Introduction

Comparisons between European and North American sport leagues have occurred over the years. Rather than list the entire literature, we refer the reader to the overviews in Fort (2000), Barros, Ibrahimo, and Szymanski (2002), Fort and Fizel (2004), and Sandy, Sloane, and Rosentraub (2004). The topics have ranged from fan differences across countries, to organizational and business model comparisons, to objective functions of team management, to competitive balance.

As with Chapter 26, we attempt to bring these comparisons down to the essential elements when analyzing what has come to be called Rottenberg’s (1956) invariance principle. In other words, the chapter provides theoretical insights into attempts to alter competitive balance using revenue sharing, talent drafts, and payroll caps. Player reserve systems (the reserve clause in North American leagues and transfer restrictions in European leagues) are also examined. However, the so-called “luxury tax” in Major League Baseball (actually titled the competitive balance tax in that league’s Collective Bargaining Agreement with players) is not examined as it is specific to that league. In this chapter, however, the differences in investment in talent provide the central focus of analysis. Differences in objective functions are handled similarly, however, as North American leagues are treated as having clubs pursuing profit maximization, while European leagues are treated as having clubs pursuing utility maximization and win maximization.

The focus is on model predictions compared to actual outcomes, and any differences between North America and Europe. The chapter proceeds by specifying a simple general model of talent markets and profit maximization for North American leagues. The recent evolution of utility maximization models in European leagues is extended to that case. A brief overview and assessment of competitive balance arrangements in North America and Europe is presented where relevant.

North American leagues (profit maximization)

The objective in this section of the chapter is to present a simple general model of a league of profit-maximizing teams. The level of generality needs to satisfy that there is an underlying
International sport league comparisons

talent market; all outcomes are consistent with a Nash non-cooperative setting,\(^1\) and both closed and open talent markets can be considered.\(^2\) We present the simple general model and then seek theoretical insights about Rottenberg’s invariance principle relative to revenue sharing for open and closed leagues. Brief notes on empirical findings about revenue sharing are included, as are extensions to other talent market impositions such as the draft, free agency, and payroll caps.

A simple general model

The intent is to have a model sufficient to include characteristics of the elasticity of talent, that is the degree of its availability for hire, within a Nash non-cooperative setting. A two-team league is modeled to facilitate comparison with the current literature, recognizing the limitations of that choice. An extremely important feature of the model is the additional dimensionality of choice implied in modeling sport league using Nash equilibria. As Winfree and Fort (2011) illustrate, within the literature there has been a traditional focus on fixed talent supply in an essentially perfectly competitive (Walrasian) setting. Whilst Nash equilibria have been introduced into the literature, with the implication that talent choices can be made independently of other teams (i.e. that talent supply is not fixed), this means that the concept of a fixed talent supply is lost to the analysis as talent is viewed as elastically supplied. To resolve this issue, Winfree and Fort (2011) separate out the investment in talent of individual teams and the impact of overall talent in the league on competitive balance. In this sense the authors distinguish between the levels of talent in teams and the investment in talent of teams. Consequently a Nash equilibrium approach to modeling does not have to imply a trade-off with respect to specifying the market for talent. Winfree and Fort (2011) show that this two-dimensional approach, a talent investment leading to team talent levels, follows a strict suggestion in Szymanski (2004), relates to the original work in the area (by Quirk and El Hodiri, El Hodiri and Quirk, Fort and Quirk, summarized and extended in Fort, 2007, and Fort and Quirk, 2007), and is an advance over previous work (Easton and Rockerbie, 2005; Vrooman, 2007; Chang and Sanders, 2009).

In the model, choosing talent is actually an investment in talent either through purchasing talent at a given level in a market or spending resources to develop talent. Let \(z_1\) for team one and \(z_2\) for team two represent investment in talent for each team. This investment may include payroll, minor league development, training facilities or other investments that affect talent. In turn, let \(t_1(z_1, z_2)\) and \(t_2(z_1, z_2)\) be the actual talent result for the two teams depending on the talent investment across the league. Our interpretation of the Nash insight is that neither team believes that their talent investment impacts the investment of the other team.

Whether the talent market is closed or open is depicted in the actual talent outcome. In a closed league, talent supply is fixed (completely inelastic), that is, \(t_1 + t_2 = T\). In turn, this implies that \(\frac{\partial t_1}{\partial z_1} = \frac{\partial t_1}{\partial z_2} = \frac{\partial t_2}{\partial z_1} = \frac{\partial t_2}{\partial z_2} = 0\). In an open league (the limiting extreme is a perfectly elastic supply of talent), either team can increase its use of talent without impacting any other team’s level of talent, that is, \(\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_1} = 0\).

Winning, ‘\(w\)’, then depends on the team’s own talent as specified in the following contest functions: \(w_1 = w_1(t_1(z_1, z_2), t_2(z_1, z_2))\) and \(w_2 = w_2(t_1(z_1, z_2), t_2(z_1, z_2))\). Let \(R_i\) be team \(i\)’s
revenue. These revenues are modeled dependent on the team’s own quality (winning percent): 
\( R_1(w_1(t_1(z_1, z_2), t_2(z_2, z_1))) \) and \( R_2(w_2(t_2(z_2, z_1), t_1(z_1, z_2))) \). If \( \alpha \) is the proportion of revenue that is shared then team profits are \(^3\):

\[
\pi_1 = (1 - \alpha) R_1 + \alpha R_2 - z, \quad \pi_2 = (1 - \alpha) R_2 + \alpha R_1 - z 
\]

(28.1)

First-order conditions for a profit maximum for team 1 and team 2, respectively, are

\[
\frac{\partial \pi_1}{\partial z_1} = (1 - \alpha) \left[ \frac{dR_1}{dw_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) + \frac{dR_1}{dt_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) \right] + \alpha \left[ \frac{dR_2}{dw_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) + \frac{dR_2}{dt_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) \right] - 1 = 0
\]

(28.2)

\[
\frac{\partial \pi_2}{\partial z_2} = (1 - \alpha) \left[ \frac{dR_2}{dw_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) + \frac{dR_2}{dt_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) \right] + \alpha \left[ \frac{dR_1}{dw_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) + \frac{dR_1}{dt_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) \right] - 1 = 0
\]

(28.3)

The winning percent adding up constraint must also hold, \( w_1 + w_2 = 1 \).

Imposing Nash conjectures, \( \frac{dz_2}{dz_1} = \frac{dz_1}{dz_2} = 0 \), (28.2) and (28.3) become, respectively:

\[
(1 - \alpha) \left[ \frac{dR_1}{dw_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) + \alpha \left( \frac{dR_2}{dw_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) + \frac{dR_2}{dt_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) \right) - 1 = 0
\]

(28.4)

\[
(1 - \alpha) \left[ \frac{dR_2}{dw_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) + \alpha \left( \frac{dR_1}{dw_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) + \frac{dR_1}{dt_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) \right) - 1 = 0
\]

(28.5)

and \( w_1 + w_2 = 1 \). Setting (28.4) and (28.5) equal to one, and then equal to each other, the talent investment equilibrium is:

\[
(1 - \alpha) \left[ \frac{dR_1}{dw_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) + \alpha \left( \frac{dR_2}{dw_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) + \frac{dR_2}{dt_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) \right) \right] = (1 - \alpha) \left[ \frac{dR_2}{dw_2} \left( \frac{\partial t_2}{\partial t_2} + \frac{\partial t_2}{\partial z_2} \right) + \alpha \left( \frac{dR_1}{dw_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) + \frac{dR_1}{dt_1} \left( \frac{\partial t_1}{\partial t_1} + \frac{\partial t_1}{\partial z_1} \right) \right) \right]
\]

(28.6)

In addition, \( w_1 + w_2 = 1 \) implies \( \frac{\partial w_1}{\partial t_1} = -\frac{\partial w_2}{\partial t_2} \) and \( \frac{\partial w_2}{\partial t_2} = -\frac{\partial w_1}{\partial t_1} \). Substituting, we rewrite (28.6) as:

390
International sport league comparisons

This is clearly a Nash equilibrium since satisfaction of (28.7) has each team taking the
talent choice of the other team as given in their independent profit-maximizing choice of
talent investment; each plays their best response.

The results in (28.5) are also “general” in the sense that no functional forms are imposed
on either revenues or the underlying contest for this non-cooperative portrayal and no refer-
ence is made specifically to either open or closed leagues.

Ultimately, one of the distinguishing outcomes for this equilibrium is the presence or
absence of Rottenberg’s (1956) invariance principle – the distribution of talent in equilibrium
is invariant with respect to revenue sharing. Essentially, the invariance principle holds if the
equilibrium with revenue sharing in (28.7) with \( \alpha > 0 \) is the same as the equilibrium without
revenue sharing, that is, with \( \alpha = 0 \). With \( \alpha = 0 \), (28.7) becomes:

\[
(1-\alpha) \left\{ \frac{dR_1}{dw_1} \left[ \frac{\partial w_1}{\partial t_1} \frac{\partial t_1}{\partial z_1} - \frac{\partial w_2}{\partial t_2} \frac{\partial t_2}{\partial z_1} \right] \right\} + \alpha \left\{ \frac{dR_2}{dw_2} \left[ \frac{\partial w_2}{\partial t_2} \frac{\partial t_2}{\partial z_2} - \frac{\partial w_1}{\partial t_1} \frac{\partial t_1}{\partial z_2} \right] \right\} = 0.
\]

Expanding and rearranging terms in (28.7) to isolate the left-hand side of (28.8), the
invariance principle holds with \( \alpha > 0 \), when:

\[
\frac{dR_1}{dw_1} \left[ \frac{\partial w_1}{\partial t_1} \frac{\partial t_1}{\partial z_1} - \frac{\partial w_2}{\partial t_2} \frac{\partial t_2}{\partial z_1} \right] - \frac{dR_2}{dw_2} \left[ \frac{\partial w_2}{\partial t_2} \frac{\partial t_2}{\partial z_2} - \frac{\partial w_1}{\partial t_1} \frac{\partial t_1}{\partial z_2} \right] = 0.
\]

Further substituting equation (28.8) into equation (28.9), the conditions for the invariance
principle are given by

\[
\frac{\partial w_1}{\partial t_1} \left( \frac{\partial t_1}{\partial z_1} + \frac{\partial t_1}{\partial z_2} \right) = \frac{\partial w_2}{\partial t_2} \left( \frac{\partial t_2}{\partial z_1} + \frac{\partial t_2}{\partial z_2} \right).
\]

Assuming that both teams have identical contest success functions and talent generating
functions, we observe the following about the invariance principle, at the level of generality
in (28.10). Assume that team 1 is in the larger-revenue market and there are diminishing
marginal returns to talent, \( \frac{\partial^2 w_1}{\partial t_1^2} < 0 \). With team two in the larger-revenue market, these
assumptions imply that if both teams increase their investment in talent by one unit, the effect
will be larger on team two in the smaller-revenue market, that is \( \frac{\partial t_1}{\partial z_1} + \frac{\partial t_1}{\partial z_2} < \frac{\partial t_2}{\partial z_1} + \frac{\partial t_2}{\partial z_2} \). Since
it would require a particular \( \frac{\partial w_1}{\partial t_1} > \frac{\partial w_2}{\partial t_2} \) to satisfy (28.10) in this case, then the invariance
principle will not hold in general. This proposition can also be analyzed with the further restrictions on the elasticity of talent, that is, examining closed versus open talent markets.

**Revenue sharing in closed and open leagues**

The typical assumption is that North American leagues are “closed,” that is, talent supply is fixed (completely inelastic). With the recent increase in the use of imported talent from international leagues in baseball, basketball, and hockey, perhaps this completely closed league idea is most relevant for the NFL. However, since extensive talent importing is a fairly recent phenomenon, the closed market remains insightful in all cases.

As noted in the last subsection, the assumption that the talent market is closed implies that $t_1 + t_2 = T$. In turn, this implies that $\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_2}$ and $\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_2}$. Substituting $\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_2}$ and

$\frac{\partial t_2}{\partial z_2} = -\frac{\partial t_1}{\partial z_1}$

into (28.10) and rearranging terms yields:

$$\left(\frac{\partial w_1}{\partial t_2} + \frac{\partial w_1}{\partial t_1}\right) = \frac{\partial w_2}{\partial t_2} + \frac{\partial w_2}{\partial t_1}$$

that is, a Nash, non-cooperative, two-team, closed league equilibrium exhibits the invariance principle if $\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_2}$.

In equilibrium, the invariance principle holds only if the marginal product of talent investment in the actual creation of talent is the same across teams. Thus, in a closed league, the distribution of talent (and, thus, the invariance principle) depends on the relationship between the investment in talent and the accumulation of talent. (Winfree and Fort, 2011, show how previous work in the area generated the invariance principle relative to (28.11) and we do not go through that here.)

Open leagues are defined by $\frac{\partial t_1}{\partial z_2} = \frac{\partial t_2}{\partial z_1} = 0$ and (28.10) becomes:

$$\frac{\partial w_1}{\partial t_1} = \frac{\partial w_2}{\partial t_2}$$

(28.12)

Now, even if $\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_2}$ in equilibrium, (28.11) would still require $\frac{\partial w_1}{\partial t_1} = \frac{\partial w_2}{\partial t_2}$ for the invariance principle to hold. This would be true of the trivial case where teams are completely balanced in the first place or for an arbitrary imposition on these marginal products that sets them equal to each other but, generally, the invariance principle does not hold for open leagues.

Interestingly, the data on the distribution of talent, primarily the Noll-Scully “ratio of standard deviation” of end-of-season winning percents, either fail to reject the invariance principle or show that balance worsened in some North American leagues with the imposition of revenue sharing and with any increase in sharing. Fort (2011) shows this for the two North American leagues with extensive sharing, Major League Baseball and the National Football League. The implications for the theory above are pretty clear, sticking with closed leagues since revenue-sharing impositions happened in MLB prior to the modern rise of imported talent and the NFL does not really import any talent. This empirical observation
International sport league comparisons

suggests that the relationship between the investment in talent and the accumulation of talent is really not quite as complex as suggested for closed leagues in (28.11). For the closed talent market, that the invariance principle holds with respect to revenue sharing is consistent with

$$\frac{\partial t_1}{\partial z_1} = \frac{\partial t_2}{\partial z_2},$$

that is, equilibrium in the talent investment market is characterized by equal marginal product of investment in talent across all teams. Of course, it would be preferable to model the talent market more extensively and approach the issue directly in that market.

**Other talent market changes and impositions**

The theoretical impacts of remaining changes and impositions in the talent market are not derived here. Fort and Quirk (1995) used a somewhat restrictive version of the simple general theory, above, to portray the impacts of the draft, free agency, and payroll caps in the closed market case. While in this chapter analysis is focused on the closed market assessment of these other impositions, the interested reader can also relate the assessment to the open market case using Vrooman (2007). This model is close enough to the general open league case and the two-team league diagram would prove insightful for that case.

Rottenberg (1956) actually only applied the logic of the invariance principle to the draft and free agency (Fort, 2005). The reverse-order-of-finish draft dominates North American professional sport and has the lowest-finishing teams choosing incoming talent first. That talent signs its first contract with the drafting team and must follow the contract for a specific number of years after that (e.g. six years in baseball). Essentially, the owner argument has always been that a reverse-order-of-finish draft should equalize talent across the league since poor teams command better incoming talent than their market revenues should support. Rottenberg’s logic, instead, is that as long as player contracts can be bought and sold and players must follow their contracts, then the draft will not change the distribution of talent. The draft itself does not alter the value of talent anywhere in the league so all it does is rearrange the value of talent away from players and toward smaller-revenue market owners. Rottenberg’s insight proves insightful with respect to the empirical evidence in the North American case (Fort, 2011, Chapters 29 and 31). Drafts have never improved competitive balance in a statistically significant way.

It is commonly accepted that Rottenberg essentially emphasized the logic of the invariance principle to the case of free agency. Fort (2005) points out that, at least in print, this was the sport version of the weak form of the Coase Theorem a few years prior to its publication by the Nobel Prize winner. Again, as long as contracts can be freely bought and sold and players must follow their contract, then free agency cannot change the distribution of talent. All that happens is the value of talent is reallocated from players to owners and players go to their highest valued use across the league. The evidence supports this case that competitive balance did not change in a statistically significant way with the advent of free agency in any North American league (Fort, 2011, Chapter 31).

Payroll caps were first examined for North American leagues by Fort and Quirk (1995) (the various restrictions in that model are covered in Szymanski, 2004, Fort and Winfree, 2009, and Winfree and Fort, 2011). At that time, the National Basketball Association cap had been in force for quite some time but the National Football League cap was quite new. Since then, the National Hockey League has also added a cap. By forcing a disequilibrium situation, a strictly “hard cap” with equal spending by all teams forces an equilibrium at 0.500 for all teams and, analytically, improves balance. Fort and Quirk (1995) also show that
larger-revenue market owners wish to buy more talent than allowed under the cap and smaller-revenue owners want to sell it to them in this disequilibrium so that leagues have an enforcement problem. In addition, leagues have never really embraced the “hard cap” approach and real-world caps are full of exclusions. The data (Fort and Quirk, 1995; Fort, 2011, Chapter 29), however, show that instead of converging to the official cap, actual payrolls have always been highly dispersed around the official cap and balance actually worsened significantly after the imposition of the National Basketball Association cap. Balance also did not change at all in the NFL with its cap, and though balance did improve in the NHL this was only after two years following the imposition of its cap.

**European leagues (utility and win maximization)**

**Organizational structure**

European team sports are organized as a pyramid of non-profit amateur clubs with a professional apex. This apex may comprise public or private limited companies, or member-owned collectives. However, the main organizational feature is one of vertical promotion and relegation between leagues based on merit. From a property rights perspective, these member associations in the professional game, or amateur games, differ significantly from the North American professional teams that are usually organized as commercial firms. The owner of such a firm has all decision rights, including the right to sell the team and to make appropriate profits. The only exception in the North American major leagues is the Green Bay Packers, who are governed as a community-owned non-profit organization.

This governance structure is similar to the European clubs that are organized as member associations. These member associations are democratically governed. Each member usually has one vote. Since there are no residual claimants within these member associations, and because in the European leagues there is a need to seek promotion and to avoid relegation for a level of competition, it is usually assumed that the clubs seek to maximize utility rather than profit. This utility maximization may translate into a variety of specific objectives because the clubs are usually open to new members who pursue their own objectives. For example, FC Barcelona, for example, has more than 170,000 members, Bayern Munich more than 150,000 members, and Sport Lisboa e Benfica more than 200,000 members. In addition, as non-profit organizations, the clubs were traditionally subsidized by their communities and therefore responsible to a large number of stakeholders.

In the last few decades, many clubs have changed their governance structures. Some clubs, such as Manchester United, Liverpool and FC Copenhagen, have become more commercially driven and managed like their North American counterparts. Others have incorporated their professionalization and subsequent developments without ceding control by the member associations. Most German clubs have chosen such a hybrid structure due to the so-called “50+1” rule that stipulates that 50 percent plus one vote of an incorporated German football club must be controlled by the club’s member association. The logic behind the 50+1 restriction is to ensure the integrity of professional football by avoiding a situation in which anybody could exercise control over more than one professional team. At the same time, however, this restriction results in a rather peculiar governance structure within the football corporations. Even if a business tycoon such as Roman Abramovich acquired all outstanding shares of a German football team, he would still control less than 50 percent of the votes.

The respective national associations organize the pyramid of competitions between the clubs at the national level. These national associations are organized as democratic governing
bodies that try to coordinate the objectives of all stakeholders within their constitution. Across national associations, supra-national associations coordinate competition. The European Football Association (UEFA), an association of national associations, organizes international competition within Europe. UEFA organizes European club competitions like the UEFA Champions League and the UEFA Europe League for the teams meeting certain sportive qualification criteria. Teams who qualify for these international competitions continue to compete in their national leagues.

The openness of the European team sport industry has a number of consequences. First, market entry is possible at any time. Through promotion and relegation stronger teams constantly replace weaker teams. Second, the non-profit orientation as well as the absence of residual claimants in many organizations enables the sport to generate revenues from a large variety of sources. For example, the governance structures are highly attractive for sponsors and donors. As a result, top European clubs have generated significantly higher revenues than their North American counterparts in recent years. Third, the sport is responsible to a wide variety of stakeholders, who are able to voice their interests as members of democratically organized governing bodies. Fourth, institutional arrangements, such as revenue sharing, salary caps, and free agency, have to be considered within the heterogeneous context of the European sport pyramid.

Revenue sharing in European sport leagues

Revenue sharing is less common in Europe than in North America. The home team, for example, usually keeps gate revenues. In most leagues, the league markets television rights collectively and the generated income is distributed according to market size or sportive success. In the German Bundesliga, for example, television rights are marketed by the league and distributed according to each club’s position in the league table, with the top team earning the largest share. In the Champions League, television revenues are also marketed collectively by UEFA and distributed according to a formula, which includes the club’s sportive success as well as the size of the television market of the club’s country of origin. In Spain, the two largest clubs, Real Madrid and FC Barcelona, sell the rights to broadcast their home games individually whereas the other clubs market their television rights collectively. Some leagues, such as the Scottish Premier League and the English Premier League, use so-called parachute payments to help clubs which are relegated to a lower league. In 2011, the Premier League, for example, guarantees relegated clubs annual payments of £12 million for four years. The purpose of these parachute payments is to enable clubs to financially survive relegation to the next division. These parachute payments are a special form of revenue sharing, in which top-division clubs share television revenues with relegated clubs.

In addition to these peculiarities, the main difference between European and North American sport leagues with respect to the effect of revenue sharing on competitive balance results from the clubs’ objective functions and the elasticity of talent supply. To highlight the effect of revenue sharing in an open league with utility-maximizing clubs and elastic talent supply, we follow Dietl et al. (2011b) and use a standard revenue function from the literature (see Szymanski, 2003; Kesenne, 2005, 2007). It should be noted that in an open league with elastic talent supply, we do not have to differentiate between talent investment and actual talent level because $\frac{\partial t}{\partial z_1} = \frac{\partial t}{\partial z_2} = 0$, $\frac{\partial t}{\partial z_1} = \frac{\partial t}{\partial z_2} = 0$, and $\frac{\partial z}{\partial z_1} = \frac{\partial z}{\partial z_2} = 0$. According to Winfree and Fort (2011), these conditions imply that choosing talent investment is the equivalent of choosing talent.7
The revenue of club \( i = 1, 2 \) is then given by

\[
R_i(z_i, z_j) = m_i w_i(z_i, z_j) - \frac{b}{2} w_i(z_i, z_j)^2, \tag{28.13}
\]

where \( b > 0 \) characterizes the effect of competitive balance on club revenues and \( m_i > 0 \) represents the market size or drawing potential of club \( i \). The win percentage \( w_i \) of club \( i \) is characterized by the Tullock contest-success function (CSF), which is the most widely used functional form of a CSF in sporting contests:

\[
w_i(z_i, z_j) = \frac{z_i}{z_i + z_j}.
\]

Equation (28.13) shows that club \( i \)'s revenues initially increase with winning until the maximum is reached for \( w'_i \). By increasing the win percentage above \( w'_i \) club \( i \)'s revenues start to decrease because excessive dominance by one team is detrimental to club revenues. This reflects the uncertainty of outcome hypothesis; the higher \( b \) is, the more important is competitive balance and the sooner revenues start to decrease due to the dominance by one team.

The objective function of club \( i \) is given by a weighted sum of one’s own profits and wins:

\[
u_i(z_i, z_j) = \pi_i(z_i, z_j) + \gamma_i w_i(z_i, z_j)
\]

\[
= (1 - \alpha) R_i(z_i, z_j) + \alpha R_j(z_i, z_j) - z_i + \gamma_i w_i(z_i, z_j)
\tag{28.14}
\]

where \( \gamma_i \geq 0 \) is the “win preference,” which characterizes the weight club owner \( i \) puts on winning in the objective function. A higher parameter \( \gamma_i \) thus reflects that club owner \( i \) becomes more win-oriented and less profit-oriented. Note that two dimensions of heterogeneity exist in the model. On the one hand, clubs differ with respect to their market size, and on the other hand, clubs differ regarding their win preference.

Each club \( i = 1, 2 \) maximizes its objective function \( u_i \), yielding the following first-order conditions:

\[
\frac{\partial u_i(z_i, z_j)}{\partial z_i} = \left( 1 - \alpha \right) \frac{\partial R_i}{\partial w_i} - \alpha \left( \frac{\partial R_j}{\partial w_j} + \gamma_i \right) \frac{\partial w_i}{\partial z_i} - 1
\]

\[
= \left( 1 - \alpha \right) (m_i - b) - \alpha m_j + \frac{bz_j}{z_i + z_j} + \gamma_i \left( \frac{z_j}{z_i + z_j} \right)^2 - 1 = 0
\tag{28.15}
\]

Regarding the effect of revenue sharing on club revenues, the partial derivative of club \( i \)'s marginal after-sharing revenue \( MR_i \), with respect to the revenue-sharing parameter \( \alpha \) is given by

\[
\frac{\partial MR_i}{\partial \alpha} = \frac{z_j}{(z_i + z_j)^2} [b - (m_i + m_j)].
\]

A higher degree of revenue sharing has a positive effect on club \( i \)'s marginal revenue if \( b > m_i + m_j \), while it has a negative effect on marginal revenue if \( b < m_i + m_j \). Revenue sharing has no effect on marginal revenue for \( b = m_i + m_j \).

Dietl et al. (2011b) show that if \( b > m_i + m_j \), more revenue sharing increases the amount of talent hired by each club and produces a more balanced league if the league is not fully balanced in equilibrium. In the case that revenue sharing has a positive effect on marginal revenue for both clubs, it enhances incentives to invest in playing talent. It follows that both clubs will increase the amount of talent hired in equilibrium. Hence, Dietl et al. (2011b) identify a new effect of revenue sharing, called the “sharpening effect.”
In the presence of the sharpening effect, a revenue-sharing arrangement proves to be an efficient instrument for improving competitive balance in an unbalanced league. If the large-market club is the dominant team in equilibrium, then the positive effect of revenue sharing on marginal revenue is stronger for the underdog (i.e., small-market club) than for the dominant team (i.e., large-market club) due to the logit formulation of the CSF. As a consequence, the sharpening effect of revenue sharing is more pronounced for the underdog than for the dominant team, because the marginal impact on the dominant team’s revenues of an increase in talent investment by the underdog is greater than the marginal impact on the underdog’s revenues of an increase in talent investment by the dominant team. As a result, the small-market club will increase its investment level relatively more than the large-market club such that the league becomes more balanced through revenue sharing.

If, however, the small-market club is the dominant team in equilibrium, then the positive effect of revenue sharing on marginal revenue is stronger for the large-market club than for the small-market club. In this case, the sharpening effect of revenue sharing is stronger for the large-market club. Again, the underdog (in this case, the large-market club) will increase its investment level relatively more than the dominant team (in this case, the small-market club) such that the league becomes more balanced through revenue sharing.

If the league is already perfectly balanced (i.e., both clubs have equal playing strength in equilibrium), the (marginal) sharpening effect of revenue sharing is equally strong for both clubs. As a consequence, both clubs will marginally increase their investment level at an equal rate and competitive balance will not be altered through revenue sharing such that the invariance proposition holds.

If $b < m_i + m_j$, more revenue sharing reduces the amount of talent hired by each club and produces a less balanced league. That is, in this case, the well-known dulling effect of revenue sharing is present. Finally, if $b = m_i + m_j$, more revenue sharing has no effect on equilibrium investments and on competitive balance such that the invariance proposition holds.

**Free agency (Bosman ruling)**

Traditionally, employment relations in European professional team sport were not only regulated by employment law and by the contracts between clubs and players, but also by a transfer system which was imposed on all employment relations within a team sport by the governing bodies of the respective sports. In football, for example, the European Football Association (UEFA) restricted the number of foreign players per team and the International Football Association (FIFA) prohibited players from signing for a new club without the consent of their former club. This transfer restriction even applied to out-of-contract players until the so-called Bosman ruling of the European Court of Justice in 1995. Jean-Marc Bosman, a Belgian player, went to court after his former club vetoed his transfer to a French club. The court ruled that this transfer restriction does not comply with the principle of free movement of workers within the European Union. As a result, all European sport governing bodies had to change their transfer systems to comply with EU principles. The Bosman ruling had a similar effect to the Seitz decision in Major League Baseball in the USA, which eliminated the reserve clause and led to free agency in American baseball. In addition, the Bosman ruling also prohibited national sport leagues within Europe from discriminating against players form other EU countries by imposing quotas on foreign players.

Players and clubs reacted to the Bosman ruling by significantly extending the average duration of player contracts (Simmons, 1997). This extension of contract durations can be interpreted as an attempt on both sides to opt back into the old transfer system, which applied
to all in-contract players. In 2001, the European Union reacted to the contract extensions by limiting the maximal contract durations to five years. In what is known as the “Monti system” after European Commissioner Mario Monti, the football governing bodies further had to adapt their regulatory framework known as the FIFA transfer rules to a whole set of new requirements. 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 the link between freedom of movement and market power is clear, it can be questioned whether salaries will ultimately be driven up by the reforms. There may be more than one channel of influence between the reforms and the salaries. For example, Antonioni and Cubbin (2000) analyzed the economic effect of the Bosman ruling and found 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.

Dietl et al. (2008b) looked at the employment relation in football from a different perspective. They developed a model which captures an important and widely overlooked aspect of this employment relation, namely the allocation of risk. 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 appropriate to treat productivity variations as a manifestation of risk. Moreover, on average, the career duration of a professional football player is very short compared with other labor markets. According to Frick et al. (2007), more than one-third of all players “disappear” again after their first season and only one career out of twelve lasts for ten years or 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 was 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.”

De facto labor law in most European countries makes long-term employment contracts asymmetrically incomplete since it is possible to legally bind employers to fulfill 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” that has allowed players to commit to fulfilling their contracts and 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
International sport league comparisons

of contracts, which is exactly what happened in the industry. 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 of risk more difficult. In their model, Dietl et al. (2008b) show that risk-averse players may lose from the reforms because they would benefit from a conversion of risky future income into risk-less current income under the “shadow of the transfer system.”

Payroll caps

Although in the last decade European sport leagues have achieved an economic and financial potential comparable to that of the North American major leagues, it has not yet followed those leagues’ examples of introducing payroll cap mechanisms. Presumably, this reluctance is not caused by the dangers of competitive imbalance and financial instability being unknown among the stakeholders of European sport leagues. Rather, the opposite seems to be the case. For example, the “Independent European Sport Review” (Arnaut, 2006), an expert report based on a process of intensive consultation with the most important stakeholder groups of European football, leaves no doubt that the general perception is that competitive balance in European club football is declining and that a large number of clubs have stumbled into a massive financial crisis and are accumulating ever-increasing debt.

As Dietl et al. (2011a) have shown, the reasons for this reluctance in Europe to introduce payroll cap mechanisms are structural because the labor relations approach employed by the American major leagues is not feasible within the European association-governed pyramid. Sport associations cannot be compared with the team owners in a North American major league, which represent the demand side of the respective labor market. Instead, associations are conceived as democratic governing bodies, which aim to integrate all of the important stakeholders in a certain geographic region including the players and fans. At the European level, the different political and market conditions of every sport nation create additional stakeholder diversity. It follows that decision-making processes concerning the introduction of payroll caps will be much more complicated in the European association-governed pyramid, as the interests of various stakeholders need to be properly balanced.

A payroll cap system would have to take into account the significant market heterogeneity within the European sport leagues, which encompasses all national and pan-European competitions through a system of promotion and relegation. The American system of an absolute capped payroll amount applicable to all clubs is not discussed in the European model because the revenue differentials between clubs of a certain division in different countries are significant. Taking into account that the cost of administering a specific absolute cap for every league in European would be prohibitive, the only workable solution in the European context seems to be a percentage-of-revenue payroll cap.

Unsurprisingly, all discussions among the stakeholders of European sport leagues focus on this relative capping strategy. For example, a small fraction of European football clubs, known as G–14 and established as an interest group of 18 prominent clubs of European football, brought up the issue of salary cost controls in 2004. The members of G–14 had planned to limit their salary expenditures to 70 percent of audited club turnover from the 2005–6 season onwards. At the same time, the minimum allowable amount for total staff costs of each member was set at €30 million. According to the G–14 plan, their statutory auditors should carry out verification of the clubs’ compliance with these principles. However, the G–14 plan has never been put into practice and G–14 dissolved in January 2008, when the new European Club Association (ECA) was founded under the auspices of UEFA.
Representing all stakeholder groups of a particular sport, sport associations perform regulatory functions normally reserved to the state. Because the scope for autonomous regulatory activity by the sport governing bodies is limited by national and EU law, it is a priori unclear whether a particular payroll cap mechanism in Europe falls under the margin of discretion granted to the associations by the European Union. As the previous interferences of EU institutions in the regulatory activities of sports associations show, the sports governing bodies will have to prove that their proposal of a salary control system is doing more than, for example, just improving the financial situation of clubs. It is well recognized that sound club financials play an important role in avoiding incomplete seasons and maintaining the integrity of football. Clubs operating on the verge of bankruptcy are more inclined to engage in illegal practices like, for example, money laundering, match fixing, and tax fraud, which harm the image of the whole industry. However, the history of interventions shows that the EU institutions will assess a salary control system from a much broader social welfare perspective, which is not restricted to the improvement of financial stability alone, but at the same time aims to secure fair treatment of players and consumers.

Dietl et al. (2011a) analyzed the effects of a percentage-of-revenue payroll cap \( \delta \) in an open league with win-maximizing clubs and elastic talent supply. They incorporate the specific European perspective into their model by assuming that a social planner must approve any regulation proposed by the league governing body, taking into account the effect of a payroll cap on all parties in the regulatory scheme: that is, clubs, fans and players. As a result, the objective function of the social planner (social welfare) is given by the sum of player salaries \( PS \) and consumer surplus \( CS \) that can be written as \( W(z_1, \ldots, z_n) = PS(z_1, \ldots, z_n) + CS(z_1, \ldots, z_n) \).

The decision of the league governing body is subjected to approval of the social planner in order to reflect the situation in European football. The social planner will accept a payroll cap proposed by the league governing body only if its introduction does not negatively affect social welfare compared to social welfare \( W^* \) in the benchmark case. The benchmark is an unregulated league, i.e., a league without a payroll cap (\( \delta = 1 \)) and represents the current situation in European soccer, where no payroll caps exist and UEFA demands a balanced budget.

The objective function of the league governing body depends not only on aggregate consumer (fan) surplus, aggregate player salaries and club surplus, but also on aggregate club profits, reflecting the league’s concern for financial sustainability. The integration of club profits is motivated by the growing evidence cited by UEFA of a financial crisis spreading throughout the European football leagues. Many European clubs face serious financial difficulties – some have even gone bankrupt. UEFA has repeatedly argued that sound club finances play an important role in avoiding incomplete seasons and maintaining the integrity of football.

It follows that the league governing body has the same objectives as the social planner but in addition has a concern for financial stability in the league. The objective function of the league governing body is defined by \( L(\delta) = PS(\delta) + CS(\delta) + \gamma \Pi(\delta) = W(\delta) + \gamma \Pi(\delta) \), where \( \gamma \geq 0 \) denotes the weight that the league authority puts on club profits. This weight depends on the financial situation of the league and increases with the degree of financial distress. In the case that the league governing body is concerned with financial stability (i.e. \( \gamma = 0 \)), the objective functions of the league governing body and the social planner coincide. League quality is now defined as \( q(z_1, \ldots, z_n) = \theta \Pi(z_1, \ldots, z_n) + CB(z_1, \ldots, z_n) \), where \( \theta > 0 \) allows the relative importance of the two components of league quality to shift. Thus, the parameter \( \theta \) can be interpreted as reflecting the fans’ relative preference for aggregate talent.

The problem of the league governing body consists of maximizing \( L(\delta) \) under the constraint that social welfare is not lower than in the benchmark \( W^* \) without a payroll cap. Formally, the league governing body solves the maximization problem given by

\[
\max_{\delta \in [\delta^m, 1]} L(\delta) \text{ s.t. } W(\delta) \geq W^*.
\]
By analyzing the constraint maximization problem of the league governing body, Dietl et al. (2011a) derive the following results:

(i) If the fans’ preference for aggregate talent is low with \( \theta \in [\theta^{\text{min}}, \theta^{\text{''}}] \), then no payroll cap will be implemented: that is \( \delta^* = 1 \). The social planner will never approve a payroll cap set by the league governing body if the fans have a relatively low preference for aggregate talent (i.e. \( \theta < \theta^{\text{''}} \)). In this situation, a payroll cap would inevitably lower social welfare because the beneficial impact of the payroll cap on competitive balance will result in a loss in player salaries and potentially in a loss in consumer surplus, as the unrestricted league is already rather balanced. Nevertheless, the league governing body would propose a payroll cap if financial distress is severe enough, that is, the weight \( \gamma \) on aggregate club profits is sufficiently high. However, the social planner will always veto this proposal. That is, even though the league governing body might want to introduce a payroll cap, the social planner will not tolerate this cap.

(ii) If the fans’ preference for talent is sufficiently high with \( \theta > \theta^{\text{''}} \), then a payroll cap will be implemented according to \( \delta^* = \begin{cases} \delta_1^* & \text{if } \theta \in (\theta^{''}, \theta^{'''}) \text{ and } \gamma > \gamma' \smallskip \delta_2^* & \text{otherwise.} \end{cases} \) The proposal of the league governing body to introduce payroll caps can pass the social welfare test if the fans’ preference for talent is sufficiently high (i.e. \( \theta > \theta^{'''}) \). In such a situation, the competitive imbalance in the league is so high that the social planner also favors a payroll cap. If the fans’ preference for talent increases even more and passes another threshold, i.e., \( \theta > \theta^{'''}, \) then the social planner always approves the league governing body’s proposal and a payroll cap \( \delta^* = \delta_2^* \) will be implemented. In this case, the optimal payroll cap from the point of view of the league governing body always lies in the interval of feasible payroll caps that yields a higher social welfare value than in benchmark case. Hence, the objectives of the league governing body and the social planner are sufficiently aligned. The same is true if \( \theta \in (\theta^{''}, \theta^{'''}) \) and the weight attached to club profits is small. However, if the league governing body puts too much emphasis on club profits (i.e. \( \gamma > \gamma' \)), the league governing body wants to implement a payroll cap that would be detrimental from a social welfare perspective, as players would suffer unduly. In this case, the league governing body will only be able to introduce the strictest possible payroll cap that still appeases the social planner, i.e. \( \delta^* = \delta_1^* > \delta_2^* \). Even though the objective function of the league governing body increases, social welfare remains unaltered compared with the benchmark case because consumer surplus increases at the expense of player salaries.

Conclusions

As the theory of sport leagues develops, it becomes increasingly clear over time that Rottenberg’s (1956) invariance principle only holds in precisely derived situations. Moving to any level of generality, even the simple generality of the models in this chapter, makes the invariance principle a matter of the interaction between investment in talent and actual talent accumulation to the level put into play during games. For revenue sharing, it is interesting that the outcomes in North American leagues typically support the invariance principal, suggesting that those very precisely derived conditions may actually hold. It is also the case in North American leagues that the talent distribution appears invariant with respect to drafts and free agency. Of course, much more work remains to be done before such a statement is even close to definitive.
Perhaps the second overall observation is not so much in the competitive balance outcomes associated with choices made in North American or European leagues, but rather in the observation that the two choose such different approaches to essentially the same problem. By allowing a bit of flexibility in terms of objective function, part of the difference is driven by the North American pursuit of profit and the European pursuit of utility (and in some cases wins). However, clearly the overriding source of the difference in these choices is simply the fundamental difference in organizational structure of the leagues themselves. For example, subject to the rigors of collective bargaining, it is straightforward for North American leagues to choose and institute payroll caps. However, give the wide variety of organizational jurisdiction, as well as the openness of European talent markets compared to those in North America, such caps are administratively simply not possible in Europe.

While form and structure must, therefore, invariably differ in how European leagues pursue competitive balance compared to their North American counterparts, there are at least lessons for each to learn from the others.

Notes

1 This means that the clubs do not form binding commitments to one another in their interactions. Further, equilibrium in the league (market) is based on clubs recognizing that gains cannot be made from further unilateral changes in strategy because of the likely response of other clubs. This approach to modeling sport leagues thus recognizes that clubs engage in a game-theoretic way.

2 In the former case one might think of the talent stock being fixed, such as in the US in which the sports are broadly country specific, with limited international migration. In the latter case, one might think of European soccer, which faces an international labour market. The general theoretical principle, however, concerns the availability of talent, and this might vary across sports, and within countries depending on talent availability and production. These issues are addressed much more fully later in the chapter.

3 Remembering that ‘z’ represents the cost of talent for teams through their investment.

4 For discussions about the clubs’ objective function, see Kesenne (2000), Fort and Quirk (2004) and Vrooman (2007, 2008).

5 At the same time, many European leagues have changed their governance structure and have adopted an organizational form similar to their North American counterparts that have been organized since their beginning in a cooperative-like manner. Based on a comparative institutional analysis, Dietl et al. (2009) explain why European leagues have moved away from a contractual towards a cooperative form of governance.

6 Based on a contest model, Dietl et al. (2008a) show that a system of promotion and relegation enhances the incentives of sport clubs to “overinvest” in playing talent.

7 Note that in an open league with profit-maximizing clubs, even if \( \frac{\partial t^i}{\partial z_1} = \frac{\partial t^j}{\partial z_2} \), equation (28.12) would still require \( \frac{\partial w^1_i}{\partial t^i} = \frac{\partial w^2_j}{\partial t^j} \) for the invariance proposition to hold. This would be true for the trivial case where teams are completely balanced in the first place but, generally, the invariance proposition does not hold for open leagues with profit-maximizing clubs.

8 The logit CSF was generally introduced by Tullock (1980) and was subsequently axiomatized by Skaperdas (1996) and Clark and Riis (1998). See Dietl et al. (2008a) and Fort and Winfree (2009) for studies of the CSF’s discriminatory power in sporting contests.

9 As in Rascher (1997) and Kesenne (2007), we refer to this objective function as the utility function of a club. Sloane (1971) was the first to suggest that the owner of a sport club actually maximizes utility, which may include inter alia playing success and profits.

10 Note that the integration of a win preference parameter \( \gamma \) for club \( i \) allows that the case, in which revenue sharing has a positive effect on marginal revenue, is a feasible equilibrium outcome. Without a win preference parameter, the parameter constellation \( b > m_i + m_j \) would not constitute an equilibrium.
11 Note that this “sharpening effect” of revenue sharing has the opposite effect of the “dulling effect.” The dulling effect describes the well-known result in sport economics that revenue sharing reduces the incentive to invest in playing talent (see Szymanski and Késenne, 2004).

12 Note that if clubs maximize wins, club surplus depends on the respective win percentages of the clubs. The clubs’ wins represent a zero-sum game and therefore enter the objective function of the social planner only as a constant.

13 A cornerstone of the recently approved financial fair play concept is the break-even rule. Beginning in the 2012/13 season, clubs will have to balance their books and operate within their financial means. The new obligation for clubs to break even over a period of time means that they cannot repeatedly spend more than their generated revenues. For the first time in European football, clubs that repeatedly spend more than 100 percent of their revenues will be sanctioned.

14 The degree of financial distress could be measured, for instance, by the ratio of total league debt to aggregate league revenues.

15 However, the objective function of the social planner is defined over the talent investments $z$, directly, while the league governing body’s objective function is defined over the values of the policy instrument $\delta$.

References


