The Economics of Professional Team Sports Leagues

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of

Doctor of Economics

presented by

Markus Lang

from Germany

approved at the request of

Prof. Dr. Helmut M. Dietl
Prof. Dr. Egon Franck

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich herewith permits the publication of the aforementioned dissertation without expressing any opinion on its views.

Zurich, December 5, 2007

The Dean: Prof. Dr. H.P. Wehrli

Contents

List of Figures

Acknowledgements

This dissertation was conducted in the context of the research project 'Overinvestment Problems in the European Football Leagues' and was financed by the *Swiss National Science Foundation*.

I want to express my special gratitude to my supervisors Helmut Dietl and Egon Franck. I have been fortunate to have mentors like them. Helmut Dietl has always given me great freedom to pursue independent work, and at the same time the guidance not to lose track of my objectives. From the very first day at his chair, he laid out clearly the theoretical and methodological foundations of the research project. He was always approachable and willing to discuss every issue. He also made it possible for me to attend conferences around the world which helped me to exchange ideas and to receive input. His patience and generosity are greatly appreciated. My co-supervisor Egon Franck has been always there to listen and give advice. I want to thank him especially for his kindness and encouragements. I am very obliged to both of my supervisors since their ideas, comments and suggestions on early draft versions of the papers contained in this dissertation were invaluable in improving their quality.

I am also very grateful to all the people from our faculty who provided feedback and support at different phases of my research: Martin Grossmann, Martin Lutzenberger, Tariq Hasan and Leif Brandes. For his valuable comments and suggestions very special thanks go to Alexander Rathke, who is the co-author of one of the papers contained in this book.

Most importantly, this dissertation would not have been possible without the emotional and social support of my friends and family. My mother and father, to whom this dissertation is dedicated, have always supported, encouraged and cared for me no matter what I have chosen to do. They have been a constant source of love and strength to me over the years. I also want to express my deepest gratitude to the rest of my family: my brother Christian, my sister Tanja, Grandma and Grandpa. Thanks for everything.

Furthermore, I want to extend my gratitude to my longtime friends Frank, Fabian, Maja and Ole for the support they have lent me over all the years. The discussions with them always inspired and motivated me and their friendship provided me with a ready source of respite and strength. Last, but far from least, a heartfelt thank you is due to my beloved Virginie. She helps to keep my life in proper perspective and balance.

This dissertation is dedicated to my mother and father

Chapter 1

Introduction

1.1 Principal Economic Issues in Professional Team Sports Leagues

The sports business industry remains one of the largest and fastest growing industries in the US. In its annual comprehensive survey, the *Sports Business Journal* (SBJ) estimated the size of the sports industry in the US at $213 billion in 2006. That is more than twice the size of the auto industry and seven times the size of the movie industry. The sports industry's pace of growth is also illustrated by the evolution of the ratio between overall sports expenditure (for goods and services) and GDP. For example, in the early 1970's the ratio was about 0.5% in most European countries. In 1990, the ratio was already between 1% and 1.5% of GDP and has increased to nearly 2% today. Another example is the increased value of Olympic broadcasting rights. While the rights for the summer Olympics in Rome 1960 were sold for 'only' $1.2 million, the summer Olympics in Sydney 2000 attracted $1332 million in TV rights. This is equivalent to an annual growth rate of 19%. On a worldwide level, the value of the sports TV rights registered an increase of 993% between 1991 and 2001.[1]

The fact that sport has become a multi-billion dollar business does not tell the whole story. An important dimension of the sports industry's size and value is the place it occupies in society and the emotional impact it provides. According to Andreff and Szymanski (2007), the world audience for the football World Cup was 50 billion viewers in 1998 and an event like the Tour de France attracts annually 12 to 15 million spectators on to the roads. Newspapers dedicate whole sections to sports while other industries are all grouped into the business section. Sport is also the only industry

[1] Andreff and Szymanski (2007)

that has its own 24-hour television networks and radio stations. Undoubtedly, sport has become fashionable, dominating a large part of mainstream media today.

As the economic significance of the sports industry has grown, it has attracted considerable attention from economists, and sports economics as a relatively young but fast-growing subfield within the economics profession gains more and more importance. 'Since the mid-1990s there has been an explosion of publishing in the sports economics field. Several hundred articles have been published and at least a dozen of major books. [...] Sports economics may now be considered a discipline in its own rights.'[2] Rascher (1997) describes the connection between sports and economics as follows:

> The study of sports is similar to the study of economics in general. Sporting events are environments where participants respond to incentives and constraints. Athletes, coaches, owners, and even gamblers make decisions under uncertainty and outcomes are produced. The wage structure may affect how players interact with each other or how much effort they put forth. Players react to the severity and probability of punishment by breaking the rules less often. Coaches try to maximize production efficiency and minimize managerial slack. Owners' objectives are functions of winning and profits. Fans search for optimal gambling strategies. [...] Data involving sports is abundant, cheap, and generally of high quality. A professional athlete is one of the few occupations whose productivity is measured and made public on a daily basis.

Research into the application of economic concepts to sporting activities is primarily focussed on professional *team* sports to the comparative neglect of *individual* (non-team) sports such as golf, boxing, athletics, auto sports, etc.[3] Although individual sports are sometimes organized on a team basis, the teams are not generally organized into leagues ranked in line with their success over the season. The main reason why sports economists are interested in team sports is that 'professional team sports leagues are classic, even textbook, examples of business cartels' (Fort and Quirk, 1995). The important difference between sports leagues and other cartels, however, is the special nature of the former. Neale (1964) referred to this as 'the peculiar

[2] Andreff and Szymanski (2007)

[3] Some research has been conducted into individual sports. For instance, see Scully (2000) (athletics); Tenorio (2000) (boxing); Shaw and Jakus (1996) (climbing); Ehrenberg and Bognanno (1990*a*), Ehrenberg and Bognanno (1990*b*), Orszag (1994) (golf); Fernie and Metcalf (1999) (horse-racing); Maloney and Terkun (2002) (motorcycle-racing); Szymanski (2000) (Olympics); Lynch and Zax (2000), Maloney and Terkun (2000) (running).

economics of professional sports.' The essence of sports leagues is the associative character of competition. No club can improve its position in the ranks without worsening the position of other teams. This rank order contest can lead to a rat race which induces clubs to 'overinvest' in playing talent. The recent development of club finances in European soccer supports this hypothesis.[4]

An important issue to do with sports leagues is the suspense associated with a close contest and its influence on fan demand. The 'uncertainty of outcome' hypothesis going back to Rottenberg (1956) and Neale (1964) claims that spectators prefer close rather than less balanced contests. In this respect, a certain degree of (competitive) balance within the league is assumed to be necessary in order to maintain fan interest and with it the revenue stream from the sale of the product (the joint game).[5] Thus, in order to guarantee a successful competition, teams have a strong interest in the economic viability of other teams. The 'uncertainty of outcome' is the central premise by which many cross-subsidization policies have been justified. The main idea is to transfer resources from rich (large-market) clubs to poor (small-market) clubs because it is commonly argued that competition without intervention generates an equilibrium distribution which does not yield enough outcome uncertainty in the league and thereby inhibits fan demand.

But what league management policies might be undertaken to increase uncertainty of outcome and enhance competitive balance?

There are two major areas in which leagues have intervened to provide cross subsidies between clubs: interventions in the sporting labour-market (e.g. reserve clause, transfer system, salary caps) and interventions regarding the distribution of revenues between clubs (e.g. gate and national TV-sharing).

The reserve clause in the US major leagues and the restrictive pre-Bosman transfer system of the European football leagues are examples of the desire of team owners to control the market for players by establishing monopsony rights. The reserve clause was introduced in US baseball in 1887 and gave club owners an exclusive option to renew unilaterally the annual contracts of their players binding them to their clubs until release, retirement or a trade. As a consequence, a player could be tied to a club for the duration of his career by a series of annual renewals.[6] Club

[4]Chapter 2 provides a detailed analysis of the overinvestment problem and explains the mechanisms of the ruinous competition between clubs.

[5]The influence of the closeness of a league (competitive balance) on fan demand is one of the central questions in the economics of team sports. The empirical support for the 'uncertainty of outcome' hypothesis is, however, everything else than clear. See Section 1.2 for a more detailed discussion.

[6]The reserve clause was finally removed in favor of 'free agency' in 1976.

owners considered the reserve clause an important instrument to promote competitive balance since they claimed that small-market clubs are able to keep their star players only if the reserve clause is in effect. Rottenberg (1956) argued, however, that the reserve clause would have no effect on the allocation of playing talent because it cannot prevent the migration of star players to the large-market clubs. Rottenberg's claim, an application of the Coase theorem,[7] is known among sport economists as the 'invariance proposition.'

The transfer system in the European football leagues before the Bosman verdict in 1998 was a particular form of labour-market restriction on player mobility similar to the reserve clause system in the North American leagues. The crucial effect of the transfer system in Europe was the creation of a unilateral property right for the clubs over the services of players. As a consequence of the transfer system the players were not able to leave their current club and sign with another club without the current club's explicit consent. A transfer was only possible if the new club paid a transfer fee claimed by the old club. Similar to the reserve clause, the main reason for adopting such a restrictive transfer system was the idea of maintaining competitive balance by protecting the small-market clubs against competition for players from the large-market clubs. Moreover, small-market clubs which had trained and developed young talented players should receive a compensation when their players moved to large-market clubs.[8]

Another intervention in the labour-market adopted in team sports leagues are salary caps. A salary cap limits the amount a club can spend on player salaries, either as a per-player limit or a total limit for the team's roster. Salary caps were unnecessary in the era of the reserve clause since players could not negotiate with another club without the permission of the current club, but this changed after the abolition of the reserve clause. With the introduction of free agency in almost any professional sports league, 'bidding wars' for the best players were commonly observed. This situation led to an explosion in player salaries[9] and gave the advantage to the more affluent large-market clubs of luring quality talent away from their poor competitors. In order to limit the danger that all high-talented players would end up in the rich clubs, which would deteriorate competitive balance, salary caps were regarded as a counterpart to free agency. Moreover, salary caps should prevent the leagues

[7] The Coase theorem states that under the assumption of perfect competition, perfect information and the absence of transaction costs, the efficient outcome will occur regardless of the initial distribution of ownership rights (see Coase, 1960). Note that Rottenberg's article preceded the publication of the Coase theorem by four years.

[8] Chapter 5 provides a detailed discussion about the role of transfer restrictions in the European football leagues by comparing the pre-Bosman, Bosman and Monti transfer systems.

[9] Quirk and Fort (1992).

from financial self-destruction. In the 1984-85 season, the North American National Basketball Association (NBA) was the first professional sports league to introduce a salary cap allowing 53% of the league revenues to be spent on player salaries. This translated into a maximum payroll of $3.6 million for each team. Other sports leagues followed the example of the NBA. Nowadays, salary caps are in effect in professional team sports league all around the world. In North America, the National Hockey League, the Canadian Football League, the National Football League, the National Basketball Association and the Arena Football League have installed salary caps. In Australia, the Australian Football League, the National Rugby League and A-League Soccer have implemented salary caps to regulate their labour-markets. In Europe, salary caps are in effect in the Guiness Premiership in rugby union and the Super League in rugby league. In European soccer, there are currently intensive discussions to introduce salary caps. The leading clubs, organized as the so-called G-14, plan to limit annual team salaries to 70% of revenues.[10]

The other main form of cross-subsidization policy targets the redistribution of revenues among teams. In this respect, the sharing of gate and national TV revenues plays an important role and has long been accepted as an exemption from antitrust law. In its simplest form gate revenue-sharing allows the visiting club to retain a share of the home club's gate revenues. The basic idea is to redistribute revenues from large-market clubs with a high drawing potential to small-market clubs with a low drawing potential in order to offset inequalities between clubs. The current revenue-sharing schemes differ widely among professional sports leagues all over the world. The most prominent is possibly that operated by the National Football League (NFL) where the visiting club secures 40% of the designated stadium income. In 1876, the Major League Baseball (MLB) introduced a 50-50 split of gate receipts that was reduced over time. Since 2003, all clubs in the American League put 34% of their locally-generated revenue (gate, concession, television, etc.) into a central pool which is then divided equally among clubs. In the Australian Football League (AFL) gate receipts were split evenly between the home and the visiting team. This 50-50 split was finally abolished in 2000. In Europe, there is less gate revenue-sharing. The soccer leagues have adopted various forms of gate revenue-sharing in their history. In England until the early 1980s up to 20% of the gate receipts were given to the visiting teams in league matches. In the German soccer league (DFL), the home team receives 94% of the gate receipts with the other 6% going to the league.[11]

[10] Chapter 3 provides a theoretical model of a professional team sports league and analyzes the effect of salary caps on social welfare.

[11] Chapter 4 analyzes the welfare effect of gate revenue-sharing in a theoretical model of a professional team sports league.

1.2 Review of Selected Literature

The academic interest in the field of sports economics starts with Rottenberg's seminal article in the *Journal of Political Economy* in 1956 analyzing the baseball players' labour-market. He is the first to define the nature of the product of a sports league and to stress the importance of uncertainty of outcome. Moreover, he questions the effectiveness of arrangements like the reserve clause to influence the distribution of playing talent between teams.

In 1964, Walter Neale describes the peculiarities of professional sport. His article can be considered as the starting-point of the theory of professional sports leagues because it differentiates market competition and sporting competition. Unlike in most industries, the participants (teams) need each other since they produce the league competition as a joint product. It is not possible to produce any output without the assistance of the other participants: clubs do not only compete but also cooperate with each other to produce individual matches and a viable league competition.

Sloane (1971) is the first to apply economic analysis to European team sports. He distinguishes the European model of sport from the North American model and challenges Neale's claim that the league rather than the individual club is the relevant 'firm' (decision-making unit). Moreover, by considering the objectives of professional sports clubs, he claims that contrary to the US situation, the clubs' objective in Europe is not to maximize profits but rather to maximize 'utility' subject to a break-even constraint. He suggests a utility function with playing success, average attendance, profit and health of the league as arguments.

El-Hodiri and Quirk (1971) and Quirk and El-Hodiri (1974) formalize the insights developed in the early literature in the first general economic model of a sports league. Their model is based on a dynamic decision-making mathematical framework involving the reserve clause, player wages, player drafts and revenue-sharing. In their model, an equalization of playing talent will occur in the long run if the supply of playing talent is constant. The convergence process is unaffected by the initial allocation of talent but will speed up with a higher depreciation rate and fewer clubs.

Canes (1974) shows that improvements in team quality have important negative external effects which may induce clubs to over-employ athletic talent. He suggests the need for institutional mechanisms such as revenue-sharing, reserve clauses and player drafts in order to 'counteract the incentive to overinvest in team quality.' Externalities are also discussed by Demmert (1973) who suggests that the increase of skill levels above the clubs' average will result in diseconomies being external to the particular club and internal to the league.

Scully (1974) analyzes the relationship between the performance of individual players in US baseball and their wages. He shows that the reserve clause gave rise to exploitation in the sense that the players' wages were lower than the players' contributions to their clubs' revenues.[12]

In a theoretical model, Atkinson et al. (1988) show that revenue-sharing stimulates an optimal distribution of playing talent among clubs. They test their model empirically for the NFL and find that revenue-sharing has in fact the predicted properties.

Fort and Quirk (1995) discuss different cross-subsidization schemes in a profit maximizing model of a professional team sports league. Theoretically, they deduce that revenue-sharing does not influence the distribution of playing talent but lowers player salaries.

Vrooman (1995, 2000) develops a model where both the revenue and the cost function depend on market size and winning percentage. He suggests that sharing of winning-elastic revenue leaves competitive balance unaltered whereas sharing of winning-inelastic revenue does improve competitive balance. Surprisingly, his model predicts that a salary cap 'promotes competitive *imbalance* within the league.'

Following Tullock's (1980) rent seeking-contest,[13] economists start to analyze team sports in the framework of a contest. Contests are games in which several contestants compete with each other for a prize.[14] Each contestant can spend money or exert (costly) effort in order to increase the possibility of winning the prize. One main feature of a contest model is the so-called 'contest success function' (CSF) which maps efforts into probabilities for the different contestants of winning the prize. The most widely-used functional form of a CSF in sporting contests is the logit CSF.[15] The contest theory is fertile ground for sports economists, providing them with useful instruments to analyze the incentive effects in team sports leagues. See Szymanski (2003) who provides a survey article regarding the application of contest theory to

[12] In order to explain the high earnings of sport stars it is useful to follow developments elsewhere in economics: see e.g. Rosen (1981) and Lazear and Rosen (1981). The former examines the wages of 'superstars' in film and music, whereas the second examines the high earnings of CEOs in large companies.

[13] The simple Tullock model has been extended in various ways (for a collection of relevant articles see e.g. Lockard and Tullock, 2001): *inter alia* different valuations of the prize, asymmetric players, sequential play, cooperative behavior and dynamic games have been considered.

[14] The first approaches in contest theory are made by Lazear and Rosen (1981), Green and Stokey (1983) and Nalebuff and Stiglitz (1983). For a summary of the theory see the textbooks by Lazear (1995), Lazear (1998) and Baron and Kreps (1999).

[15] The logit CSF equals the ratio of each club's talent investment to total talent investment. The logit CSF is generally introduced by Tullock (1980) and subsequently is axiomatized by Skaperdas (1996) and Clark and Riis (1998). An alternative functional form would be the probit CSF (e.g. Dixit, 1987) and the difference-form CSF (e.g. Hirshleifer, 1989).

sports economics.

The two main pillars in sports economics, however, the 'uncertainty of outcome hypothesis' and the 'invariance proposition,' are controversial topics among economists. No consensus has arisen so far regarding the impact of competitive balance on fan attendance and the impact of revenue-sharing on the distribution of playing talent in a team sports league.

Sports economists differentiate two measures for the 'uncertainty of outcome' or competitive balance: match uncertainty and championship uncertainty. The former is the most widely-used approach in the literature, measures the uncertainty of outcome of an individual match in a league and tests whether the degree of uncertainty has a significant influence on attendance at the match. It is remarkable that empirical studies have problems in finding evidence in support of the uncertainty of outcome hypothesis given the importance that is placed on this argument in policy-making and anti-trust cases. For instance, see Szymanski (2003) who summarizes a total of 22 studies. He affirms that 'ten offer clear support for the uncertainty of outcome hypothesis, seven offer weak support and five contradict it.' A similar conclusion is drawn by Downward and Dawson (2000), who suggest that 'the evidence suggests that uncertainty of outcome has been an overworked hypothesis in explaining the demand for professional sports.'[16]

The other measure - championship uncertainty - meaning that one or a small number of teams dominate the championship (within a single season or between seasons) has attracted surprisingly little attention. Notable exceptions are Schmidt and Berri (2001) and Humphreys (2002), who only find weak support for the uncertainty of outcome hypothesis.

The 'invariance proposition' as another hallmark of sports economics was taken for granted for a long time. Here also, however, in the last decades a controversy has emerged. The 'invariance proposition' is first challenged by Daly and Moore (1981). Further studies show that in the profit-maximizing scenario, the 'invariance proposition' fails to hold when fans care about the relative and absolute quality of a given team: see e.g. Atkinson et al. (1988), Marburger (1997), Késenne (2000*a*). If the objective of clubs is such that they maximize utility, Rascher (1997), Késenne (2000*a*) and Késenne (2005) come to the same conclusion. However, on the assumption that only the own winning percentage of the home team affects club revenue, revenue-sharing will not influence competitive balance in the profit-maximizing scenario: see e.g. El-Hodiri and Quirk (1971), Quirk and El-Hodiri (1974), Fort and Quirk (1995)

[16]See also Borland and MacDonald (2003).

and Rascher (1997). The confusion in the literature stems from the different approaches (objectives of clubs, preference of fans), the different models (2-club, n-club league model) and the different methodology (fixed-, non-fixed supply of talents) used in these works.

1.3 Structure of the Book

This book contains a collection of four independent research papers on the economics of professional team sports leagues.

In Chapter 2 we will present a basic model of talent investment in a team sports league. This model may give the reader a deeper understanding of the associative character of competition. No team can improve its position in the standings without worsening the position of other teams. Since sporting and economic success are usually correlated, this rank order contest may result in destructive competition. Our analysis can explain why clubs are caught in a typical prisoners' dilemma type of equilibrium and why clubs tend to 'overinvest' in playing talent. The model will incorporate typical features of team sports leagues such as an endogenous league prize, additional exogenous prizes and a system of promotion and relegation.

In Chapters 3 and 4 we will extend our analysis by incorporating with the clubs and the players also the consumers (fans). We will derive fan demand from a general utility function by assuming that a fan's willingness to pay depends on the quality of the match/league. This approach will enable us to provide a full-fledged welfare analysis of a sports league. In Chapter 3 we take the currently intensive discussions of the leading European football clubs to introduce salary caps as a motivation to analyze their welfare effects. Financial disparity and spiralling wages have triggered this debate in Europe. Using a game-theoretical model of a league consisting of both small- and large-market clubs we will investigate the effect of salary caps on competitive balance, the aggregate salary payments and social welfare. In Chapter 4 we analyze the welfare effect of one of the most common means of redistribution in a sports league - gate revenue-sharing. In its simplest form, gate revenue-sharing allows the visiting club to retain a share of the home club's gate revenues. By means of a contest model with two asymmetric clubs, we will analyze the effect of alternative gate revenue-sharing arrangements on the investment behavior, competitive balance, club profits and social welfare.

In Chapter 5 we turn our attention towards the labour-market of the European football leagues and will analyze the role of transfer restrictions. Transfer restrictions have a long tradition in professional football, but have come under heavy attack in recent years. The transfer system imposed by the football governing bodies on employment relations made sure that a player could not leave his current club and sign with another club without the current club's explicit consent. The 1995 Bosman judgement of the European Court of Justice, declaring football players to be free agents after expiration of their contracts, and the 2001 intervention of the European Commission, which, among other things, limited contract durations in football, can be interpreted as the two major steps towards restricting the application of the transfer system. We will develop a theoretical model which captures an important and widely overlooked aspect of the employment relation in professional football: the allocation of risk. Players and clubs alike do not know how the productivity of a player will develop in future periods. If risk is the key driver behind the performance uncertainty of football players then there is a potential for value creation in this industry. Risk-averse players could buy insurance against future income uncertainty when contracting with risk-neutral clubs. Our model will analyze whether risk-averse football players really benefit from less restrictive transfer systems.

The book concludes in Chapter 6 with a short discussion and a summary of the main results.

Chapter 2

The Overinvestment Problem in Team Sports Leagues

2.1 Introduction

One of the economic peculiarities of professional team sports is the associative character of competition.[1] No club (team) can improve its position in the standings without worsening the position of other teams. If sportive and economic success are correlated, which they usually are, this rank order contest may result in destructive competition. In this chapter we show that the dissipation of the league revenue in a team sports league arises from 'overinvestment' in playing talent as a direct consequence of the ruinous competitive interaction between clubs. This 'overinvestment' problem aggravates if the discriminatory power of the contest function increases, revenue-sharing decreases, and the size of an additional exogenous prize increases. We further show that clubs invest more when they play in an open compared with a closed league. Moreover, the 'overinvestment' problem within open leagues increases with the revenue differential between leagues.

Before proceeding with the model, we will give a short overview of the related literature: The first academic analyzes of the economics of sports were presented by Rottenberg (1956), Neale (1964) and Canes (1974). In his seminal article Rottenberg studies the structural characteristics of the markets in which professional sports teams operate. Neale describes the peculiarities of team sports leagues which are characterized by their mutual interdependence. The participants need each other because it is impossible to produce any output without the assistance of the other producers.

[1] This chapter is based on a paper which was published in the *Scottish Journal of Political Economy* (see Dietl et al. (2008*a*)). Reprinted with permission.

Canes shows that improvements in team quality have important negative external effects, which may induce clubs to over-employ athletic talent. He suggests the need for institutional mechanisms such as revenue-sharing, reserve clauses and player drafts in order to 'counteract the incentive to overinvest in team quality.' The precise rationale for this tendency to 'overinvest' in team quality, however, remains obscure, since the specific nature of competitive interaction between the clubs in a league is not explicitly addressed. El-Hodiri and Quirk (1971) formalize the insights developed in the early literature in the first general economic model of a sports league, based on a dynamic decision-making mathematical framework. Fort and Quirk (1995), Vrooman (1995) and Vrooman (2000) update this framework, however, without explicitly modelling competition and interaction among the clubs. Whitney (1993) is the first to formalize ruinous competitions within sports leagues using a labour-market model. He suggests that the market for star athletes could be subject to 'destructive competition' which drives some participants out of the market even though it is inefficient for them to leave. The recent sports economics literature has suggested to model a team sports league by making use of contest theory which reflects the noncooperative nature of such leagues. Szymanski (2003) applies Tullock's (1980) rent-seeking contest to find the optimal design of sports leagues. However, he does not explicitly address the problem of 'overinvestment' in his model. Dietl and Franck (2000) and Dietl et al. (2003) are first to model the 'overinvestment' problem based on contest theory. Our model substantially extends their analysis by providing an integrated framework based on contest theory which allows to study the strategic interaction in a league with clubs competing for an endogenously-determined league prize. In this respect we are able to explain the tendency to 'overinvest' in playing talent as a direct consequence of the ruinous competitive interaction between clubs. We also model the effects that typical features of team sports leagues, such as exogenous prizes and promotion and relegation, have on talent investments.

The remainder of this chapter is organized as follows: In Section 2.2 we present our basic model of a league with profit-maximizing clubs competing for an endogenously-determined league prize. Section 2.3 considers a league in which an exogenously-given prize is offered to the winner of the championship in addition to the endogenous league prize. Section 2.4 provides a two-period dynamic two-league model that incorporates a system of promotion and relegation. Finally, Section 2.5 concludes the chapter.

2.2 Basic Model

The following contest model describes the investment behavior of profit maximizing clubs which are organized as public limited companies in a team sports league. The league consists of n clubs where each club $i \in J = \{1, .., n\}$ invests independently a certain amount t_i in playing talent. We assume that talent is measured in perfectly divisible units that can be hired in a competitive market for talent at a constant wage rate c per unit. Hence, club i's investment costs $C(t_i)$ for talent are given by $C(t_i) = ct_i$.

The league's total revenue $R(t_1, t_2, ..., t_n)$ is assumed to be a concave function of aggregate investments in playing talent, given by

$$R(t_1, t_2, ..., t_n) := \left(\sum_{i=1}^{n} t_i \right)^{\frac{1}{2}}.$$

This function reflects the fact that with raising investments in playing talent, e.g. better players, the league becomes more attractive for fans or TV-broadcasters and therefore the league income increases. In addition to this, we assume that investments in playing talent have decreasing returns to scale. Furthermore, we are considering a league with a revenue-sharing arrangement, i.e. also the defeated clubs receive a certain amount of the league revenue. In our model the share of the endogenously-determined league prize $R(t_1, .., t_n)$ which is awarded to the winner of the championship is given by the parameter $\alpha \in [\frac{1}{2}, 1]$, while $\frac{1-\alpha}{n-1}$ is assumed to be the share of the endogenous league prize received by each of the defeated clubs.[2] The limiting case $\alpha = 1$ describes a 'winner-takes-all' league whereas $\alpha = \frac{1}{2}$ describes a league with full revenue-sharing in which all clubs get the same share of the league revenue, independent of on-field success.

The probability of club i's success (i.e. win percentage of club i) is a function, called 'contest success function' (CSF), and maps the vector $(t_1, ., t_n)$ of talent investment into probabilities for each club. By applying the logit approach, the win percentage for club $i \in J$ in this imperfectly discriminating contest is derived as

$$w_i^{\gamma}(t_1, .., t_n) = \frac{t_i^{\gamma}}{\sum_{j=1}^{n} t_j^{\gamma}}.$$

Given that the win percentages must sum up to unity, we obtain the adding-up constraint $\sum_{j=1}^{n} w_j^{\gamma}(t_1, .., t_n) = 1$. The derivative of the win percentage with respect to

[2]For reasons of simplicity, we have assumed that each of the defeated $(n-1)$ clubs receives the same share of the remaining league's revenue.

own talent investments is given as

$$\frac{\partial w_i^\gamma}{\partial t_i} = \frac{\gamma t_i^{\gamma-1}\left(\sum_{j=1,j\neq i}^n t_j^\gamma\right)}{\left(\sum_{j=1}^n t_j^\gamma\right)^2}.$$

The parameter $\gamma > 0$, the so-called 'discriminatory power' of the CSF, measures how easily money buys on-field success. With other words, γ determines the ease of affecting the probability of winning the championship by a certain level of talent investment and specifies how much impact the club's own investments in playing talent have on its winning probability. The parameter γ also reflects the importance of luck or coincidence in a game. Luck plays a less important role in sports with high scores or a high frequency of matches. As γ increases, the marginal costs of influencing the probability of success decreases, i.e. the probability of winning the championship increases for the club i with the highest level of talent investment t_i and differences in talent investments affect the winning probability in a stronger way. In the limiting case where γ goes to infinity, we would have a so-called 'all-pay auction,' i.e. a perfectly discriminating contest, where the club with the highest talent investment wins the prize with probability one. However, for a sports league this is not a realistic assumption since the club with the highest investment in playing talent cannot be certain of winning the championship race. If each club invests the same amount in playing talent, the probability of winning equals $\frac{1}{n}$ for each club. In case that no club is willing to invest a positive amount in talents, i.e. $t_i = 0 \ \forall i \in J$, the corresponding probability is then defined as $w_i^\gamma(0,..,0) := \frac{1}{n}$. Furthermore, it is straightforward to verify that the CSF of club i is an increasing function in the club's own investments t_i and a decreasing function in the other clubs' investments $t_j \ \ (j \neq i \in J)$.[3]

We start our analysis by considering the league's optimum which serves as a benchmark case. The league's optimal level of talent investments maximizes the social surplus of the clubs and is defined as

$$(t_1^{LO},..,t_n^{LO}) = \arg\max_{(t_1,.,t_n)}\left(R(t_1,.,t_n) - c\sum_{i=1}^n t_i\right).$$

Solving this maximization problem yields[4]

$$t_i^{LO} = \frac{1}{4c^2 n}, \quad i \in J. \tag{2.1}$$

[3]Formally, $\frac{\partial w_i^\gamma}{\partial t_i} = \frac{\gamma t_i^{\gamma-1}(\sum_{j=1,j\neq i}^n t_j^\gamma)}{(\sum_{j=1}^n t_j^\gamma)^2} > 0$ and $\frac{\partial w_i^\gamma}{\partial t_j} = -\frac{\gamma t_i^\gamma t_j^\gamma}{t_j(\sum_{j=1}^n t_j^\gamma)^2} < 0$.

[4]Since clubs are symmetric we only consider the symmetric optimum.

The terminologies 'overinvest' and 'underinvest' are defined as situations in which a club invests in equilibrium more and less, respectively, than in the league optimum.

The expected payoff for club $i \in J$ is determined by the following (expected) profit function

$$\begin{aligned} E(\Pi_i) &= w_i^\gamma(t_1, .., t_n)\alpha R(t_1, .., t_n) \\ &+ (1 - w_i^\gamma(t_1, .., t_n))\frac{1-\alpha}{n-1}R(t_1, .., t_n) - C(t_i) \\ &= \left(\alpha\frac{t_i^\gamma}{\sum_{j=1}^n t_j^\gamma} + \frac{1-\alpha}{n-1}\frac{\sum_{j=1,j\neq i}^n t_j^\gamma}{\sum_{j=1}^n t_j^\gamma}\right)\left(\sum_{j=1}^n t_j\right)^{\frac{1}{2}} - ct_i. \end{aligned} \tag{2.2}$$

The expected payoff of club i depends on the probability of winning w_i^γ multiplied by the share α of the endogenous league prize $R(t_1, ., t_n)$ awarded to the winner, plus the probability of losing $(1 - w_i^\gamma)$ multiplied by the share $\frac{1-\alpha}{n-1}$ of the endogenous league prize $R(t_1, ., t_n)$ awarded to each of the defeated clubs, minus the investment costs $C(t_i)$ in playing talent.

The club-owners choose an investment level of playing talent such that expected profits are maximized, i.e. club i solves $\max_{t_i} E(\Pi_i)$, where $E(\Pi_i)$ is given by equation (2.2). Hence, the FOC for club $i \in J$ is derived as[5]

$$\frac{\partial E(\Pi_i)}{\partial t_i} = \frac{\alpha n - 1}{n-1}\left(\frac{\partial w_i^\gamma}{\partial t_i}R + w_i^\gamma\frac{\partial R}{\partial t_i}\right) + \frac{1-\alpha}{n-1}\frac{\partial R}{\partial t_i} - c = 0.$$

By solving this system of (implicitly defined) reaction functions, we obtain the following equilibrium expected investment level for club $i \in J$:[6]

$$t_i^* = \frac{(1 + 2\gamma(\alpha n - 1))^2}{4c^2n^3}. \tag{2.3}$$

In the symmetric equilibrium (2.3) the clubs realize identical strictly positive investment levels and therefore obtain an equal probability of $\frac{1}{n}$ to receive the share α of the endogenously-determined league prize $R(t_1^*, .., t_n^*) = \frac{1+2\gamma(\alpha n-1)}{2cn}$. Furthermore, the equilibrium investments $(t_1^*, .., t_n^*)$ in playing talent generate costs for each club i given by $C(t_i^*) = \frac{(1+2\gamma(\alpha n-1))^2}{4cn^3}$.

Plugging these investment levels into the (expected) profit function (2.2) yields the equilibrium expected payoff $E(\Pi_i^*)$ for club $i \in J$ as

$$E(\Pi_i^*) = \frac{(1 + 2\gamma(\alpha n - 1))(2n - 1 - 2\gamma(\alpha n - 1))}{4cn^3}.$$

[5] It is straightforward to verify that the second-order conditions for a maximum are satisfied.

[6] Note that the size n of the league is fixed since it is assumed to be exogenously-given.

The existence of an equilibrium in pure strategies depends on the discriminatory power γ of the CSF and the parameter α of the revenue-sharing arrangement:

Lemma 2.1.
The existence of a Nash equilibrium in pure strategies is guaranteed if (i) the discriminatory power γ is restricted to $0 < \gamma \leq \bar{\gamma}(\alpha) := \frac{2n-1}{2(\alpha n-1)}$ or (ii) the revenue-sharing parameter α is restricted to $\frac{1}{2} \leq \alpha \leq \bar{\alpha}(\gamma) := \frac{2\gamma+2n-1}{2\gamma n}$.

Proof. See Appendix 2.6.1. □

If $\gamma > \bar{\gamma}(\alpha)$ or $\alpha > \bar{\alpha}(\gamma)$, then the FOCs and SOCs fail to characterize the global maximum.[7] Moreover, note that $\gamma \leq \bar{\gamma}(\alpha)$ is equivalent to $\alpha \leq \bar{\alpha}(\gamma)$.

The 'ratio of revenue dissipation,' denoted D, measures the degree of dissipation of the league revenue and is defined in our model as[8]

$$D(\alpha, \gamma) := \frac{NS^{LO} - NS^*}{NS^{LO}} = \frac{(1 - n + 2\gamma(\alpha n - 1))^2}{n^2}.$$

The terms $NS^{LO} := R(t_1^{LO}, .., t_n^{LO}) - c\sum_{j=1}^n t_j^{LO}$ and $NS^* := R\left(t_1^*, .., t_n^*\right) - c\sum_{j=1}^n t_j^*$ characterize the net surplus at the league optimum and the Nash equilibrium, respectively. The higher the ratio $D(\alpha, \gamma)$, the higher the degree of dissipation in the league. Note that if α or γ are bigger than the threshold values $\alpha^*(\gamma) := \frac{2\gamma+n-1}{2\gamma n}$ and $\gamma^*(\alpha) := \frac{n-1}{2(\alpha n-1)}$, then D is an increasing function in α and γ. Moreover, the ratio D is within the interval $[0, 1]$ since we have assumed that $\gamma \leq \bar{\gamma}(\alpha)$ and $\alpha \leq \bar{\alpha}(\gamma)$.[9]

The next proposition provides comparative statics for the equilibrium investments $(t_1^*, .., t_n^*)$:

Proposition 2.1.
The equilibrium investments t_i^* of club $i \in J$ increase if (i) the discriminatory power γ of the CSF increases, i.e. money buys on-field success more easily, (ii) the share α of the league prize awarded to the winner increases, i.e. the league's revenue is distributed more unequally, or (iii) marginal costs c for talent investments decrease.

---

[7] The existence of Nash equilibria in the Tullock contest is discussed in the rent-seeking literature e.g. in Lockard and Tullock (2001). The case of mixed-strategies in a discrete choice set is analyzed by Baye et al. (1994).

[8] Note that in the rent-seeking literature the ratio D is called 'ratio of rent dissipation.' See for instance Chung (1996).

[9] Formally, $\frac{\partial D(\alpha,\gamma)}{\partial \alpha} = \frac{4\gamma(1-n+2\gamma(\alpha n-1))}{n} > 0 \Leftrightarrow \alpha > \frac{2\gamma+n-1}{2\gamma n}$. In this case it is also guaranteed that $\frac{\partial D(\alpha,\gamma)}{\partial \gamma} = \frac{4(\alpha n-1)(1-n+2\gamma(\alpha n-1)}{n^2} > 0$. Moreover, $\lim_{\alpha \to \alpha^*} D(\alpha, \gamma) = \lim_{\gamma \to \gamma^*} D(\alpha, \gamma) = 0$ and $\lim_{\alpha \to \bar{\alpha}} D(\alpha, \gamma) = \lim_{\gamma \to \bar{\gamma}} D(\alpha, \gamma) = 1$.

Proof. Straightforward. □

From this proposition we derive the following results:

ad (i) If γ is bigger than the threshold value $\gamma^*(\alpha) = \frac{n-1}{2(\alpha n-1)}$, then each club 'overinvests' in playing talent, i.e. $t_i^* > t_i^{LO}$, and the degree of revenue dissipation in the league increases with γ.[10] Intuitively, this is clear: If smaller differences in playing talent have a stronger impact on the probability of success, then the clubs have a stronger incentive for higher talent investments. If the discriminatory power γ equals the other threshold value $\bar{\gamma}(\alpha) = \frac{2n-1}{2(\alpha n-1)}$, then the net surplus NS^* at the Nash equilibrium amounts to zero and the ratio of dissipation $D(\alpha, \gamma)$ reaches its maximum of one. In this case the clubs dissipate the whole league revenue through their investment behavior.

ad (ii) Similarly, if α is bigger than the threshold value $\alpha^*(\gamma) = \frac{2\gamma+n-1}{2\gamma n}$, then each club 'overinvests' in playing talent and the degree of dissipation of the league revenue increases with α. Moreover, revenue dissipation is maximal, i.e. the ratio $D(\alpha, \gamma)$ amounts to one, if the parameter α equals the other threshold value $\bar{\alpha}(\gamma) = \frac{2\gamma+2n-1}{2\gamma n}$. In this case the net surplus NS^* at the Nash equilibrium is zero. We conclude that less revenue-sharing induces the clubs to increase their investments in playing talent and therefore contributes to aggravate the 'overinvestment' problem. The result that a bigger spread between first and second prize leads to higher equilibrium efforts is well-known in contest theory and follows from the stronger incentives to win.

ad (iii) Even though marginal costs influence the equilibrium investments, altering marginal costs does not affect the dissipation of the league revenue since the ratio of dissipation $D(\alpha, \gamma)$ is independent of c. Hence, marginal costs have no influence on the 'overinvestment' problem.

Summarizing the results derived above yields that if (a) the discriminatory power γ of the CSF is within the interval $(\gamma^*, \bar{\gamma}] = (\frac{n-1}{2(\alpha n-1)}, \frac{2n-1}{2(\alpha n-1)}]$ or (b) the parameter α of the revenue-sharing arrangement is within the interval $(\alpha^*, \bar{\alpha}] = (\frac{2\gamma+n-1}{2\gamma n}, \frac{2\gamma+2n-1}{2\gamma n}]$ existence of a Nash equilibrium is guaranteed in which each club 'overinvests' in playing talent and therefore dissipates a part of the league's revenue. However, the increase of the investment level in playing talent does not affect the winning-probability in equilibrium since the clubs simultaneously increase their investments and will end up with identical equilibrium investments. The same relative performance among the clubs could be obtained at the league optimum. Formally,

[10]However, in a league with full revenue-sharing, i.e. $\alpha = 0.5$, each club will invest less in equilibrium than in the league optimum, independent of the discriminatory power γ. Clearly, if each club gets the same share of the league's revenue, irrespective of field success, incentives to invest in playing talents are low.

$w_i^\gamma(t_1^*,..,t_n^*) = w_i^\gamma(t_1^{LO},..,t_n^{LO}) = \frac{1}{2}$. Even though the clubs would be better off if they agreed upon the investment level in the league optimum, this solution does not characterize a feasible equilibrium strategy because of strategic interaction. That is, the league optimum cannot be sustained without cooperation. Starting at the league optimum t_i^{LO}, club i has an incentive to increase its investments in talents since this behavior raises the probability of winning the share of the endogenous league prize awarded to the winner. The other clubs, however, have the same incentives and therefore the clubs are caught in a typical prisoners' dilemma type of equilibrium. As a result, each club will enter in a ruinous competition leading to the symmetric Nash equilibrium where the clubs 'overinvest' in playing talent, with no relative gain in performance compared with the league optimum.

2.3 The Effect of an Additional Exogenous League Prize

We assume that our league now offers an exogenously-given prize besides the endogenously determined league prize. This exogenous league prize is a proxy for all sorts of performance-related revenues like sponsorship contracts and the secure monetary value of qualifying for international competition.[11] The exogenous league prize, denoted P, is solely awarded to the winner of the championship while the endogenous league prize $R(t_1,..,t_n) = (\sum_{i=1}^n t_i)^{\frac{1}{2}}$ is again distributed among the clubs according to the revenue-sharing arrangement from Section 2.2. For the sake of simplicity, we henceforth assume that the discriminatory power γ of the CSF equals unity, i.e. the probability of success for club $i \in J$ is given by $w_i(t_1,..,t_n) = \frac{t_i}{\sum_{j=1}^n t_j}$. The expected profit of club $i \in J$ is now given by

$$\begin{aligned} E(\Pi_i) &= w_i(t_1,..,t_n)(\alpha R(t_1,..,t_n) + P) \\ &\quad +(1-w_i(t_1,..,t_n))\frac{1-\alpha}{n-1}R(t_1,..,t_n) - C(t_i), \\ &= \frac{t_i}{\sum_{j=1}^n t_j}\left(\alpha\left(\sum_{i=1}^n t_i\right)^{\frac{1}{2}} + P\right) + \frac{1-\alpha}{n-1}\frac{\sum_{j\neq i}^n t_j^\gamma}{\sum_{j=1}^n t_j^\gamma}\left(\sum_{i=1}^n t_i\right)^{\frac{1}{2}} - ct_i, \end{aligned}$$

[11] For example, in the European football leagues the clubs compete against each other also for the right to participate in international competition like the UEFA Champions League. The participation in the Champions League guarantees participants a minimum number of matches at the group stage and therefore secure revenue.

yielding the following FOC of profit-maximization:[12]

$$\frac{\partial E(\Pi_i)}{\partial t_i} = \frac{1}{n-1}\left((1-\alpha)\frac{\partial R}{\partial t_i} + (2n-1)\left(\frac{\partial R}{\partial t_i}w_i + \frac{\partial w_i}{\partial t_i}R\right)\right) + \frac{\partial w_i}{\partial t_i}P - c = 0.$$

By solving this system of reaction functions, we determine the Nash equilibrium for club $i \in J$ as[13]

$$t_i^*(P) = \frac{(n-1)}{cn^2}P + \frac{(2\alpha n-1)}{8c^2n^3}\left((2\alpha n-1) + \left((2\alpha n-1)^2 + 16cn(n-1)P\right)^{\frac{1}{2}}\right) \quad (2.4)$$

and derive the following results:

Proposition 2.2.
(i) The equilibrium investments $t_i^*(P)$ of club $i \in J$ increase in the exogenous prize P.
(ii) The equilibrium investments and the ratio of dissipation are higher in a league that offers an exogenous prize besides an endogenous prize compared with a league that offers an endogenous prize only, i.e. $t_i^*(P) > t_i^*$ and $D(P) > D$ $\forall P > 0$.

Proof. See Appendix 2.6.2. □

The proposition shows that in a league that offers an additional exogenous prize awarded to the winner of the championship the clubs are induced to spend more on playing talent and the 'overinvestment' problem is aggravated compared with a league that offers an endogenous prize only. The potential extra prize P generates additional financial incentives that encourages clubs to gamble on success by 'overinvesting' in playing talent in the hope of gaining admission to lucrative international competition and therefore to compensate their expenditures. Even though expected profits are non-negative, such a strategy is risky since the clubs cannot be sure of receiving the prize.

[12] Note that the second-order conditions for a maximum are satisfied.

[13] Formally, we obtain two equilibria. However, the negative one can be ruled out since it does not constitute an interior solution.

2.4 The Effect of Promotion and Relegation

In this section we provide a model in which the leagues are organized hierarchically in ascending divisions, offering a system of promotion and relegation. At the end of each season the worst performing clubs in each division are relegated to the next lower division and are replaced by the best performing clubs from that division. In order to analyze how a system of promotion and relegation affects the investment behavior of clubs, we will incorporate such a system in our league model by considering an open winner-takes-all league, i.e. a league without a revenue-sharing arrangement but which is open to promotion and relegation. Our dynamic model covers two periods and consists of two divisions denoted division A and division B, with each division containing two clubs. The time dimension becomes relevant now, because current investment behavior depends on the expected future profits as well as current profits. In other words, the prospect of promotion and relegation affects the first-period investments in playing talent. For the sake of simplicity, we assume that the revenue of each division is exogenously-given with R_A and R_B denoting the prize of division A and B, respectively. Division A is considered as the top-flight division which offers a higher prize than the second division B, i.e. $R_A > R_B$.

We assume that club 1 and club 2 start in period one in division A competing for the first-division prize R_A. The first-period champion receives the prize R_A, remains in division A and competes in period two against the promoted club from division B. The defeated club from division A receives nothing, is relegated to the second division and competes in the second period against the defeated club from division B. Club 3 and club 4 start in the first period in division B and compete for the second-division prize R_B. The first-period champion receives the prize R_B, is promoted to division A and competes in period two against the first-period champion of division A. The defeated club from division B receives nothing, remains in the division and competes in the second period against the relegated club from division A.

The investments in playing talent of club $\mu \in J = \{1, 2, 3, 4\}$ in period $t \in \{1, 2\}$ are denoted $t_{\mu,t}$ generating costs $C(t_{\mu,t}) = t_{\mu,t} \ \forall \mu \in J$. That is, marginal costs are now normalized to one. The expected profits of club μ, which competes against club ν in period t in division k, are denoted $E(\Pi^{t,k}_{\mu,\nu}) \ \forall \mu, \nu \in J$. Again, we assume that the discriminatory power γ of the CSF equals unity such that the probability that club $\mu \in J$ wins against club $\nu \in J$ in period $t \in \{1, 2\}$ is given by

$$w^t_{\mu,\nu}(t_{\mu,t}, t_{\nu,t}) = \frac{t_{\mu,t}}{t_{\mu,t} + t_{\nu,t}}.$$

Since it is assumed that the division prize R_k is won by one of the two clubs in the corresponding division $k \in \{A, B\}$ with certainty, it must be the case that $w^t_{\mu,\nu} = (1 - w^t_{\nu,\mu})$. For notational clarity, we exclusively use the subscripts $i, j \in \{1, 2\}$ to characterize the division A clubs 1 and 2, while the subscripts $r, s \in \{3, 4\}$ stand for the division B clubs 3 and 4. The superscript k denotes the division, with $k \in \{A, B\}$ and t stands for the period, with $t \in \{1, 2\}$.

In the top-flight division A, expected first-period profits $E(\Pi^{1,A}_{i,j})$ of club i and j are given by

$$E(\Pi^{1,A}_{i,j}) = w^1_{i,j}(R_A + E(\Pi^{2,A}_{i,r})) + (1 - w^1_{i,j})E(\Pi^{2,B}_{i,s}) - t_{i,1}. \tag{2.5}$$

With probability $w^1_{i,j}$ club i wins against club j in period one and obtains the first-division prize R_A. Club i then remains in division A, competes in period two against the promoted club r from division B and receives an expected second-period payoff of $E(\Pi^{2,A}_{i,r})$. With probability $(1 - w^1_{i,j})$ club i loses against club j and is relegated to division B without receiving a prize in period one. In the second period, club i competes against the defeated club s of division B and obtains an expected second-period payoff of $E(\Pi^{2,B}_{i,s})$.

In the second division expected first-period profits $E(\Pi^{1,B}_{r,s})$ of club r and s are given by

$$E(\Pi^{1,B}_{r,s}) = w^1_{r,s}(R_B + E(\Pi^{2,A}_{r,i})) + (1 - w^1_{r,s})E(\Pi^{2,B}_{r,j}) - t_{r,1}. \tag{2.6}$$

With probability $w^1_{r,s}$ club r is successful against club s in period one and receives the division B prize R_B. Club r is then promoted to division A and obtains an expected payoff of $E(\Pi^{2,A}_{r,i})$ in period two. With probability $(1 - w^1_{r,s})$ club r loses against club s in period one and stays in division B, receiving in period two an expected payoff of $E(\Pi^{2,B}_{r,j})$.

Following the logic of backward induction, we first determine expected profits $E(\Pi^{2,k}_{i,s})$ for club i and expected profits $E(\Pi^{2,k}_{r,j})$ for club r in division k of the subgame beginning in period two. Since clubs are assumed to be symmetric, it is irrelevant for the division A club i against which division B club r it will compete in the second period in division k and vice versa. Formally, it holds $E(\Pi^{2,k}_{i,3}) = E(\Pi^{2,k}_{i,4})$ and $E(\Pi^{2,k}_{r,1}) = E(\Pi^{2,k}_{r,2})$. Hence, the expected payoffs in period two are given by

$$E(\Pi^{2,k}_{i,s}) = w^2_{i,s}R_k - t_{i,2} \quad \text{and} \quad E(\Pi^{2,k}_{r,j}) = w^2_{r,j}R_k - t_{r,2}.$$

By deriving the respective FOCs and solving the system of reaction functions, we determine the equilibrium investment levels $t^o_{i,2}$ and $t^o_{r,2}$ besides the equilibrium-payoffs

$E^o(\Pi^{2,k}_{i,s})$ and $E^o(\Pi^{2,k}_{r,j})$ in the second period for club i and club r, respectively, as[14]

$$t^o_{i,2} = t^o_{r,2} = \frac{R_k}{4} \quad \text{and} \quad E^o(\Pi^{2,k}_{i,s}) = E^o(\Pi^{2,k}_{r,j}) = \frac{R_k}{4}.$$

In an open league, each of the four clubs invests in period two $\frac{R_k}{4}$ in playing talent and receives an expected payoff of $\frac{R_k}{4}$, dependent in which division k it competes. Plugging the second-period expected payoffs $E^o(\Pi^{2,k}_{i,s})$ and $E^o(\Pi^{2,k}_{r,j})$ into the first-period profit functions (2.5) and (2.6), respectively, yields

$$\begin{aligned} E(\Pi^{1,A}_{i,j}) &= w^1_{i,j}(R_A + \frac{R_A}{4}) + (1 - w^1_{i,j})\frac{R_B}{4} - t_{i,1}, \\ E(\Pi^{1,B}_{r,s}) &= w^1_{r,s}(R_B + \frac{R_A}{4}) + (1 - w^1_{r,s})\frac{R_B}{4} - t_{r,1}. \end{aligned}$$

By deriving the corresponding FOCs and solving the system of reaction functions, we determine the first-period equilibrium investments $t^o_{i,1}$ and $t^o_{r,1}$ besides the expected profits $E^o(\Pi^{1,A}_{i,j})$ and $E^o(\Pi^{1,B}_{r,s})$ for club $i \in \{1,2\}$ and club $r \in \{3,4\}$, respectively, as

$$t^o_{i,1} = \frac{1}{16}(5R_A - R_B) \quad \text{and} \quad E^o(\Pi^{1,A}_{i,j}) = \frac{1}{16}(5R_A + 3R_B), \tag{2.7}$$
$$t^o_{r,1} = \frac{1}{16}(R_A + 3R_B) \quad \text{and} \quad E^o(\Pi^{1,B}_{r,s}) = \frac{1}{16}(R_A + 7R_B). \tag{2.8}$$

Note that the division A clubs 1 and 2 spend more on playing talent in the first period than the division B clubs 3 and 4 but they also receive a higher expected payoff.[15]

As a reference point, we now calculate the respective investment levels and payoffs in a closed league, i.e. in a league where it is not possible to be promoted or relegated from one division to another. In a closed league the first-period expected profits of the division k clubs μ and ν are given by:

$$E(\Pi^{1,k}_{\mu,\nu}) = w^1_{\mu,\nu}(R_k + E(\Pi^{2,k}_{\mu,\nu})) + (1 - w^1_{\mu,\nu})E(\Pi^{2,k}_{\mu,\nu}) - C_\mu(t_{\mu,1}) \tag{2.9}$$

with $k = A$ if $\mu,\nu \in \{1,2\}$ and $k = B$ if $\mu,\nu \in \{3,4\}$. With probability $w^1_{\mu,\nu}$ the division k club μ wins against club ν in period one, obtains the division k prize R_k and competes in period two again with club ν for the prize R_k, receiving an expected payoff of $E(\Pi^{2,k}_{\mu,\nu})$. With probability $(1 - w^1_{\mu,\nu})$ club μ is defeated by club ν in period one, receives nothing and plays in the second period again against club

[14] Note that the superscript o stands for 'open' league whereas c stands for 'closed' league.

[15] $t^o_{i,1} > t^o_{r,1} \Leftrightarrow R_A > R_B$ and $E^o(\Pi^{1,A}_{i,j}) > E^o(\Pi^{1,B}_{r,s}) \Leftrightarrow R_A > R_B$.

ν in division k, obtaining an expected payoff of $E(\Pi^{2,k}_{\mu,\nu})$.

For the subgame beginning in period two, expected profits $E(\Pi^{2,A}_{i,j})$ for the division A clubs 1 and 2 and expected profits $E(\Pi^{2,B}_{r,s})$ for the division B clubs 3 and 4, respectively, are given by

$$E(\Pi^{2,A}_{i,j}) = w^2_{i,j}R_A - t_{i,2} \quad \text{and} \quad E(\Pi^{2,B}_{r,s}) = w^2_{r,s}R_B - t_{r,2},$$

yielding the following second-period equilibrium investments and payoffs in division A and B, respectively:

$$t^c_{i,2} = \frac{R_A}{4},\ E^c(\Pi^{2,A}_{i,j}) = \frac{R_A}{4} \quad \text{and} \quad t^c_{r,2} = \frac{R_B}{4},\ E^c(\Pi^{2,B}_{r,s}) = \frac{R_B}{4}.$$

By plugging the equilibrium-payoffs $E^c(\Pi^{2,A}_{i,j})$ and $E^c(\Pi^{2,B}_{r,s})$ into (2.9) and computing the corresponding FOCs we derive the first-period equilibrium investments and payoffs in division A and B as

$$t^c_{i,1} = \frac{R_A}{4} \quad \text{and} \quad E^c(\Pi^{1,A}_{i,j}) = \frac{R_A}{2}, \qquad (2.10)$$

$$t^c_{r,1} = \frac{R_B}{4} \quad \text{and} \quad E^c\left(\Pi^{1,B}_{r,s}\right) = \frac{R_B}{2}. \qquad (2.11)$$

Comparison of the first-period investment levels (2.7) with (2.10) in division A and (2.8) with (2.11) in division B, respectively, yields the following results:

Proposition 2.3.
In an open league, the aggregate first-period investments in both divisions are higher than the respective investments in a closed league. This difference increases if the spread between the division A prize and the division B prize augments.

Proof. See Appendix 2.6.3. □

This proposition shows that clubs compete more intensively in an open league than in a closed league in the first period. In the second period, however, the investment levels in an open league and in a closed league are equal in both divisions, i.e. $t^o_{\mu,2} = t^c_{\mu,2}$ $\forall \mu \in J$. In an open league the prospect of promotion as an additional reward for clubs in the second division and the threat of relegation for clubs in the top division both induce an increase of talent investments in the first period, compared with a closed league. We conclude that under a system of promotion and relegation the incentives to improve team quality by investing a higher amount in playing talent are enhanced since clubs obtain financial benefits from promotion and suffer financial penalties from relegation.

Moreover, the larger the difference between division A and division B in terms of revenues, the bigger the difference between the first-period investments in an open and a closed league. Hence, each club will spend more on playing talent in an open league, if the promotion from division B to division A becomes more lucrative and the relegation from A to B more 'costly' (in terms of reduced revenues).

2.5 Conclusion

In the past decade, many clubs in the European football leagues were able to increase total revenues because of higher broadcasting receipts, bigger crowds, sponsorship and a more professional approach to merchandizing. According to Deloitte and Touche (2004), the combined revenue generated by the top divisions of Europe's 'Big Five' leagues[16] increased by 190%: from approximately €1.9 billion in the season 1995/96 to €5.6 billion in the season 2002/03. Manchester United, the world's richest club, even augmented its turnover from €25 million in 1990 to €188 million in 2001, an increase by more than 750%.[17] However, at the same time there is growing evidence of a financial crisis spreading throughout the European football leagues. Striking examples are Italy's *Serie A* and Spain's *Primera Division:* The *Serie A* clubs accumulated total losses of €1.2 billion in the period from 1995/96 up to 2002/03, with 84% of theses losses sustained from 2000/01-2002/03.[18] In the *Primera Division* the total amount of debt in 2003 amounted to €1.6 billion.[19] Many European clubs face serious financial difficulties. Some even went bankrupt. Examples illustrating this general tendency are numerous: The Serie A club *AC Fiorentina* went bankrupt in 2002 and was relegated to the third Italian league. A court declared *AC Parma* insolvent in April 2004 with €310 million in debt. Furthermore *Lazio Rome* is €310 million in debt and *AS Rome* €300 million. Presently in England *Leeds United* faces serious financial problems and is near bankruptcy. In Spain *FC Barcelona* and *FC Valencia* are seriously in debt with €250 million and €125 million, respectively.[20] In the Bundesliga *Borussia Dortmund* is near insolvency after making a loss of €67 million in 2004 and being in debt with €118 million.[21] In Switzerland *Servette Genf* was declared insolvent in February 2005; after *FC Lugano* and *Lausanne Sports* in 2002 this is already the

[16] The 'Big Five' leagues in Europe are: *Premier League* (England, 20 clubs), *Ligue1* (France, 20 clubs), *Bundesliga* (Germany, 18 clubs), *Primera Division* (Spain, 20 clubs) and the *Serie A* (Italy, 18 clubs).

[17] Economist (2002).

[18] Deloitte and Touche (2004).

[19] El Pais, 28th of August 2002.

[20] Kicker, 12th of January 2004.

[21] Annual report 2004 of *Borussia Dortmund.*

third club to go bankrupt. The Czech club *Bohemians Prague* could only be bailed out because fans donated more than €100'000.

How can this 'paradox of rising revenues and declining profits' be explained? A first explanation stresses inadequate club constitutions. As organizations without residual claimants, traditional clubs are more likely to behave as win-maximizers. Having no ownership stakes in the operation and, at the same time, lacking genuine owners as monitors, club managers have discretion to maximize individual utility through sportive success. The chance to privatize a part of the fame and glamor derived from sporting success while socializing the inherent financial risks creates strong incentives to invest too much in playing talent. However, a closer look at the real situation in professional team sports shows the limitation of this constitutional explanation. The paradox of raising revenues and declining profits persists even in leagues where clubs have been transformed into capitalistic corporations with profit-maximizing owners. Obviously the problem must have deeper roots.

Based on the analysis of a theoretical league model with profit-maximizing clubs competing for a league prize, we have tried to deal with these roots. The analysis has shown that the tendency to 'overinvest' in playing talent leading to the dissipation of the league's revenue is a direct consequence of the ruinous competition between the clubs. The following factors enhance the incentives to 'overinvest' and therefore to dissipate the league's revenue:

- a stronger correlation between talent investments and league performance,
- a more unequal distribution of the league's revenue,
- an additional exogenous prize (e.g. participation to international competition) awarded to the winner of the domestic championship,
- a system of promotion and relegation, and
- an increased inequality between first and second division of a domestic league.

2.6 Appendix

2.6.1 Proof of Lemma 2.1

The existence of a Nash equilibrium in pure strategies is guaranteed if each club receives non-negative equilibrium-payoffs, i.e. $E(\Pi_i^*) \geq 0 \quad \forall \gamma \in (0, \bar{\gamma}]$ and $i \in J$.

ad (i) We derive $E(\Pi_i^*) \geq 0 \;\; \forall \gamma \in [\bar{\gamma}_1, \bar{\gamma}_2]$ with $\bar{\gamma}_1 = -\frac{1}{2(\alpha n-1)}$ and $\bar{\gamma}_2 = \frac{2n-1}{2(\alpha n-1)}$. Since γ is assumed to be strictly positive we conclude $\bar{\gamma}_1 < 0$ and thus we can concentrate on the interval $(0, \bar{\gamma}_2]$. Hence by restricting the discriminatory power γ to $0 < \gamma \leq \bar{\gamma}(\alpha) := \frac{2n-1}{2(\alpha n-1)}$, we obtain non-negative equilibrium-payoffs and therefore the existence of the Nash equilibrium.

ad (ii) Similarly,[22] we derive $E(\Pi_i^*) \geq 0 \;\; \forall \alpha \in [\bar{\alpha}_1, \bar{\alpha}_2]$ with $\bar{\alpha}_1 = \frac{1}{n} - \frac{1}{2\gamma n}$ and $\bar{\alpha}_2 = \frac{1}{n} + \frac{2n-1}{2\gamma n}$. Since α is assumed to be bigger or equal $\frac{1}{2}$ we conclude $\bar{\alpha}_1 < \frac{1}{2}$ and thus we can concentrate on the interval $[\frac{1}{2}, \bar{\alpha}_2]$. Hence, by restricting α to $\frac{1}{2} \leq \alpha \leq \bar{\alpha}(\gamma) := \frac{2\gamma+2n-1}{2\gamma n}$ proves the claim.

2.6.2 Proof of Proposition 2.2

ad (i) We compute $\frac{\partial t_i^*(P)}{\partial P} = \frac{n-1}{n}\left(1 + \frac{2\alpha n-1}{\sqrt{16cn(n-1)P+(2\alpha n-1)^2}}\right) > 0$. Thus, by increasing the exogenous prize P each club is induced to spend more on playing talent.

ad (ii) The equilibrium investments in a league that offers an endogenous prize only are given by equation (2.3) as $t_i^* = \frac{(2\alpha n-1)^2}{4c^2n^3}$, $i \in J$. Note that we have set γ equal unity. By comparing these investments with the corresponding equilibrium investments $t_i^*(P)$ given by (2.4) in a league that offers an exogenous prize besides an endogenous prize proves the claim that $t_i^*(P) > t_i^*$.

The additional exogenous prize, however, has no influence on the league optimum which is still given, as in the basic model in Section 2.2, by $t_i^{LO} = \frac{1}{4nc^2}$. Hence, the net surplus at the league optimum is given by $NS^{LO}(P) = \frac{1}{4c} + P$. We derive that due to the additional exogenous prize, the corresponding ratio of dissipation $D(P) := \frac{NS^{LO}(P)-NS^*(P)}{NS^{LO}(P)}$ is higher than the ratio $D = \frac{NS^{LO}-NS^*}{NS^{LO}}$ of a league with an endogenous prize only. This claim can be straightforward proved by noting that $NS^*(P)$ is a decreasing function in P.

[22] Note that $\gamma \leq \bar{\gamma}(\alpha)$ is equivalent to $\alpha \leq \bar{\alpha}(\gamma)$.

2.6.3 Proof of Proposition 2.3

In an open league the division A clubs 1 and 2 realize an investment level of $t^o_{1,1} = t^o_{2,1} = \frac{1}{16}(5R_A - R_B)$ in the first period. This investment level lays above the first-period investment level $t^c_{1,1} = t^c_{2,1} = \frac{R_A}{4}$ of the respective clubs in a closed league since we have assumed that $R_A > R_B$. The same holds true for the division B clubs 3 and 4: the first-period talent investments $t^o_{3,1} = t^o_{4,1} = \frac{1}{16}(R_A + 3R_B)$ in an open league are higher than the respective investment levels $t^c_{3,1} = t^c_{4,1} = \frac{R_B}{4}$ in a closed league. Hence, the first-period aggregate investment level in both divisions is higher in an open league than the respective level in a closed league.

Moreover, the difference between the first-period investments in playing talent in an open and a closed league is given for both divisions by $t^o_{\mu,1} - t^c_{\mu,1} = \frac{1}{16}(R_A - R_B) \forall \mu \in J$. The difference $t^o_{\mu,1} - t^c_{\mu,1}$ becomes larger if the spread between the division prize R_A and R_B increases.

Chapter 3

The Welfare Effect of Salary Caps

3.1 Introduction

A salary cap is a limit on the amount of money a club can spend on player salaries.[1] The cap is usually defined as a percentage of average annual revenues and limits the club's investment in playing talent. Since most leagues compute their caps on the basis of the revenues of the preceding season, the cap is actually a fixed sum. In 2006, for example, the National Football League (NFL) had a salary cap of approximately 102 million US dollars per team.

The North-American National Basketball Association (NBA) was the first league to introduce a salary cap for the 1984-85 season.[2] Today, salary caps are in effect in professional team sports all around the world. In North America, the National Hockey League,[3] the Canadian Football League, the National Football League, the National Basketball Association and the Arena Football League have installed salary caps. In Australia, the Australian Football League, the National Rugby League and A-League Soccer have implemented salary caps to regulate their labour-markets. In Europe, salary caps are in effect in the Guiness Premiership in rugby union and the Super League in rugby league. In European soccer, there are currently intensive discussions to introduce salary caps. The leading clubs, organized as the so-called G-14, planned to limit annual team salaries to 70% of revenues.[4]

[1] This chapter is based on Dietl et al. (2007).

[2] See Staudohar (1999).

[3] A lockout in 2004-05 resulted, for the first time, in the loss of an entire season in the National Hockey League. The main point of contention was that club owner insisted on the introduction of a salary cap to have cost certainty (Staudohar, 2005).

[4] See Késenne (2003) for an analysis.

From an economic perspective, salary caps are often regarded as a collusive agreement of wealthy owners to use their monopoly power to transfer player rents back to ownership.[5] Nevertheless, salary caps are not illegal in the US because they are the result of a freely negotiated collective bargaining agreement between the players' union and the league, represented by their governing body. The stated rationale for salary caps focuses on two main objectives: increasing competitive balance and maintaining financial stability. The concern for competitive balance describes one of the most important peculiarities of professional team sports:[6] It is a widely held belief that a certain degree of uncertainty about the outcome is necessary to ensure an entertaining competition.[7] Salary caps prevent large-market clubs from becoming too dominant by helping small-market clubs to keep star players who would otherwise be attracted by higher salary offers from large-market clubs. Fort and Quirk (1995) consider an enforceable salary cap as the only effective device to maintain 'financial viability' and improve competitive balance.

In Europe, the leading football clubs cited the protection of the financial future of the game as the main reason for their attempts to introduce a salary cap. Many clubs are facing financial ruin after gambling on spiralling wages. Owing to its contest structure, professional team sports carries the risk that its clubs overinvest in playing talent. Salary caps prevent clubs from overinvesting in playing talent.

Both arguments have been discussed in the economic literature. According to Rottenberg (1956) clubs would not voluntarily bid themselves into bankruptcy and diminishing returns to talent will guarantee at least some level of competitive talent. Whitney (1993), on the other hand, shows that the market for star athletes in professional team sports is subject to destructive competition - a process which drives some clubs into bankruptcy. According to Whitney, club managers will, on average, overspend on talent that turns their team into a contender, i.e. they will overinvest in star players. The recent development of club finances in European soccer supports Whitney's hypothesis.

Késenne (2000*b*) develops a two-team model consisting of a large- and a small-market club and shows that a payroll cap, defined as a fixed percentage of league revenue divided by the number of teams, will improve competitive balance as well as the distribution of player salary within the league. Moreover, he shows that profits of both the small- and the large-market club will increase.

The effect of salary caps on consumers (fans) has not been analyzed in the literature.

[5] See e.g. Vrooman (1995, 2000).

[6] Going back to Rottenberg (1956) and Neale (1964).

[7] For a survey and discussion, see Szymanski (2003) and Borland and MacDonald (2003).

This model tries to fill this gap. We present a complete analysis of social welfare incorporating the effect of salary caps on clubs, players and fans. Based on a game-theoretical model of a league consisting of both small- and large-market clubs, we show that salary caps will increase competitive balance and decrease the aggregate level of talent within the league. The resulting effect on social welfare is counter-intuitive and depends on the relative preference of fans for aggregate talent and for competitive balance. A salary cap that binds only for large-market clubs will increase social welfare if fans prefer aggregate talent despite the fact that the salary cap will result in lower aggregate talent. If fans prefer competitive balance, on the other hand, any binding salary cap will reduce social welfare.

The remainder of this chapter is organized as follows. Section 3.2 outlines the basic model. In Section 3.3 we introduce salary caps into the model and distinguish different regimes depending on whether the salary cap is binding or not. Section 3.4 compares the aggregate salary payments, competitive balance and social welfare between the regimes. Finally, Section 3.5 concludes.

3.2 Model Specification

The following model describes the impact of a salary cap on social welfare in a professional team sports league consisting of n (an even number) profit-maximizing clubs. The league generates total revenues according to a league demand function. The league revenue is then split among the clubs that differ with respect to their bargaining power. We assume that there are two types of clubs, 'large-market' clubs with strong bargaining power and 'small-market' clubs with weak bargaining power. In order to maximize profits each club independently invests in playing talent. We regard the salary payment of each club as an investment in talent where the maximum amount that each club can invest in playing talent is defined by the salary cap.

League demand depends on the quality of the league q and is derived as follows:[8] We assume a continuum of fans that differ in their willingness to pay for a league with quality q. Every fan k has a certain preference for quality that is measured by θ_k. The fans θ_k are assumed to be uniformly distributed in $[0,1]$, i.e. the measure of potential fans is one. The net-utility of fan θ_k is specified as $\max\{\theta_k q - p, 0\}$. At price p the fan that is indifferent between consuming the league product or not is given by $\theta^* = \frac{p}{q}$. Hence, the measure of fans that purchase at price p is $1-\theta^* = \frac{q-p}{q}$. The league demand function is therefore given by $d(p,q) = 1 - \frac{p}{q}$. Note that league

[8] Our approach is similar to Falconieri et al. (2004) but we use a different quality function.

demand increases in quality, albeit with a decreasing rate, i.e. $\frac{\partial d}{\partial q} > 0$ and $\frac{\partial^2 d}{\partial q^2} < 0$. By normalizing all other costs (e.g. stadium and broadcasting costs) to zero, league revenue is simply $LR = p \cdot d(p,q)$. The league will then choose the profit-maximizing price $p^* = \frac{q}{2}$.[9] Given this profit-maximizing price, league revenue depends solely on the quality of the league

$$LR = \frac{q}{4}.$$

Following Falconieri et al. (2004) we assume that league quality depends on the level of the competition, as well as the suspense associated with a close competition (competitive balance).[10] The level of the competition is measured by the aggregate talent within the n-club league. We assume that the marginal effect of the salary payment (talent investment) on the level of the competition T is positive but decreasing,

$$T(t_1,..,t_n) := \phi \sum_{j=1}^{n} t_j - \left(\sum_{j=1}^{n} t_j \right)^2.$$

This is guaranteed in our model if $\frac{\partial T}{\partial t_i} > 0 \Leftrightarrow \sum_{j=1}^{n} t_j < \frac{\phi}{2}$ and $\frac{\partial^2 T}{\partial t_i^2} < 0$ which will always be satisfied in equilibrium. Competitive balance CB is measured as minus the variance of salary payments

$$CB(t_1,..,t_n) := -\frac{1}{n} \sum_{j=1}^{n} \left(t_j - \bar{t}_n \right)^2 \text{ with } \bar{t}_n = \frac{1}{n} \sum_{j=1}^{n} t_j.$$

Note that a lower variance of salary payments by the n clubs implies a closer competition and therefore a higher degree of competitive balance. League quality is now defined as

$$q(t_1,..,t_n) := \mu T(t_1,..,t_n) + (1-\mu) CB(t_1,..,t_n). \tag{3.1}$$

The parameter $\mu \in (0,1)$ represents how much the fans weight aggregate talent and competitive balance. Given aggregate salaries $\sum_{j=1, j\neq i}^{n} t_j$ of the other $(n-1)$ clubs, league quality increases in club i's salary payment t_i until a threshold value $t_i'(\mu)$, i.e. $\frac{\partial q}{\partial t_i} > 0 \Leftrightarrow t_i < t_i'(\mu)$. Since fans have at least some preference for competitive balance excessive dominance by one club causes the quality to decrease.[11]

[9] Note that the optimal price is increasing in quality, i.e. $\frac{\partial p^*}{\partial q} > 0$.

[10] According to Szymanski (2003) fan demand depends not only on the level of the competition and competitive balance but also on the 'likelihood of the home team's success.' Taking home team winning into consideration would result in an asymmetric quality function but would not alter our basic findings. For the sake of simplicity, we abstract from home team winning.

[11] Note that the threshold value $t_i'(\mu)$ beyond which league quality decreases in club i's salary payments is an increasing function of the preference parameter μ because an increase in μ implies an increase in the preference for aggregate talent.

League revenues are split between the two types of clubs according to their bargaining power. For the sake of simplicity, we assume that half of the n clubs have strong bargaining power and half of them have weak bargaining power. Each of the strong, large-market clubs receives a fraction $\frac{m_l}{n/2}$ of league revenues and each of the weak, small-market clubs receives a fraction $\frac{m_s}{n/2}$ of league revenues, with

$$m_l > m_s \text{ and } m_l + m_s = 1.$$

We denote J_l and J_s as the set of large-market and small-market clubs, respectively, i.e. $J = \{1, .., n\} = J_l \cup J_s$.

The profit function $\Pi_i(t_1, .., t_n)$ of club $i \in J$ is given by revenue minus salary payments,

$$\Pi_i(t_1, .., t_n) = \frac{m_\delta}{2n} \left(\mu\phi \sum_{j=1}^{n} t_j - \mu \left(\sum_{j=1}^{n} t_j \right)^2 - \frac{1-\mu}{n} \sum_{j=1}^{n} \left(t_j - \overline{t}_n \right)^2 \right) - t_i, \qquad (3.2)$$

with $\delta = l$ for $i \in J_l$ and $\delta = s$ for $i \in J_s$.

Social welfare is given by the sum of aggregate consumer (fan) surplus, aggregate club profit and aggregate player salaries. Aggregate consumer surplus CS corresponds to the integral of the demand function $d(p, q)$ from the equilibrium price $p^* = \frac{q}{2}$ to the maximal price $\overline{p} = q$ which fans are willing to pay for quality q ,

$$CS = \int_{p^*}^{\overline{p}} d(p, q) dp = \int_{\frac{q}{2}}^{q} \frac{q-p}{q} dp = \frac{q}{8}.$$

Summing up aggregate consumer surplus, aggregate club profit and aggregate salary payments, social welfare is derived as

$$W(t_1, .., t_n) = \frac{3}{8} q(t_1, .., t_n). \qquad (3.3)$$

Note that salary payments do not directly influence social welfare because salaries merely represent a transfer from clubs to players. As a consequence, social welfare depends only on the quality of the league.

3.3 Salary Caps in a Profit-Maximizing League

Following Késenne (2000*b*), we introduce a salary cap into our model, which limits the total amount a club can spend on player salaries. The size of the salary cap, which is the same for each club, is based on the total league revenue in the previous season, divided by the number of clubs in the league. Therefore, the salary cap is exogenously-given in the current season.

Clubs choose salary levels such that profits (3.2) are maximized subject to the salary cap constraint.[12] That is, salary payments t_i must not exceed the threshold *cap* given by the salary cap. The maximization problem for club $i \in J$ is

$$\max_{t_i} \left\{ \frac{m_\delta}{2n} \left(\mu\phi \sum_{j=1}^{n} t_j - \mu \left(\sum_{j=1}^{n} t_j \right)^2 - \frac{1-\mu}{n} \sum_{j=1}^{n} \left(t_j - \bar{t}_n \right)^2 \right) - t_i \right\},$$

$$\text{subject to } t_i \leq cap,$$

with $\delta = l$ for $i \in J_l$ and $\delta = s$ for $i \in J_s$.

The corresponding first-order conditions are

$$\begin{aligned} \frac{m_\delta}{2n} \left(\mu \left(\alpha - 2\sum_{j=1}^{n} x_j \right) - \frac{2(1-\mu)}{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right) \right) - (1+\lambda_i) &\leq 0, \\ x_i \left(\frac{m_\delta}{2n} \left(\mu \left(\alpha - 2\sum_{j=1}^{n} x_j \right) - \frac{2(1-\mu)}{n} \left(x_i - \frac{1}{n} \sum_{j=1}^{n} x_j \right) \right) - (1+\lambda_i) \right) &= 0 \\ x_i - cap &\leq 0 \\ \lambda_i (x_i - cap) &= 0, \end{aligned} \tag{3.4}$$

where λ_i denotes the Lagrange multiplier for club $i \in J$ with $\delta = l$ for $i \in J_l$ and $\delta = s$ for $i \in J_s$.[13] To characterize the equilibrium, we have to distinguish different regimes depending on whether the salary cap is binding or not.

3.3.1 Regime A: Salary cap is ineffective for all clubs

In this section, we assume that the salary cap is ineffective for all clubs, i.e. we consider the benchmark case that no (effective) salary cap exists.

In regime A, the equilibrium salary payments (talent investments) are computed from

[12] For a discussion about the clubs' objective function see e.g. Sloane (1971) and Késenne (2000*a*).

[13] It is easy to show that the second-order conditions for a maximum are satisfied.

(3.4) as [14]

$$
\begin{aligned}
t_i^A &= \frac{\phi}{2n} - \frac{m_l(1-\mu(1+n^2)) + m_s(1+\mu(n^2-1))}{2m_l m_s(1-\mu)\mu} =: t_l^A \; \forall i \in J_l, \\
t_j^A &= \frac{\phi}{2n} - \frac{m_l(1+\mu(n^2-1)) + m_s(1-\mu(1+n^2))}{2m_l m_s(1-\mu)\mu} =: t_s^A \; \forall j \in J_s.
\end{aligned} \tag{3.5}
$$

For (3.5) to hold, in the following we restrict μ to $(\underline{\mu}, \overline{\mu})$.[15] The equilibrium salary payments show that all large-market (small-market) clubs choose the same salary level t_l^A (t_s^A). Note that without a binding salary cap the large-market clubs invest more in playing talent in equilibrium than the small-market clubs because the marginal revenue of talent investments is higher for these clubs. Thus, we are in regime A if in equilibrium the salary cap does not bind for the large-market clubs, i.e. if $cap \in I^A = [t_l^A, \infty)$.

In regime A, the aggregate level of salary payments $S^A = \sum_{j=1}^n t_j^A$ and competitive balance CB^A are given by

$$
S^A = \frac{\phi}{2} - \frac{n}{2\mu m_l m_s} \quad \text{and} \quad CB^A = -\left(\frac{n^2(m_l - m_s)}{2(1-\mu)m_l m_s}\right)^2. \tag{3.6}
$$

Note that S^A (CB^A) is increasing (decreasing) in μ. That is, the higher the preference of fans for aggregate talent is, the higher are aggregate salaries and the more unbalanced is the league. The opposite holds if fans have a high preference for competitive balance.

Plugging the equilibrium salary payments (3.5) into equation (3.3) for social welfare yields the following level of total welfare in regime A

$$
W^A = \frac{3}{32}\left(\mu\phi^2 - \left(\frac{1}{\mu}\left(\frac{n}{m_l m_s}\right)^2 + \frac{1}{(1-\mu)}\left(\frac{n^2(m_l - m_s)}{m_l m_s}\right)^2\right)\right).
$$

[14] We denote the salary payments of club $i \in J$ in regime A with t_i^A. Analogous for regime B and C.

[15] For μ very close to zero or one the optimal choice for some clubs is zero. Since we are not interested in a situation where clubs are not participating, we choose to restrict the range of μ to ensure positive equilibrium investments. Formally, we compute $(\underline{\mu}, \overline{\mu})$ as $\underline{\mu} = \frac{1}{2} - \frac{n^3(m_l - m_s) - n + ((n^3(m_s - m_l) + n + \phi m_l m_s)^2 - 4n\phi m_l m_s)^{1/2}}{2\phi m_l m_s}$ and $\overline{\mu} = \frac{1}{2} + \frac{n^3(m_s - m_l) + n + ((n^3(m_s - m_l) + n + \phi m_l m_s)^2 - 4n\phi m_l m_s)^{1/2}}{2\phi m_l m_s}$.

3.3.2 Regime B: Salary cap is only effective for large-market clubs

In this section, we assume that the salary cap is only effective for the large-market clubs. That is, the salary cap constraint is only binding for club i with $i \in J_l$.

In regime B, the equilibrium salary payments (talent investments) are computed from (3.4) as

$$\begin{aligned} t_i^B &= cap =: t_l^B \ \forall i \in J_l, \\ t_j^B &= \frac{n(\phi\mu m_s - 2n)}{m_s(1+\mu(n^2-1))} + cap\frac{1-\mu(n^2+1)}{1+\mu(n^2-1)} =: t_s^B \ \forall j \in J_s. \end{aligned} \tag{3.7}$$

Thus, we are in regime B if $cap \in I^B = (cap', t_l^A)$ with $cap' := \frac{\phi}{2n} - \frac{1}{\mu m_s}$. This condition guarantees that in equilibrium the small-market clubs invest less than cap. Otherwise the salary cap constraint would be binding for all clubs and regime C would be effective.

We now analyze how variations of the salary cap affect the clubs' optimal choice of salary payments. A more restrictive salary cap, i.e. a lower value of cap, induces the large-market clubs to decrease their salary payments in equilibrium, i.e. $\frac{\partial t_l^B}{\partial cap} > 0$. However, the effect on the small-market clubs' investment level is ambiguous since

$$\frac{\partial t_s^B}{\partial cap} = \frac{1-\mu(n^2+1)}{1+\mu(n^2-1)} \begin{cases} > 0 \text{ if } \mu \in \left(\underline{\mu}, \frac{1}{n^2+1}\right), \\ = 0 \text{ if } \mu = \frac{1}{n^2+1}, \\ < 0 \text{ if } \mu \in \left(\frac{1}{n^2+1}, \overline{\mu}\right). \end{cases}$$

Hence, a more restrictive salary cap induces the small-market clubs to decrease their salary payments in equilibrium if $\mu \in \left(\underline{\mu}, \frac{1}{n^2+1}\right)$ and to increase their salary payments in equilibrium if $\mu \in \left(\frac{1}{n^2+1}, \overline{\mu}\right)$.[16] As a consequence, the higher the fans' preference for aggregate talent, the less talent is lost through a more restrictive salary cap.

What is the intuition for this result? The tightening of the salary cap has two effects on the investment incentives of the small-market clubs. On the one hand a more restrictive cap lowers the salary payments by the large-market clubs and therefore enhances the incentive of the small clubs to pay higher salaries in order to 'compensate' for the decrease in aggregate talent.[17] On the other hand the incentive to improve competitive balance is weakened. If μ is relatively high, i.e. fans have a high preference for aggregate talent, then the first effect dominates the second effect and the

[16] Note that in equilibrium the small-market clubs never compensate the reduction of talent by the large-market clubs due to the salary constraint.

[17] Remember that quality is concave in aggregate talent.

small-market clubs increase their salary payments in equilibrium. If μ is relatively low, i.e. the fans have a high preference for competitive balance, then the incentive to improve competitive balance is lowered by the salary cap restriction so much that the small-market clubs will lower their salary payments in equilibrium. Finally, if $\mu = \frac{1}{n^2+1}$ then both effects exactly balance each other out.

The level of aggregate salary payments and competitive balance in regime B are given by

$$\begin{aligned} S^B =& \frac{n(1-\mu)}{1+\mu(n^2-1)} cap + \frac{n^2(\phi\mu m_s - 2n)}{2m_s(1+\mu(n^2-1))} \quad \text{and} \\ CB^B =& -\left(\frac{n(2n+\mu m_s(2n\cdot cap - \phi)}{2m_s(1+\mu(n^2-1))}\right)^2. \end{aligned} \tag{3.8}$$

Since $\frac{\partial t_l^B}{\partial cap} > \frac{\partial t_s^B}{\partial cap}$ a more restrictive salary cap will increase competitive balance and decrease aggregate salaries in regime B.

Social welfare in regime B is given by

$$W^B(cap) = \frac{3n(-n\mu(1-\mu)cap^2 + \phi\mu(1-\mu)cap)}{8(1+\mu(n^2-1))} + \frac{3n(n\phi^2\mu^2 m_s^2 - 4n^3)}{32m_s^2(1+\mu(n^2-1))}.$$

and is maximized if the salary cap is fixed at

$$cap^B_{\max} = \frac{\phi}{2n}. \tag{3.9}$$

In this case, social welfare is

$$W^B\left(\frac{\phi}{2n}\right) = \frac{3\phi^2\mu}{32} - \frac{3n^4}{8m_s^2(1+\mu(n^2-1))}.$$

Note that the welfare maximizing level of the salary cap $cap^B_{\max}$ need not necessarily lie within the interval of feasible salary caps I^B. If $\mu \in (\underline{\mu}, \mu']$ with

$$\mu' := \frac{1}{1+n^2(m_l - m_s)} \tag{3.10}$$

then the welfare maximizing level of the salary cap is not an element of the interval of feasible salary caps, i.e. $cap^B_{\max} \notin I^B$. Whereas if $\mu \in (\mu', \overline{\mu})$ then $cap^B_{\max} \in I^B$.[18] We defer the discussion of the implications to Section 3.4.

[18] See Appendix 3.6.1 for a derivation of condition (3.9) and (3.10).

3.3.3 Regime C: Salary cap is effective for all clubs

In this section, we assume that the salary cap is binding for the large-market and the small-market clubs. In this case, the equilibrium salary payments are simply given by

$$t_i^C = cap \text{ for all } i \in J. \tag{3.11}$$

We are in regime C if $cap \in I^C = (0, cap']$. Total salary payments are equal to $n \cdot cap$ and the competition is completely balanced. Social welfare in regime C is given by

$$W^C(cap) = \frac{3n}{8}\mu(-n \cdot cap^2 + \phi cap).$$

3.4 Comparison of the Regimes

By comparing the aggregate salary payments and competitive balance in regime A, B and C, we derive the following proposition:

Proposition 3.1.
(i) The level of competitive balance is decreasing in cap*, i.e.* $CB^C \geq CB^B \geq CB^A$.
(ii) The level of aggregate salaries is increasing in cap*, i.e.* $S^A \geq S^B \geq S^C$.

Proof. This result follows directly from (3.5), (3.7) and (3.11) and the definitions of I^k, $k \in \{A, B, C\}$. □

This proposition shows that the introduction of a salary cap has the expected effect of increasing competitive balance and decreasing aggregate salaries.

By comparing social welfare in regime A, B and C, we establish the following proposition:

Proposition 3.2.
(i) If $\mu \in (\underline{\mu}, \mu']$*, i.e. fans prefer competitive balance, then an effective salary cap is always detrimental to social welfare.*
(ii) If $\mu \in (\mu', \overline{\mu})$*, i.e. fans prefer aggregate talent, then social welfare is maximized in regime* B.

Proof. See Appendix 3.6.2. □

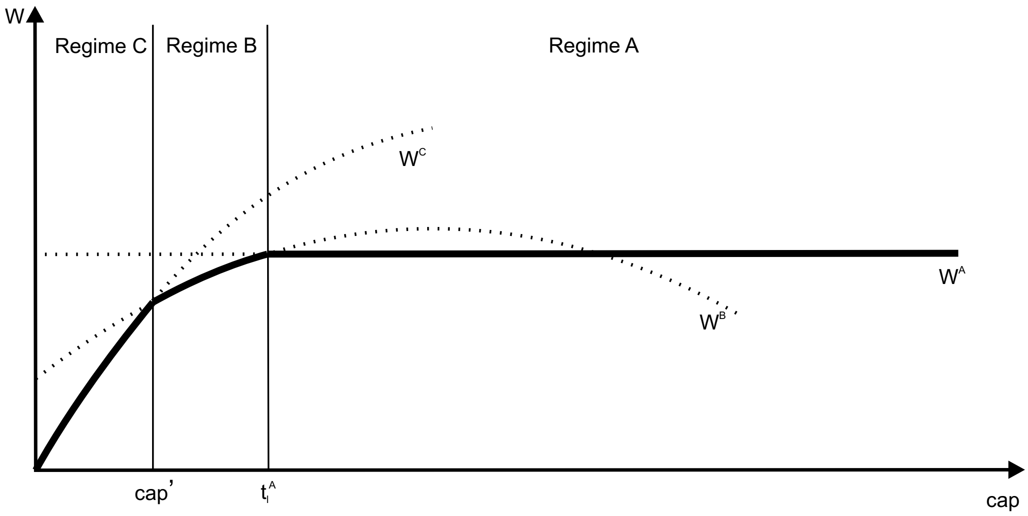

Figure 3.1: Effect of Salary Caps on Social Welfare - Fans Prefer Competitive Balance

To see the intuition behind Proposition 3.2 consider Figure 3.1 and Figure 3.2. The figures plot social welfare as a function of the salary cap for the case that fans prefer competitive balance (Figure 3.1) and the case that fans prefer aggregate talent (Figure 3.2). In both figures, the dashed line shows the hypothetical levels of social welfare in the different regimes while the bold line depicts the actual attainable levels of social welfare. Remember that regime A is only effective for $cap \geq t_l^A$, regime B for $cap' < cap < t_l^A$ and regime C for $cap \leq cap'$. Also remember that a salary cap decreases aggregate talent in favor of a more even competition.

Figure 3.1 shows the case in which fans prefer competitive balance, i.e. $\mu \in (\underline{\mu}, \mu')$. The figure shows that the introduction of a binding salary cap decreases social welfare in regime B compared with regime A. This counter-intuitive result is because of the fact that the unrestricted equilibrium in case of a high preference for competitive balance is already characterized by a high level of competitive balance and a low level of aggregate talent.[19] At these equilibrium levels, the marginal benefit of increased competitive balance through the salary cap is small, while the marginal loss due to a decrease in aggregate talent is high. Additionally, for low μ less talent is lost through a more restrictive salary cap. As a consequence, there is no need to additionally increase competitive balance since the loss in aggregate talent outweighs the gains from a more even competition.

However, this changes as μ increases to $\mu > \mu'$, i.e. fans prefer aggregate talent. Figure 3.2 depicts this situation. Here, the unrestricted equilibrium is characterized by a relatively high level of aggregate talent and a low level of competitive balance. In

[19]Remember that S^A (CB^A) is increasing (decreasing) in μ.

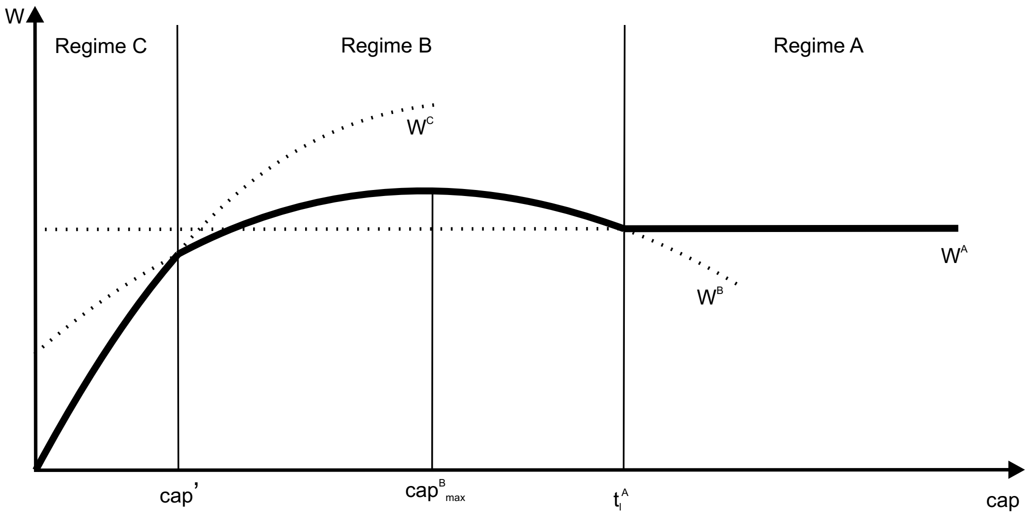

Figure 3.2: Effect of Salary Caps on Social Welfare - Fans Prefer Aggregate Talent

this case, a binding salary cap for the large-market clubs will increase social welfare in regime B compared with regime A because the marginal benefit of increased competitive balance overcompensates the marginal loss because of a decrease in aggregate talent. This is true until the welfare maximum is attained at $cap = \frac{\phi}{2n}$. Beyond that threshold, social welfare starts to decrease again, as the loss in talent cannot be overcompensated by the increase in competitive balance.

Imposing a stricter salary cap than cap' (implementing regime C) can never be optimal from a social point of view because the resulting loss in aggregate talent is not compensated by a positive effect on competitive balance, as the competition is already perfectly balanced.[20]

3.5 Conclusion

Salary caps are employed within professional team sports leagues all over the world. Conventional wisdom suggests that they are a collusive effort of club owners to control labour costs. Based on this assumption most economists would predict that salary caps decrease social welfare. Based on a game-theoretical model of a league consisting of small- and large-market clubs, we have shown that a salary cap may increase or decrease social welfare depending upon the fans' valuation of competitive balance and

[20]Note that if the fans' preference for aggregate talent increases beyond another threshold $\mu'' = \frac{3m_l+m_s}{3m_l+m_s+n^2(m_l-m_s)}$, i.e. $\mu \in (\mu'', \overline{\mu})$, then social welfare can also be higher in regime C than in regime A, although the welfare maximum is also reached in regime B. See Figure 3.3 at the end of the appendix for a graphical illustration of this situation.

aggregate talent. A salary cap that binds only for large-market clubs will increase social welfare if fans prefer aggregate talent despite the fact that the salary cap will result in lower aggregate talent. If fans prefer competitive balance, on the other hand, any binding salary cap will reduce social welfare. In any case, a binding cap will increase competitive balance and will help to keep salary costs under control. Moreover, we can show that if salary caps are beneficial for social welfare they also increase club profits.[21] Therefore clubs will never oppose salary caps which have a positive effect on social welfare. However, caution is necessary since there exists a range of the preference parameter μ within which club profits increase and social welfare decreases through the introduction of a salary cap.[22] These results suggest that salary caps need not be a collusive effort but can be an important mechanism to increase social welfare within professional team sports leagues.

[21] The analysis of club profits is similar to the analysis of social welfare.

[22] Formally: If the fans relative preference for aggregate talent is in the interval $(\widehat{\mu}', \mu')$ with $\widehat{\mu}' := \frac{n-4m_l m_s}{n^3(m_l-m_s)+n-4m_l m_s}$ then the introduction of a salary cap will be beneficial for the clubs and detrimental to social welfare.

3.6 Appendix

3.6.1 Derivation of conditions (3.9) and (3.10)

We compute

$$\frac{\partial W^B(cap)}{\partial cap} = \frac{3}{8}\left(\frac{n(\phi - 2n \cdot cap)(1-\mu)\mu}{1+\mu(n^2-1)}\right) > 0 \Leftrightarrow cap < cap^B_{\max} = \frac{\phi}{2n}.$$

However, the welfare maximizing salary cap $cap^B_{\max}$ need not necessarily be within the interval of feasible salary caps $I^B = (cap', t^A_l)$ with $cap' = \frac{\phi}{2n} - \frac{1}{\mu m_s}$ in regime B. We derive

$$t^A_l \leq cap^B_{\max} \Leftrightarrow \mu \leq \mu' := \frac{1}{1+n^2(m_l - m_s)} \in (\underline{\mu}, \overline{\mu}).$$

Hence, if $\mu \in (\underline{\mu}, \mu']$ then $cap^B_{\max} \notin I^B$ and a more restrictive salary cap, i.e. a lower variable cap, will decrease social welfare $W^B(cap)$ in regime B.

However, if $\mu \in (\mu', \overline{\mu})$ then the welfare maximizing salary cap $cap^B_{\max}$ is in the interval of feasible salary caps I^B, i.e. $cap^B_{\max} \in I^B$. In this case the effect of a more restrictive salary cap on social welfare depends crucially on the size of the salary cap. Formally, we derive

$$\frac{\partial W^B(cap)}{\partial cap} > 0 \; \forall cap \in (cap', cap^B_{\max}) \text{ and } \frac{\partial W^B(cap)}{\partial cap} < 0 \; \forall cap \in (cap^B_{\max}, t^A_l).$$

3.6.2 Proof of Proposition 3.2

This proof consists of three parts. In part (1) we compare regime A and B, in part (2) regime A and C and in part (3) regime B and C with respect to social welfare. Remember that regime A is only effective for $cap \geq t^A_l$, regime B for $cap' < cap < t^A_l$ and regime C for even tighter salary caps, $cap \leq cap'$.

(1) By comparing social welfare in regime A and B, we derive

$$W^A \leq W^B(cap) \Leftrightarrow cap \in [cap^{AB}_1, cap^{AB}_2],$$

where

$$\begin{aligned}
cap^{AB}_1 &= \frac{\phi}{2n} - \frac{m_l(1-\mu(1+n^2)) + m_s(1+\mu(n^2-1))}{2m_l m_s(1-\mu)\mu}, \\
cap^{AB}_2 &= \frac{\phi}{2n} + \frac{m_l(1-\mu(1+n^2)) + m_s(1+\mu(n^2-1))}{2m_l m_s(1-\mu)\mu}.
\end{aligned}$$

Note that cap_1^{AB} is exactly the equilibrium investment level of the large-market clubs in regime A, i.e. $cap_1^{AB} = t_l^A$.

We now analyze whether a salary cap from the interval $[cap_1^{AB}, cap_2^{AB}]$, for which social welfare is higher in regime B than in regime A, is part of the interval I^B of feasible salary caps in regime B. We derive

$$cap_1^{AB} < cap_2^{AB} \Leftrightarrow \mu < \mu' := \frac{1}{1 + n^2(m_l - m_s)} \in (\underline{\mu}, \overline{\mu}).$$

(1a) If $\mu \in (\underline{\mu}, \mu')$ then $\nexists$ $cap' \in [cap_1^{AB}, cap_2^{AB}]$ such that $cap' \in I^B$. That is, we cannot find a salary cap out of the interval $[cap_1^{AB}, cap_2^{AB}]$ which is also included in the interval I^B of feasible salary caps for regime B. Hence,

$$W^A > W^B(cap) \; \forall cap \in (cap', cap_1^{AB}) = I^B.$$

This shows that an effective salary cap is always detrimental to social welfare because social welfare is higher in regime A than in regime B.

(1b) If $\mu = \mu'$ then $W^A = W^B(cap) \Leftrightarrow cap = \frac{\phi}{2n} = t_l^A$. Since $I^B = (cap', t_l^A)$ we also conclude that $W^A > W^B(cap) \; \forall cap \in I^B$.

(1c) If $\mu \in (\mu', \overline{\mu})$ then $cap_1^{AB} > cap_2^{AB}$. In this case, we have to analyze if cap_2^{AB} is in the interval of feasible salary caps I^B. We derive

$$cap_2^{AB} \leq cap' \Leftrightarrow \mu \geq \mu'' := \frac{3m_l + m_s}{3m_l + m_s + n^2(m_l - m_s)} \in (\mu', \overline{\mu}).$$

(i) If $\mu \in (\mu', \mu'')$ then $cap' < cap_2^{AB}$ and thus the interval $[cap_2^{AB}, cap_1^{AB}]$ is a subset of the interval I^B. In this case the size of the salary cap determines whether social welfare is higher in regime A or B. More precisely, $W^A \geq W^B(cap)$ $\forall cap \in (cap', cap_2^{AB}]$ and $W^A < W^B(cap)$ $\forall cap \in (cap_2^{AB}, t_l^A)$

(ii) If $\mu \in [\mu'', \overline{\mu})$ then $cap' \geq cap_2^{AB}$ and thus social welfare in regime B is higher than in regime A independent of the size of the salary cap, i.e. $W^A < W^B(cap)$ $\forall cap \in I^B$.

Moreover, note that social welfare is maximized in regime B if the salary cap is fixed at $cap_{\max} = \frac{\phi}{2n}$.

(2) By comparing social welfare in regime A and C, we derive

$$W^A \leq W^C(cap) \Leftrightarrow cap \in [cap_1^{AC}, cap_2^{AC}],$$

where

$$
\begin{aligned}
cap_1^{AC} &= \frac{\phi}{2n} - \frac{\sqrt{(n^2\mu m_l m_s)^2(1-\mu)(1-\mu+\mu n^2(m_l-m_s)^2)}}{2(n\mu m_l m_s)^2(1-\mu)},\\
cap_2^{AC} &= \frac{\phi}{2n} + \frac{\sqrt{(n^2\mu m_l m_s)^2(1-\mu)(1-\mu+\mu n^2(m_l-m_s)^2)}}{2(n\mu m_l m_s)^2(1-\mu)}.
\end{aligned}
$$

We derive that $cap_1^{AC} < cap_2^{AC}$ and $cap_2^{AC} > cap'$, i.e. cap_2^{AC} is not in the interval of feasible salary caps I^C for regime C.

Similar to (1) we analyze whether a salary cap from the interval $[cap_1^{AC}, cap_2^{AC}]$, for which social welfare is higher in regime C than in regime A, is part of the interval $I^C = (0, \frac{\phi}{2n} - \frac{1}{\mu m_s}]$ of feasible salary caps in regime C. We derive

$$
cap_1^{AC} \leq cap' \Leftrightarrow \mu \geq \mu'' := \frac{3m_l + m_s}{3m_l + m_s + n^2(m_l - m_s)}.
$$

(2a) If $\mu \in (\underline{\mu}, \mu'')$ then $cap_1^{AC} > cap'$. In this case cap_1^{AC} is not in the interval of feasible salary caps I^C for regime C and thus we derive that social welfare is higher in regime A than in regime C, i.e. $W^A > W^C(cap)\ \ \forall cap \in I^C$.

(2b) If $\mu \in [\mu'', \overline{\mu})$ then $cap_1^{AC} \leq cap'$. In this case the size of the salary cap determines whether social welfare is higher in regime A or C. More precisely, $W^A > W^C(cap)$ for all $cap \in (0, cap_1^{AC})$ and $W^A < W^C(cap)$ for all $cap \in (cap_1^{AC}, cap']$. Note that for $cap = cap'$ holds $W^A = W^C(cap)$.

Moreover, we derive that in regime C social welfare would also be maximized if the salary cap was fixed at $cap_{\max} = \frac{\phi}{2n}$. However, this welfare maximizing salary cap is never part of the interval I^C.

(3) By comparing social welfare in regime B and C, we derive

$$
W^B(cap) \leq W^C(cap) \Leftrightarrow cap \in [cap_1^{BC}, cap_2^{BC}],
$$

where

$$
cap_1^{BC} = \frac{\phi}{2n} - \frac{1}{\mu m_s} \text{ and } cap_2^{BC} = \frac{\phi}{2n} + \frac{1}{\mu m_s}.
$$

Note that $cap_1^{BC} = cap'$. Moreover, regime C is only effective if $cap \in (0, cap']$. This directly implies that social welfare in regime C can never be higher than in regime B. As a consequence, implementing a sufficiently strict salary cap (i.e. $cap \leq cap'$) such that regime C is effective, will always decrease social welfare.

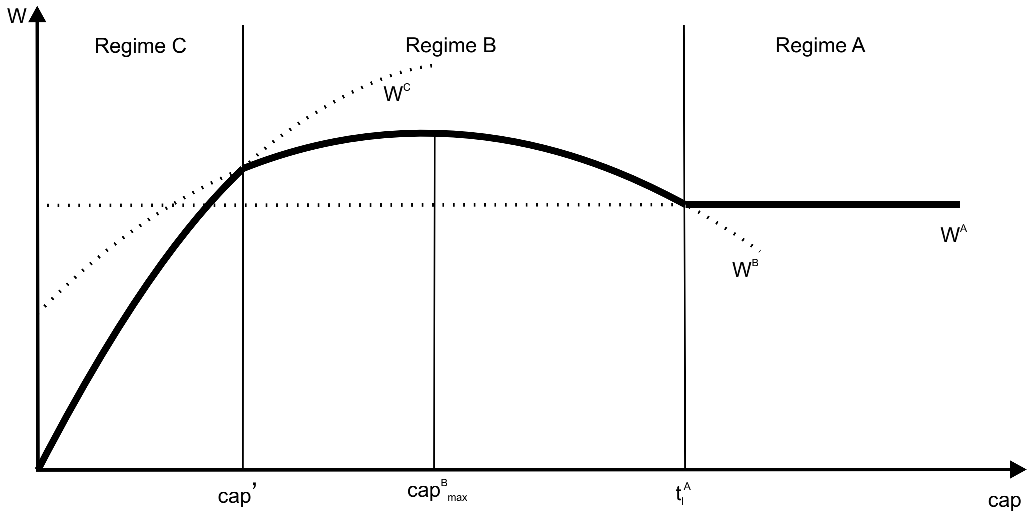

Figure 3.3: Effect of Salary Caps on Social Welfare for $\mu \in (\mu'', \overline{\mu})$

Figure 3.3 depicts the situation in which aggregate talent is even more important than in Figure 3.2, i.e. $\mu \in (\mu'', \overline{\mu})$.[23]

[23]Remember that the dashed line shows the hypothetical levels of social welfare in the different regimes while the bold line depicts the actual attainable levels of social welfare.

Chapter 4

The Welfare Effect of Gate Revenue-Sharing

4.1 Introduction

According to the 'uncertainty of outcome' hypothesis a certain degree of (competitive) balance is necessary to maintain a successful sporting contest.[1] One of the most common means of improving competitive balance within a professional sports league is gate revenue-sharing. In its simplest form gate revenue-sharing allows the visiting club to retain a share of the home club's gate revenues.

The current revenue-sharing arrangements differ widely among professional leagues all over the world. In 1876 the Major League Baseball (MLB) introduced a 50-50 split of gate receipts that was reduced over time. Since 2003 all clubs in the American League have put 34% of their locally-generated revenue (gate, concession, television, etc.) into a central pool which is then divided equally among clubs. The current revenue-sharing arrangement of the National Football League (NFL) secures the visiting team 40% of the gate receipts (revenues from luxury boxes, parking and concessions are excluded from this sharing arrangement). In the Australian Football League (AFL) gate receipts were split evenly between the home and the visiting team. This 50-50 split was finally abolished in 2000. In Europe there is less gate revenue-sharing. The soccer leagues have adopted various forms of gate revenue-sharing in their history. In England until the early 1980s up to 20% of the gate receipts were given to the visiting teams in league matches. In the German soccer league (DFL), the home team receives 94% of the gate receipts with the other 6% going to the league. Gate revenue-sharing

[1] This chapter is based on a paper which was published in *Contemporary Economic Policy* (see Dietl and Lang (2008)). Reprinted with permission.

is quite common in most Cup competitions with a knock-out system. In addition some leagues have adopted other means of increasing competitive balance, such as salary caps, rookie draft systems and luxury taxes.

The effect of gate revenue-sharing on competitive balance has been challenged by the so-called 'invariance proposition,' which states that revenue-sharing does not affect the distribution of talent between clubs. The 'invariance proposition' has remained highly controversial even up until today and no consensus has emerged so far. Most of the existing controversy on the effect of revenue-sharing on competitive balance stems from the different approaches, the different models and the different methodology used in the literature.

El-Hodiri and Quirk (1971), Fort and Quirk (1995) and Rascher (1997) conclude that revenue-sharing will not affect the distribution of talent between profit-maximizing clubs on the assumption that only the win percentage of the home team affects club revenue.[2] Vrooman (1995) shows that the sharing of winning-elastic revenue does not affect competitive balance whereas the sharing of winning-inelastic revenue improves competitive balance. Atkinson et al. (1988) challenge the 'invariance proposition' by showing that revenue-sharing can improve competitive balance if clubs maximize profits. In their model Atkinson et al. adopt a pool sharing arrangement and a club revenue function that depends on a team's own performance and on the performance of all other teams. Their result is supported by Marburger (1997) and Késenne (2000*a*)), who build their models on the assumption that fans care about the relative and absolute quality of teams, and Rascher (1997) and Késenne (2000*a*), who both consider an objective function which includes the maximization of wins ('utility maximization').

The most counter-intuitive result is presented by Szymanski and Késenne (2004). From a model of two profit-maximizing clubs and a club revenue function that depends on the relative quality of the home team, they show that gate revenue-sharing decreases competitive balance. This result is driven by the so-called 'dulling effect.' The dulling effect describes the well-known result in sports economics that revenue-sharing reduces the incentive to invest in playing talent. This dulling effect is stronger for the small-market club than for the large-market club. In effect, the difference in talent investments between both clubs increases.

In our opinion, the major drawback of the literature analyzing the effect of revenue-sharing is the implicit assumption that competitive balance is socially desirable. On

[2] Note that Fort and Quirk (1995) derive this result on the assumption that only gate revenues are shared. Moreover, they argue that the sharing of locally-generated television revenues can improve competitive balance when teams earn revenue from the gate and from local television contracts.

this assumption the revenue-sharing arrangement that maximizes competitive balance is optimal. We show that this assumption does not hold true in our model. Maximizing competitive balance does not maximize social welfare. In particular, we derive club-specific demand, revenue and profit functions from a general fan utility function and develop a contest model of a sports league with heterogeneous clubs. From the (consumer) utility and (club) profit functions we are able to analyze the welfare effects of alternative gate revenue-sharing arrangements. We arrive at two counter-intuitive results. First, we derive that gate revenue-sharing decreases competitive balance. Second, we show that a lower degree of competitive balance compared with the non-cooperative league equilibrium actually yields a higher level of social welfare and club profits. Combining the two results, we conclude that gate revenue-sharing increases both social welfare and club profits in our model.

Of course, our analysis is not the first to integrate consumer preferences in economic models of sports leagues. To our knowledge Cyrenne (2001) was the first to model explicitly consumer preferences and perform a welfare analysis in a sports league. He develops a quality-of-play model which captures consumer preferences and shows under which conditions the clubs' demand for talented players are strategic complements or substitutes. Falconieri et al. (2004) investigate the conditions under which the collective sale of broadcasting rights is preferred from a social welfare point of view compared with their sale individually by teams. In their model, they derive the demand and the price for a match with a given quality via consumer preferences. To the best of our knowledge, however, we are the first to analyze the welfare effects of gate revenue-sharing on the basis of a general fan utility function.

The remainder of this chapter is organized as follows: Section 4.2 outlines the model. In Section 4.3 we investigate the non-cooperative equilibrium and in Section 4.4 the social welfare optimum and league optimum. Section 4.5 presents a comparison between the outcomes. Section 4.6 concludes the chapter.

4.2 Model Specification

In this chapter, we consider a league with two asymmetric clubs, i.e. $J = \{1, 2\}$. The league awards an exogenously-given league prize P to the winner of the championship. As in Chapter 2, the exogenous league prize is a proxy for all sorts of performance-related revenues like sponsorship contracts and the secure monetary value of qualifying for international competition. In addition to this performance-related exogenous league prize, each club i generates its own revenues R_i stemming from gate receipts of the match played at the ground of club i against club j. These revenues R_i are assumed to depend on the gate price p_i and club i's fan demand $d(m_i, p_i, q_i)$ for the match between club i and club j. Fan demand in turn depends on the market size m_i (or drawing power) of the club i, the gate price p_i and the quality q_i of the match between club i and club j. Moreover, the gate revenues from the home and away match are distributed among clubs according to a (gate) revenue-sharing arrangement with $\alpha \in [\frac{1}{2}, 1]$ characterizing the share assigned to the home team. Note that a high parameter α represents a league with a low degree of redistribution, i.e. $\alpha = 1$ characterizes a league without revenue-sharing, while $\alpha = 1/2$ characterizes a league with full revenue-sharing.

In order to derive the price and the fan demand for a match with quality q_i we follow the approach in Falconieri et al. (2004): we assume a continuum of fans that differ in their willingness to pay for a match between club i and club j with a given quality q_i. Every fan k has a certain preference for match quality which is measured by θ_k. For simplicity, the fan types θ_k are assumed to be uniformly distributed in $[0, 1]$, i.e. the measure of potential fans amounts to 1. The net-utility of fan k with type θ_k is specified as

$$\max\{\theta_k q_i - p_i, 0\}.$$

By assuming an interior solution at price p_i the fan type which is indifferent is given by $\theta^* = \frac{p_i}{q_i}$. Hence the measure of fans that purchase at p_i is derived as $1-\theta^* = \frac{q_i - p_i}{q_i}$. By assuming that each club has a certain market size or drawing potential given by $m_i > 0$, the aggregate demand function for club $i = 1, 2$ is now defined as

$$d(m_i, p_i, q_i) := m_i \frac{q_i - p_i}{q_i} = m_i \left(1 - \frac{p_i}{q_i}\right).$$

Note that match quality q_i has a positive, but decreasing, marginal effect on demand, i.e. $\frac{\partial d}{\partial q_i} > 0$ and $\frac{\partial^2 d}{\partial^2 q_i} < 0$. Moreover, the market size or drawing potential m_i has a positive effect on demand, i.e. $\frac{\partial d}{\partial m_i} > 0$. For a given set of parameters (p_i, q_i), the club with the larger market size or drawing potential generates higher demand. Without loss of generality, we assume throughout this chapter that club 1 is the 'large-market club' and club 2 the 'small-market club' with $m_1 \geq m_2$. By normalizing the costs of hosting a match to zero, we find that gate revenues are derived as $R_i = p_i d(m_i, p_i, q_i)$. The club maximizes the revenues R_i and thus fixes the price of a match with quality q_i to $p_i^* = \frac{q_i}{2}$. Hence, gate revenues of club $i = 1, 2$ are derived as

$$R_i = \frac{m_i}{4} q_i.$$

In accordance with the literature we assume that the match quality q_i depends on two factors: the probability of club i's success and the uncertainty of outcome.[3]

We measure the probability of club i's success by the win percentage w_i of this club. The win percentage is characterized by the contest-success function (CSF) and depends on the proportion of playing talent hired by each club. As in Chapter 2, the win percentage is characterized by the logit contest-success function (CSF) and depends on the proportion of playing talent hired by each club, yielding the following win percentage of club i:[4]

$$w_i(t) = \frac{t_i}{t_i + t_j} \quad (i, j = 1, 2, \ i \neq j).$$

Given that the win percentages must sum up to unity, we obtain the adding-up constraint: $w_j = 1 - w_i$. In our model, we allow that the supply of talent may be fixed or elastic. Furthermore, we adopt the usual 'Contest-Nash conjecture' $\frac{dt_i}{dt_j} = 0$ and compute the derivative of the win percentages as $\frac{\partial w_i}{\partial t_i} = \frac{t_j}{(t_i+t_j)^2}$.[5] Owing to the adding-up constraint we derive

$$\frac{\partial w_i}{\partial t_i} = -\frac{\partial w_j}{\partial t_i}. \tag{4.1}$$

The uncertainty of outcome is measured by the competitive balance in the league. Following Hoehn and Szymanski (1999) and Szymanski (2003), we specify competitive balance as $w_i w_j$.

In order to deduce explicitly the gate revenues, we use the following specific formulation of the quality q_i from a match played between club i and club j:[6]

$$q_i(w_i, w_j) := \mu w_i + (1-\mu) w_i w_j \quad (i, j = 1, 2, \ i \neq j) \tag{4.2}$$

with $\frac{\partial q_i}{\partial w_i} = 1 - 2(1-\mu)w_i$. The parameter $\mu \in [0, 1]$ represents the weight in the quality function between fans' preference for 'own team winning' and for competitive

[3]For the sake of simplicity, we abstract from the possibility that match quality also depends on aggregate talent. This is a restrictive assumption but can be justified by a focus on North America, where all available talent plays in the major leagues.

[4]For the sake of simplicity, we have set the discriminatory power γ of the CSF to unity.

[5]According to Szymanski (2004) 'it makes no sense to talk of any conjectural variation other than zero.' Only the Contest-Nash conjectures are consistent with the concept of Nash-equilibrium in a static game. Moreover, note that the assumption of fixed or elastic supply only affects the equilibrium price of talent in our model.

[6]Note that this specification of the quality function differs fundamentally from the quality function used in Falconieri et al. (2004). Moreover, we will see below that the gate revenues which are derived from our specification of the quality function give rise to the revenue functions widely used in the sports economic literature.

balance. When $\mu = 1$ then the match quality only depends on the win percentage of the home team, while when $\mu = 0$ the match quality only depends on the degree of competitive balance. If the relative preference for 'own team winning' is equal or bigger than $1/2$, then the match quality increases in the win percentage of the home team for all $w_i \in [0, 1]$. Whereas, if the relative preference for 'own team winning' is smaller than $1/2$, then match quality increases in the win percentage if $w_i < \frac{1}{2(1-\mu)} \leq 1$.

With this specification of the quality function, gate revenues R_i of club $i = 1, 2$ are now given by

$$R_i = \frac{m_i}{4} q_i = \frac{m_i}{4}(w_i - (1 - \mu)w_i^2). \tag{4.3}$$

This club-specific revenue function is consistent with the revenue functions used e.g. in Hoehn and Szymanski (1999), Szymanski (2003) and Szymanski and Késenne (2004). In contrast, however, with the articles quoted, we have derived our revenue function from consumer preferences and thus are able to perform a welfare analysis with respect to gate revenue-sharing.

4.3 Non-Cooperative Equilibrium

In this section we consider the competitive equilibrium in the league. Both clubs participate in a non-cooperative game and choose independently a level of talent in order to maximize (expected) profits.[7] As in Chapters 2 and 3, we assume that talent is measured in perfectly divisible units that can be hired in a competitive market for talent at a constant wage rate c per unit. Hence, club i's investment costs $C(t_i)$ for talent are given by $C(t_i) = ct_i$. The expected payoff of club $i = 1, 2$ is determined by the following (expected) profit function:

$$\begin{aligned} E(\Pi_i) &= w_i P + \alpha R_i + (1 - \alpha) R_j - C(t_i) \\ &= w_i P + \alpha \left(\frac{m_i}{4}(w_i - (1 - \mu)w_i^2) \right) + (1 - \alpha) \left(\frac{m_j}{4}(w_j - (1 - \mu)w_j^2) \right) - ct_i. \end{aligned}$$

with $i, j = 1, 2,\ i \neq j$. With probability w_i club i wins the championship given club i's and club j's investment level t_i and t_j, respectively, and receives the exogenous league prize P. From the home match club i obtains share α of the gate revenues

[7] The clubs in the US major leagues are commonly considered to be profit maximizers whereas in Europe clubs are usually considered to be win maximizers. The situation in Europe is changing, however, as many examples (Manchester United and Liverpool) demonstrate. For a discussion about the clubs' objective function see e.g. Sloane (1971), Késenne (2000*a*), Fort and Quirk (2004) and Késenne (2006).

$R_i = \frac{m_i}{4}(w_i - (1-\mu)w_i^2)$ and from the away match share $(1-\alpha)$ of the gate revenues $R_j = \frac{m_j}{4}(w_j - (1-\mu)w_j^2)$. The investment costs are determined by ct_i.

The corresponding first-order conditions are derived as

$$\begin{aligned}
\frac{\partial E(\Pi_1)}{\partial t_1} &= \frac{\partial w_1}{\partial t_1}P + \alpha\frac{\partial R_1}{\partial w_1}\frac{\partial w_1}{\partial t_1} + (1-\alpha)\frac{\partial R_2}{\partial w_2}\frac{\partial w_2}{\partial t_1} - c = 0, \\
\frac{\partial E(\Pi_2)}{\partial t_2} &= \frac{\partial w_2}{\partial t_2}P + \alpha\frac{\partial R_2}{\partial w_2}\frac{\partial w_2}{\partial t_2} + (1-\alpha)\frac{\partial R_1}{\partial w_1}\frac{\partial w_1}{\partial t_2} - c = 0.
\end{aligned}$$

By combining the first-order conditions and using the adding-up constraint (4.1), we obtain

$$\left(P + \alpha\frac{\partial R_1}{\partial w_1} - (1-\alpha)\frac{\partial R_2}{\partial w_2}\right)\frac{\partial w_1}{\partial t_1} = \left(P + \alpha\frac{\partial R_2}{\partial w_2} - (1-\alpha)\frac{\partial R_1}{\partial w_1}\right)\frac{\partial w_2}{\partial t_2}$$

and compute

$$\frac{\frac{\partial w_2}{\partial t_2}}{\frac{\partial w_1}{\partial t_1}} = \frac{t_1^*}{t_2^*} = \frac{P + \alpha\frac{\partial R_1}{\partial w_1} - (1-\alpha)\frac{\partial R_2}{\partial w_2}}{P + \alpha\frac{\partial R_2}{\partial w_2} - (1-\alpha)\frac{\partial R_1}{\partial w_1}}. \tag{4.4}$$

We are not able to solve explicitly for the equilibrium investments (t_1^*, t_2^*). Instead, we establish the following relationship which must hold in equilibrium between club 1's and club 2's investment level t_1^* and t_2^*, respectively:

$$t_1^* = \psi(\alpha)t_2^*.$$

In the following lemma, we specify the function $\psi(\alpha)$ and derive some useful properties of it by assuming that the exogenous league prize P is sufficiently high:

Lemma 4.1.
(i) The function $\psi(\alpha)$ which describes the relationship in the non-cooperative equilibrium between t_1^* and t_2^* is given by[8]

$$\psi(\alpha) = \frac{1}{2\lambda_2}\left(\rho + \sqrt{\rho^2 + 4\lambda_1\lambda_2}\right)$$

with $\rho := (1-2\alpha(1-\mu))(m_1 - m_2), \lambda_1 := 4P + \alpha m_1 - (1-\alpha)(2\mu - 1)m_2$ and $\lambda_2 := 4P + \alpha m_2 - (1-\alpha)(2\mu-1)m_1$.
(ii) $\psi(\alpha)$ is an increasing function in revenue-sharing, i.e. a decreasing function in the parameter α such that $\frac{\partial\psi(\alpha)}{\partial\alpha} < 0$.
(iii) $\psi(\alpha)$ is equal to or larger than unity, i.e. $\psi(\alpha) = 1 \Leftrightarrow \mu = 0 \vee m_1 = m_2$ and $\psi(\alpha) > 1 \Leftrightarrow \mu > 0 \wedge m_1 > m_2 \ \ \forall\alpha \in [\frac{1}{2}, 1]$.

[8]Note that ψ is a function of $(\alpha, \mu, m_1, m_2, P)$. For notational clarity we only write $\psi(\alpha)$.

Proof. See Appendix 4.7.1. □

Unless otherwise stated, we assume in the subsequent analysis that fans besides competitive balance also care about their own team winning, i.e. $\mu \in (0,1)$ and clubs are heterogeneous with respect to their market size, i.e. $m_1 > m_2$.

The win percentages in the non-cooperative equilibrium of club 1 and club 2 can be expressed in terms of $\psi(\alpha)$ by

$$w_1^*(\alpha) = \frac{\psi(\alpha)}{\psi(\alpha)+1} \text{ and } w_2^*(\alpha) = \frac{1}{\psi(\alpha)+1}$$

with the derivatives given by

$$\frac{\partial w_1^*(\alpha)}{\partial \alpha} = \frac{\partial \psi(\alpha)}{\partial \alpha} \frac{1}{(\psi(\alpha)+1)^2} \text{ and } \frac{\partial w_2^*(\alpha)}{\partial \alpha} = -\frac{\partial \psi(\alpha)}{\partial \alpha} \frac{1}{(\psi(\alpha)+1)^2}.$$

The next proposition summarizes the main results in this section:

Proposition 4.1.
(i) The investment level of the large-market club 1 is higher than the investment level of the small-market club 2.
(ii) Equilibrium investments decrease in revenue-sharing.
(iii) The win percentage of club 1 (club 2) increases (decreases) in revenue-sharing.
(iv) Revenue-sharing reduces competitive balance and produces a more unbalanced league.

Proof. See Appendix 4.7.2. □

A direct consequence of part (i) is that in the non-cooperative equilibrium, the large-market club 1 is the dominant team yielding a win percentage of more than $\frac{1}{2}$ and the small-market club is the subordinate team yielding a win percentage of less than $\frac{1}{2}$, independent of the revenue-sharing arrangement. The reason for this result is that the marginal impact of an additional win on gate revenues is higher for the large-market club than for the small-market club. Moreover, the difference in win percentages between club 1 and club 2 in the non-cooperative equilibrium is given by $w_1^*(\alpha) - w_2^*(\alpha) = \frac{\psi(\alpha)-1}{\psi(\alpha)+1} > 0$.

Part (ii) reflects the 'dulling effect' of revenue-sharing. The dulling effect describes the well-known result in sports economics that revenue-sharing reduces the incentive to invest in playing talent. This result follows from the fact that the marginal benefit of own investment has to be shared with the other club through the revenue-sharing arrangement.

Part (iii) states that a higher degree of redistribution in the league (more revenue-sharing) yields a higher probability of the large-market club and a lower probability of the small-market club to win the championship. This shows that the dulling effect is stronger for the small-market club than for the large-market club. The reason for this result is a form of 'free-riding.' When gate revenues are shared, the clubs' investment behavior is such that they take into account the impact of their investment on gate revenues for both their home games and their away games. Owing to the logit formulation of the contest success function, the (positive) marginal impact on the large-market club's gate revenues of a decrease in talent investments by the small-market club is greater than the (positive) marginal impact on the small-market club's gate revenues of a decrease in talent investments by the large-market club. As a consequence, the small-market club will reduce its investment level more than the large-market club.

Part (iv) stating that revenue-sharing decreases competitive balance represents the central result in this section and proves to be counter-intuitive. League authorities established restrictive arrangements such as revenue-sharing in order to improve competitive balance. The basic idea of revenue-sharing was to redistribute revenues form the rich (large-market) clubs to the poor (small-market) clubs because the non-cooperative equilibrium was assumed to produce a level of competitive balance that is too low. The branch of theoretical literature challenging the 'invariance proposition' and stating that revenue-sharing improves competitive balance is in line with this argumentation.[9] Our analysis, however, reveals that revenue-sharing has the opposite effect on competitive balance.[10] The intuition behind this result is the following: Part (ii) and part (iii) of this proposition have revealed that the dulling effect of revenue-sharing is stronger for the small-market club than for the large-market club. Since the large-market club is the dominant team and the small-market club the subordinate team (see part (i)), a higher degree of revenue-sharing increases the difference between the clubs' win percentages in equilibrium. This produces a more unbalanced league and thus decreases competitive balance.

[9]See e.g. Atkinson et al. (1988), Marburger (1997) and Késenne (2000*a*).

[10]Our result is sustained by Szymanski and Késenne (2004). Moreover, Szymanski (2004) comes to the same result by assuming that the supply of talent is fixed. This shows that revenue sharing can lead to a more unbalanced league in fixed talent supply models and flexible talent supply models.

4.4 Social Welfare Optimum and League Optimum

Social welfare is given by the sum of aggregate consumer (fan) surplus, aggregate club profit and total player utility.

Aggregate consumer surplus is computed by summing up the consumer surplus from fans of club 1 and club 2. The consumer surplus CS_i from fans of club $i = 1, 2$ in turn corresponds to the integral of the demand function $d(m_i, p_i, q_i)$ from the equilibrium price $p^* = \frac{q}{2}$ to the maximal price $\overline{p}_i = q_i$ which fans are willing to pay for a match with quality q_i:

$$CS_i = \int_{p_i^*}^{\overline{p}_i} d(m_i, p_i, q_i) dp_i = \int_{\frac{q_i}{2}}^{q_i} m_i \frac{q_i - p_i}{q_i} dp_i = \frac{m_i}{8} q_i.$$

Aggregate club profit is derived by summing up the profits of club 1 and club 2:

$$\Pi(t_1, t_2) = P + \frac{m_1}{4} q_1(t_1, t_2) + \frac{m_2}{4} q_2(t_1, t_2) - c \cdot (t_1 + t_2).$$

Note that the league optimum is characterized by the maximum of aggregate club profit.

If we assume that the players' utility corresponds to their salary, total players' utility is given by the aggregate salary payments $ct_1 + ct_2$ in the league.

Addition of aggregate consumer surplus, aggregate club profit and aggregate salary payments, produces social welfare as

$$\begin{aligned} W &= P + \frac{3}{8}(m_1 q_1 + m_2 q_2) \\ &= P + \frac{3}{8}(m_1(w_1 - (1-\mu) w_1^2) + m_2(w_2 - (1-\mu) w_2^2)). \end{aligned} \tag{4.5}$$

The players' utility does not influence social welfare since the only costs faced by the clubs are player salaries. That is, player salaries are only a transfer from clubs to players. Moreover, social welfare is independent of the revenue-sharing parameter α since the aggregate club profit is independent of α. Hence, social welfare only depends on the market size m_i, the match quality q_i and the exogenous league prize P.

In the following proposition we maximize aggregate club profit and social welfare and derive the corresponding optimal win percentages:

Proposition 4.2.
The welfare optimal and league optimal win percentages coincide and are given for $\kappa \in \{LO, WO\}$ *by*

$$w_1^\kappa = \frac{m_1 + m_2(1-2\mu)}{2(m_1+m_2)(1-\mu)} > \frac{1}{2} \text{ and } w_2^\kappa = \frac{m_1(1-2\mu) + m_2}{2(m_1+m_2)(1-\mu)} < \frac{1}{2}. \tag{4.6}$$

Proof. See Appendix 4.7.3. □

The proposition shows that the relative performances in the welfare optimum and the league optimum coincide in our model. The absolute level of talent investment need not, however, coincide. A league planner who wants to maximize joint club profits will choose the minimal necessary investment level consistent with (4.6).[11] This investment level chosen by the league planner will necessarily maximize social welfare since only the relative level of talent investment between both clubs is crucial for social welfare. The reverse does not hold true. Not every welfare optimal investment level maximizes joint club profits. In other words, a continuum of investment levels consistent with (4.6) maximizes social welfare whereas there is a unique investment level consistent with (4.6) which maximizes aggregate club profit.

Similar to the non-cooperative equilibrium, we further conclude that in the welfare optimum and the league optimum the large-market club is the dominant team and the small-market club the subordinate team with $w_1^\kappa > \frac{1}{2}$ and $w_2^\kappa < \frac{1}{2}$. The corresponding difference between the win percentages in the welfare/league optimum is given by $w_1^\kappa - w_2^\kappa = \frac{(m_1-m_2)\mu}{(m_1+m_2)(1-\mu)} > 0$ with $\kappa \in \{LO, WO\}$. The difference increases in the preference parameter μ of the quality function. In other words, if fans care more for their own team winning, then the welfare/league optimal degree of competitive balance decreases. Furthermore, the welfare/league optimal win percentage of club i increases in its own market size m_i and decreases in the market size m_j of the other club j.[12] A bigger market size of the large (small) market club causes $w_1^\kappa - w_2^\kappa$ to increase (decrease) and thus a more unbalanced (balanced) league becomes desirable from the perspective of a league planner and from a social welfare point of view.

[11] One can think of this minimal necessary investment level as the minimal amount which has to be invested in order to maintain the league's operation.

[12] Since $\frac{\partial w_i^\kappa}{\partial m_i} = \frac{m_j\mu}{(m_i+m_j)^2(1-\mu)} > 0$ and $\frac{\partial w_i^\kappa}{\partial m_j} = -\frac{m_i\mu}{(m_i+m_j)^2(1-\mu)} < 0$ for $\mu \in (0,1)$ and $\kappa \in \{LO, WO\}$.

4.5 Comparison of the Outcomes

So far, our analysis has shown that the dulling effect of revenue-sharing in the non-cooperative case is stronger for the small-market club than for the large-market club. As a consequence, increased revenue-sharing reduces competitive balance and produces a more unbalanced league. But how does revenue-sharing influence social welfare and aggregate club profit?

By comparing the non-cooperative equilibrium with the welfare optimum and the league optimum, we derive the following results:

Proposition 4.3.
(i) The league is more unbalanced in the welfare/league optimum compared with the non-cooperative equilibrium.
(ii) Revenue-sharing increases social welfare and aggregate club profit.

Proof. See Appendix 4.7.4 □

According to part (i), the difference between the clubs' win percentages is larger in the welfare/league optimum than in the non-cooperative equilibrium. Thus, from the perspective of a league planner and from a social welfare point of view, the degree of competitive balance in the non-cooperative equilibrium is too high. A more imbalanced league is desirable. In the non-cooperative equilibrium, the small-market club wins too often and the large-market club does not win often enough. Formally, $w_1^\kappa > w_1^*(\alpha) > \frac{1}{2}$ and $w_2^\kappa < w_2^*(\alpha) < \frac{1}{2}$ $\forall \alpha \in [\frac{1}{2}, 1]$ and $\kappa \in \{LO, WO\}$. This is a surprising result since it is usually argued that if playing talent can be freely traded in the market the outcome will be such that the large-market club obtains too much talent and the small-market club too little talent. The proposition shows, however, that the distribution of playing talent in the non-cooperative equilibrium is still too balanced. As a consequence, measures that decrease, not increase competitive balance will increase social welfare and aggregate club profit in our league. In this respect, gate revenue-sharing proves to be an appropriate measure of decreasing competitive balance and increasing social welfare and club profit.

What is the intuition behind this result? Each club imposes a negative externality through own talent investments on the other club's expected revenue. Because of the asymmetric market size, the small-market club imposes a larger externality on the large-market club than vice versa.[13] None of the clubs, however, internalizes

[13] This is because of the fact that the increase in revenue for a given increase in win percentage is higher for the large-market club than for the small-market club.

this negative externality. As a consequence, in the non-cooperative equilibrium, the marginal revenue of talent is equalized between the two clubs, but not the marginal revenue of a win. More precisely, the marginal revenue of a win is larger for the large-market club than for the small-market club. As a consequence, a decrease in the win percentage of the small-market club and an increase in the win percentage of the large-market club in the non-cooperative equilibrium results in higher social welfare and larger club profits. The maximum degree of competitive 'imbalance' and therefore the highest levels of social welfare and club profit are obtained in a league with full revenue-sharing $(\alpha = \frac{1}{2})$.

Moreover, the consumers/fans also benefit from a higher degree of revenue-sharing in the league since the aggregate consumer surplus is also maximized for the welfare optimal win percentages (w_1^{WO}, w_2^{WO}).[14] Hence, analogous to social welfare, revenue-sharing increases the aggregate consumer surplus and thus benefits consumers.

In the following corollary, we determine under which conditions the social optimum and the non-cooperative equilibrium coincide:

Corollary 4.1.
Social welfare is maximized in the non-cooperative equilibrium and the league is perfectly balanced with $w_i^{WO} = w_i^*(\alpha) = \frac{1}{2} \;\; \forall \alpha \in [\frac{1}{2}, 1] \;\; (i = 1, 2)$ *if at least one of the following conditions is satisfied:*
(i) Clubs are homogeneous with respect to their market size $(m_1 = m_2 = m)$.
(ii) Fans only care for competitive balance $(\mu = 0)$.

Proof. See Appendix 4.7.5. □

In a league of homogeneous clubs (case (i)), both clubs invest the same amount in the non-cooperative equilibrium obtaining a perfectly balanced league. In this case, the symmetric investment level in the non-cooperative equilibrium is given by $t_1^*(\alpha) = t_2^*(\alpha) = \frac{1}{4c}(P + \frac{m(2\alpha-1)\mu}{4})$.[15] Social welfare is derived as $W = P + \frac{3}{8}m(q_1 + q_2)$ and is maximized for each symmetric investment level in a perfectly balanced league. In a league in which fans only care for competitive balance (case (ii)), the symmetric investment level in the non-cooperative equilibrium is given by $t_1^*(\alpha) = t_2^*(\alpha) = \frac{P}{4c}$. In this case, the match quality is equal for both clubs (i.e. $q_1 = q_2 = q$) and is maximized

[14] It is straightforward to prove this claim. Compare $CS_1 + CS_2 = \frac{1}{8}(m_1 q_1 + m_2 q_2)$ with social welfare $W = P + \frac{3}{8}(m_1 q_1 + m_2 q_2)$ and note that the exogenous league prize P does not influence the maximization problem.

[15] Note that the assumption of fixed or elastic supply affects the equilibrium price of talent. Moreover, note that in a league of full revenue sharing ($\alpha = 1/2$), equilibrium investments are independent of the club's drawing potential m and the preference parameter μ.

for each symmetric investment level. Since social welfare $W = P + \frac{3}{8}q(m_1 + m_2)$ is proportional to the total quality, W is maximized in a perfectly balanced league.[16]

4.6 Conclusion

In this chapter we have developed a theoretical model of a team sports league based on contest theory in order to study the welfare effect of alternative gate revenue-sharing arrangements. We have derived club-specific demand and revenue from a general fan utility function by assuming that a fan's willingness to pay depends on the fan-type, win-percentage of the home team and competitive balance. Using this approach, we are able to extend the literature by providing an integrated framework which analyzes the effect of gate revenue on social welfare. The existing literature is focused on the effect of revenue-sharing on competitive balance and implicitly assumes that competitive balance is socially desirable without explaining the underlying assumptions regarding consumer preferences.

Our analysis challenges the 'invariance proposition' by showing that gate revenue-sharing decreases competitive balance and produces a more unbalanced league. This result is driven by the dulling effect of revenue-sharing. The dulling effect is revealed to be stronger for the small-market club than for the large-market club. Moreover, we have shown that a lower degree of competitive balance than in the non-cooperative league equilibrium yields a higher level of social welfare and aggregate club profit. Combining both results, we have concluded that in order to increase social welfare and club profits, arrangements which decrease, not increase, competitive balance should be implemented. In this respect, gate revenue-sharing proves to be an appropriate measure of decreasing competitive balance and increasing social welfare.

[16]Note that the league optimal win percentages in case (i) and (ii) are also given by $w_1^{LO} = w_2^{LO} = \frac{1}{2}$. As already mentioned, however, the investment level in the league optimum is given by an infinitesimally small amount consistent with $w_1^{LO} = w_2^{LO} = \frac{1}{2}$ and therefore does not coincide with the welfare optimum.

4.7 Appendix

4.7.1 Proof of Lemma 4.1

ad (i) Equation (4.4) is given by $\frac{t_1}{t_2} = \frac{P+\alpha\frac{\partial R_1}{\partial w_1}-(1-\alpha)\frac{\partial R_2}{\partial w_2}}{P+\alpha\frac{\partial R_2}{\partial w_2}-(1-\alpha)\frac{\partial R_1}{\partial w_1}}$, we compute

$$\frac{t_1}{t_2} = \frac{4P(t_1+t_2)+\alpha m_1(t_1(2\mu-1)+t_2)-(1-\alpha)(t_1+t_2(2\mu-1))}{4P(t_1+t_2)+\alpha m_2(t_1+t_2(2\mu-1))-(1-\alpha)(t_1(2\mu-1)+t_2)}. \tag{4.7}$$

By arranging (4.7) such that $t_1 = \psi(\alpha)t_2$, we formally obtain two solutions for the function $\psi(\alpha)$ which characterizes the relationship between t_1 and t_2 in the non-cooperative equilibrium:

$$\psi_1(\alpha) = \frac{1}{2\lambda_2}\left(\rho + \sqrt{\rho^2+4\lambda_1\lambda_2}\right) \text{ and } \psi_2(\alpha) = \frac{1}{2\lambda_2}\left(\rho - \sqrt{\rho^2+4\lambda_1\lambda_2}\right)$$

with $\rho := (1-2\alpha(1-\mu))(m_1-m_2), \lambda_1 := 4P+\alpha m_1-(1-\alpha)(2\mu-1)m_2$ and $\lambda_2 := 4P+\alpha m_2-(1-\alpha)(2\mu-1)m_1$. However, the negative solution $\psi_2(\alpha)$ can be ruled out because in case of a sufficiently high exogenous league prize P it yields negative equilibrium-payoffs and therefore does not ensure the existence of a Nash equilibrium in pure strategies.

We will show that $\psi_2(\alpha)$ always yields a negative solution:
To prove this claim we assume that the exogenous league prize P is sufficiently high with $P > P^* := \frac{(1-\alpha)(2\mu-1)m_2-\alpha m_1}{4}$. In this case we obtain $\lambda_1 > 0$ and $\lambda_2 > 0$ for all $\mu \in [0,1]$.

(a1) Suppose $\mu \in [0, \frac{1}{2}]$. Let $\alpha \leq \frac{1}{2(1-\mu)}$ then $\rho > 0$ and we derive $\psi_1(\alpha) > 0$ and $\psi_2(\alpha) < 0$. Let $\alpha > \frac{1}{2(1-\mu)}$ then $\rho < 0$ and we derive $\psi_1(\alpha) > 0$ and $\psi_2(\alpha) < 0$.

(b1) Suppose $\mu \in (1/2, 1]$ then $\rho \geq 0$ $\forall \alpha \in [1/2, 1]$. Also in this case we derive $\psi_1(\alpha) > 0$ and $\psi_2(\alpha) < 0$.

From (a1) and (b1) we derive that only the positive solution $\psi_1(\alpha)$ yields non-negative equilibrium-payoffs and thus ensures the existence of a Nash equilibrium in pure strategies.[17]

ad (ii) We claim that $\psi(\alpha)$ is an increasing function in revenue-sharing, i.e. a decreasing function in the parameter α such that $\frac{\partial\psi(\alpha)}{\partial\alpha} < 0$ $\forall \alpha \in [\frac{1}{2}, 1]$. It suffices to show that: $s(\alpha) := \frac{\partial\psi(\alpha)}{\partial\alpha} < 0$ $\forall \alpha \in [\frac{1}{2}, 1]$.

(a2) $s(\alpha)$ is a continuous function for all $\alpha \in \mathbb{R}$

[17] Note that in the subsequent analysis we write $\psi(\alpha)$ instead of $\psi_1(\alpha)$.

(b2) There exists only one α where $s(\alpha) = 0$: $s(\alpha) = 0 \Leftrightarrow \alpha^* = \frac{m_1(2\mu-1)-4P}{m_1(2\mu-1)+m_2}$. We derive that α^* is smaller than $\frac{1}{2}$ for all $P > 0$ if $\mu \in [0, \frac{1}{2}]$ and for all $P > P^{**} := \frac{m_1(2\mu-1)-m_2}{4}$ if $\mu \in (\frac{1}{2}, 1]$.

(c2) Evaluation of the function $s(\alpha)$ for $\alpha > \alpha^*$ yields $s(\alpha) < 0$. E.g. evaluation of $s(\alpha)$ for $\alpha = \frac{1}{2}$ yields $s(\frac{1}{2}) = -\frac{8\mu^2(m_1-m_2)(m_1(2P+m_2(1-\mu))+2m_2P)}{8P+m_1(1-\mu)+m_2(1-\mu)} < 0$ for $m_1 > m_2$ and $\mu > 0$.

From (a2), (b2) and (c2) we conclude that the continuous function $s(\alpha)$ is always smaller than zero on the compact interval $\alpha \in [\frac{1}{2}, 1]$ and thus $\psi(\alpha)$ is an increasing function in revenue-sharing, i.e. $\frac{\partial \psi(\alpha)}{\partial \alpha} < 0 \;\; \forall \alpha \in [\frac{1}{2}, 1]$. This proves the claim.

ad (iii) We claim that $\psi(\alpha) = 1 \Leftrightarrow \mu = 0 \vee m_1 = m_2$ and $\psi(\alpha) > 1 \Leftrightarrow \mu > 0 \wedge m_1 > m_2$ $\forall \alpha \in [\frac{1}{2}, 1]$.

It is straightforward to show that $\psi(\alpha) = 1 \Leftrightarrow \mu = 0 \vee m_1 = m_2$ which proves the first part of the claim.

In the next step, we set $\mu > 0 \wedge m_1 > m_2$ and prove that $\psi(\alpha) > 1 \;\; \forall \alpha \in [\frac{1}{2}, 1]$. It suffices to show that $r(\alpha) := \psi(\alpha) - 1 > 0 \;\; \forall \alpha \in [\frac{1}{2}, 1]$.

(a3) $r(\alpha)$ is a continuous function for all $\alpha \in \mathbb{R}$.

(b3) There exists only one $\alpha \in \mathbb{R}$ where $r(\alpha) = 0$: $r(\alpha) = 0 \Leftrightarrow \alpha^* = \frac{m_1(2\mu-1)-4P}{m_1(2\mu-1)+m_2}$. Analogous to above, we derive that α^* is smaller than $\frac{1}{2}$ for all $P > 0$ if $\mu \in [0, \frac{1}{2}]$ and for all $P > P^{**}$ if $\mu \in (\frac{1}{2}, 1]$.

(c3) Evaluation of the function $r(\alpha)$ for $\alpha > \alpha^*$ yields $r(\alpha) > 0$. For example, evaluation of $r(\alpha)$ for $\alpha = \frac{1}{2}$ yields $r(\frac{1}{2}) = 2\mu(m_1 - m_2) > 0$ for $m_1 > m_2$ and $\mu > 0$.

From (a3), (b3) and (c3) we derive that the continuous function $r(\alpha)$ is always larger than zero on the compact interval $\alpha \in [\frac{1}{2}, 1]$ and thus $\psi(\alpha)$ is always larger than unity on the same interval, i.e. $\psi(\alpha) > 1 \;\; \forall \alpha \in [\frac{1}{2}, 1]$. This proves the claim.

4.7.2 Proof of Proposition 4.1

We assume that fans care besides competitive balance also for own team winning $(\mu > 0)$ and clubs are heterogeneous with respect to their market size $(m_1 > m_2)$.

ad (i) We claim that in the non-cooperative equilibrium the investment level of the large-market club 1 is higher than the investment level of the small-market club 2, i.e. $t_1^*(\alpha) > t_2^*(\alpha) \ \ \forall\alpha \in [\frac{1}{2}, 1]$. It suffices to show that

$$t_1^*(\alpha) > t_2^*(\alpha) \Leftrightarrow w_1^*(\alpha) > w_2^*(\alpha) \ \forall\alpha \in [\frac{1}{2}, 1].$$

The win percentages in equilibrium are given by $w_1^*(\alpha) = \frac{\psi(\alpha)}{\psi(\alpha)+1}$ and $w_2^*(\alpha) = \frac{1}{\psi(\alpha)+1}$. Hence, $\frac{w_1^*(\alpha)}{w_2^*(\alpha)} = \psi(\alpha)$ and according to Lemma 4.1, we know that $\psi(\alpha) > 1 \ \ \forall\alpha \in [\frac{1}{2}, 1]$. We conclude $w_1^*(\alpha) > w_2^*(\alpha)$ and thus obtain $t_1^*(\alpha) > t_2^*(\alpha)$ which proves the claim.

ad (iii) We claim that the win percentage of the large (small) market club 1 (club 2) is an increasing (decreasing) function in revenue-sharing, i.e. decreases (increases) in the parameter α such that $\frac{\partial w_1^*(\alpha)}{\partial\alpha} < 0$ and $\frac{\partial w_2^*(\alpha)}{\partial\alpha} > 0 \ \ \forall\alpha \in [\frac{1}{2}, 1]$. From Lemma 4.1 we know that $\frac{\partial\psi(\alpha)}{\partial\alpha} < 0 \ \ \forall\alpha \in [\frac{1}{2}, 1]$ and derive

$$\frac{\partial w_1^*(\alpha)}{\partial\alpha} = \frac{\partial\psi(\alpha)}{\partial\alpha}\frac{1}{(\psi(\alpha)+1)^2} < 0 \text{ and } \frac{\partial w_2^*(\alpha)}{\partial\alpha} = -\frac{\partial\psi(\alpha)}{\partial\alpha}\frac{1}{(\psi(\alpha)+1)^2} > 0.$$

This proves the claim.

ad (iv) We claim that revenue-sharing reduces competitive balance and produces a more unbalanced league. In other words, a lower parameter α of the revenue-sharing arrangement increases the difference $w_1^*(\alpha) - w_2^*(\alpha) = \frac{\psi(\alpha)-1}{\psi(\alpha)+1}$ between the win percentages of club 1 and club 2. To prove the claim it suffices to show that $\frac{\partial(w_1^*(\alpha)-w_2^*(\alpha))}{\partial\alpha} < 0$. Since club 1 is the dominant team and club 2 is the subordinate team, we obtain $w_1^*(\alpha) - w_2^*(\alpha) > 0$ and compute

$$\frac{\partial(w_1^*(\alpha) - w_2^*(\alpha))}{\partial\alpha} = \frac{\partial\psi(\alpha)}{\partial\alpha}\frac{2}{(\psi(\alpha)+1)^2}.$$

From Lemma 4.1, we know that $\frac{\partial\psi(\alpha)}{\partial\alpha} < 0$ and thus $\frac{\partial(w_1^*(\alpha)-w_2^*(\alpha))}{\partial\alpha} < 0$. Hence, decreasing the parameter α of the revenue-sharing arrangement, i.e. more revenue-sharing in the league, increases the difference between the win percentages of club 1 and club 2 and produces a more unbalanced league. This proves the claim.

ad (ii) We claim that equilibrium investments decrease in revenue-sharing, i.e. increase in the parameter α such that $\frac{dt_1^*}{d\alpha} > 0$ and $\frac{dt_2^*}{d\alpha} > 0$. To prove this claim, we follow Szymanski and Késenne (2004) and derive the total differential of the first-order conditions $\frac{\partial E(\Pi_1)}{\partial t_1} = 0$ and $\frac{\partial E(\Pi_2)}{\partial t_2} = 0$ which yields

$$\begin{aligned} \frac{\partial^2 E(\Pi_1)}{\partial t_1^2} dt_1 + \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2} dt_2 + \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial \alpha} &= 0, \\ \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial t_1} dt_1 + \frac{\partial^2 E(\Pi_2)}{\partial t_2^2} dt_2 + \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial \alpha} &= 0. \end{aligned}$$

This system of equations can also be written as

$$\begin{bmatrix} \frac{\partial^2 E(\Pi_1)}{\partial t_1^2} & \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2} \\ \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial t_1} & \frac{\partial^2 E(\Pi_2)}{\partial t_2^2} \end{bmatrix} \begin{bmatrix} dt_1 \\ dt_2 \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial \alpha} \\ -\frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial \alpha} \end{bmatrix} d\alpha.$$

Applying Cramer's Rule we derive

$$\frac{dt_1}{d\alpha} = \frac{\frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2} \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial \alpha} - \frac{\partial^2 E(\Pi_2)}{\partial t_2^2} \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial \alpha}}{\frac{\partial^2 E(\Pi_1)}{\partial t_1^2} \frac{\partial^2 E(\Pi_2)}{\partial t_2^2} - \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2} \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial t_1}}. \tag{4.8}$$

According to the stability condition in Dixit (1986) the denominator of equation (4.8) is assumed to be positive. A sufficient condition for stability is therefore

$$\frac{\partial^2 E(\Pi_1)}{\partial t_1^2} \frac{\partial^2 E(\Pi_2)}{\partial t_2^2} > \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2} \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial t_1},$$

since the second-order conditions $\frac{\partial^2 E(\Pi_1)}{\partial t_1^2}$ and $\frac{\partial^2 E(\Pi_2)}{\partial t_2^2}$ are negative, given the assumptions about the revenue function. Moreover, we compute

$$\begin{aligned} \frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial \alpha} &= \left(\frac{\partial R_1}{\partial w_1} + \frac{\partial R_2}{\partial w_2} \right) \frac{\partial w_1}{\partial t_1} > 0, \\ \frac{\partial^2 E(\Pi_2)}{\partial t_2 \partial \alpha} &= \left(\frac{\partial R_2}{\partial w_2} + \frac{\partial R_1}{\partial w_1} \right) \frac{\partial w_2}{\partial t_2} > 0, \end{aligned}$$

for all $w_i \in [0,1]$ if $\mu \in (\frac{1}{2}, 1]$ and for all $w_i < \frac{1}{2(1-\mu)} < 1$ if $\mu \in [0, \frac{1}{2}]$.

The expression $\frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2}$ characterizes the slope of club 1's reaction function. Since club 1 is the large-market club its reaction function slopes upward and therefore we obtain $\frac{\partial^2 E(\Pi_1)}{\partial t_1 \partial t_2} > 0$. Hence, also the numerator of (4.8) is positive and we derive that $\frac{dt_1^*}{d\alpha} > 0$.

From part (iv) of this proposition we know that revenue-sharing reduces competitive balance. Now, if revenue-sharing induces club 1 to reduce its investment level then it must also be the case for club 2, i.e. $\frac{dt_2^*}{d\alpha} > 0$. This proves the claim.

4.7.3 Proof of Proposition 4.2

We assume that fans care besides competitive balance also for own team winning $(\mu > 0)$ and clubs are heterogeneous with respect to their market size $(m_1 > m_2)$. We first derive the social welfare optimum and then the league optimum.

a) Social welfare optimum

Social welfare W is given by

$$W = P + \frac{3}{8}(m_1 q_1 + m_2 q_2) = P + \frac{3}{8}(m_1(w_1 - (1-\mu)w_1^2) + m_2(w_2 - (1-\mu)w_2^2)).$$

The corresponding first-order conditions are computed as[18]

$$\begin{aligned}\frac{\partial W}{\partial t_1} &= \frac{3t_2(m_1(t_2 - t_1(1-2\mu)) + m_2(t_2(1-2\mu) - t_1))}{8(t_1+t_2)^3} = 0,\\ \frac{\partial W}{\partial t_2} &= \frac{3t_1(m_1(t_1(1-2\mu) - t_2) + m_2(t_1 - t_2(1-2\mu))}{8(t_1+t_2)^3} = 0.\end{aligned}$$

Since we are not able to explicitly solve for the welfare optimal investment levels (t_1^{WO}, t_2^{WO}), we establish similar to Lemma 4.1 the following relationship which must hold in the welfare optimum:

$$t_1^{WO} = \psi^{WO} t_2^{WO} \text{ with } \psi^{WO} = \frac{m_1 + m_2(1-2\mu)}{m_1(1-2\mu) + m_2}.$$

By assuming that $\mu < \frac{m_1+m_2}{2m_1}$ we guarantee an interior solution. In this case, the corresponding win percentages $w_i^{WO} = \frac{t_i^{WO}}{t_i^{WO}+t_j^{WO}}$ in the welfare optimum are given by

$$w_1^{WO} = \frac{m_1 + m_2(1-2\mu)}{2(m_1+m_2)(1-\mu)} \text{ and } w_2^{WO} = \frac{m_1(1-2\mu) + m_2}{2(m_1+m_2)(1-\mu)}.$$

It is straightforward to show that $w_1^{WO} > \frac{1}{2}$ and $w_2^{WO} < \frac{1}{2}$ for all $m_1 > m_2$ and $\mu > 0$. This proves the claim.

Moreover, we claim that each investment level (t_1, t_2) which satisfies $t_1^{WO} = \psi^{WO} t_2^{WO}$ maximizes social welfare: We define $(t_1^{(k)}, t_2^{(k)})$ as a sequence which is consistent with $t_1^{WO} = \psi^{WO} t_2^{WO}$. For example, define $t_2^{(k)} := \frac{1}{k}$ and $t_1^{(k)} := \frac{m_1+m_2(1-2\mu)}{m_1(1-2\mu)+m_2}\frac{1}{k}$. We derive that $\frac{\partial W(t_1^{(k)}, t_2^{(k)})}{\partial t_1} = 0$ and $\frac{\partial W(t_1^{(k)}, t_2^{(k)})}{\partial t_2} = 0$ for all $k \in \mathbb{N}$. Moreover, $w_1(t_1^{(k)}, t_2^{(k)}) =$

[18] The second-order conditions for a maximum are satisfied.

$\frac{m_1+m_2(1-2\mu)}{2(m_1+m_2)(1-\mu)}$ and $w_2(t_1^{(k)}, t_2^{(k)}) = \frac{m_1(1-2\mu)+m_2}{2(m_1+m_2)(1-\mu)}$. Hence, $(t_1^{(k)}, t_2^{(k)})$ maximizes social welfare for all $k \in \mathbb{N}$ and the claim is proved.

b) League Optimum

In order to maximize aggregate club profit $\Pi(t_1, t_2)$, a league planner has to solve the following maximization problem:[19]

$$\max_{(t_1,t_2)}\{P + \frac{m_1}{4}q_1 + \frac{m_2}{4}q_2 - c\cdot(t_1+t_2)\}.$$

Analogous to a), we derive the following relationship which must hold in the league optimum:

$$t_1^{LO} = \psi^{LO} t_2^{LO} \text{ with } \psi^{LO} = \frac{m_1 + m_2(1-2\mu)}{m_1(1-2\mu)+m_2}. \tag{4.9}$$

Hence, the league optimal win percentages are given by

$$w_1^{LO} = \frac{m_1 + m_2(1-2\mu)}{2(m_1+m_2)(1-\mu)} > \frac{1}{2} \text{ and } w_2^{LO} = \frac{m_1(1-2\mu)+m_2}{2(m_1+m_2)(1-\mu)} < \frac{1}{2} \tag{4.10}$$

and coincide with the welfare optimal win percentages. However, in contrast to the welfare optimum, not every investment level (t_1, t_2) which satisfies $t_1^{LO} = \psi^{LO} t_2^{LO}$ maximizes the aggregate club profit. This is because of the fact that aggregate costs $c\cdot(t_1+t_2)$ are now included in aggregate profits. As a consequence, an infinitesimal small amount consistent with $t_1^{LO} = \psi^{LO} t_2^{LO}$ (such that (4.10) is satisfied) maximizes aggregate club profit. To see this, consider a monotone decreasing sequence $(t_1^{(k)}, t_2^{(k)})$ with limit 0 such that $t_1^{(k)} = \psi^{LO} t_2^{(k)}$ for all $k \in \mathbb{N}$. Hence, $(t_1^{(k)}, t_2^{(k)})$ satisfies (4.10) and thus maximizes aggregate gate revenues $\frac{m_1}{4}q_1 + \frac{m_2}{4}q_2$ for all $k \in \mathbb{N}$. Moreover, aggregate club profit $\Pi(t_1, t_2)$ can be increased by decreasing the investment level, i.e. $\Pi(t_1^{(k+1)}, t_2^{(k+1)}) > \Pi(t_1^{(k)}, t_2^{(k)})$. Without restrictions on the minimal amount of talent which has to be invested, the league planner would spend in the league optimum an infinitesimal small amount still consistent with (4.9). However, in a league in which a minimal amount $T > 0$ of talent investment is necessary in order to maintain the league's operation, a league planner who wants to maximize aggregate club profit will exactly invest this minimal amount such that $t_1^{LO} = \psi^{LO} t_2^{LO}$ and $T = t_1^{LO} + t_2^{LO}$.

[19] We assume that the league planner has no influence on the equilibrium price $p_i^* = \frac{q_i}{2}$ and hence acts as a price taker.

4.7.4 Proof of Proposition 4.3

We assume that fans care besides competitive balance also for own team winning ($\mu > 0$) and clubs are heterogeneous with respect to their market size ($m_1 > m_2$).

ad (i) We claim that our two-club league is more unbalanced in the welfare optimum and the league optimum compared with the non-cooperative equilibrium independent of the revenue-sharing parameter α, i.e. $|w_1^\kappa - w_2^\kappa| > |w_1^*(\alpha) - w_2^*(\alpha)| \;\; \forall \alpha \in [\frac{1}{2}, 1]$ and $\kappa \in \{WO, LO\}$.

We define $g(\alpha) := w_1^\kappa - w_1^*(\alpha)$ and derive the following properties of $g(\alpha)$:

(a) $g(\alpha)$ is a continuous function for all $\alpha \in \mathbb{R}$.

(b) There exists only one $\alpha \in \mathbb{R}$ where $g(\alpha) = 0$:

$$g(\alpha) = 0 \Leftrightarrow \alpha^{**} = \frac{1}{2} - \frac{(m_1 + m_2)P}{m_1 m_2 \mu} < \frac{1}{2}.$$

(c) Evaluation of the function $g(\alpha)$ for $\alpha > \alpha^{**}$ yields that $g(\alpha) > 0$. For example, evaluation of $g(\alpha)$ for $\alpha = \frac{1}{2}$ and $\kappa \in \{WO, LO\}$ yields

$$g(\frac{1}{2}) = w_1^\kappa - w_1^*(\frac{1}{2}) = \frac{4\mu P(m_1 - m_2)}{(1-\mu)(m_1 + m_2)(8P + (1-\mu)(m_1 + m_2))} > 0.$$

From (a),(b) and (c) we derive that the continuous function $g(\alpha)$ is always larger than zero on the compact interval $\alpha \in [\frac{1}{2}, 1]$ and thus $w_1^\kappa > w_1^*(\alpha) \;\; \forall \alpha \in [\frac{1}{2}, 1]$ and $\kappa \in \{WO, LO\}$. Moreover, we know that the large-market club 1 is the dominant team, i.e. $w_1^*(\alpha) > \frac{1}{2} \;\; \forall \alpha \in [\frac{1}{2}, 1]$. By using the adding-up constraint: $w_j = 1 - w_i$ we conclude that $1 - w_1^\kappa = w_2^\kappa < w_2^*(\alpha) = 1 - w_1^*(\alpha) < \frac{1}{2} \;\; \forall \alpha \in [\frac{1}{2}, 1]$ and $\kappa \in \{WO, LO\}$. Hence, the following inequality holds true:

$$w_1^\kappa - w_2^\kappa > w_1^*(\alpha) - w_2^*(\alpha) > 0 \;\forall \alpha \in [\frac{1}{2}, 1] \text{ and } \kappa \in \{WO, LO\}.$$

This proves the claim.

ad (ii) Part (i) of this proposition has shown that the league is more unbalanced in the welfare and league optimum than in the non-cooperative equilibrium. This implies $w_1^\kappa > w_1^*(\alpha) > \frac{1}{2}$ and $w_2^\kappa < w_2^*(\alpha) < \frac{1}{2} \;\; \forall \alpha \in [\frac{1}{2}, 1]$ and $\kappa \in \{WO, LO\}$. A more imbalanced league is socially desirable and also desirable from the league planner's point of view. Moreover, according to Proposition 4.1, the win percentage of the large (small) market club 1 (club 2) is an increasing (decreasing) function in revenue-sharing. Thus, by decreasing the parameter α (more gate revenue-sharing), the win percentage of the large-market club 1 increases and the win percentage of the small-market club 2 decreases. This causes the degree of competitive balance to decrease which in turn increases social welfare and aggregate club profit (due to the fact that the welfare optimal and league optimal win percentages are approached). Social welfare and aggregate club profit increase until the maximal level of revenue-sharing is reached in a league with full revenue-sharing $(\alpha = \frac{1}{2})$.

4.7.5 Proof of Corollary 4.1

We claim that social welfare is maximized in the non-cooperative equilibrium and the league is perfectly balanced if (i) clubs are homogeneous with respect to their market size ($m_1 = m_2$) or (ii) the fan's preference is such that they only care for competitive balance ($\mu = 0$).

If $m_1 = m_2$ or $\mu = 0$ we derive that

(i) in the non-cooperative equilibrium holds $\psi(\alpha) = 1$ and thus the corresponding win percentages are given by $w_1^*(\alpha) = w_2^*(\alpha) = \frac{1}{2}$ $\forall \alpha \in [\frac{1}{2}, 1]$.

(ii) in the social optimum the win percentages are given by $w_1^{WO} = w_2^{WO} = \frac{1}{2}$ according to (4.6).

Comparing (i) and (ii) proves the claim.

Chapter 5

The Role of Transfer Restrictions in Professional Football

5.1 Introduction

Employment relations in football are governed by a set of distinct institutional mechanisms: contracts between players and clubs, employment law and a regulatory framework known as the transfer system enforced by the football governing bodies (FIFA and the national associations).[1]

The crucial effect of the transfer system is the creation of a unilateral property right for the clubs over the services of players. As a consequence of the transfer system the players are not able to leave their current club and sign with another club without the current club's explicit consent. The football governing bodies enjoy a certain degree of freedom to self-regulate as sport is considered to differ from other industries because of well-known peculiarities (Neale, 1964). Until 1995 the football authorities were able to impose the transfer system on all employment relations in football. Players out-of-contract as well as players in-contract required the permission of their current club before signing with another club. In this sense all employment in football was governed by the 'shadow of the transfer system' and clubs only agreed to release players conditional on receiving adequate remuneration through a transfer fee.

Since 1995 the ability of the football governing bodies to apply the transfer system has been restricted in two major steps. In December 1995 the European Court of Justice

[1] This chapter is based on a paper which was published in the *European Journal of Law and Economics* (see Dietl et al. (2008*b*)). Reprinted with permission.

issued its famous Bosman verdict,[2] which ruled that the transfer system could no longer be applied to out-of-contract players. As a consequence, players now become free agents after expiration of their contracts and their former employer has no right to demand transfer remuneration if they sign with new clubs.

Finally in 2001, the European Commission further restricted the ability of the football governing bodies to self-regulate the employment relations of football. In what is known as the 'Monti system' after Commissioner Mario Monti,[3] the football governing bodies had to adapt their regulatory framework known as the FIFA transfer rules to a whole set of new requirements.[4] The Bosman verdict changed the situation in that the transfer system remained applicable to in-contract players only. However, clubs and players were still free to deliberately place employment relations under the 'shadow of the transfer system' by excluding the advent of contract expiration through extended contract durations. By limiting contract durations the Monti system rendered this avoidance strategy more difficult.

The standard interpretation of these restrictions in the application of the transfer system stresses the increased freedom of movement for players, which translates into a relative gain in market power and therefore into higher salaries. While we do not deny the link between freedom of movement and market power, we question that the salaries will ultimately be driven up by the reforms. There may be more than one channel of influence between the reforms and the salaries. Our model looks at the employment relation in football from a different perspective. We develop a model which captures an important and widely overlooked aspect of this employment relation: the allocation of risk. The basic intuition of our approach can be stated as follows. Players and clubs alike do not know how the productivity of a player will develop in future periods. Given that players perform in public and taking into account the importance of reputation effects, pride and career concerns in sport it seems unlikely that players should shirk on effort. Instead, it seems more adequate to treat

[2] The Bosman verdict had the following background. In 1990, the contract of Jean Marc Bosman, a professional football player, with his Belgian club R.C. Liegois expired. After the club offered him a new contract worth only 25% of his former contract, Bosman wanted to transfer to the French club U.S. Dunkerque. According to the transfer system of the International Football Association, however, the Belgian Football Association had to send Bosman's registration certificate to the French Football Association before Bosman was eligible to play for U.S. Dunkerque. Since R.C. Liegois was not satisfied with the transfer payment offered by U.S. Dunkerque, the Belgian Football Association withheld Bosman's registration certificate. As a result, Bosman could not play for U.S. Dunkerque and took his case to the courts.

[3] The new FIFA transfer rules were adopted after more than two years of discussions between the European Commission - in particular, Commissioner Mario Monti - the European Football Association (UEFA), and FIFA.

[4] We will only focus on one aspect of the Monti system in our model, the limitation of contract durations in football to a maximum of five years.

productivity variations as a manifestation of risk. Moreover, on average, the career duration of a professional football player is very short compared with other labour-markets. According to Frick (2006) 'more than one third of all players 'disappear' again after their first season and only one career out of twelve lasts for 10 years and more.' During this short career duration, the high performance uncertainty creates strong incentives for the player to buy insurance against income uncertainty.

If risk is the key driver behind the performance uncertainty of football players then there is an obvious potential for value creation in this industry. Risk-averse players could buy insurance against future income uncertainty when contracting with risk-neutral clubs, which have the possibility to diversify the risk of productivity variations within their portfolio of players and also through diversified ownership structures. However, if the player turns out to be more productive in the course of time than assumed when writing down the initial contract, he has incentives to renegotiate the contract. The same holds for the club if the player turns out to be a 'bad risk.'

The third institutional mechanism governing employment relations in football comes into play here, labour law. De facto labour law in most European countries makes long-term employment contracts asymmetrically incomplete since it is possible to legally bind employers to fulfil long-term contracts but it is practically impossible to bind the employee. There is no 'shadow of the law' that prevents players from accepting better job offers. Since 'good risk' players would therefore renegotiate the contract and receive wages reflecting their marginal productivity, clubs would be left with all the 'bad risks.' Given this assumption, clubs cannot offer value creating insurance services. In this context the transfer system imposed by the governing bodies of football works as a surrogate which makes insurance contracts complete. 'Good risk' players know that they will have to pay for the insurance, be it through the transfer fee or by continuing to play for a salary below marginal productivity. It is the 'shadow of the transfer system' which allowed players to commit to fulfilling their contracts. It is the 'shadow of the transfer system' which enabled the efficient allocation of risk in this industry.

The Bosman verdict restricted the 'shadow of the transfer system' to the market for in-contract players. However, it provided freedom for players and clubs to voluntarily position their transactions under the 'shadow of the transfer system' by extending the duration of contracts, which is exactly what happened in the industry.[5] The Monti system makes it more difficult to position transactions under the 'shadow of the transfer system' by limiting contract durations, thereby making the efficient allocation

[5]Feess et al. (2004) show that after the Bosman verdict the average contract length has increased considerably, e.g. in the german Bundesliga from 2.43 to 2.91 years.

of risk more difficult. In our model we show that risk-averse players may lose from the reforms since they would benefit from a conversion of risky future income into risk-less current income under the 'shadow of the transfer system.'

The remainder of this chapter is organized as follows. Section 5.2 reviews the literature. Section 5.3 specifies the model. In Section 5.4 we analyze the role of transfer restriction. Section 5.5 characterizes the relationship between the 'shadow of the transfer system' and the pre-Bosman, Bosman and Monti transfer system. Finally, Section 5.6 concludes.

5.2 Literature

Rottenberg (1956) presents the first economic analysis of transfer restrictions in professional team sports. He describes the mandatory lifelong tie of a player to his original club in U.S. Major League Baseball combined with the club's right to demand transfer compensations from other clubs in case that the player transfers as the result of the league's market power. According to Rottenberg, these labour-market restrictions preclude players from earning salaries equal to their marginal productivity.[6] Since new clubs cannot offer an in-contract player more than his marginal productivity minus transfer compensations to the old club, the player is not able to bargain his salary up to his marginal productivity.

Our model differs from this view. We show that the existence of transfer restrictions combined with the right to demand transfer compensations does not mean that players are worse off or that any kind of market power is exerted upon them. To the contrary, our model highlights that the players' loss in ex post bargaining power is compensated by an increase in ex ante bargaining power.

According to our knowledge, Rottenberg was also the first to conclude that the right to demand transfer payments does not result in an inefficient allocation of playing talent. If football contracts are incomplete with respect to transfer fees, a player's current club can always renegotiate the transfer fee in order to maximize profits by transferring the player to the club where he is most productive. Carbonell-Nicolau and Comin (2005) recently provided empirical evidence for the claim that football contracts are incomplete with respect to transfer fees. Based on a data set with information about football contracts, transfer payments, and several measures of a players' value in the Spanish Primera Division for the three seasons from 1999/00-

[6] Similar arguments are presented by Demmert (1973) and Scully (1974).

2001/02. Carbonell-Nicolau and Comin show that the player's contractually specified transfer fee has a large positive effect on the new club's total cost of hiring the player.

Burguet et al. (2002) show that transfer restrictions are a common feature in labour-markets in which a worker's (invariant) productivity is unknown ex ante but can be observed by outsiders after the worker has signed a contract and works for an incumbent firm. In these markets, ex post competition for workers is likely to be vigorous and outsiders can earn positive rents by signing workers with the desired productivity characteristics. Transfer restrictions allow the pair incumbent firm-worker to expropriate at least some of the outsiders' rents. In their model, transfer restrictions have no efficiency effect. Without transfer restrictions no firm would be willing to sign a worker with unknown productivity characteristics. Workers would have to work without a wage before their productivity becomes common knowledge. Transfer restrictions only affect the distribution of profits between incumbent firm, worker and outsiders. This result of Burguet et al. is due to their assumption that worker productivity is invariant over time. We believe that a football player's productivity (playing strength) varies significantly during his career and, more importantly, these variations cannot be predicted. There are many players who were believed to become superstars, but were never able to meet expectations. Similarly, there are at least as many players who became much better players than initially predicted by experts. Once we introduce unknown productivity variations over a player's career, risk allocation becomes a crucial feature of welfare considerations and transfer restrictions are no longer efficiency neutral.

Based on the bargaining model of Burguet et al.,[7] Feess and Muehlheusser (2003) argue that the prohibition of transfer restrictions reallocates bargaining power from a player's current club to potentially new clubs. This reallocation of bargaining power reduces the current club's incentive to invest in the player's human capital because the current club has to bear the investment costs without being able to appropriate all investment benefits if the player transfers to a new club.

Antonioni and Cubbin (2000) analyze the economic effect of the Bosman ruling. Based on empirical evidence and the theory of real options, Antonioni and Cubbin argue that the Bosman ruling had little effect on player salaries, investment in human capital and transfer activity. They attribute the rise in salaries to increasing television revenues. According to Antonioni and Cubbin, a club's incentive to invest in training players is not impaired, because the club can always exercise its option to sell a player before

[7] Burguet et al. (2002) and Feess and Muehlheusser (2003) model the bargaining process as a simultaneous Nash bargaining game in which the player simultaneously bargains with his old and his new club. The Nash bargaining solution in each individual bargaining game serves as the threat point of the other bargaining game.

his contract expires. At the same time, no club will wait until the contract of a desired player has expired so that the player becomes available for free because no club will take the risk to lose the desired player to a rival club who does not wait until the contract has expired. Like Burguet et al. (2002) and Feess and Muehlheusser (2003), Antonioni and Cubbin do not analyze the effect of transfer restrictions on the allocation of risk.

5.3 Model Specification

Our model consists of a representative player, who has a career horizon of two periods, and two representative clubs, club S and club L. The player is assumed to be risk-averse whereas the clubs are assumed to be risk-neutral since they have the possibility of diversifying the risk of productivity variations within their portfolio of players and also through diversified ownership structures.[8] The utility of the player is given by his salary whereas the utility of each club corresponds to its profit. The total expected utility (i.e. expected utility over two periods) of the risk-averse player is defined as the sum of the risk-free first-period salary and the security equivalent of the risky second-period salary. The total expected utility of the risk-neutral club is defined as the sum of the expected first-period profit and the expected second-period profit.

The player's productivity in $t \in \{1, 2\}$ is a random variable with Markov property denoted S_t at club S and L_t at club L. To abstract from moral hazard problems, we assume that the player's productivity in each period is exogenous.[9] It follows a stochastic process characterized by the binomial tree model presented in Figure 5.1. The player's productivity in each period either increases in club S by a fixed amount, $s > 0$, with probability $p \in (0, 1)$, or decreases by the same amount with probability $(1 - p)$. For club L this fixed amount is given by $l > 0$.[10]

With probability p the player's productivity increases, leading in club S to a first-period productivity of $S_1 = e_0 + s$. With probability $(1-p)$ the player's productivity decreases, leading in club S to a first-period productivity of $S_1 = e_0 - s$.

[8] Our main insights still hold if the club is less risk-averse than the player. Our results will also remain valid if the club is more risk-averse than the player but can diversify most of the risk by signing contracts with many players.

[9] More realistically, a player's performance is the combined result of the player's (exogenous) talent and (endogenous) effort. Nevertheless, our abstraction can be justified with two arguments. First, players have pride and try to maximize their chance of winning by providing full effort. Second, since performance of football players is perfectly observable, players who do not provide full effort will be regarded as less talented than they actually are.

[10] We refer to the time-span between $t = 0$ and $t = 1$ as 'period 1' and between $t = 1$ and $t = 2$ as 'period 2.'

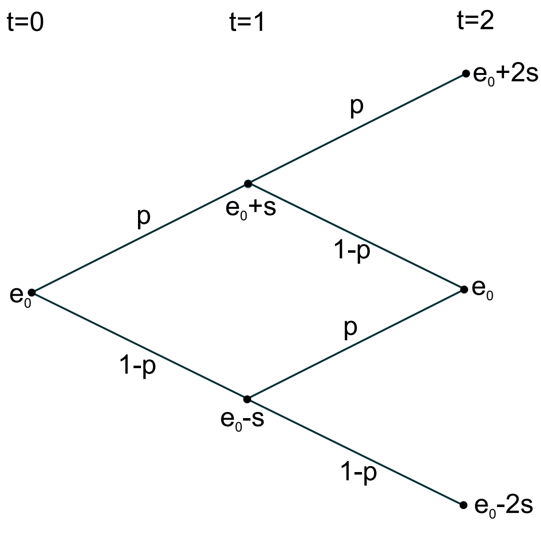

Figure 5.1: Development of the Player's Productivity in Club S

In the event that the player's productivity has increased (decreased) during period 1, the player's productivity will increase during period 2 with probability p, leading in club S to a second-period productivity of $S_2 = e_0 + 2s$ ($S_2 = e_0$). With probability $(1-p)$ the player's productivity will decrease during period 2, leading in club S to a second-period productivity of $S_2 = e_0$ ($S_2 = e_0 - 2s$).[11]

The probability p is assumed to be common knowledge. Without loss of generality, we assume throughout this analysis that $s < l$. Hence, we can interpret club S as a 'small-market' club where variations of the player's productivity only cause a low productivity alteration. Club L then is a 'large-market' club where variations of the player's productivity cause a high productivity alteration. Moreover, we call a player with $p > 1/2$ ($p \leq 1/2$) 'high-talented' ('low-talented').

At the beginning of each period the player and the two clubs have common expectations about the player's productivity (playing strength) in this and future periods. During each period the player and the clubs observe the player's current playing strength. From this information both will update their expectation regarding the player's productivity in future periods. The terms $E_t[S_{t+1}|S_t]$ and $E_t[L_{t+1}|L_t]$ denote the expected value of S_{t+1} and L_{t+1} based on the information available in t (before S_{t+1} and L_{t+1} are revealed) conditional on the player's productivity S_t and L_t.[12] In $t = 0$, the player's productivity is assumed to be common knowledge and given for both clubs by $e_0 > 0$, i.e. $S_0 = L_0 = e_0$.[13]

[11]Analogous for club L with l instead of s.

[12]For notational clarity we write in the subsequent analysis $E_t[S_{t+1}]$ and $E_t[L_{t+1}]$ instead of $E_t[S_{t+1}|S_t]$ and $E_t[L_{t+1}|L_t]$.

[13]The information set at $t = 0$ does not have to be empty. Before starting a professional career

In $t = 0$, the player's expected first-period productivities $E_0[S_1]$ in club S and $E_0[L_1]$ in club L are computed as

$$E_0[S_1] = e_0 + s(2p-1) \quad \text{and} \quad E_0[L_1] = e_0 + l(2p-1).$$

Moreover, in $t = 0$ both clubs have expectations about the player's second-period productivity, denoted $E_0[S_2]$ for club S and $E_0[L_2]$ for club L, which are given by

$$E_0[S_2] = e_0 + 2s(2p-1) \quad \text{and} \quad E_0[L_2] = e_0 + 2l(2p-1).$$

With $s < l$ and $p < 1/2$ it gives: $E_0[S_1] > E_0[L_1]$ and $E_0[S_2] > E_0[L_2]$. Hence, in $t = 0$ a low-talented player is expected to be more productive in both periods in the small-market club S than in the large-market club L. The reverse is true for a high-talented player.

In $t = 1$, the player, club S, and club L observe the player's current productivity and update their expectation regarding his productivity in period 2. If the player's productivity has increased during period 1 we denote the expected second-period productivity in $t = 1$ at club S with $E_1[S_2^+]$ and at club L with $E_1[L_2^+]$. In the other case, we write $E_1[S_2^-]$ and $E_1[L_2^-]$. The expected second-period productivities are computed as

$$\begin{aligned} E_1[S_2^+] &= e_0 + 2sp & \text{and} \quad E_1[L_2^+] = e_0 + 2lp, \\ E_1[S_2^-] &= e_0 + 2s(p-1) & \text{and} \quad E_1[L_2^-] = e_0 + 2l(p-1). \end{aligned}$$

In order to guarantee a positive expected second-period productivity for all $p \in (0,1)$ we assume: $e_0 > 2l(p-1)$.

If the player's productivity has increased (decreased) during period 1, then in $t = 1$, each type of player is expected to be more productive at club L (club S) than at club S (club L). Formally,

$$E_1[S_2^+] < E_1[L_2^+] \text{ and } E_1[S_2^-] > E_1[L_2^-] \ \forall p \in (0,1). \tag{5.1}$$

The clubs compete for the player by offering contracts which specify the number of periods the player will play for the club and the salary paid by the club to the player in each of the respective periods. We distinguish two regimes:

In Section 5.4.1 we consider short-term contracts in a restricted transfer system where

a player usually played in minor or youth leagues. The market can form its expectation regarding a rookie player's productivity based on the player's past performance.

all employment is governed by the 'shadow of the transfer system.' We assume that the contract between the player and his initial club expires after period 1, but the player cannot transfer to a new club without the permission of his initial club. In this case, the initial club has the right to demand an unlimited transfer fee from the other club for the player. If the initial club is not satisfied with the amount offered by the other club, the initial club has the right to prevent the player from transferring to the other club. This right gives the initial club strong bargaining power, because it enables this club to prevent a transfer by demanding an exorbitantly high transfer fee. The other club, however, cannot be forced to pay any amount demanded by the initial club. The new club is free to withdraw its offer if it cannot reach an agreement with the initial club regarding the transfer fee and with the player regarding the player's second-period salary.

In Section 5.4.2 we consider short-term contracts in an unrestricted transfer system, i.e. without the 'shadow of the transfer system.' The contract between the player and his initial club expires after period 1 and the player is free to sign a contract with another club without the permission of his initial club. Moreover, the initial club does not receive any transfer fee.

5.4 The Role of Transfer Restrictions

5.4.1 Short-term Contracts in the 'Shadow of the Transfer System'

We model the bargaining process in $t = 0$ between the player and the clubs concerning the player's first-period salary as a pair of simultaneous negotiations in Nash bargaining fashion: one for each club vis-à-vis the player, using as threat points in each negotiation what each expects from the other.[14] This bargaining model captures the cooperative situation between the clubs and the player on the one hand and the non-cooperative situation between the two clubs on the other hand (both clubs compete against each other by offering contracts to the player). Moreover, the two clubs and the player take into account that the 'shadow of the transfer system' prevents the out-of-contract player from signing a valid contract with another club without the permission of his current club. Besides the relevant threat points, we have to compute the player's and club's total expected utility.

[14] Note that this approach is similar to Burguet et al. (2002).

Formally, we have to distinguish two cases: (a) the player signs a short-term contract with club S in $t = 0$ and (b) the player signs a short-term contract with club L in $t = 0$.[15]

We proceed by assuming that the player has signed a short-term contract which specifies a first-period salary of $\underline{w}_{r,1}^{S}$ with club S in $t = 0$:[16]

(i) If the player's productivity has increased during period 1 (which happens with probability p), we know by equation (5.1) that each type of player will achieve a higher expected second-period productivity at the large-market club L compared with the small-market club S. According to the Coase theorem, the player will transfer to club L since the player is then allocated efficiently.

But how will the (expected) productivity gain that is generated through the transfer be divided between the player, club S and club L?

In contrast with the bargaining game in $t = 0$, where a solution concept which captured the partial non-cooperative nature of the game was needed, we now have a cooperative bargaining situation between all three parties since we know ex ante that the grand coalition will form. Thus, we need now a solution concept that captures the cooperative nature of the bargaining game between the three parties. The Shapley value is a appropriate solution concept in this case since it describes a reasonable or fair way to allocate the gains realized by cooperation between three or more parties. Each party then receives its contribution from the (expected) productivity gain obtained by the grand coalition.

In the following lemma we determine each party's contribution to the player's transfer from club S to club L in $t = 1$:

Lemma 5.1.
The Shapley values determine the outcome of the cooperative bargaining game as follows: $\frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+]$ *(player),* $\frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+]$ *(club S) and* $\frac{1}{3}(E_1[L_2^+] - E_1[S_2^+])$ *(club L).*

Proof. See Appendix 5.7.1 □

According to the lemma the player will receive at club L a second-period salary of

$$\underline{w}_{r,2}^{L+} = \frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+].$$

[15] In this section we will only analyze case (a). The other case (b) is postponed to Appendix 5.7.6.

[16] Note that the subscript r stands for 'restricted' transfer system.

Club S receives as a transfer fee T^S its contribution to the coalition determined by its Shapley value and therefore realizes an expected second-period profit of $E_1[\pi_{r,2}^{S+}] = \frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+] = T^S$. Analogously, club L receives its Shapley value and realizes an expected second-period profit of $E_1[\pi_{r,2}^{L+}] = \frac{1}{3}(E_1[L_2^+] - E_1[S_2^+])$.

(ii) If the player's productivity has decreased during period 1 (which happens with probability $(1-p)$), we know by equation (5.1) that each type of player will achieve a higher expected second-period productivity at his initial club S compared with the other club L. Hence, in $t=1$ club L will not place any offer for the player since it knows that it cannot reach an agreement with club S regarding the transfer fee and with the player regarding the player's salary. Without a competing offer from club L, the player will stay at club S in $t=1$ and the player's reservation wage falls to zero. The player's second-period salary $\underline{w}_{r,2}^{S-}$ is now determined by the negotiations only between club S and the player. It is appropriate, therefore, to apply the Nash bargaining solution to derive the outcome of this bargaining process. Club S's utility is given by its expected second-period profit $E_1[S_2^-] - \underline{w}_{r,2}^{S-}$ whereas the player's utility is given by the salary $\underline{w}_{r,2}^{S-}$ he will receive at club S. The threat points of club S and the player both amount to zero. Formally, we compute

$$\underline{w}_{r,2}^{S-} = \arg\max_{\underline{w}_{r,2}^{S-}} \left\{ (E_1[S_2^-] - \underline{w}_{r,2}^{S-} - 0)(\underline{w}_{r,2}^{S-} - 0) \right\} = \frac{1}{2}E_1[S_2^-].$$

Hence, the player will earn a second-period salary of $\underline{w}_{r,2}^{S-} = \frac{1}{2}E_1[S_2^-]$, club S expects a second-period profit of $E_1[\pi_{r,2}^{S-}] = E_1[S_2^-] - \underline{w}_{r,2}^{S-} = \frac{1}{2}E_1[S_2^-]$ and club L will earn $E_1[\pi_{r,2}^{L-}] = 0$.

We can now determine the total expected utility in $t=0$ of club S and the player, respectively:

Total expected utility $E_0[u_r^S]$ of the risk-neutral club S is given by the expected first-period profit $E_0[\pi_{r,1}^S]$ plus the expected second-period profit $pE_1[\pi_{r,2}^{S+}] + (1-p)E_1[\pi_{r,2}^{S-}]$. We compute

$$E_0[u_r^S] = (E_0[S_1] - \underline{w}_{r,1}^S) + pT^S + (1-p)\frac{1}{2}E_1[S_2^-].$$

Total expected utility $E_0[\underline{u}_r^P]$ of the risk-averse player is given by

$$E_0[\underline{u}_r^P] = \underline{w}_{r,1}^S + E_0[\underline{w}_{r,2}] - \frac{1}{2}\tau V[\underline{w}_{r,2}], \tag{5.2}$$

where τ measures the degree of the player's risk-aversion. The higher τ, the more risk-averse is the player. In the first period the player receives $\underline{w}_{r,1}^S$ with certainty.

Since the second-period salary is risky, we use the security equivalent as the player's expected second-period utility, where the expected second-period salary $E_0[\underline{w}_{r,2}]$ and the variance $V[\underline{w}_{r,2}]$ of the second-period salary are given by

$$\begin{aligned} E_0[\underline{w}_{r,2}] &= p\underline{w}_{r,2}^{L+} + (1-p)\underline{w}_{r,2}^{S-} = p(\frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+]) + (1-p)(\frac{1}{2}E_1[S_2^-]), \\ V[\underline{w}_{r,2}] &= p(\underline{w}_{r,2}^{L+})^2 + (1-p)(\underline{w}_{r,2}^{S-})^2 - (E_0[\underline{w}_{r,2}])^2. \end{aligned}$$

The threat points of the simultaneous negotiations in Nash bargaining fashion in $t=0$ are derived as follows: with probability $(1-p)$ each type of player will achieve a higher expected second-period productivity at the small-market club S. In the case where the player has signed a short-term contract with club L in $t=0$ he will transfer to club S in $t=1$. Club S will then receive its contribution to the coalition determined by its Shapley value. Thus the threat point of club S, denoted d^S, is given by $(1-p)\frac{1}{3}(E_1[S_2^-] - E_1[L_2^-])$. The player's threat point, denoted $\overline{d}^P$, is determined by the player's total expected utility $E_0[\overline{u}_r^P]$ that he could achieve by playing at club L.

The pair of simultaneous negotiations in Nash bargaining fashion in $t=0$ concerning the player's first-period salary, denoted $(\underline{w}_{r,1}^S, \overline{w}_{r,1}^L)$, are formally given by:[17]

$$\begin{aligned} \underline{w}_{r,1}^S &= \arg\max \left\{(E_0[u_r^S] - d^S)(E_0[\underline{u}_r^P] - \overline{d}^P)\right\} \\ \overline{w}_{r,1}^L &= \arg\max \left\{(E_0[u_r^L] - d^L)(E_0[\overline{u}_r^P] - \underline{d}^P)\right\} \end{aligned} \tag{5.3}$$

All relevant information is available to solve this problem and to specify the player's first-period salary $\underline{w}_{r,1}^S$ at club S and $\overline{w}_{r,1}^L$ at club L:

Lemma 5.2.
The first-period salaries $\underline{w}_{r,1}^S$ and $\overline{w}_{r,1}^L$ of the player are computed as

$$\begin{aligned} \underline{w}_{r,1}^S &= \frac{2}{3}E_0[S_1] + \frac{1}{3}E_0[L_1] + pT^S \\ &\quad +(1-p)(T^L - \frac{1}{2}E_1[\pi_{r,2}^{S-}]) + \frac{\tau}{6}(V[\underline{w}_{r,2}] - V[\overline{w}_{r,2}]), \end{aligned} \tag{5.4}$$

$$\begin{aligned} \overline{w}_{r,1}^L &= \frac{1}{3}E_0[S_1] + \frac{2}{3}E_0[L_1] + (1-p)T^L \\ &\quad +p(T^S - \frac{1}{2}E_1[\pi_{r,2}^{L+}]) + \frac{\tau}{6}(V[\overline{w}_{r,2}] - V[\underline{w}_{r,2}]). \end{aligned} \tag{5.5}$$

Proof. See Appendix 5.7.2 □

[17] See Appendix 5.7.6 for a detailed derivation of the player's total expected utility $E_0[\overline{u}_r^P]$, club L's total expected utility $E_0[u_r^L]$ and the relevant threat points d^L and $\underline{d}^P$.

In $t = 0$, the risk-averse player will sign a contract with the club where he maximizes his total expected utility:

Corollary 5.1.
(i) A low-talented player will sign a contract with the small-market club S independent of his degree of risk-aversion, i.e. if $p \leq \frac{1}{2}$ then $E_0[\underline{u}_r^P] > E_0[\overline{u}_r^P]$ $\forall \tau > 0$.
(ii) A high-talented player will sign a contract with the large-market club L if his risk-aversion is sufficiently low and with the small-market club S if his risk-aversion is sufficiently high, i.e. if $p > \frac{1}{2}$ then $E_0[\underline{u}_r^P] < E_0[\overline{u}_r^P]$ $\forall \tau < \widetilde{\tau}(p, s, l)$.

Proof. See Appendix 5.7.3 □

The corollary shows that a low-talented, risk-averse player maximizes his total expected utility at the small-market club S independent of his degree of risk-aversion whereas a high-talented player only maximizes his total expected utility at the large-market club L if his risk-aversion is sufficiently low. Intuitively this is clear: A low-talented player will play for the club where variations of his productivity only generate a low productivity alteration (club S), whereas a high-talented player will play for the club where variations of his productivity generate a high productivity alteration (club L). If, however, the risk-aversion of a high-talented player becomes sufficiently high, then this player will also prefer to play for the club where variations of his productivity only generate a low productivity alteration.

5.4.2 Short-term Contracts without the 'Shadow of the Transfer System'

Similar to $t = 0$ in Section 5.4.1, the bargaining process concerning the player's salary in each of the respective periods is modelled via a pair of simultaneous negotiations, one for each club vis-à-vis the player, using as threat points in each negotiation what each expects from the other. Without the 'shadow of the transfer system' the initial club, however, *cannot* be sure either to hold the player in period 2 and obtain the player's second-period productivity or to transfer the player and receive a transfer fee. The club's expectations in $t = 0$ regarding the player's second-period productivity therefore amount to zero. Similarly, the player *cannot* be sure to either stay in period 2 at his initial club and receive a second-period salary from this club or to be transferred and obtain his Shapley value as a second-period salary from the new club. These circumstances influence the bargaining process in $t = 0$ insofar as now the player's and the club's expected utilities in the Nash product of the Nash bargaining solution only involve the first period. We now determine the player's and club's expected utilities in each period:

If the player signs a contract with club $Z \in \{S, L\}$ in $t \in \{0,1\}$, which specifies a salary of $w^Z_{u,t+1}$, the expected utility in $t \in \{0,1\}$ of the risk-neutral club is given by[18]

$$\begin{aligned} E_t[u^S_{u,t+1}] &= E_t[S_{t+1}] - w^S_{u,t+1} \text{ (for club } S\text{)}, \\ E_t[u^L_{u,t+1}] &= E_t[L_{t+1}] - w^L_{u,t+1} \text{ (for club } L\text{)}. \end{aligned}$$

The player's expected one-period utility in $t \in \{0,1\}$ is given by

$$E_t[\underline{u}^P_{u,t+1}] = w^S_{u,t+1} \text{ and } E_t[\overline{u}^P_{u,t+1}] = w^L_{u,t+1}.$$

We derive the relevant threat points as follows: In $t \in \{0,1\}$, club S's threat point is zero, whereas the threat point of the player is determined by the expected one-period utility $E_t[\overline{u}^P_{u,t+1}]$ that he could achieve at club L. Analogously for the other Nash bargaining solution.

Formally, the pair of simultaneous negotiations in $t \in \{0,1\}$ is given by

$$\begin{aligned} w^S_{u,t+1} &= \arg\max \left\{ (E_t[u^S_{u,t+1}] - 0)(E_t[\underline{u}^P_{u,t+1}] - E_t[\overline{u}^P_{u,t+1}]) \right\}, \\ w^L_{u,t+1} &= \arg\max \left\{ (E_t[u^L_{u,t+1}] - 0)(E_t[\overline{u}^P_{u,t+1}] - E_t[\underline{u}^P_{u,t+1}]) \right\}. \end{aligned} \tag{5.6}$$

The solution to this problem is derived in the following lemma:

Lemma 5.3.
The player's salary $w^S_{u,t+1}$ and $w^L_{u,t+1}$ in $t \in \{0,1\}$ are computed as

$$w^S_{u,t+1} = \frac{2}{3}E_t[S_{t+1}] + \frac{1}{3}E_t[L_{t+1}] \text{ and } w^L_{u,t+1} = \frac{1}{3}E_t[S_{t+1}] + \frac{2}{3}E_t[L_{t+1}]. \tag{5.7}$$

Proof. Straightforward. □

We derive $w^S_{u,1} > w^L_{u,1}$ for all $p \in (0, \frac{1}{2}]$ and $w^S_{u,1} < w^L_{u,1}$ for all $p \in (\frac{1}{2}, 1)$.[19] As a consequence, a low (high) talented player will sign a short-term contract with the small-market (large-market) club in $t = 0$. The intuition is similar to that of Corollary

[18] Note that the subscript u stands for 'unrestricted' transfer system.

[19] It holds: $w^S_{u,1} > w^L_{u,1} \Leftrightarrow E_0[S_1] > E_0[L_1]$. Since $l > s$, we derive $E_0[S_1] > E_0[L_1] \Leftrightarrow p \in (0, \frac{1}{2})$ and $E_0[S_1] < E_0[L_1] \Leftrightarrow p \in (\frac{1}{2}, 1)$. Without loss of generality, we assume that the small-market club S contracts a low-talented player with $p = \frac{1}{2}$.

5.1. According to Lemma 5.3 the low-talented player then receives a first-period salary of

$$\underline{w}^S_{u,1} = \frac{2}{3}E_0[S_1] + \frac{1}{3}E_0[L_1].$$

whereas, the high-talented player receives a first-period salary of

$$\overline{w}^L_{u,1} = \frac{1}{3}E_0[S_1] + \frac{2}{3}E_0[L_1].$$

We now analyze the situation in $t = 1$:

If the player's productivity has decreased during period 1, then according to equation (5.1) each type of player will achieve a higher second-period productivity at the small-market club S, i.e. $E_1[S_2^-] > E_1[L_2^-]$, which implies $w^S_{u,2} > w^L_{u,2}$. The player will therefore sign a short-term contract with club S and receive, according to Lemma 5.3, a second-period salary of [20]

$$w^{S-}_{u,2} = \frac{2}{3}E_1[S_2^-] + \frac{1}{3}E_1[L_2^-].$$

If the player's productivity has increased during period 1, then in $t = 1$ each type of player will sign a short-term contract with club L and receive according to Lemma 5.3 a second-period salary of

$$w^{L+}_{u,2} = \frac{1}{3}E_1[S_2^+] + \frac{2}{3}E_1[L_2^+].$$

Under short-term contracts without the 'shadow of the transfer system,' the player cannot be sure in $t = 0$ either to stay in period 2 at his initial club and receive a second-period salary from this club or to be transferred and obtain his Shapley value as a second-period salary from the new club. Nevertheless, the player can form expectations in $t = 0$ about his utility over the two periods. Depending on his type, the player will receive $w^S_{u,1}$ or $w^L_{u,1}$ in the first period with certainty. With probability p, the (low- and high-talented) player will sign a contract in $t = 1$ at club L and receive $w^{L+}_{u,2}$. With probability $(1 - p)$ the (low- and high-talented) player will sign a contract in $t = 1$ at club S and receive $w^{S-}_{u,2}$. Hence, the player's expected second-period salary $E_0[w_{u,2}]$ and the variance $V[w_{u,2}]$ of the second-period

[20]We can omit the underline and the upperline, since the second-period salary is equal for each type of player.

salary are determined by[21]

$$
\begin{aligned}
E_0[w_{u,2}] &= p w_{u,2}^{L+} + (1-p) w_{u,2}^{S-} \\
&= p(\frac{1}{3}E_1[S_2^+] + \frac{2}{3}E_1[L_2^+]) + (1-p)(\frac{2}{3}E_1[S_2^-] + \frac{1}{3}E_1[L_2^-]), \\
V[w_{u,2}] &= p(w_{u,2}^{L+})^2 + (1-p)(w_{u,2}^{S-})^2 - (E_1[w_{u,2}])^2.
\end{aligned}
$$

In $t=0$, total expected utility, denoted $E_0[\underline{u}_u^P]$ for a low-talented player and $E_0[\overline{u}_u^P]$ for a high-talented player, is analogous to Section 5.4.1 computed as

$$
\begin{aligned}
E_0[\underline{u}_u^P] &= w_{u,1}^S + E_0[w_{u,2}] - \frac{1}{2}\tau V[w_{u,2}], \\
E_0[\overline{u}_u^P] &= w_{u,1}^L + E_0[w_{u,2}] - \frac{1}{2}\tau V[w_{u,2}].
\end{aligned}
$$

5.4.3 In *versus* Out of the 'Shadow of the Transfer System'

In this section we compare the player's salary under a short-term contract in the 'shadow of the transfer system' with the respective salary under a short-term contract without the 'shadow of the transfer system.' Moreover, we show that a risk-averse player benefits from the 'shadow of the transfer system.'

Proposition 5.1.
Under a short-term contract in the 'shadow of the transfer system' a high-talented, risk-averse player receives a higher first-period salary combined with a lower (expected) second-period salary compared with a short-term contract without the 'shadow of the transfer system.' The same holds true for a low-talented player whose risk-aversion is sufficiently low. Formally,

$$
\begin{aligned}
&\textit{(i) Low-talented player} : \quad \underline{w}_{r,1}^S > \underline{w}_{u,1}^S \ \forall \tau < \tau^* \ \textit{and} \ E_0[\underline{w}_{r,2}] < E_0[w_{u,2}], \\
&\textit{(ii) High-talented player} : \quad \overline{w}_{r,1}^L > \overline{w}_{u,1}^L \ \forall \tau > 0 \ \textit{and} \ E_0[\overline{w}_{r,2}] < E_0[w_{u,2}].
\end{aligned}
$$

Proof. See Appendix 5.7.4 □

The proposition shows that for a high-talented player and a low-talented player (whose risk-aversion is sufficiently low), the risk-free first-period salary under a short-term contract in the 'shadow of the transfer system' is higher than the respective salary without the 'shadow of the transfer system.' The opposite holds true for the expected

[21] Note that the expected second-period salary and variance are equal for a low- and a high- talented player.

second-period salary of a (low- or high-talented) player since it is higher without the 'shadow of the transfer system.' The intuition behind this result is as follows: The 'shadow of the transfer system' gives the player an instrument to commit himself successfully not to renege on the insurance deal since the club can be sure that the player only leaves its portfolio in $t = 1$ if the transfer fee exceeds the expected profit that the player could achieve by staying at the club in period 2. As a consequence a risk-neutral club can partially insure its risk-averse player against income uncertainty by transforming a part of the player's risky future (second-period) salary in risk-free current (first-period) salary.

In the next proposition we show that a risk-averse player benefits from the 'shadow of the transfer system' and therefore from a more restrictive transfer system:

Proposition 5.2.
Total expected utility of a low- and high-talented, risk-averse player is higher under a short-term contract in the 'shadow of the transfer system' than under a short-term contract without the 'shadow of the transfer system.' Formally,

$$\begin{aligned} \textit{(i) Low-talented player} \quad &: \quad E_0[\underline{u}_r^P] > E_0[\underline{u}_u^P] \ \forall \tau > 0, \\ \textit{(ii) High-talented player} \quad &: \quad E_0[\overline{u}_r^P] > E_0[\overline{u}_u^P] \ \forall \tau > 0. \end{aligned}$$

Proof. See Appendix 5.7.5 □

The above proposition shows that a risk-averse player benefits from the 'shadow of the transfer system' since a risk-averse player prefers a higher current salary combined with a lower (expected) future salary to a lower current salary combined with a higher but uncertain future salary.

5.5 The 'Shadow of the Transfer System' in the pre-Bosman, Bosman and Monti Transfer System

As a point of departure we have to take into account the fact that labour law cannot be employed in reality to prevent an employee from accepting superior alternative job offers. In addition to this let us assume a situation where the 'shadow of the transfer system' does not exist. This means that the weak or inexistent 'shadow of the law' is supplemented by an inexistent 'shadow of the transfer system.' Will a Pareto efficient contract that creates value by enabling risk-averse players to buy insurance against future income uncertainty from risk-neutral clubs be feasible in this setting?

Given the inexistent (or at least very weak) external enforcement system, the insurance deal between player and club will only work if the contract written down in period one is time-consistent. The insurance deal is to be regarded as a series of one-period contracts. After each period the parties re-calculate the terms of the next one-period contract taking into account the information available at the beginning of the respective period. In the absence of both the 'shadow of the transfer system' and the 'shadow of the law' a Pareto efficient contract enabling risk-averse players to buy insurance from risk-neutral clubs is unlikely to be achieved. Nothing prevents a 'good risk' player whose productivity turns out to be underestimated in the course of his career to use external offers in order to bid his salary to a level reflecting marginal productivity. Why should a 'good risk' player still agree to pay the 'insurance fee' established in the original contract in this setting? Why should clubs offer value-creating insurance services if they are not able to appropriate any of this value because of a regulatory environment leaving them with all the 'bad risks?'

How does the 'shadow of the transfer system' change this situation? Let us first assume the pre-Bosman world. All employment relations in football are governed by the 'shadow of the transfer system' in this world. Players out-of-contract as well as players in-contract require the permission of their current club in order to be able to sign with another club. The 'shadow of the transfer system' works as a surrogate which makes insurance contracts complete. Let a 'good risk' player whose productivity has been underestimated receive an external transfer offer. Player and club will of course re-calculate their deal taking into account the new information available. However the 'good risk' player cannot defect on the insurance deal. As was shown in Section 5.4.1 the club can be sure that the player only leaves its portfolio at the beginning of the next period if the transfer fee exceeds the expected profit that the player could contribute by staying with the club in the future. In the 'shadow of the transfer system' the Pareto efficient contract is time-consistent. Although contracts may be renegotiated every period in the pre-Bosman world on the basis of new information available, these renegotiations cannot be used to defect on the insurance deal. Enabling the player to commit to the insurance deal the 'shadow of the transfer system' allows clubs to transfer risky future income in risk-less current income and make risk-averse players better off as has been shown in Section 5.4.3. Seen from this insurance perspective, contract duration is not important in the pre-Bosman world since the 'shadow of the transfer system' effectively links one-period contracts to a time-consistent series. Even if the actual contract of the player expires at the end of the current period his promise not to use an external offer in order to defect on the 'insurance fee' remains perfectly credible in the 'shadow of the transfer system.'

The Bosman verdict transforms the potential promise of a player not to defect on the insurance deal after the expiration of his contract into cheap talk. The 'shadow of the transfer system' only continues to provide credibility to player promises given within the time-span of valid contracts. Contract duration becomes a crucial variable for the functioning of the insurance market. By expanding contract duration employment relationships can be deliberately taken under the 'shadow of the transfer system' where commitments to honor the insurance deals work. This is exactly what happened between clubs and players on a perfectly voluntary basis. Despite the fact that the players had the choice to become free agents outside the 'shadow of the transfer system' and the clubs had the choice to sign these free agents, the bulk of all transfer activity took place within the 'shadow of the transfer system.' Clubs and players restored the pre-Bosman situation on the insurance market by expanding contract durations. In terms of the model this intuition is captured by switching from Section 5.4.2 back to Section 5.4.1.[22] A long-term contract which covers the player's career horizon of two periods specifies a salary for each period, given by (w_1, w_2). In our model, a long-term contract is equivalent to the short-term contract in the 'shadow of the transfer system' described in Section 5.4.1, where the first-period salary w_1 is given by $w_1 = \underline{w}^S_{r,1}$ or $w_1 = \overline{w}^L_{r,1}$, dependent of the player's type. The second-period salary w_2 is given for a low-talented player by $w_2 = \underline{w}^{S-}_{r,2}$ or $w_2 = \underline{w}^{L+}_{r,2}$ (for a high-talented player by $w_2 = \overline{w}^{S-}_{r,2}$ or $w_2 = \overline{w}^{L+}_{r,2}$), dependent of the development of the player's productivity during period 1. By signing a long-term contract, the club can be sure to either hold the player and obtain the player's second-period productivity or to transfer the player and receive a transfer fee. This expected second-period productivity increases the player's productivity in $t = 0$ and should also be incorporated in the contractually specified first-period salary of a long-term contract. The calculation of the player's first-period salary in Section 5.4.1 effectively incorporates the player's expected second-period productivity. Furthermore, the second-period salary of a long-term should reflect the player's expected development during period 1, which is the case in Section 5.4.1.

The Monti system limits the voluntary attempt of clubs and players to take employment relations under the 'shadow of the transfer system.' The maximum duration of contracts is 5 years. Contracts that are signed up the 28th birthday of the player are protected against unilateral breach for the first three years. Contracts that are

[22] In other words, in our model the Bosman transfer system can always resemble the pre-Bosman world by adjusting the contract length accordingly. Note that this does not hold the other way round.

signed thereafter are only protected for two years. The 'shadow of the transfer system' will only work for these three respectively two years. After the 'protected period' only a 'scattered shadow of the transfer system' will be effective. The transfer fee for players-in-contact shall reflect whether contracts are broken in the 'protected period.' No transfer fee can be charged for players out-of-contract. Our model captures the intuition behind the changes from the Bosman to the Monti world by a switch from Section 5.4.1 to 5.4.2. Risk-averse players will find it more difficult to put up insurance deals with clubs. Clubs will face greater problems to convert risky future income in risk-less current income in a world where players cannot make longer-term commitments to honor insurance deals. Outside the remaining small 'shadow of the transfer system' their promises are bound to be cheap talk.

5.6 Conclusion

Transfer restrictions have a long tradition in professional football, but came under heavy attack in recent years. In this chapter we have analyzed whether a risk-averse player really benefits from less restrictive transfer systems. Given that the player's productivity varies significantly during his career and taking into account that these variations cannot be predicted, the allocation of risk becomes a crucial feature. Our model, which captures this important aspect of employment relations in football, has revealed that a risk-averse player benefits from 'the shadow of the transfer system' and therefore from a more restrictive transfer system. Under the pre-Bosman transfer system, clubs could partially insure their players against income uncertainty by transforming a part of the player's risky future salary in risk-free current income. A risk-averse player prefers a higher current salary combined with a lower (expected) future salary to a lower current salary combined with a higher but uncertain future salary. The Bosman transfer system, which is equivalent to the pre-Bosman transfer system in terms of our model, did not change this situation since clubs and players voluntary restored the pre-Bosman world by expanding contract durations. However, by limiting the maximal contract duration to 5 years the insurance deal does not work anymore in the new Monti transfer system. As a consequence, the Monti transfer system can be considered as an impediment to Pareto efficient risk-allocation in the football industry. If risk can be considered as the basic source of productivity variations in football, the failure of the insurance market imposed by the free movement philosophy of the European institutions might impose a high price to be paid by the labour force in this industry.

5.7 Appendix

5.7.1 Proof of Lemma 5.1

We will show how the (expected) productivity gain that is generated through the player's transfer from club S to club L in $t = 1$ will be divided between the player, club S and club L:

The Shapley value gives each member i of a coalition C her expected contribution, where the expectation is taken over all coalitions to which i might belong. Formally, party i's share of the pie is given by

$$\sum_{C|i\in C} \frac{(c-1)!(n-c)!}{n!}\,(v(C) - v(C/\{i\})\,,$$

where $c = |C|$ is the number of parties in coalition C, n is the total number of parties bargaining, $v(C)$ is the surplus produced by coalition C, and $v(C/\{i\})$ is the surplus produced by coalition C without party i.

First, we compute the share of the player: Without the player, neither of the two clubs can generate any surplus. Together with the player, club S can generate a surplus of $v(\{P,S\}) = E_1[S_2^+]$. The respective probability of this coalition is $1/6$. The player and club L cannot generate any surplus, because they need the consent of club S, i.e. $v(\{P,L\}) = 0$. The coalition of club S and club L cannot generate any surplus, i.e. $v(\{S,L\}) = 0$. The coalition of the two clubs and the player will generate a surplus of $v(\{P,S,L\}) = E_1[L_2^+]$. The respective probability of this coalition is $1/3$. As a result, the player's Shapley value is $\frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+]$.

Club S's situation is symmetric to the player's. Accordingly, club S's Shapley value is $\frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+]$.

Club L needs the grand coalition to generate a surplus of $E_1[L_2^+]$. Without club L, club S and the player can generate a surplus of only $E_1[S_2^+]$. Hence, club L's Shapley value is $(E_1[L_2^+] - E_1[S_2^+])$.

5.7.2 Proof of Lemma 5.2

The player's first-period salary $\underline{w}_{r,1}^S$ and $\overline{w}_{r,1}^L$ are determined by the simultaneous negotiations (5.3) which are modelled in Nash bargaining fashion, one for each club vis-à-vis the player. Deriving the corresponding FOC and solving for $\underline{w}_{r,1}^S$ and $\overline{w}_{r,1}^L$, respectively, yields

$$\begin{aligned}
\underline{w}^S_{r,1} &= \frac{1}{4}(2E_0[S_1] + (1-p)E_1[L_2^-] + pE_1[L_2^+] + \tau(V[\underline{w}_{r,2}] - V[\overline{w}_{r,2}]) + 2\overline{w}^L_{r,1}),\\
\overline{w}^L_{r,1} &= \frac{1}{4}(2E_0[L_1] + (1-p)E_1[S_2^-] + pE_1[S_2^+] + \tau(V[\overline{w}_{r,2}] - V[\underline{w}_{r,2}]) + 2\underline{w}^S_{r,1}).
\end{aligned}$$

By solving this system of equations and using the fact that $E_1[\pi^{S+}_{r,2}] = \frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+] = T^S$, $E_1[\pi^{L+}_{r,2}] = \frac{1}{3}(E_1[L_2^+] - E_1[S_2^+])$, $E_1[\pi^{S-}_{r,2}] = \frac{1}{3}(E_1[S_2^-] - E_1[L_2^-])$ and $E_1[\pi^{L-}_{r,2}] = \frac{1}{3}E_1[S_2^-] + \frac{1}{6}E_1[L_2^-] = T^L$, we derive (5.4) and (5.5).

5.7.3 Proof of Corollary 5.1

We claim that in $t = 0$, total expected utility of a low-talented player is higher at club S than at club L. The reverse is shown to hold true for a high-talented player with a sufficiently low risk-aversion.

If the player signs a short-term contract in $t = 0$ at club S, then according to (5.2) the player's total expected utility is given by $E_0[\underline{u}^P_r] = \underline{w}^S_{r,1} + E_0[\underline{w}_{r,2}] - \frac{1}{2}\tau V[\underline{w}_{r,2}]$. In the other case, the player's total expected utility is given by $E_0[\overline{u}^P_r] = \overline{w}^L_{r,1} + E_0[\overline{w}_{r,2}] - \frac{1}{2}\tau V[\overline{w}_{r,2}]$ according to (5.8). In order to prove our claim, we have to show that $E_0[\underline{u}^P_r] > E_0[\overline{u}^P_r]$ for $p \leq \frac{1}{2}$ and $E_0[\underline{u}^P_r] < E_0[\overline{u}^P_r]$ for $p > \frac{1}{2}$. We define $f(\tau) := E_0[\underline{u}^P_r] - E_0[\overline{u}^P_r]$ and compute

$$\begin{aligned}
E_0[\underline{u}^P_r] - E_0[\overline{u}^P_r] &= 0 \Leftrightarrow \tau = \widetilde{\tau}(p,s,l) := \frac{18(2p-1)}{(1-p)p(s(5-4p) + l(1+4p))},\\
\frac{\partial(E_0[\underline{u}^P_r] - E_0[\overline{u}^P_r])}{\partial \tau} &= \frac{(l-s)}{54}(1-p)p(s(5-4p) + l(1+4p)) > 0, \forall p \in (0,1).
\end{aligned}$$

We derive that if $\tau > \widetilde{\tau}(p,s,l)$, then $E_0[\underline{u}^P_r] > E_0[\overline{u}^P_r]$.

Let $p \in (0, \frac{1}{2}]$, then $\widetilde{\tau}(p,s,l) \leq 0$ and hence $E_0[\underline{u}^P_r] > E_0[\overline{u}^P_r] > 0 \;\; \forall \tau > 0$. That is, a low-talented, risk-averse player $(p \leq \frac{1}{2})$, independent of his risk-aversion, realizes a higher total expected utility by signing a contract with club S in $t = 0$. Note that $\widetilde{\tau} = 0$ for $p = \frac{1}{2}$.

Let $p \in (\frac{1}{2}, 1)$, then $\widetilde{\tau}(p,s,l) > 0$ and hence $E_0[\underline{u}^P_r] < E_0[\overline{u}^P_r] \;\; \forall \tau < \widetilde{\tau}(p,s,l)$. That is, a high-talented, risk-averse player $(p > \frac{1}{2})$ with a sufficiently low risk-aversion $(\tau < \widetilde{\tau}(p,s,l))$ realizes a higher total expected utility by signing a contract with club L in $t = 0$.

5.7.4 Proof of Proposition 5.1

(i) We claim that the risk-free first-period salary under a short-term contract in the 'shadow of the transfer system' is higher than the respective salary without the 'shadow of the transfer system' for a high-talented player. The same holds true for a low-talented player whose risk-aversion is sufficiently low. Formally, we show $\underline{w}^S_{r,1} > \underline{w}^S_{u,1}$ if $\tau < \tau^*(p, e_0, s, l)$ and $\overline{w}^L_{r,1} > \overline{w}^L_{u,1}$ for all $\tau > 0$.

For a low-talented player, we derive

$$\underline{w}^S_{r,1} > \underline{w}^S_{u,1} \Leftrightarrow pT^S + (1-p)(T^L - \frac{1}{2}E_1[\pi^{S-}_{r,2}]) + \frac{\tau}{6}(V[\underline{w}_{r,2}] - V[\overline{w}_{r,2}]) > 0$$

and compute

$$\underline{w}^S_{r,1} - \underline{w}^S_{u,1} = 0 \Leftrightarrow \tau = \tau^*(p, e_0, s, l) := \frac{9(3e_0 - (1-p)(4s+2l))}{(l-s)(1-p)p(s(5-4p) + l(1+4p))}.$$

We deduce that $\tau^*(p, e_0, s, l) > 0$ since we assumed $e_0 > 2l(1-p)$ and $l > s$. Moreover,

$$\frac{\partial(\underline{w}^S_{r,1} - \underline{w}^S_{u,1})}{\partial \tau} = -\frac{(l-s)}{54}(1-p)p(s(5-4p) + l(1+4p)) < 0.$$

Hence, if $\tau < \tau^*(p, e_0, s, l)$, then $\underline{w}^S_{r,1} > \underline{w}^S_{u,1}$. A low-talented risk-averse player whose risk-aversion is sufficiently low realizes a higher risk-free first-period salary in the 'shadow of the transfer system' than without the 'shadow of the transfer system.'

For a high-talented player, we derive

$$\overline{w}^L_{r,1} > \overline{w}^L_{u,1} \Leftrightarrow (1-p)T^L + p(T^S - \frac{1}{2}E_1[\pi^{L+}_{r,2}]) + \frac{\tau}{6}(V[\overline{w}_{r,2}] - V[\underline{w}_{r,2}])$$

and compute

$$\begin{aligned} \overline{w}^L_{r,1} - \overline{w}^L_{u,1} \quad &\Leftrightarrow \quad \tau = -\tau^*(p, e_0, s, l) < 0, \\ \frac{\partial(\overline{w}^L_{r,1} - \overline{w}^L_{u,1})}{\partial \tau} \quad &= \quad \frac{(l-s)}{54}(1-p)p(s(5-4p) + l(1+4p)) > 0. \end{aligned}$$

We deduce that if $\tau > 0 > \tau^*(p, e_0, s, l)$, then $\overline{w}^L_{r,1} > \overline{w}^L_{u,1}$. A high-talented, risk-averse player, independent of his risk-aversion, realizes a higher risk-free first-period salary in the 'shadow of the transfer system' than without the 'shadow of the transfer system.' This proves the claim.

(ii) We claim that the expected second-period salary under a short-term contract in the 'shadow of the transfer system' is lower than the respective salary without the 'shadow of the transfer system' for both, a low and a high-talented player.

We derive that $E_0[\underline{w}_{r,2}] < E_0[w_{u,2}]$ and $E_0[\overline{w}_{r,2}] < E_0[w_{u,2}]$ hold if the following inequalities are fulfilled:

$$\begin{aligned} p(\frac{1}{6}E_1[S_2^+] + \frac{1}{3}E_1[L_2^+]) + (1-p)(\frac{1}{6}E_1[S_2^-] + \frac{1}{3}E_1[L_2^-]) &> 0, \\ p(\frac{1}{3}E_1[S_2^+] + \frac{1}{6}E_1[L_2^+]) + (1-p)(\frac{1}{3}E_1[S_2^-] + \frac{1}{6}E_1[L_2^-]) &> 0. \end{aligned}$$

This proves the claim, since all terms are positive.

5.7.5 Proof of Proposition 5.2

We claim that both a low- and a high-talented, risk-averse player benefits from the 'shadow of the transfer system.' In order to prove the claim, we show that total expected utility of each type of risk-averse player is higher under a short-term contract in the 'shadow of the transfer system' than under a short-term contract without the 'shadow of the transfer system.' Formally, we show that $E_0[\underline{u}_r^P] > E_0[\underline{u}_u^P]$ and $E_0[\overline{u}_r^P] > E_0[\overline{u}_u^P]$ for all $\tau > 0$.

(i) We compute for a low-talented player:

$$\begin{aligned} E_0[\underline{u}_r^P] - E_0[\underline{u}_u^P] &= 0 \Leftrightarrow \tau = 0, \\ \frac{\partial(E_0[\underline{u}_r^P] - E_0[\underline{u}_u^P])}{\partial\tau} &= \frac{1}{54}(1-p)p(2s^2(13-8p) + 4sl(11-p) + l^2(11+20p)). \end{aligned}$$

We derive $\frac{\partial(E_0[\underline{u}_r^P] - E_0[\underline{u}_u^P])}{\partial\tau} > 0, \forall p \in (0, 1/2]$ and thus if $\tau > 0$, then $E_0[\underline{u}_r^P] > E_0[\underline{u}_u^P]$.

(ii) We compute for a high-talented player:

$$\begin{aligned} E_0[\overline{u}_r^P] - E_0[\overline{u}_u^P] &= 0 \Leftrightarrow \tau = 0, \\ \frac{\partial(E_0[\overline{u}_r^P] - E_0[\overline{u}_u^P])}{\partial\tau} &= \frac{1}{54}(1-p)p(s^2(31-20p) + 4sl(10+p) + 2l^2(5+8p)). \end{aligned}$$

We derive $\frac{\partial(E_0[\overline{u}_r^P] - E_0[\overline{u}_u^P])}{\partial\tau} > 0, \forall p \in (1/2, 1)$ and thus if $\tau > 0$, then $E_0[\overline{u}_r^P] > E_0[\overline{u}_u^P]$. For a high-talented player must additionally hold $\tau < \widetilde{\tau}(p, s, l)$, since in Corollary 5.1 we have restricted the risk-aversion of a high-talented player in order to guarantee that the player signs a contract with club L in $t = 0$. This proves the claim.

5.7.6 Derivation of the player's and club L's total expected utility in $t = 0$

This appendix contains the derivation of the player's and club L's total expected utility for the case that the player has signed a short-term contract with club L in $t = 0$.

(i) If the player's productivity has increased during period 1 (which happens with probability p), we know by equation (5.1) that each type of player will achieve a higher expected second-period productivity at his initial club L compared with the other club S. In $t = 1$, club S will not place any offer and the player will therefore stay at club L. Without a competing offer from club S, the player will stay at club L in $t = 1$ and the player's reservation wage falls to zero. Similar to Section 5.4.1, the player's second-period salary is now determined by the negotiations only between club L and the player via Nash bargaining:

$$\overline{w}_{r,2}^{L+} = \arg\max_{\overline{w}_{r,2}^{L+}} (E_1[L_2^+] - \overline{w}_{r,2}^{L+} - 0)(\overline{w}_{r,2}^{L+} - 0) = \frac{1}{2}E_1[L_2^+].$$

Club L then expects a second-period profit of $E_1[\pi_{r,2}^{L+}] = E_1[L_2^+] - \overline{w}_{r,2}^{L+} = \frac{1}{2}E_1[L_2^+]$ and club S will earn $E_1[\pi_{r,2}^{S+}] = 0$.

(ii) If the player's productivity has decreased during period 1 (which happens with probability $(1-p)$), we know by equation (5.1) that each type of player will achieve a higher expected second-period productivity at club S compared with club L. According to the Coase theorem the player will be transferred from club L to club S. Similar to Section 5.4.1 the following lemma determines each party's contribution to the player's transfer from club L to club S in $t = 1$:

Lemma 5.4.
The Shapley values determine the outcome of the cooperative bargaining game as follows: $\frac{1}{3}E_1[S_2^-] + \frac{1}{6}E_1[L_2^-]$ *(player),* $\frac{1}{3}(E_1[S_2^-] - E_1[L_2^-])$ *(club S) and* $\frac{1}{3}E_1[S_2^-] + \frac{1}{6}E_1[L_2^-]$ *(club L).*

Proof. Analogous to Lemma 5.1. □

Thus, the player will receive at club S a second-period salary of

$$\overline{w}_{r,2}^{S-} = \frac{1}{3}E_1[S_2^-] + \frac{1}{6}E_1[L_2^-].$$

Club L receives as a transfer fee T^L and realizes an expected second-period profit of $E_1[\pi_{r,2}^{L-}] = \frac{1}{3}E_1[S_2^-] + \frac{1}{6}E_1[L_2^-] = T^L$. Club S similarly obtains its Shapley value and realizes an expected second-period profit of $E_1[\pi_{r,2}^{S-}] = \frac{1}{3}(E_1[S_2^-] - E_1[L_2^-])$.

Analogous to Section 5.4.1, in $t = 0$, club L's total expected utility, denoted $E_0[u_r^L]$, is given by

$$E_0[u_r^L] = E_0[\pi_{r,1}^L] + pE_1[\pi_{r,2}^{L+}] + (1-p)E_1[\pi_{r,2}^{L-}],$$

with an expected first-period profit of $E_0[\pi_{r,1}^L] = E_0[L_1] - \overline{w}_{r,1}^L$.

In $t = 0$, the player's total expected utility, denoted $E_0[\overline{u}_r^P]$, is given by

$$E_0[\overline{u}_r^P] = \overline{w}_{r,1}^L + E[\overline{w}_{r,2}] - \frac{1}{2}\tau V[\overline{w}_{r,2}], \tag{5.8}$$

where the expected second-period salary $E_0[\overline{w}_{r,2}]$ and the variance $V[\overline{w}_{r,2}]$ of the second-period salary are given by

$$\begin{aligned} E_0[\overline{w}_{r,2}] &= p\overline{w}_{r,2}^{L+} + (1-p)\overline{w}_{r,2}^{S-}, \\ V[\overline{w}_{r,2}] &= p(\overline{w}_{r,2}^{L+})^2 + (1-p)(\overline{w}_{r,2}^{S-})^2 - (E[\overline{w}_{r,2}])^2. \end{aligned}$$

Similar to Section 5.4.1, the threat points of the simultaneous negotiations in Nash bargaining fashion in $t = 0$ are derived as follows: with probability p each type of player will achieve a higher expected second-period productivity at the large-market club L. In case that the player has signed a short-term contract with club S in $t = 0$ he will transfer to club L in $t = 1$. Club L will then receive its contribution to the coalition determined by its Shapley value. Thus club L's threat point, denoted d^L, is given by $p\frac{1}{3}(E_1[L_2^+] - E_1[S_2^+])$. The player's threat point, denoted $\underline{d}^P$, is determined by the player's total expected utility $E_0[\underline{u}_r^P]$ that he could achieve by playing at the other club S.

Chapter 6

Summary

In this book we have provided four theoretical models of team sports leagues that addressed different issues in the economics of sports. In the introduction, we referred to the importance of the sports industry and the place it occupies in today's society. In this regard, it is therefore not surprising that economists have discovered the sports industry as a research field. The recent explosion of publishing in the sports economics field underlines this trend. Next, we have argued that owing to the peculiarities of team sports leagues a certain degree of competitive balance is necessary to ensure a successful competition. In this respect, we have specified the most important league policy measures which were/are implemented to provide cross subsidies among teams. The first chapter ends with a review of selected literature in the field of sports economics and an outline of the book.

Based on an integrated framework, Chapter 2 has provided a general model of talent investment in a team sports league. We applied contest theory to analyze the competitive interaction between clubs and to give a rationale as to why clubs tend to 'overinvest' in playing talent. We have shown that dissipation of the league revenue arises from 'overinvestment' in playing talent as a direct consequence of the ruinous competitive interaction between clubs. Moreover, the model may help to explain the 'paradox of rising revenues and declining profits' encountered in the European Football Leagues. The analysis has shown that the overinvestment problem increases if the discriminatory power of the contest function increases, revenue-sharing decreases, and the size of an additional exogenous prize increases. We have further deduced that clubs invest more when they play in an open, as opposed to a closed, league. Moreover, the overinvestment problem within open leagues increases with the revenue differential between leagues.

In Chapter 3 we have introduced salary caps in a game-theoretical model of a team sports league and have analyzed their effect on social welfare. Through our model, we have shown that salary caps increase competitive balance and help to keep salary costs under control. The resulting effect on social welfare is counter-intuitive and depends on the relative preference of fans for aggregate talent and for competitive balance. A salary cap that binds only large-market clubs will increase social welfare if fans prefer aggregate talent, despite the fact that the salary cap will result in lower aggregate talent. If fans prefer competitive balance, on the other hand, any binding salary cap will reduce social welfare. Moreover, we have shown that if salary caps are beneficial for social welfare they also increase club profits. Therefore clubs will never oppose salary caps which have a positive effect on social welfare.

In Chapter 4 we have studied the welfare effect of gate revenue-sharing. Using our contest model we have arrived at two counter-intuitive results: First, we have challenged the 'invariance proposition' by showing that gate revenue-sharing reduces competitive balance and thus produces a more unbalanced league. This result is driven by the dulling effect of revenue-sharing which is revealed to be stronger for the small-market than for the large-market club. Second, we have shown that a lower degree of competitive balance compared with the non-cooperative league equilibrium actually yields a higher level of social welfare and club profits. Combining both results, we have concluded that gate revenue-sharing increases social welfare and club profits. This result may have important policy implications.

In Chapter 5 we have investigated the role of transfer restrictions in the European labour-market for professional football players. Given that the player's productivity varies significantly during his career, and taking into account that these variations cannot be predicted, the allocation of risk becomes a crucial feature. Our model, which captures this important aspect of employment relations in football, has revealed that a risk-averse player benefits from a more restrictive transfer system. We have found that under the pre-Bosman regime clubs can partially insure their players against income uncertainty by transforming a part of the player's risky future salary in risk-free current income. As a result, a risk-averse player benefits from the 'shadow of the transfer system,' since he prefers a higher current salary combined with a lower (expected) future salary to a lower current salary combined with a higher but uncertain future salary.

Further research should be dedicated to the generalization of the presented models since they build upon some restrictive assumptions. Caution is therefore necessary in inferring direct policy implications from our results. Nevertheless, this book can guide policy-makers, who can gain a deeper understanding of the functionality of team sports leagues from a theoretical point of view.

Bibliography

Andreff, W. and Szymanski, S. (2007), Introdution: Sport and Economics, *in* W. Andreff and S. Szymanski, eds, 'Handbook on the Economics of Sport', Edward Elgar Publishing Ltd, Cheltenham, UK, pp. 1–8.

Antonioni, P. and Cubbin, J. (2000), 'The Bosman Ruling and the Emergence of a Single Market in Soccer Talent', *European Journal of Law and Economics* **9**, 157–173.

Atkinson, S., Stanley, L. and Tschirhart, J. (1988), 'Revenue Sharing as an Incentive in an Agency Problem: An Example from the National Football League', *The RAND Journal of Economics* **19**, 27–43.

Baron, J. and Kreps, D. (1999), *Strategic Human Resources: Frameworks for General Managers*, Wiley, New York.

Baye, M., Kovenock, D. and de Vries, C. (1994), 'The Solution to the Tullock Rent-Seeking Game When $R > 2$: Mixed-Strategy Equilibrium and Mean Dissipation Rates', *Public Choice* **81**, 363–380.

Borland, J. and MacDonald, R. (2003), 'The Demand for Sports', *Oxford Review of Economic Policy* **19**, 478–502.

Burguet, R., Caminal, R. and Matutes, C. (2002), 'Golden Cages for Showy Birds: Optimal Switching Costs in Labour Markets', *European Economic Review* **46**, 1153–1185.

Canes, M. E. (1974), The Social Benefits of Restrictions in Team Quality, *in* R. G. Noll, ed., 'Government and the Sports Business', Brooking Institution, Washington D.C., pp. 81–113.

Carbonell-Nicolau, O. and Comin, D. (2005), 'Testing Out Contractual Incompleteness: Evidence from Soccer'. NBER Working Paper No. W11110 at New York University.

Chung, T. (1996), 'Rent-Seeking Contest When the Prize Increases with Aggregate Efforts', *Public Choice* **87**, 55–66.

Clark, D. and Riis, C. (1998), 'Contest Success Functions: An Extension', *Economic Theory* **11**, 201–204.

Coase, R. (1960), 'The Problem of Social Cost', *The Journal of Law and Economics* **3**, 1–44.

Cyrenne, P. (2001), 'A Quality-of-Play Model of a Professional Sports League', *Economic Inquiry* **39**, 444–452.

Daly, G. and Moore, W. (1981), 'Externalities, Property Rights and the Allocation of Resources in Major League Baseball', *Economic Inquiry* **19**, 77–95.

Deloitte and Touche (2004), 'Annual Review of Football Finance'.

Demmert, H. (1973), *The Economics of Professional Team Sports*, Lexington Books, Lexington.

Dietl, H. and Franck, E. (2000), 'Effizienzprobleme in Sportligen mit Gewinnmaximierenden Kapitalgesellschaften - eine Modelltheoretische Untersuchung', *Zeitschrift für Betriebswirtschaft* **70**, 1157–1175.

Dietl, H., Franck, E. and Lang, M. (2008*a*), 'Overinvestment in Team Sports Leagues - A Contest Theory Model', *Scottish Journal of Political Economy* **55**, 353–368.

Dietl, H., Franck, E. and Lang, M. (2008*b*), 'Why Football Players May Benefit from the "Shadow of the Transfer System"', *European Journal of Law and Economics, forthcoming* .

Dietl, H., Franck, E. and Roy, P. (2003), 'Ueberinvestitionsprobleme in einer Sportliga', *Betriebswirtschaftliche Forschung und Praxis* **5**, 528–540.

Dietl, H. and Lang, M. (2008), 'The Effect of Gate Revenue-Sharing on Social Welfare', *Contemporary Economic Policy, Prepublished April 22, 2008; DOI:10.1111/j.1465-7287.2007.00090.x* .

Dietl, H., Lang, M. and Rathke, A. (2007), 'The Effect of Salary Caps in Professional Team Sports on Social Welfare', *ISU Working Paper Series No. 72* .

Dixit, A. (1986), 'Comparative Statics for Oligopoly', *International Economic Review* **27**, 107–122.

Dixit, A. (1987), 'Strategic Behavior in Contests', *American Economic Review* **77**, 891–898.

Downward, P. and Dawson, A. (2000), *The Economics of Professional Team Sports*, Routledge, London.

Economist (2002), 'For Love or Money', *Economist* **363(8275)**, 7.

Ehrenberg, R. and Bognanno, M. (1990*a*), 'Do Tournaments Have Incentive Effects?', *Journal of Political Economy* **13**, 1307–1324.

Ehrenberg, R. and Bognanno, M. (1990*b*), 'The Incentive Effect of Tournaments Revisited: Evidence from the European PGA Tour', *Industrial and Labor Relations Review* **43**, 74S–88S.

El-Hodiri, M. and Quirk, J. (1971), 'An Economic Model of a Professional Sports League', *The Journal of Political Economy* **79**, 1302–1319.

Falconieri, S., Palomino, F. and Sákovics, J. (2004), 'Collective versus Individual Sale of Television Rights in League Sports', *Journal of the European Economic Association* **5**, 833–862.

Feess, E., Frick, B. and Muehlheusser, G. (2004), 'Legal Restrictions on Buyout Fees: Theory and Evidence from German Soccer'. IZA Discussion Paper No. 1180.

Feess, E. and Muehlheusser, G. (2003), 'Transfer Fee Regulations in European Football', *European Economic Review* **47**, 645–668.

Fernie, S. and Metcalf, D. (1999), 'It's Not What You Pay, It's the Way That You Pay It, and That's What Gets Results: Jockeys' Pay and Performance', *Labour* **13**, 385–411.

Fort, R. and Quirk, J. (1995), 'Cross-Subsidization, Incentives, and Outcomes in Professional Team Sports Leagues', *Journal of Economic Literature* **33**, 1265–1299.

Fort, R. and Quirk, J. (2004), 'Owner Objectives and Competitive Balance', *Journal of Sports Economics* **5**, 20–32.

Frick, B. (2006), The Labour Market for Soccer Players in Germany: Career Duration in a Competitive Environment, *in* C. Jeanreneaud and S. Késenne, eds, 'Economics Applied to Sports: Five Case Studies', Neuchatel: Éditions CIES, pp. 45–62.

Green, J. and Stokey, N. (1983), 'A Comparison of Tournaments and Contests', *Journal of Political Economy* **91**, 349–364.

Hirshleifer, J. (1989), 'Conflict and Rent-Seeking Success Functions: Ratio vs. Difference Models of Relative Success', *Public Choice* **63**, 101–112.

Hoehn, T. and Szymanski, S. (1999), 'The Americanization of European Football', *Economic Policy* **14**, 204–240.

Humphreys, B. (2002), 'Alternative Measures of Competitive Balance in Sports Leagues', *Journal of Sports Economics* **3**, 133–148.

Késenne, S. (2000*a*), 'Revenue Sharing and Competitive Balance in Professional Team Sports', *Journal of Sports Economics* **1**, 56–65.

Késenne, S. (2000*b*), 'The Impact of Salary Caps in Professional Team Sports', *Scottish Journal of Political Economy* **47**, 422–430.

Késenne, S. (2003), 'The Salary Cap Proposal of the G-14 in European Football', *European Sport Management Quarterly* **3**, 120–128.

Késenne, S. (2005), 'Revenue Sharing and Competitive Balance - Does the Invariance Proposition Hold?', *Journal of Sports Economics* **6**, 98–106.

Késenne, S. (2006), 'The Win Maximization Model Reconsidered: Flexible Talent Supply and Efficiency Wages', *Journal of Sports Economics* **7**, 416–427.

Lazear, E. (1995), *Personnel Economics*, MIT Press, Cambridge, MA.

Lazear, E. (1998), *Personnel Economics for Managers*, Wiley, New York.

Lazear, E. and Rosen, S. (1981), 'Rank-Order Tournaments as Optimal Labor Contracts', *Journal of Political Economy* **89**, 841–864.

Lockard, A. and Tullock, G. E. (2001), *Efficient Rent-Seeking: Chronicle of an Intellectual Quagmire*, Kluwer Academic Publisher, Boston.

Lynch, J. and Zax, J. (2000), 'The Rewards to Running: Prize Structure and Performance in Professional Road Racing', *Journal of Sports Economics* **1**, 323–340.

Maloney, M. and Terkun, K. (2000), 'The Response of Workers to Wages in Tournaments: Evidence from Foot Races', *Journal of Sports Economics* **1**, 99–123.

Maloney, M. and Terkun, K. (2002), 'Road Warrior Booty: Prize Structures in Motorcycle Racing', *Contributions to Economic Analysis and Policy* **1**, 1–18.

Marburger, D. R. (1997), 'Gate Revenue Sharing and Luxury Taxes in Professional Sports', *Contemporary Economic Policy* **15**, 114–123.

Nalebuff, B. J. and Stiglitz, J. E. (1983), 'Prizes and Incentives: Towards a General Theory of Compensation and Competition', *The Bell Journal of Economics* **14**, 21–43.

Neale, W. (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* **78**, 1–14.

Orszag, J. (1994), 'A New Look at Incentive Effects and Tournaments', *Economics Letters* **46**, 77–88.

Quirk, J. and El-Hodiri, M. (1974), The Economic Theory of a Professional Sports League, *in* R. G. Noll, ed., 'Government and the Sports Business', Brooking Institution, Washington D.C., pp. 33–80.

Quirk, J. and Fort, D. (1992), *Pay Dirt: The Business of Professional Team Sports*, Princeton University Press, Princeton.

Rascher, D. (1997), A Model of a Professional Sports League, *in* W. Hendricks, ed., 'Advances in Economics of Sport', Vol. 2, JAI Press, Greenwich, pp. 27–76.

Rosen, S. (1981), 'The Economics of Superstars', *American Economic Review* **71**, 845–858.

Rottenberg, S. (1956), 'The Baseball Players' Labor Market', *Journal of Political Economy* **64**, 242–258.

Schmidt, M. and Berri, D. (2001), 'Competitive Balance and Attendance: The Case of Major League Baseball', *Journal of Sports Economics* **2**, 145–167.

Scully, G. (1974), 'Pay and Performance in Major League Baseball', *American Economic Review* **64**, 915–930.

Scully, G. (2000), 'Diminishing Returns and the Limit of Athletic Performance', *Scottish Journal of Political Economy* **47**, 456–470.

Shaw, W. and Jakus, P. (1996), 'Travel Cost Models of the Demand for Rock Climbing', *Agricultural and Resource Economics Review* **25**, 133–142.

Skaperdas, S. (1996), 'Contest Success Functions', *Economic Theory* **7**, 283–290.

Sloane, P. (1971), 'The Economics of Professional Football: The Football Club as a Utility Maximizer', *Scottish Journal of Political Economy* **17**, 121–146.

Staudohar, P. (1999), Salary Caps in Professional Team Sports, *in* C. Jeanrenaud and S. Késenne, eds, 'Competition Policy in Professional Sports: Europe After the Bosman Case', Standaard Edition, Antwerpen, Belgium.

Staudohar, P. (2005), 'The Hockey Lockout of 2004-05', *Monthly Labor Review* **128**, 23–29.

Szymanski, S. (2000), 'The Market for Olympic Gold Medals', *World Economics* **1**, 1–8.

Szymanski, S. (2003), 'The Economic Design of Sporting Contests', *Journal of Economic Literature* **41**, 1137–1187.

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* **5**, 111–126.

Szymanski, S. and Késenne, S. (2004), 'Competitive Balance and Gate Revenue Sharing in Team Sports', *The Journal of Industrial Economics* **52**, 165–177.

Tenorio, R. (2000), 'The Economics of Professional Boxing Contracts', *Journal of Sports Economics* **1**, 363–384.

Tullock, G. (1980), Efficient Rent-Seeking, *in* J. Buchanan, R. Tollison and G. Tullock, eds, 'Toward a Theory of Rent Seeking Society', University Press, Texas, pp. 97–112.

Vrooman, J. (1995), 'A General Theory of Professional Sports Leagues', *Southern Economic Journal* **61**, 971–990.

Vrooman, J. (2000), 'The Economics of American Sports Leagues', *Scottish Journal of Political Economy* **47**, 364–398.

Whitney, J. (1993), 'Bidding Till Bankrupt: Destructive Competition in Professional Team Sports', *Economic Inquiry* **31**, 100–115.

Curriculum Vitae

Markus Lang

Date of Birth:	September 13, 1978
Citizenship:	German

Education:

Since 12/2004	Doctoral Studies in Management and Economics, University of Zurich
Since 01/2007	Swiss Program for Beginning Doctoral Students in Economics, Study Center Gerzensee
08/2005	Advanced Doctoral Course in Economics, Cooperative Game Theory, Study Center Gerzensee
09/2001- 07/2002	Studies abroad at the University of Valencia, Spain Department of Applied Mathematics and Economics
10/1998-12/2004	Studies of Applied Mathematics, University of Ulm, Germany Degree: Dipl.-Math.oec (M.A. equivalent), Grade: 1,3
1988-1997	Albertus Gymansium (High School), Lauingen, Germany