahrer? Eine ökonomische Talentbewertung, *Perspektiven der Wirtschaftspolitik*, Vol. 9, No. 4, pp. 486-512.

Young, H. P. (1975), Social Choice Scoring Functions, *SIAM Journal of Applied Mathematics*, Vol. 28, No. 4, pp. 824-838.

7 The search for optimal Competitive Balance in Formula One

Thomas Krauskopf, Martin Langen, Björn Bünger

Abstract: In this paper an analysis of the determinants of attractiveness of the Formula One is attempted. Therefore the concept of competitive balance will be explained and applied. The principal item of this analysis will be the result that there is an optimal level of competitive balance which maximizes the attractiveness. This concept will be applied through an empirical analysis of the determinants of the German Formula One number of TV viewers. It is shown that the influence of competitive balance is not totally clear. A too high level of competitive balance seems to be as detrimental as a too low level of competitive balance.

Keywords: Sports Economics, Formula 1, competitive balance, attractiveness of sports

JEL Classification: L83 (Sports)

1 Introduction

One of the main determinants of interest of the general public in sports is competitive balance. In addition to more formal criteria like the proverbial level playing field, the uncertainty of outcome is likely to be a major determinant of whether a discipline attracts millions of viewers or is just one of several side issue in Monday's local newspaper. The dominance of a single protagonist, from this point of view, is likely to exert an adverse effect on the number of TV viewers. At the same time, however, TV viewers like heroes and superstars which stand out from a crowd of competitors, and they like underdogs who unexpectedly win.

The main issue of this paper is how these two seemingly contradictory effects influence the attractiveness of motor racing, namely Formula One racing.

Formula One is a sport which is especially popular in Europe and South America. It is not researched in such a way as other big "American Sports" like Basketball, Baseball or American Football (Vrooman, 1995), although it is the biggest and most valuable annual worldwide sport event (Deloitte, 2008).[1] An interesting economic aspect of Formula One is the way the competition is being organized: In Formula One, we can observe, different from Basketball, Baseball or American Football, a direct simultaneous competition among competitors. Another interesting economic aspect of Formula One is that it aspects of both individual and team sports. On the one hand, it is the individual driver who attracts the attention of the audience's. On the other hand, identification with a team seems to play an important role for the audience. Therefore, Formula One combines elements of team sports and individual sports. Consequently, we are interested in analyzing the determinants of drivers' performance.

---

[1] Of course, there are some worldwide sport events which have more spectators like the Fifa Football World Cup, the Fifa Football Euro Championship or the Olympic Games, but these events are not organized annually. The NFL Superbowl also causes a high and often higher worldwide TV interest than a single Formula One race, but it only takes place once a year.

Our research builds on earlier research of the Formula One by Spenke and Beilken (2000) and Mastrmarco and Runkel (2004).Spenke and Beilken use the database of Formula One races to analyze data analysis tool. Mastromarco and Runkel analyze the effects of rule changes in Formula One and their impact on competitive balance. We also draw on research by Kipker (2003), who describes the determinants of competitive balance which are supposed to influence the suspense of competition. He describes three levels of competitive balance, uncertainty of race outcome, uncertainty of championship outcome and the absence of long-term dominance. The viewers interest is measured by the number of TV viewers. Kipker also describes the role played by superstars in Formula One.[2] Finally, our research builds on earlier research on the theory of sport leagues (Vrooman, 1995) and on research on the determinant of demand for sports (Borland and Macdonald, 2003). While the previous papers focus on competitive balance as a factor supposed to be instrumental for the attractiveness of competition, we argue that competitive balance is not the only factor influencing attractiveness. To this end, we shall describe a theory that renders it possible to trace out the optimal level of competitive balance. Our main assumption is that both a very low and a very high level of competitive balance are detrimental for the attractiveness of Formula One. The economic intuition motivating this assumption is that a sport is attractive for TV viewers if a championship is very close and competition is balanced. We, thus, show that this balanced competition is not the only competition-related determinant of attractiveness of Formula One. Especially duels at the top of the overall standing also play an important role. We structure the remainder of our paper as follows. In Chapter 2, we give a short overview of the theoretical assumptions underlying our theory, and we explain which factors determine the attractiveness of Formula One. We also explain how we quantify the attractiveness of Formula One. In Chapter 3, we present the data we used for our analyses. In Chapter 4, we

2 For a general exhibition of superstar effects, see Rosen (1981).

describe the empirical method and report our results. In Chapter 5, we conclude.

2 The theory of attractiveness of Formula One

The theoretical approach we will investigate is quite simple. The main point of our description is that there has to be an optimal level of competitive balance which maximizes the viewer's interest. We assume that Formula One is not attractive if all drivers have the same chances to win. In this case there would be no surprise if an unknown driver won a race or the championship. Thus a main condition for attractiveness is that there are real surprises like the victory of an outsider or the failure of a champion. On the other hand it is assumed that it is not really attractive if there is a driver who dominates the whole competition because this could lead to a situation where the championship is determined too early in the season. It is not our aim to show the exact optimum of competitive balance which maximizes the attractiveness. All we want to show is that this optimum could exist by analyzing the following two working hypotheses:

Hypothesis 1: A higher level of competitive balance leads to less attractiveness.
Hypothesis 2: A duel at the top leads to more attractiveness.

These preliminary hypotheses will be specified later. If the analysis of both hypotheses results in a non-rejection we can assume that our considerations are correct.

But before we can answer the question what influences the level of attractiveness of a sport it has to be clarified how to measure the attractiveness of a sport. Here it is suggested to operationalize attractiveness by viewer interest namely the number of TV viewers (Kipker, 2003). This measure has certain advantages: it is quantitative and in principle easy to measure. Of course, there are other possible indicators for attractiveness like personal preferences or tickets sold for seats at the circuit. In particular personal preferences could indicate the attractiveness much more accurate than the number of TV viewers. However, the problem is that it is hardly possible to directly collect data on preferences. An advantage of the number of tickets sold would be that they can be surveyed even more easily than the number of TV viewers. But as the number of seats at the circuits is limited this data is truncated.[3] This is the essential advantage of the number of TV viewers as a measurement because the number of viewers is quasi unlimited. In Germany, e. g., there are ca. 35 million television sets for 80 million citizens (GEZ, n. d.). In addition, no Formula One event ever came near a full use of the capacity, i.e., a viewer level around 90 percent. Thus, virtually everyone who wants to watch a Formula One race can watch it.

To answer the question how to find parameters that influence the attractiveness we suggest the concept of competitive balance (Utt and Fort, 2002). The level of competitive balance is high if there are similar chances to win for all participants. Referred to Formula One this means that all drivers have the same chance to win a race and the championship. The opposite is true for a low level of competitive balance. However, if you take into account that there are different Formula One cars with different engines which are driven by persons of different talents the assumption of same or just similar chances to win has to be doubted. To quantify competitive balance the Gini coefficient which is a concentration index is a very useful tool. It is mostly used in the analysis of team sports like baseball, basketball or American Football (Utt and Fort, 2002). The application for individual sports like Formula One, however, is rare. However, in this case the application of the Gini coefficient seems to be meaningful as the drivers score different points for ending a race in different places.

---

3 Thus, you cannot really estimate the common interest because you do not know how many tickets could be sold. One possibility to cope with this problem could be to include the shadow market prices for tickets. This procedure, though, would negate the advantages of easy access to data.

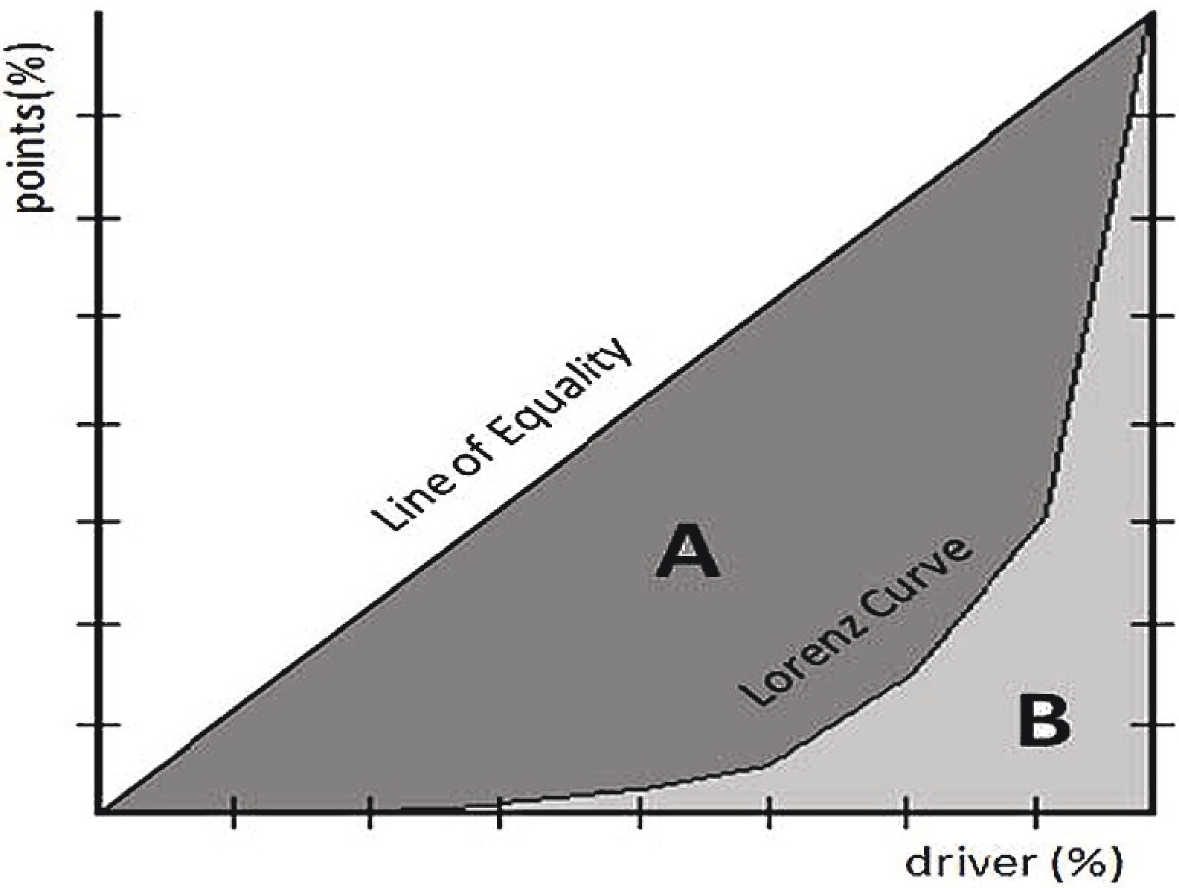

Figure 1: Graphical illustration of the Gini coefficient

In Figure 1 the concept of the Gini coefficient is illustrated in a graphic way. The denomination of the axes is the percentage of drivers and the percentage of points, which are gained by the drivers. The so-called Line of Equality and the Lorenz Curve are also depicted. The Line of Equality shows the hypothetical situation when all drivers gain the same amount of points. The Lorenz Curve shows the actual concentration of the points scored. Because the points, which are gained by the drivers in every race, are accumulated through the season there would be a situation of perfect equality or balance if 10 percent of all drivers accumulated 10 percent of all points and e.g. 50 percent of all drivers have 50 percent of all points. But in a situation where 50 percent of all drivers only have 10 percent of all points there is a smaller level of balance. This disparity can now be quantified by the Gini coefficient. The Gini coefficient is constructed by the ratio of area A to the sum of area A and B. A high Gini coefficient implies a high disparity which in turn represents a small competitive balance.[4] If a high level of competitive balance attracts more spectators than a smaller level, small values of the Gini coefficient are an indicator for attractiveness. An overview about the exact approach and the used scoring vector will be given in the following chapter.

A second index which can be used as a parameter influencing the attractiveness of Formula One is the relative distance between the first and the second driver in the overall standings. This distance signifies the competitive balance at the top of the ranking. If the difference between the first and the second driver becomes too large there is the problem that the sport could become boring. We assume that a very high relative distance between the first and the second driver in the overall standings reduces the spectator's interest because the exciting aspect of a duel of these two drivers disappears. Relative distance between first and second driver and not absolute distance between them is used as the amount of points and also the magnitude of the differences btween drivers increases during the season. Thus, the absolute distance after the third race and the last race, e. g. cannot be compared meaningfully.

3 The data employed in the analysis

In this chapter an overview about the data used in the analysis will be given. First, we will have a look at the number of TV viewers. As it can be seen in Figure 2 the viewer levels of the German TV network RTL for the seasons 1992 until the season 2009 are employed. Germany is one of the countries with the highest number of Formula One supporters and the exclusive use of German data is due to data availability.

The abscissa is scaled by the number of races starting with the first race in 1992. It is not useful to scale the abscissa in years because the seasons are not evenly distributed over the years. In addition, time between two races

4 The range of the Gini coefficient is between 0 and 1 where 0 means total equality while 1 denotes total concentration.

is not equidistant. But mostly the season starts in March and ends in October with two-week intervals between races. The number of races has varied in the different seasons and has increased over the years.

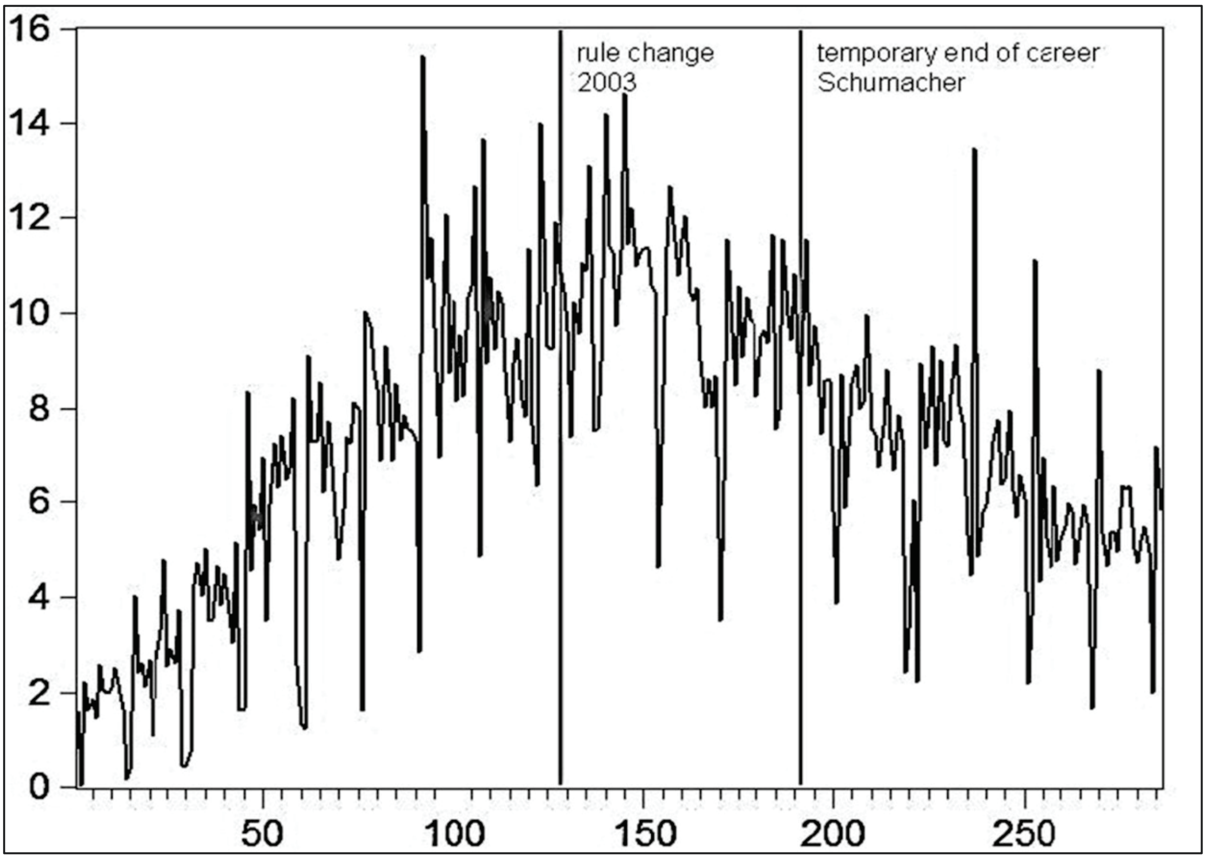

Figure 2: Formula One number of viewers of German RTL (March 1992 – November 2009, Media Control, 2010)

The ordinate is scaled with the number of RTL-TV viewers in million persons. Two important points in time are marked by the vertical lines. The first line marks the rule change at the beginning of the season in 2003 when the score vector was changed while the second line stands for the point in time when Michael Schumacher temporarily finished his Formula One career at the end of the 2006 season (October 2006). These two points in time are highlighted because they will be analyzed in the empirical chapter later on. The rule change, which will be explained more precisely in the next chapter, consisted of an enlargement of the point ranks.

The number of ranks where points could be gained was increased from 6 to 8 and the amount of scored points was also enlarged.[5]
As it can be seen in Figure 2 there are huge differences in the viewing levels between two consecutive races. One of the reasons for this is that there are races in American time zones, European time zones and Asian-Pacific time zones (compare Kipker, 2003). Thus, some races take place at times which are not convenient to European TV viewers because some Asian races start in the early morning. This fact will be taken into account in the empirical part of this survey. Another point to mention is that the first race of every season is not included in the analysis. The reason for this approach will be explained later on.

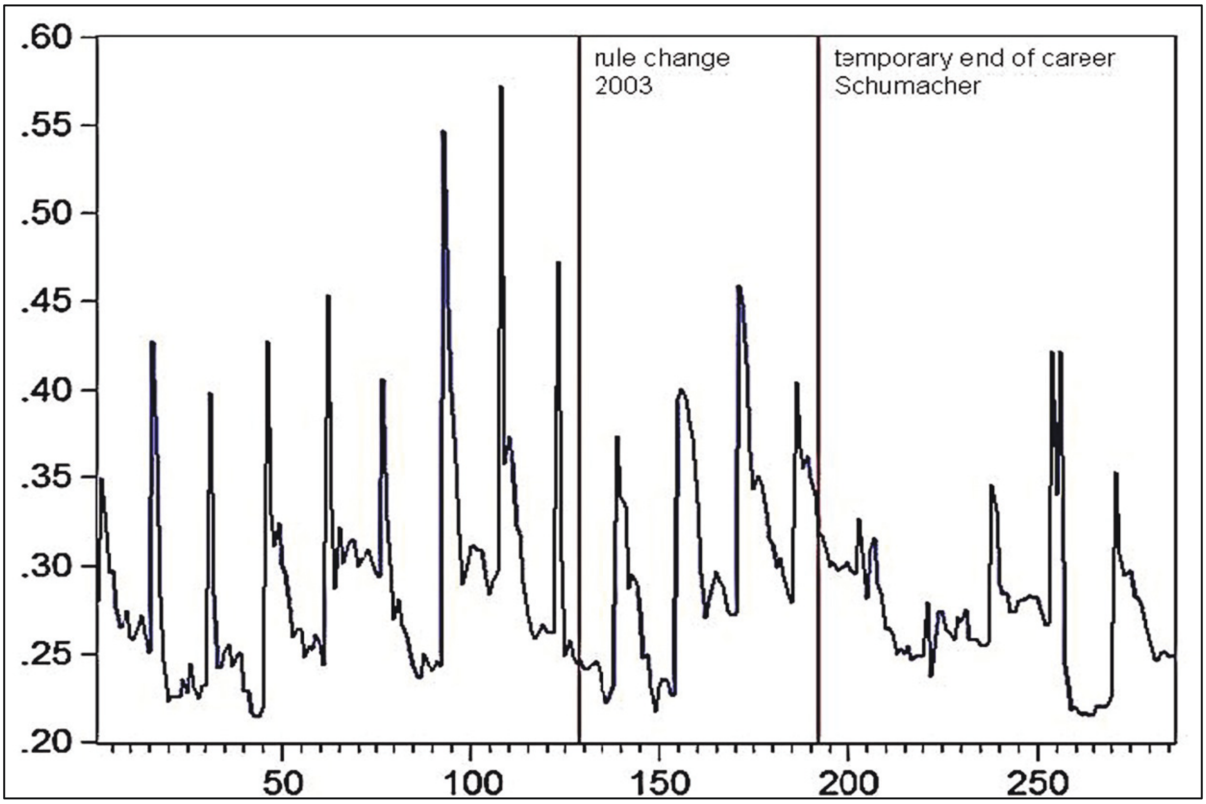

Figure 3: Values of the Gini-coefficent

[5] Thus, the scoring vector representing these rules changed from s_1 = (10; 6; 4; 3; 2; 1) to s_2 = (10; 8; 6; 5; 4; 3; 2; 1).

Figure 3 shows the Gini coefficient of the cumulated points achieved. This corresponds to the Gini coefficient calculated on the basis of the overall standings. An important point to mention is that the actual used scoring vector s2 was not employed here because in the original scoring vectors only the first six places and since 2003 only the first eight places are considered. The different number of results endowed with championship points before and after 2003 further impedes the comparison of these different scoring vectors. This leads to the approach to replace the vectors s1 and s2 by a modified vector s3 = (22; 21; … ; 2; 1) what means that the driver who wins a race gets 22 points, the second driver 21 and so on.[6] Thereby every ranking is considered and the level of competitive balance can be analyzed for all drivers. Another important fact is that the scoring vector is independent of rule changes, e. g. the modification in 2003. As in Figure 2 the points in time when the rule changed and Michael Schumacher temporarily ended his career are marked.

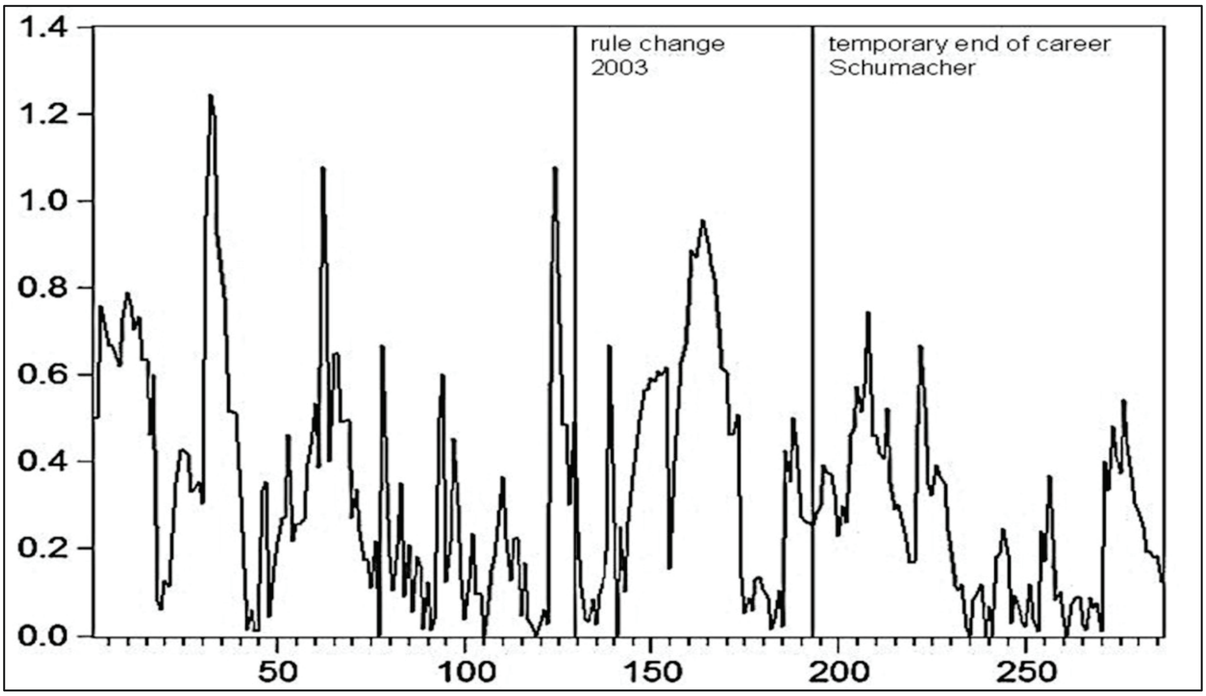

Figure 4: Relative distance between first and second driver

6 22 points were chosen as the maximum score as in most of the races included in the analysis this corresponds to the number of drivers in the races.

Figure 4 shows the time series of relative distance of the first and the second driver in the overall standings. For this analysis the original vectors are used. The reason is that this index shall give information about the tenseness of the duel. If the modified scoring vector was used instead, this tension, which is generated by the real distance in the overall standings, would not be considered. Only if the original distances in the overall standings are considered, this information can be used. As the distance between the points of the first and the second driver has to be identical after the first race of each season these races are not included in the analysis (this procedure was also followed in Figure 2 and 3).

4 Empirical Analysis

With regards to the competitive balance one could assume that a high level of competitive balance increases the number of TV viewers because they want to watch a competition which is not dominated by only one or a few group of drivers. It could guarantee a kind of tension if every driver has similar chances to win the competition. But this would imply that no favorites or outsiders exist whose wins could be a big surprise. Superstars or losers would not exist anymore. This leads to the assumption that a too high level of competitive balance reduces the interest in this sport. That is the reason for the following hypotheses which are specifications of the working hypotheses of chapter 2.

Hypothesis 1: A higher level of competitive balance leads to a weaker viewer interest.

Referring to the relative distance between the first and the second driver we assume that a large distance decreases the suspense and therefore the attractiveness of the competition. From these thoughts the following hypothesis is derived:

Hypothesis 2: A small distance between the first and the second driver increases the interest of the TV viewers.

To analyze the hypothesis a model has to be constructed and estimated. An ordinary least squares (OLS) regression is complex enough to analyze the problem. As regressands two time series are used. In estimation a (Table 1) the number of TV viewers of the German TV network RTL (compare Kipker, 2003) are used as regressand. In estimation b (Table 2 and Table 3) the market ratio, which gives information about the viewing rate, is used. The regressors are identical in both models: a constant term (C), a trend variable (trend), a quadratic trend variable (trend^2), the Gini coefficient (Gini), the relative distance (relative distance), a so-called prime time dummy (Pt dummy), a night dummy, a Schumacher dummy and a rule change dummy. The rationale behind the use of the Gini coefficient and the relative distance were already explained above. The use of a constant term seems to be necessary because it can be assumed that there is always a basic level of viewers. The trend and the quadratic trend are used because the first inspection of the data makes a trend development appear probable. But this approach can also be explained with regards to content. As it can be seen few lines below the Schumacher hype is covered by the Schumacher dummy. Next to Schumacher there was also an increasing presence of German Formula One drivers in these years. Between 1992 and 1997 the number of German Formula One drivers increased from 1 to 3. In 1994 Heinz-Harald Frentzen and in 1997 Ralf Schumacher, Michael Schumacher's brother, joined the Formula One (Formula One Administration Ltd 2009).

The prime time dummy is used for those races which are shown in German TV at the prime time between 06:00 p.m. and 12:00 p.m. in the evening. These are races taking place in North or South America. The value of this dummy is 1 if the race takes place in America and otherwise 0. The night dummy is very similar to the prime time dummy. It is 1 if the race takes place in East or Southeast Asia or in Pacific regions. It is called the night dummy because it applies to races which live TV coverage takes place in the early morning hours in Germany between 04:00 a.m. and 08:00 a.m. The Schumacher dummy is used because there was a Schumacher-hype in Germany during Michael Schumacher's career until 2006. Through the use of this dummy we hope to isolate these special Schumacher-effects (Kipker, 2003). The rule change dummy is used because we want to analyze the effects of the rule changes at the beginning of the year 2003.[7]

Table 1: Estimation a[8]

dependent variable	number of TV viewers
independent variables	**Coefficient**
C	-1.478480***
trend	0.106740***
trend^2	-0.000289***
Gini	7.118657***
relative distance	-0.840806**
Pt dummy	1.351560***
night dummy	-3.462045***
Schumacher dummy	0.938303**
rule change dummy	-2.396442***
F-statistic	182.0259***
Adjusted R-squared	0.835565

*** 1% level of significance, ** 5% level of significance, * 10% level of significance.

The F-statistic is very high and significant (s. Table 1). Although the R-squared is very high (0.83) this result should not be overrated because some dummy variables are used which automatically increase the R-squared. All coefficients are significant on a 5 percent level and with exception of the Schumacher dummy and the coefficient of the relative distance even significant on a 1 percent level. The constant variable C has a negative coefficient, which seems to be unusual when interpreting the term as a constant basis level of TV viewers. But this result should also

7 Langen, Krauskopf and Bünger (2010) have shown that the modification of rules in Formula One can play an important role for the outcome of the championship.

8 Heteroscedasticity consistent estimators are used.

not be overinterpreted because the values of the other coefficients are relatively high. Therefore a positive total of viewers can easily be reached.

The coefficient of the trend variable is positive while the coefficient of the quadratic trend variable is negative. That shows that a quadratic trend of viewer interest exists with an increase in the beginning and a decrease in later periods. Most dummy variables can be interpreted with ease. The prime time dummy has a positive coefficient and the night dummy a negative coefficient. That shows that the viewer interest increases respectively decreases if the race is shown at German prime time respectively at night. Also the positive coefficient of the Schumacher dummy is not remarkable. An interesting result is the negative coefficient of the rule change dummy, especially if one presumes that the rule modification in 2003 was intended to increase the attractiveness of Formula One.

The Gini coefficient has a high positive value. This means that a high Gini coefficient leads to higher viewers interest. As a high Gini coefficient means a low competitive balance this result can be interpreted in the following way: The viewers do not want to have a high level of competitive balance. Instead they want to have a competition which is not balanced.

Hypothesis 1 cannot be rejected for this reason. The coefficient of the relative distance is negative. This means that a high relative difference between the first and the second driver leads to low viewer interest. For this reason hypothesis 2 cannot be rejected either. These two results seem to contradict each other. But this contradiction can be explained. On the one hand the spectators do not want to have a competition which is too balanced and on the other hand they do not want to have a competition without an interesting duel at the top. This shows that an absolute competitive balance is as detrimental as a too low level of competitive balance in the top region of the competition.

After analyzing the number of TV viewers the viewing rate, which is the relative number of TV viewers, is taken into consideration (Table 2). This analysis of the viewing rate leads to similar results as the analysis of the number of TV viewers with one important difference: The Schumacher dummy loses its statistical significance.

Table 2: Estimation b1[9]

dependent variable	viewing rate
independent variables	**Coefficient**
C	21.73453***
trend	0.414048***
trend^2	-0.001229***
Gini	20.37742***
relative distance	-6.083538***
Pt dummy	-15.99070***
night dummy	11.78425***
Schumacher dummy	1.471344
rule change dummy	-11.07776***
F-statistic	125.4253***
Adjusted R-squared	0.777414

Table 3: Estimation b2[10]

dependent variable	viewing rate
independent variables	**Coefficient**
C	22.89844***
trend	0.424239***
trend^2	-0.001291***
Gini	19.37797**
relative distance	-5.485605***
Pt dummy	-15.91289***
night dummy	11.87955***
rule change dummy	-10.31951***
F-statistic	143.3755***
Adjusted R-squared	0.777627

*** 1% level of significance, ** 5% level of significance, * 10% level of significance.

After eliminating this variable the F-statistic increases which shows that the Schumacher dummy seems to be dispensable for this specification (Table 3). Without the Schumacher dummy it can be seen that the two most important variables can be interpreted in the same way as in estimation a (Table 1). One interesting difference is the inversion of the signs for the prime time dummy and the night dummy. The coefficient of the prime

9 Heteroscedasticity consistent estimators are used.
10 Heteroscedasticity consistent estimators are used.

time dummy becomes negative while the coefficient of the night dummy becomes positive. This can be explained by the relatively high respectively low TV consumption at prime time respectively at night. If some Formula One consumers watch TV at a time when just few other people watch TV this will increase the relative viewing levels. At prime time this effect runs in the opposite direction. This is a phenomenon analyzed by Kipker (2003).

In consequence the results of estimation b2 (Table 3), i.e., employing relative viewer levels instead of absolute amounts, do not lead to the rejection of the two hypotheses, either. They therefore confirm the conclusions drawn from the first estimation.

5 Conclusion

In this paper possibilities to measure the attractiveness of Formula One are discussed and analyzed how this attractiveness is influenced. It was assumed that the competitive balance could play an important role for the attractiveness of Formula One. To quantify this, two different indices for assumed influence on the attractiveness were constructed. These were the Gini coefficient and the relative distance between the first and the second driver. To measure the level of attractiveness the German TV viewer levels were used. The estimation of this approach was done by using an OLS regression with Heteroscedasticity consistent estimators. Through this OLS it could be shown that the viewers do not have an absolute clear preference concerning the competitive balance. On the one hand there is the wish to see a thrilling duel at the top and on the other hand a high level of competitive balance is not interesting for the viewers. Furthermore, we could show that there has been a quadratic trend over time in the viewer levels. Another interesting result was that it could be shown that the Schumacher effect only plays a role if you use the number of TV viewers. Once we use the viewing rate the Schumacher effect does not play any role. This result of an optimal level of competitive balance which maximizes the viewer's interests differs from the results of previous studies which mostly show that a high level of competitive balance maximizes the attractiveness of a sport. With our paper we can refute these assumptions concerning the competitive balance. This finding could be the basis for further empirical or theoretical work. In these further studies it could attempt to construct a more detailed theoretical model, which identifies the exact level of optimal competitive balance.

References

Borland, J, R. Macdonald (2003), Demand for sport, *Oxford Review of Economic Policy* Vol. 19, No. 4, pp. 478-502.

Deloitte (2008), F1 Grand Prix are the world's most valuable annual sporting events, *press release*, http://www.deloitte.com/view/en_MK/mk/press/mk-press-releaes-en/press-release/2b5b483cc320e110VgnVCM100000ba42f00aRCRD.htm, [Accessed 14th September 2009].

Formula One Administration Ltd (2009), http://www.formulaone.com [Accessed 28th March 2010].

GEZ (no date), Geschäftsbericht 2008, http://www.gez.de/e160/e161/e1248/gb2008.pdf, [Accessed 21th November 2009].

Langen, M., T. Krauskopf, B. Bünger (2010), The Election of a World Champion, *CAWM-Discussion Paper 39*, University of Münster.

Kipker, I. (2003), Determinanten der kurzfristigen TV-Nachfrage in der Formel 1 – Superstar- vs. Spannungseffekte und Implikationen für die Wettbewerbspolitik, In: H. M. Dietl (Eds.), *Globalisierung des wirtschaftlichen Wettbewerbs im Sport,* pp. 85-104, Schorndorf, Hofmann-Verlag.

Mastromarco, C., M. Runkel (2004), Rule Changes and Competitive Balance in Formula, *Munich Discussion Paper*, 2004-16.

Media Control (2010), Einschaltquoten FORMEL 1, Baden-Baden, exclusive dataset.

Rosen, S. (1981), The Economics of Superstars, *The American Economic Review*, Vol. 71, No. 5, pp. 845-858.

Rottenberg, S. (1956), The Baseball player's labor market, *The Journal of Political Economy,* Vol. 64, No. 3, pp. 242-258.

Spenke, M., C. Beilken (2000), InfoZoom - Analyzing Formula One racing results with an interactive data mining and visualization tool, Second International Conference on Data Mining 2000, 5-7 July 2000, Cambridge University.

Utt, J., J. Fort (2002), Pitfalls to Measuring Competitive Balance With Gini Coefficients, *Journal of Sports Economics*, Vol. 3, No. 4, pp. 367-373.

Vrooman, J. (1995), A General Theory of Professional Sports Leagues, *Southern Economic Journal*, Vol. 61, No. 4, pp. 971-990.

8 Conclusion

"Fußball ist deshalb spannend, weil niemand weiß, wie das Spiel ausgeht."

- Sepp Herberger, n.d.

Both, economic theory and common sense, suggest that sporting competitions should not be too one-sided. The factors leading to a specific outcome of a sport competition, their interpretation as balanced or imbalanced and the effects of more or less balanced outcomes on the demand for sporting events are the topics of this thesis. In none of these areas is there general agreement in the sports economic literature. By providing new insights in these areas of research, this thesis contributes to a deeper understanding of competitive balance in sports.

The way talent is allocated on the teams within a sport league is a substantial area of sports economic research. Within this research area, the debate about a sports team's objective is crucial. Defining the maximization of the sports team's value as the objective of team owners reduces the conflict between the proponents of profit maximization and win-percentage maximization. Sports teams create revenue based on sporting success, which itself depends on buying talented players in a kind of vivious circle. Analyzing this investment circle in models, which only consider one period, yields misleading results. In a more periodic model, players are not only the input variable in the production of sporting success, but also a highly tradable asset. Furthermore, they create higher value for the teams they play for, as they create future revenue potential. Such revenue potential encompasses, for example, the right to compete in lucrative competitions or the general fan base. Future revenue is therefore dependent on past success. Moreover, reinvesting profits in players is nothing other than an accumulation of capital. Analytically considering the value of a team leads to a convergence of profit and win-percentage maximization. Value maximization incorporates both approaches, as the difference between profit and win maximization constitutes the liquidity of the assets. Therefore a loss, even if high, should not be misinterpreted necessarily as economic failure, or overinvestment, because focusing on the flow variable of profit disregards the state variable of the asset. In perfect markets, the liquidity of the assets should not matter for a value-maximizing firm. However, market imperfections, especially information asymmetry in the talent and capital markets, might therefore force sports teams to retain more players as assets than theory would suggest.

Another argument supporting the existence of overinvestment comes from the application of the evolutionarily stable strategy equilibrium concept, although the usual approach to solve non-cooperative games is to determine the Nash equilibrium. For infinite populations, each evolutionary stable equilibrium is a Nash equilibrium as well. In sports leagues with a finite number of teams, these two equilibrium concepts do not necessarily coincide. Using evolutionary game theory instead of the standard non-cooperative game theory in a sports league can be justified by the unique characteristics of such leagues. In sports, the quality of one team is always determined in relation to its competitors. In evolutionary game theory, payoffs have to be relatively greater in the long run in order to survive. However, originating from the biological analysis of evolving animal populations, the interpretation of evolutionary stable strategies in the context of sporting contexts remains problematic. An evolutionary stable strategy is neither an optimal strategy, nor a rationally chosen one, let alone an unbeatable one. By maximizing relative fitness, an evolutionarily stable equilibrium is merely stable, because it cannot be invaded in the long run, and, as shown, can be beaten in the context of a sports league. As long as relative performance and relative payoffs do matter in a sports league, the optimal aggregate-taking strategy provides a far more intuitive interpretation of evolutionary stable strategies. By assuming the aggregate to be fixed, which, in the context of a sports league, is the total amount of investment, teams decide rationally on the level of talent they hire, which leads then to an evolutionary stable equilibrium. Furthermore, this approach shows that the debate on the appropriate equilibrium concept in a sports league has a lot in common with that on the wage elasticity of talent supply. Assuming a wage-inelastic talent supply, or a fixed supply of talent, is mathematically equivalent to assuming the aggregate level of talent to be fixed. Using the optimal aggregate-taking strategy approach, considering the stability conditions and restricting the model parameters to reasonable values, it can further be shown that no overinvestment occurs at equilibrium, neither in an evolutionarily stable, nor a Nash equilibrium.

The hypothesis that sport teams tend to overinvestment is a major issue in sports economics, as well as with the general public. Especially for economists, this perspective is somewhat surprising, as it implies that even though a sport league is a form of repeated game, teams tend to make the same mistakes again and again and do not learn. However, simply taking account of occasional major losses does not justify the conclusion that there is overinvestment. These loses are generated by teams by their own choices and whether there is a general market law compelling teams to overinvest, is far from proven. Sports teams do always have the possibility to invest nothing at all, and if they invest, this has to be financed somehow, so at least in the long run, no overinvestment should occur. In any event, the Financial Fair Play Regulations from the Union des Associations Européennes de Football (UEFA) were introduced in 2011 to prevent football clubs from spending more than they earn. As shown, a general issue in the economic literature when attempting to prove a tendency towards overinvestment, is the confusion between objectives and benchmarks. Teams that overinvest are usually assumed to be utility maximizing. Their utility function is either their own win percentage or a weighted sum of their win percentage and their revenue. The investment decision of these teams is then evaluated by comparing only monetary values. With a benchmark based on profit maximization, each investment decision resulting from utility maximization, is thus a form of overinvestment. Therefore, if it is assumed that sports teams tend to maximize their utility, it would be far more meaningful to compare their investment with a benchmark accounting for their utility, and not their revenue. As long as sports teams are able to finance their investments, even through wealthy private investors or companies, there is no need for market restrictions to be imposed on European football clubs.

Unequal revenue potential is one of the main reasons for a competitive imbalance in sports leagues. The majority of models used in the sport economics literature explain the uneven distribution of talent across teams in a sport league in terms of different market sizes. This model assumption about the difference in market size forms the foundation of competitive imbalance in sports leagues. This leads to the conclusion that a large market induces success in sport leagues, which is only partially reflected in reality. Having a large market is indeed needed to generate enough revenue to buy success. However, success is not a monotonic function of market size, rather than a function with a maximum at some point. Markets exceeding this maximum size give local rivals the opportunity to prevail within the same region. This reduces the revenue potential of large-market teams, and sport teams from smaller regions with no a strong local rival, are therefore frequently more successful. Especially in leagues which incorporate some form of promotion and relegation system, which essentially allows any team to enter the league at some point in time, endogenous markets are inevitable. As every league has a limited number of teams and every country has a specific population distribution, each league has a unique minimum market size necessary for a team to compete in it. Further empirical research could hence determine this minimum market size for different leagues, which is also crucial for defining regions in which regional competition is possible.

Beside the numerous reasons for competitive imbalance, there are also numerous dimensions to outcome uncertainty. Predictability is the opposite of suspense, with the former having a negative and the latter a positive effect on the demand for an event. However, sports event can be unpredictable in many respects. Hence, caution is required analyzing a competition and the question is answered as to whether this competition is balanced and its outcome uncertain. The most common approach in the sports economic literature analyzing round-robin tournaments, like European football leagues, is to use the standard deviation of winning percentages of one season, which is sometimes compared with the expected standard deviation of winning percentages of an ideally balanced league. However, sport events can be unpredictable in many ways. Especially in competitions where numerous single events are aggregated into one single ranking, some dimensions of uncertainty remain unreflected in this ranking. One dimension of uncertainty derives from the volatility of relative team standings from season to season. Greater similarity means more predictability, which should therefore have a negative effect on demand. Measuring the similarity of consecutive seasons captures the dynamic uncertainty of outcome. The cosine similarity measure is new to sport economics and is able to deal with problems unique to sports leagues. The method is able to deal with promotion and relegation, as well as the irregular absolute number of points awarded to the teams due to the 3-point rule. It can be shown that although static measures have yielded a slightly more balanced German Bundesliga in the last decades, the dynamic measure yields greater similarity and therefore a less balanced league. Taking account of the different dimensions of outcome uncertainty is necessary in order to depict the competitive balance of a league.

These different dimensions of outcome uncertainty also influence the mechanism by which world champions are determined in Formula One. The only possible way to aggregate multiple races into one ranking is a scoring vector, but it is not feasible to find the best approach. The many possible objectives that a scoring vector could have include creating a season which is undecided for a long time or a maximum number of different champions or according a winning probability to drivers, which is equal to the relative talent distribution. All these objectives are reasonable and together with others, aggregate outcome uncertainty of Formula One. As there is no one scoring vector that is optimal in all mentioned areas, it is necessary for a chosen scoring vector to balance the different dimensions. Adding to the complexity of finding a best scoring vector is its dependence on factors like talent distribution, which are not directly observable, for obvious reasons.

Given that each aggregation mechanism is a compromise of numerous objectives, the most important objective is certainly to ensure a high level of demand for the sport competition. With the greatest competitive balance, each competitor is equally strong, thus having the same winning probability. As shown for Formula One, this does not create necessarily the highest demand for a competition. In a completely balanced competition, the result is dependent on pure chance, as there are no differences in quality. At least in Formula One, there is a demand for an unequal talent distribution and identification with well-known drivers. Hence, even an unbalanced competition does not imply the need to improve competitive balance in order to increase the demand for such events.

Sports competitions need to be balanced in order to attract an audience. This is the uncertainty-of-outcome hypothesis, but and it is difficult to prove for sport events. In sports, a single event, season or couple of seasons can be uncertain. There may be an uneven distribution of chance to win, a group of favorites competing with outsiders and even competitors who have no chance of winning, and yet, the competition remains uncertain. Since Rottenberg (1956) stated his uncertainty-of-outcome hypothesis, many explanations of why competitions are imbalanced have been proposed, the measures for evaluating whether a competition is balanced or not are numerous and the effects of unevenly balanced competition audience demand continues to be analyzed. Whether it is necessary or possible to alter the competitive balance of a sport competition depends on numerous factors. It has been shown that the specific characteristics of a sport league and its teams exert a great influence on the competitive outcome. Although the results of a sport competition are obvious after the fact, the interpretation of whether this result derives from a competitively balanced or imbalanced competition is not so clear. Applying different measures to the same sports event can therefore reveal an improvement as well as a deterioration in the competitive balance for a sport. Without capturing all dimensions of uncertainty, the determination of the degree of competitive balance or imbalance remains vague at best. Taking account of all these dimensions is still not enough to explain demand, as the audience might favor some competitors more than others and therefore, the dominance of a favored competitor might be accepted as enough of a reason to attend. Hence, given the many dimensions of competitive balance and the contradictory effects on the demand side any authority claiming to improve competitive balance through regulatory intervention, should be particularly specific in their objectives. Provided the claimed objectives coincide with the real intentions, attempts to improve competitive balance remain ambiguous and regulation a case of trial and error, because the dimensions of competitive balance and its effects on the demand for sport are manifold.

References

Alós-Ferrer, C., Ania, A. B. (2005), The evolutionary stability of perfectly competitive behavior, *Economic Theory*, Vol. 26, No. 3, pp. 497–516.

El-Hodiri, M., J. Quirk (1971), An Economic Model of a Professional Sports League, *Journal of Political Economy,* Vol. 79, No. 6, pp. 1302-1319.

Grossmann, M. (2013), Evolutionarily stable strategies in sports contests, *Journal of Sports Economics (forthcoming),* DOI: 10.1177/1527002512470957.

Kuper S., S. Szymanski (2009), *Soccernomics,* New York, NY: Nation Books.

Rottenberg, S. (1956), The Baseball Players' Labor Market, *Journal of Political Economy*, Vol. 64, No. 3, pp. 242-258.

Sepp Herberger (n.d.), Sepp Herberger-Stiftung, Retrieved October 22, 2013, from www.sepp-herberger.de/Sepp-Herberger/Zitate.

Sloane, P. (1971), The Economics of Professional Football: The Football Club as a Utility Maximizer, *Scottish Journal of Political Economy*, Vol. 17, No. 2, pp. 121-146.

The Uncertainty-of-Outcome Hypothesis and Competitive Balance in Sports

Martin Langen

Sportliche Wettkämpfe leben davon, dass Ihr Ausgang nicht vorherbestimmt ist. Je ähnlicher sich Konkurrenten in ihrer sportlichen Stärke sind, umso weniger Gewissheit herrscht über den Ausgang des Wettbewerbs. Das Ausmaß der dem Wettkampf inhärenten Spannung ist, zumindest der Theorie nach, von entscheidender Bedeutung für die Nachfrage nach professionellem Sport. Die in diesem Band gesammelten Aufsätze untersuchen die Ursachen, Arten und Folgen unterschiedlicher Wahrscheinlichkeitsverteilungen bezüglich des Ausgangs sportlicher Wettbewerbe. Das individuelle Investitionsverhalten der Sport Teams erweist sich hierbei als ursächlich für die sportliche Unausgeglichenheit von Sport Ligen. Zudem werden die verschieden Dimensionen von Spannung in einer Sport Liga untersucht und neue Verfahren vorgestellt, diese zu messen. Darüber hinaus wird am Beispiel der Formel Eins dargestellt, wie die Regeln eines sportlichen Wettbewerbs das Ausmaß der Unsicherheit beeinflussen.

ISBN 978-3-8405-0095-4
EUR 13,50
01350
9 783840 500954