This is the final working paper version of an article that appeared in the Journal of the European Economic Association, 2006, 4(5), p. 957-987.

Transfer Fee Regulations and Player Development*

Marko Terviö
Haas School of Business
University of California, Berkeley

September 15, 2005

Abstract

This paper studies the role of transfer fees in professional sports, where players can commit to binding long-term contracts. They cannot switch clubs before their contract expires unless the old club agrees to let them go; transfer fee is the price of that agreement. Transfer fees have been defended as a necessary incentive for clubs to invest into training their young players. The apparent absence of significant training costs (compared to the level of transfer fees) has undermined this defense. We present a model without training where an industry of clubs with heterogeneous marginal revenue products for player ability and a population of players with various levels of talent and experience match. Transfer fees are needed to efficiently allocate scarce playing opportunities among players of different levels of known and potential ability. We show that total surplus is lower without transfer fees because playing time gets reallocated towards older players with less upside potential. The resulting increase in player salaries exceeds the transfer fee costs for each level of ability. (J31, J41, K12, L83)

*Acknowledgements: This paper grew out of chapter 3 of my PhD Thesis (MIT, 2003). I am grateful for useful comments by David Autor, Abhijit Banerjee, Matthew Gentzkow, Bengt Holmström, Michael Katz, David I Levine, Hideo Owan, and seminar participants at Haas, Olin, and ESEM-Madrid. I thank the Yrjö Jahnsson Foundation for financial support.

1 Introduction

The labor market institutions in professional team sports are unusual. The most striking feature is the enforceability of long-term wage contracts, which prevents players from unilaterally terminating their contracts. Even if another club were to offer much better terms to a player than his current club the latter is under no obligation to let the player go before the contract expires. A sufficient payment can of course induce the current club to release a player prematurely, i.e., to "sell" him. Such payments are known as transfer fees.[1] This paper is motivated by the recent process to regulate the transfer fee system in European football (soccer). Our purpose is to build a model to understand the role of transfer fees in the football industry and to use it to analyze the market-level effects of transfer regulations.

The market for football players started attracting much attention in the 1990s thanks to soaring transfer fees and to the landmark court case of Jean-Marc Bosman. In its 1995 ruling, the European Court of Justice ended foreign-player quotas within the European Union and the practice by which clubs could require fees even for players moving at the expiry of a contract. In the wake of the Bosman case it became widely argued that the whole transfer fee system was in breach of the EU labor regulations. The mere idea of trading in people was a cause for a lot of indignation, as is well captured by the statement of Viviane Reding, the EU Sports Commissioner: "I find it scandalous that players are being used as objects of speculation, bought and sold like commodities."[2] The push for regulation was apparently not motivated at any stage by efficiency considerations but by the desire to bring the football industry in line with others. If workers in other industries are not required to honor long-term contracts, why should professional athletes be treated differently?

The football industry defended the transfer fees as necessary for their investment in young players. A quote from Rick Parry, the CEO of Liverpool, sums up the main point of industry leaders: "My great concern is the impact of these proposals on developing young players. How can you protect the investment over the long term?"[3] The only respite for the

[1] These payments totaled £187m for English Premiership clubs in the 2002-03 season (down from a peak of £364m two years earlier). Of this £101m were paid for imports. By comparison, total income of English Premiership clubs was £1.2bn, and total wages £761m. All numbers are from *Deloitte Annual Review of Football Finance (2004).*

[2] Financial Times, August 31, 2000.

[3] Quoted in www.soccernet.com, September 6, 2000.

transfer fee system came from the general acceptance that clubs that train young players should be compensated when their players are "poached" by other, wealthier clubs. The EU Commissioner for Competition Policy, Mario Monti called for an end to "transfer systems based on arbitrarily calculated fees that bear no relation to training costs [...]."[4] Different formulas were suggested for pinning down "a reasonable price" that would replace market-determined transfer fees and presumably decrease their general level. Unsurprisingly, there was much disagreement over the details, especially how differences between countries, clubs, and other factors, should be accounted for.[5] Finally, after years of wrangling, a new EU-approved transfer system went into effect in 2003. The new system still allows for negotiable transfer fees, but in-contract players who are unhappy with their clubs' transfer fee requests can appeal to "an arbitration body with members chosen in equal numbers by players and clubs." Given the widespread unacceptability of high transfer fees, this new mechanism has the potential to put severe limitations on the fee levels—however, the full effect of the arbitration procedure on the future of transfer fees remains to be seen. Furthermore, the length of enforceable contracts was restricted to 3 years (2 years for players over the age of 28), which restricts the compensation for many transfers since the fee is in practice the price of the remaining contract.[6]

The training cost defense of transfer fees is well in line with basic contract theory, much of which studies the problems that arise from worker inability to commit to long-term contracts. Such commitment can motivate employers to provide training when young workers are unable to simply pay for it up-front. Firms will not provide enough general training if trained individuals can quit and take outside offers at their improved post-training market wage. In this sense the labor markets in professional sports are a benign anomaly, and certainly provide an instance of a very interesting institutional arrangement.[7]

The problem with the training cost defense is that the industry has been unable to show training costs that could justify the observed levels of transfer fees; and training costs certainly could never explain the huge variance in transfer fees across players.[8] Furthermore, while players' market value can increase by orders of magnitude over the course of a year or two, the largest increases take place while players are already playing professionally—

[4]Speech given at a Commission-organised conference on sports in Brussels, April 17, 2000.

[5]See the press release by FIFPro on February 26, 2002: "Calculation of Training Costs is Question Mark in New System."

[6]For more details, see the European Commission press release IP/01/314.

[7]For a recent overview of the economics of sports labor markets, see Rosen and Sanderson (2001).

[8]The current record fee of 67 million Euros was paid by Real Madrid for Zinedine Zidane in 2001 for the remaining four years of his contract with Juventus. His salary was speculated to be about 4 million Euros per year at the time (www.footballtransfers.info).

not earlier while they train at youth academies or play for junior teams. If training costs cannot explain transfer fees, then what can?

The most notable economic analyses of the transfer fee system are two papers by Feess and Muehlheusser (2003a, 2003b). Their first paper uses a worker-firm renegotiation model in which the player exerts unverifiable effort and the club invests into his general training, both of which increase the expected value of the player to a potential buyer club. They find that restrictions on contract length and fee levels lead to less training and lower payoff for the selling club, but the effect on player welfare is ambiguous because restricted contract length also leads to higher effort. In the second paper, which doesn't have effort, both the club and the player are unambiguously worse off under capped transfer fees, while a buyer club benefits. Empirical studies on transfer fees have been impeded by a lack of data on contract lengths (which is essentially the "quantity" of what is being sold) and wages. Two notable studies are by Carmichael and Thomas (1993) and Carmichael, Forrest and Simmons (1999), which provide evidence about how transfer fees relate to observable player performance and club characteristics. The first study with data on both wages and contract lengths is Feess, Frick, and Muehlheusser (2004), who find, among other things, that the Bosman ruling led to longer contracts but had no significant impact on post-transfer wages.

For the purposes of analyzing the effects of transfer fee regulations, we take the three crucial features of the industry to be the strong complementarity between talent and club size, an essentially fixed number of jobs per club, and the irreplaceability of on-the-job learning. Due to the complementarity, it makes sense for the best players to play for the clubs and leagues with the most fans. The efficient matching of a cohort of players with the clubs changes as players develop and new information becomes available. The net flow of discovered talent is from the smaller clubs and smaller leagues to the bigger, and so the net flow of transfer fees is to the opposite direction. In practice, the almost fixed number of jobs per clubs means that a single transfer can cause a chain reaction of players being bumped into other clubs, which generates a tremendous amount of turnover between clubs. However, we abstract away from the "horizontal" trades and focus on the net flows of talent and transfer fees that take place between clubs in the big and small leagues.

To focus on the problem of on-the-job learning, we assume in the model that there is no training. A change in a player's market value is due to his development as a player, which is a by-product of getting to play. We model this development as public learning about the talent of a player, but it can also be interpreted as learning-by-doing by the player if talent is defined as the player's uncertain capacity to benefit from learning opportunities. Not just talent but also the opportunities for learning are scarce, due to the scarcity of actual playing time with able co-players and opponents at the professional level.

The matching of clubs and players takes place in a competitive market, where the prices of talent (transfer fees and wages) determine the division of rents between buyers and sellers of talent. Heterogeneity of clubs implies that economic rents are possible for clubs that have higher-value use for talent, i.e., for the "big" clubs. Besides affecting efficiency, the nature of the transfer system can therefore affect the division of these rents both between and within factor owners.

Our first use of the model is to analyze the effects of completely ending the enforcement of long-term contracts. There would be no transfer fees as players could just leave at will (or at a very short notice). While a complete ban on transfer fees is no longer a near-term threat to the industry, it gives a simple illustration of the benefits of the transfer fee system. Furthermore, since long-term wage contracts are not enforceable outside professional team sports, these results are suggestive of potential welfare losses in other labor markets inasmuch as they exhibit complementarities in matching and public learning about talent. Second, we analyze the effects of early termination penalties, the level of which is capped by the regulator. (The difference between termination penalties and a transfer fee cap is that players can leave unilaterally if the new club pays the maximum penalty.) This regulation corresponds to the most salient feature of the new EU-approved system. Finally, we also analyze the effects of integrating labor markets with different club size distributions. This last analysis corresponds to the main effect of the Bosman ruling.

We find that a simple abolition of long-term contracts reduces total surplus produced by the industry and increase the salaries of all player types by more than their corresponding transfer fee cost under the unregulated system, with the cost of the highest star talent increasing the most. The abolition also shows up as an upward shift in the age distribution of professional players. We also find that early termination penalties are essentially a half-way house between abolition and no regulation, but that for moderate levels of the maximum enforceable penalty they are closer to complete abolition. As for the labor market integration, we find that it benefits the biggest clubs in big markets and (at least) the smaller clubs in less lucrative markets, but hurts the smaller clubs in big markets. Integration is also likely to increase the aggregate level of experimentation with new talent.

If it is useful to have tradable long-term contracts in sports, why don't we observe them in other professions? We believe that similar contracts could indeed enhance efficiency in other labor markets, but such contracts have typically not been enforceable—not since the end of indentured servitude. It is probably unpalatable to any modern court to take the side of an employer that wants to hold on to a worker, who has a lucrative outside offer, without even increasing the much lower contract wage. An interesting exceptional case was the

Hollywood studio system in the 1930s, where seven-year wage contracts for new actors and actresses were the norm. Studios did not actually sell the contracts, but loan-outs (at a profit to the owning studio) were common. However, this system unraveled under various court challenges in the 1940s.[9] One market where something like long-term "wage" contracts are enforceable is in the music industry, where record deals with new artists (which are not considered to be employment contracts) usually give the label the option to several future records at a predetermined compensation—so the small minority of artists who turn out to be successful end up earning less than their market value. Somewhat analogously to the football industry, the contracts with star musicians discovered at small labels are often sold to bigger labels that have a comparative advantage in mass marketing and distribution.

So why are there long-term wage contracts in sports? We believe that the main reason why binding long-term contracts have historically been enforceable in professional team sports is the peculiar nature of production, which allows the industry to enforce the contracts without help from courts. It is an industry where no firm (club) can produce on its own, as the production inherently requires cooperation from other firms. Other clubs could punish a poaching club simply by excluding it from the sporting competition. For example, UEFA routinely metes out fines for violations of the transfer rules, the most common breach being the approaching of in-contract players by other clubs. Non-compliance to fines is unheard of. Also, the ability of players is very industry-specific, so poaching from outside the industry is not a problem. Finally, transfer fees are at their most useful when there is large uncertainty about the ability of workers, and that there is aplenty in sports.

In the U.S., the exceptional ability of sports leagues to enforce long-term wage contracts (and to engage in some clearly anti-competitive practices, such as salary caps) is supported by several U.S. Supreme Court rulings.[10] It is worth pointing out that the labor market institutions in European sports are significantly different. In Europe, young players start as free agents in what is a very competitive market but are able to agree to binding long-term contracts. In North America, each major sport operates as a closed league that allocates between clubs the exclusive right to negotiate with each potential new player (using the so-called "draft" system). In response to the monopsonist leagues, players have organized and gained, for example, in baseball the right to free agency after 6 years in the league. It could be argued that North American professional athletes are to some extent "exploited" by the leagues, but similar claims do not carry over to Europe.[11]

[9]For a discussion of the rise and fall of the studio system, see Chapter 5 in Caves (2000).

[10]Baseball was granted immunity from the Sherman Antitrust Act in Federal Baseball Club v. National League (1922). See Fort and Quirk (1992).

[11]For an opposing view, see Sanderson and Siegfried (1997).

The outline of the paper is as follows. The basic model is presented in Section 2, where the outcome of the market—efficiency, turnover, and the distribution of profits and wages—is first analyzed under long-term contracts with completely market-determined transfer fees. The effects of ending long-term contracts and of early termination penalties are then analyzed in sections 3 and 4 and compared with the benchmark of unrestricted transfer fees. In Section 5 we analyze the comparative statics of integrating two markets with different club size distributions. We discuss the results and their limitations in Section 6 and conclude with a modest recommendation for the football industry.

2 The Model

The model combines two major features: matching and experimentation. There is a unit measure of firms (clubs) that each hires one worker (player) per period. Players who have never been hired before are "novices" and have unknown talent; players that have been hired before are "veterans" and their talent level is known. Each club faces a decision: whether to hire a novice or an experienced player. The equilibrating variable in the model is the fraction of clubs choosing to hire novice players.

Assumptions

1. The revenue y generated by a player of talent a at a club of size b is given by $y = ab$.
2. Every period a unit measure of potential players are born, with talent distributed according to a continuous and strictly increasing profile, $\theta[i]$, $i \in [0, 1]$.
3. Player careers last up to two periods: the talent level of a novice is unknown, but becomes public knowledge after one period of employment in the industry.
4. Players maximize their income, but cannot work for less than a reservation wage, which is normalized at zero.
5. There is a unit measure of clubs, each employing one player.
6. The club size distribution is described by a continuous and strictly increasing profile $b[i]$, $i \in [0, 1]$.
7. Clubs are infinitely lived, risk neutral, and maximize long-run average profits.
8. A new player would produce enough revenue in expectation at any club to cover his outside opportunity: $b[0] \int_0^1 \theta[j] \mathrm{d}j \geq 0$.

For the purposes of this paper, talent is defined as the capacity to generate revenue, whether through sales of tickets, television rights, or merchandise. We assume that the audience values player talent for its entertainment value on the quality of the game and not just for its effect on winning.[12] If, to the contrary, audiences only cared about seeing their favorite club win, then any investment into player development would be socially wasteful.[13]

It is assumed that clubs are inherently heterogeneous by their productivity for player talent, referred to as "club size" for brevity. The main component of club size is the number of potential viewers (which includes supporters of other clubs in the same league, to the extent that there is revenue-sharing).[14] What is the exogenous component in club size that is the source of economic rents for clubs and cannot be competed away? It is possible to enter into being a professional football club in Manchester—by starting from the lowest-tier division—but not into being Manchester United, currently the biggest football club by market value.[15] Fan loyalty acts as a very high switching cost and as a source of rents for a club with a large fan base. A club brand name with a glorious history and with a home in a large city is a unique asset that can earn rents. These rents may have been dissipated in expectation back when it was decided which club gets to occupy that lucrative niche, but that is inconsequential for the contemporaneous division of rents between clubs and players.

The model will be presented in three steps. First, the division of rents is solved, while taking the amount of experimentation and therefore the distribution of talent as fixed. Second, the supply of known talent, for a given amount of experimentation—for a given fraction of jobs held by novices—is solved. And third, these parts are tied together by solving for the stationary equilibrium fraction of novices in the benchmark case of unrestricted transfer fees.

2.1 Matching and the Division of Rents

The significance of Assumption 1 is that, of any two players, the more talented one would generate more revenue at any club, but this difference is larger the bigger the club.[16] The

[12] Szymanski (2001) finds that "match attendance appears unrelated to competitive balance" in English league football.

[13] A slightly less extreme version of this view of talent markets is developed in Frank and Cook (1995).

[14] Falconieri, Sákovics and Palomino (2004) analyze the implications of within-league sharing of broadcast revenue, which is the other major regulatory issue in European football.

[15] The market capitalization of Manchester United was £700 million in March 2005.

[16] For example, suppose b is the number of a club's supporters; talent a would then be defined as the average revenue per supporter.

crucial consequence of this complementarity is that the efficient matching of clubs and players is positively assortative: the highest talents should be matched with the biggest clubs. The prices are determined in a competitive market where buyers (clubs) and sellers (clubs or players) meet under symmetric information. The price of talent can consist of a wage, a transfer fee, or both, depending on who owns the (contractual rights to) talent. In either case, equilibrium prices must support the efficient matching. The assumptions of a continuum of clubs and continuous distributions of talent and club size guarantee that the prices are unique, i.e., there will be no match-specific rents left for bargaining. With these assumptions the static setup is basically an assignment model, such as presented in Sattinger (1993).

In this section we take it as given that the clubs obtaining their talent from the matching market are those in quantiles $i \in [z, 1]$. The profiles of talent and club size are denoted by $a[i]$ and $b[i]$ respectively, where $i \in [z, 1]$. The equilibrium price for talent $a[i]$ is denoted by $p[i]$, it is paid by club i to the owner of contractual rights to talent $a[i]$. (The price of talent would not include the reservation wage, as it is the cost of labor that must be paid to any player regardless of the level and contractual status of talent.) For this section, both the profile of talent and the resulting prices are denoted as if they were independent of the proportion of novice-hiring clubs z, although their shapes depend on z.

The equilibrium consists of prices of talent at which no club can lower its offer to its efficient match without losing him to another club and no club would like to hire another club's match at their equilibrium price. Neither buyers nor sellers can gain by making any other offers to anyone else besides the equilibrium offer to their efficient match. The condition for all clubs to want to stick to their own efficient match is

$$a[i]b[i] - p[i] \geq a[j]b[i] - p[j] \qquad \forall i, j \in [z, 1]. \tag{1}$$

Furthermore, the firms have to at least break even and the sellers must get a nonnegative price.

$$\begin{aligned} a[i]b[i] - p[i] &\geq 0 \quad \forall i \in [z, 1] & (2) \\ p[i] &\geq 0 \quad \forall i \in [z, 1] & (3) \end{aligned}$$

Inequalities (1) and (2) are mathematically analogous to incentive compatibility and participation constraints in a nonlinear pricing problem with quasi-linear utility functions and "types" $b[i]$. The prices that simultaneously fulfill the above criteria for all buyers and sellers can be found using the constraint reduction method familiar from nonlinear pricing problems. The binding constraints are those that prevent firms from wanting to hire the next lowest talent. These binding constraints define the slope of the price profile.[17]

$$p'[i] = a'[i]b[i] \tag{5}$$

Finally, by integrating the slope of the price profile, we get the equilibrium prices for talent.

$$p[i] = \int_z^i a'[j]b[j]\mathrm{d}j, \quad i \in [z, 1]. \tag{6}$$

The intercept $p[z] = 0$ is taken as given for now, but it will result in equilibrium from there being more potential players than clubs: the lowest talent hired in the matching market is as good as the highest discarded known talent so it cannot receive any rent. Thus buyers capture all of the rent at the bottom, $\pi[z] = a[z]b[z]$. The buyer's share of the rents is easily recovered from equation (6) as the leftover $\pi[i] = a[i]b[i] - p[i] = \pi[z] + \int_z^i a[j]b'[j]\mathrm{d}j$. However, buyers are not residual claimants here, the equilibrium could equally well have been derived starting from sellers' constraints.

The level and dispersion of rents depend on the dispersion of talent levels and club size.[18] The division of rents at any quantile i depends on the whole distributions of talent and club size below. Mathematically, the price of a talent of level $a[i]$ in (6) is a weighted sum of the "increments" in talent between $a[i]$ and $a[0]$, where the weights are the sizes of the clubs matched at each increment. Intuitively, it is best to have one's own competitors to be of low productivity, and to have one's equilibrium match have to compete with many nearly-as-good competitors.

As is typical in matching markets, the price of talent is not equal to the marginal productivity of talent in the usual sense. This is because a marginal increase in the talent level of a player is not the economically relevant margin, as talent cannot be moved between individuals. The right marginal question to be asked is, how would the total surplus created by the whole industry change if a player of a certain talent level were to disappear from the industry, taking into account the resulting change in the matching of remaining players and clubs in the industry? By this definition, the marginal productivity of a player is indeed equal to the corresponding price of talent.

---

[17] Regrouping the IC constraint (1) for $j = i - \varepsilon$ and dividing it by ε gives

$$\frac{p[i] - p[i-\varepsilon]}{\varepsilon} \leq \frac{(a[i]b[i] - a[i-\varepsilon])\, b[i]}{\varepsilon}. \tag{4}$$

This holds as an equality as $\varepsilon \to 0$ and, via the definition of the derivative, yields the slope of the price profile.

[18] If clubs were homogeneous, so that $b[i] \equiv \bar{b} > 0$, then rents to talent would simply be Ricardian rents: $p[i] = (a[i] - a[0])\, \bar{b}$. But then there wouldn't ever be a need to actually trade anyone.

2.2 Experimentation and the Supply of Talent

The distribution of talent in the population of potential players is fixed, but the distribution of talent in the industry depends on how many novices were hired in the previous period. In steady state some proportion z of jobs are filled with novices and the remaining $1 - z$ jobs with veteran players. Players who turn out to be above some threshold a^* "make the grade" and get to stay in the industry as veterans, while those below exit after one period. Each feasible proportion of novices $\left(z \geq \frac{1}{2}\right)$ corresponds to a different threshold level of talent a^*.

There can be no known types below population mean working in the industry since new players are always available at the lowest possible wage. Therefore novices must comprise the lowest z types in the industry by expected talent, and they are all in expectation of the mean type $\bar{\theta}$. Since there is a measure $1 - z$ of veterans who are the best of the last period's cohort of novices, they are the top $\frac{1-z}{z}$ proportion of talent in their cohort. The threshold type must therefore be the $\left(1 - \frac{1-z}{z}\right)$th quantile of the population distribution, giving the relation of the thresholds as

$$a^*(z) = \theta[2 - \frac{1}{z}]. \tag{7}$$

The talent profile of veteran players comes from spreading the truncated distribution of talents above $a^*(z)$ among the top $(1 - z)$ quantiles. The complete profile of (expected) talent in the industry combines the profile of veteran talent at the top with the novices at the bottom.

$$a[i|z] = \begin{cases} \overline{\theta} & i \in [0, z] \\ \theta[1 - \frac{1-i}{z}] & i \in (z, 1] \end{cases} \tag{8}$$

Players in $[0, z]$ are actually random draws from the whole distribution, but since firms are risk neutral they can be treated as the mean type $\overline{\theta}$.

2.3 Equilibrium with Unrestricted Transfer Fees

Because talent is complementary with club size, and novices are the least talented players by expectation to work in the industry, it must be the small clubs that match with them in equilibrium. We are looking for an equilibrium z at which all clubs in $(z, 1]$ prefer to be buyers, all clubs in $[0, z)$ prefer to be sellers, and the threshold club z is indifferent between being a buyer or a seller.

Buyer profits Buyer profits consist of the revenue generated by their match minus the cost of the corresponding transfer fee and the cost of labor. The transfer fees are the prices of talent, as determined in section 2.1 but with the profile of talent $a[i|z]$ now dependent on the endogenous proportion of novices z. Given z, the transfer fee paid by a club $i \in (z, 1]$ for its match, a player of talent $a[i|z]$, is

$$p[i|z] = \int_z^i a'[j|z]b[j]\mathrm{d}j \quad \text{for } i \in (z, 1]. \tag{9}$$

By the definition of equilibrium prices, buyer i would not want to deviate and buy any other talent besides $a[i|z]$, including the threshold talent $a^*(z) = a[z|z]$, which is available at zero transfer fee. Buyer profits are thus

$$\pi^B[i|z] = a[i|z]b[i] - p[i|z], \; i \in (z, 1]. \tag{10}$$

Seller profits Seller profits consist of the revenue generated by a novice player while playing at the club, and from the expected transfer fee revenue from players that turn out to be good enough to be sold. The total amount of transfer fees paid (per period) in the industry are

$$P(z) = \int_z^1 p[i|z]\mathrm{d}i = \int_z^1 \int_z^i a'[j|z]b[j]\mathrm{d}j\mathrm{d}i = \int_z^1 (1-i)\, a'[i|z]b[i]\mathrm{d}i, \tag{11}$$

where the last step involves a partial integration. The long-run average profits of a novice-hiring club i are then

$$\pi^S[i|z] = \overline{\theta}b[i] + \frac{1}{z}P(z). \tag{12}$$

Note that when players can commit to long-term contracts, novices agree to do so at their reservation wage. They cannot do any better since there are more potential novices then there are jobs. Thus when a seller clubs discovers a high talent and sells his remaining contract it gets the full talent rent as the transfer fee.

The solution The threshold club z must be indifferent between employing novices or veterans so the equilibrium is defined by $\pi^B[z|z] = \pi^S[z|z]$. This condition is simplified by noting that $a[z|z] = a^*(z)$ and $p[z|z] = 0$. Rearranging this equilibrium condition we get

$$\begin{aligned} & a^*(z^*)b[z^*] = \overline{\theta}b[z^*] + \frac{1}{z^*}P(z^*) \\ \Longrightarrow \quad & \left(a^*(z^*) - \overline{\theta}\right) b[z^*] = \frac{1}{z^*}P(z^*). \end{aligned} \tag{13}$$

Note that the equilibrium rehiring threshold is strictly above the population average: small clubs sacrifice some current revenue in exchange for expected transfer fees in the future. The equilibrium is unique because the left side is strictly increasing in z and changes sign within $(0, 1)$, whereas the right side is positive, strictly decreasing, and reaches zero at $z = 1$. (The proof of uniqueness is in the Appendix). Higher z means that more firms are trying to sell talent to fewer buyers, so it is intuitive that the expected price goes down.

The system with unrestricted transfer fees results in the maximization of total surplus, because we have assumed perfect competition and no externalities. As an aside, observe that transfer fees would not be needed for efficiency if novices could pay to play. (This is the simplest solution to all worker commitment problems: the worker "posts a bond.") Risk neutral novices with unconstrained credit would be willing to pay for the opportunity to play up to the expected value of second period talent rents. In equilibrium the novice wage would then be

$$w_1^* = -\frac{P(z^*)}{z^*} = -\left(a^*(z) - \overline{\theta}\right) b[z^*]. \tag{14}$$

The novices would in effect have to outbid veteran players of below $a^*(z^*)$ talent for jobs, which could be very costly if the difference between mean and threshold talent is worth a lot of revenue at the threshold club. The veteran wage would be $p[i|z^*]$, and this payoff could be very risky. For skewed distributions of talent and club size most of the expected rents come from a small chance of becoming a superstar, so even moderately risk averse novices with access to unconstrained credit would be willing to pay only a small fraction of the expected rents.

What information is reflected in the transfer fees? Both sides of the equilibrium condition (13) are equal to the expected transfer fee generated by a novice. On the left is the opportunity cost of hiring a novice: it is the near-term lost revenue from hiring a novice as opposed to the threshold talent (who would be available at zero transfer fee). On the right side is the expected benefit to the industry: having one more novice playing at a small club today increases the expected revenue generated at the big clubs in the future.

At the threshold club z the opportunity cost of hiring a novice is just equal to the expected surplus at bigger clubs in the future, but inframarginal clubs enjoy gains from trade. The level of transfer fees depends on these gains from trade: it makes economic sense for smaller clubs to do the experimentation and for big clubs to use high and proven talent. The economic cost to small clubs from hiring novices is not likely to show up in their accounts because it is a pure opportunity cost.

3 Ending Long-Term Contracts

The ending of long-term contracts would in practice amount to an abolition of transfer fees, because without long-term commitment talented players can be "poached" at will by other clubs offering higher wages. Legally this would of course just mean that the market for football players would be brought in line with other labor markets.

In the absence of transfer fees, a novice-hiring club only gets revenue from its player's current output. In terms of the market equilibrium condition (13) , the right side is replaced by zero, so that

$$\begin{aligned} \left(a^*\left(z\right)-\bar{\theta}\right)b[z] &= 0 \\ \Rightarrow a^*\left(z^0\right) &= \bar{\theta} \end{aligned} \tag{15}$$

defines the new equilibrium. In other words, any player who turns out to be better than a novice by expectation will be hired again. It is still true that the bigger clubs hire known talent and smaller clubs hire novices, since bigger clubs have higher willingness to pay for talent. The rents to veteran talent now accrue to the players and not to the clubs that discovered them. Notice that the solution z^0 is independent of the club size profile, reflecting the fact that gains from trade between clubs are not taken into account.

Proposition 1 *The ending of long-term contracts causes the proportion of novices in the industry to be decreased, and the talent levels to decrease at big clubs and to increase at a segment of mid-size clubs.*

Proof. First note from (7) that $a^*\left(z\right) = a[z|z]$ is an increasing function. Therefore $z^0 < z^*$ and $a^*\left(z^*\right) > a^*\left(z^0\right)$. Then recall from (8) that $a[i|z] = \theta[1-\frac{1-i}{z}]$ for $i \in (z, 1]$. This is clearly increasing in z for $i \in (z^*, 1)$. Thus, we have $a[i|z^*] > a[i|z^0]$ for $i \in (z^*, 1)$, so talent levels go down at the biggest clubs. As for the midsection of clubs, $i \in (z^0, z^*)$, using (8) we see that $a[i|z^0] > \bar{\theta} = a[i|z^*]$.

In other words, since a larger fraction of novices get to stay in the industry as veterans, the proportion of jobs held by novices must now be smaller. That fewer novices are hired is intuitive since it is made less profitable by the elimination of long-term contracts. Since veterans are yesterday's novices, big clubs face a reduced supply of known talents of any given level; as a result they have to make do with lower quality matches then before. On the other hand, a section of medium-productivity jobs will now be filled with veterans of known talent $\theta \in [\bar{\theta}, a^*\left(z^*\right)]$ instead of novices. These "mediocre" types are more talented than novices by expectation, but would not be employed under the transfer fee system. Figure 1 shows the profile of talent in the industry with and without transfer fees, depicted by light and dark gray lines respectively.

Proposition 2 *Distribution of realized talent levels in the industry becomes less extreme: there are fewer very good and very bad and more mediocre players.*

More exactly, there are fewer players with actual talent levels $\theta > a^*(z^*)$ and $\theta < \bar{\theta}$, and more players with talent levels between $\bar{\theta}$ and $a^*(z^*)$.

Proof. For either $z \in \{z^*, z^0\}$, there are $zF(\bar{\theta})$ "bad" players $\theta < \bar{\theta}$ and $2z(1 - F(a^*(z^*)))$ "good players" $\theta > a^*(z^*)$ in the industry; the rest are "mediocre" players. These take into account that good types exist both within the novice and the veteran cohorts, while the bad types only exist among the novice players. The result follows from $z^* > z^0$.

Proposition 3 *Total surplus is decreased, and players gain less in wages than clubs lose in profits.*

The reduction in efficiency caused by what is in effect a price ceiling at zero follows directly from the assumption of perfect competition and lack of externalities.[19] The increase in talent at mid-size clubs is not enough to compensate the decrease in talent at the big clubs, where talent is more productive. Player wages must go up because the price of known talent now accrues to the veterans (and is given by $p[i|z^0]$) instead of the clubs that initially hired them. As total surplus is lower and players gain, the clubs must be losing more than the players gain.

Proposition 4 *Price of talent is increased for every level of talent, and expected profits are lower for all clubs.*

The proof is in the Appendix. Intuitively, the price of talent is increased because the supply of known talent is decreased. This clearly makes big clubs worse off than under transfer fees. The smallest clubs are also worse off as they don't get any more transfer fee income. As for the middle clubs that switch from being novice-hiring clubs to hiring mediocre veterans, even though they are now matching with better players then before, they too must be worse off because before they would have been able to hire someone better than their current match—a threshold talent $a^*(z^*)$—at zero fee.

It is also worth noting that the end of long-term contracts means longer careers on average. As the required level of talent that players have to show to stay in the industry goes down, it becomes easier for those players who get their initial chance to enter the industry to stay in. Naturally, the amount of turnover in and out of the industry is increasing in the rehiring threshold.

[19]The equilibrium condition (13) is also the first-order condition for maximizing total surplus $Y(z) = \int_0^1 a[i|z]b[i]\mathrm{d}i$.

To sum up, the elimination of transfer fees would result in fewer highly talented players and more mediocre players in professional leagues. Due to the complementarity in production, the best players still move up and play in the biggest clubs, but now the clubs that "discovered" them will get no compensation. As before, players that turn out to be below average will not be rehired because novices are abundant and more talented by expectation. The problem is that now all players that turn out to be better than novices in expectation, no matter by how little, will be hired by some club. All clubs are worse off as a result, while salaries of players (other than novices) increase by more than the corresponding transfer fee cost of similar talents before. However, in total, players gain less than the clubs lose.

4 Early Termination Penalties

Early termination penalty is a payment that clubs can demand from players, or in effect from their new clubs, if they leave for another club while the contract is still in effect. The maximum allowable penalty is determined by the regulator (or courts). If the market price of a player turns out to be lower than the penalty, then the owner of the contract can accept to release the player at a discount. The difference between termination penalties and a transfer fee cap is that clubs are not able to retain players for whom someone is willing to pay the maximum penalty. Efficient matching is not disturbed as buyers will compete for the talents with full penalty cost by wage offers, but the incentives to hire novices are of course diminished by any cap that is sometimes binding.

Now let's solve for the market equilibrium when the maximum penalty set by the regulator is $\hat{p}$. The equilibrium is again defined by a fraction of novice-hiring clubs z and a corresponding rehiring threshold level of talent a^*. The price of talent is still determined on a matching market where competing clubs can make offers, however now if the price exceeds $\hat{p}$ then the excess will go to the player (who in the end decides which club's offer to accept). Players who are good enough to be transferred but whose market price is less than $\hat{p}$ get a discount from the full penalty but no rents for themselves.

Recall from talent rent equation (9) that the market price for talent depends on the supply of known talent z. The only difference to the equilibrium condition (13) is that total transfer payments are now only $\hat{P}(z|\hat{p}) \equiv \int_z^1 \min\{p[i|z], \hat{p}\}\mathrm{d}i$, which is clearly continuously increasing in $\hat{p}$. The equilibrium condition becomes

$$\left(a^*(z) - \overline{\theta}\right) b[z] = \frac{1}{z}\hat{P}(z|\hat{p}). \tag{16}$$

At one extreme case of $\hat{p} = 0$ this is the same as equilibrium without transfer fees, and we get $a^* = \overline{\theta}$. At the other extreme, if the maximum penalty is higher than the highest transfer fee without restrictions then the penalty has no effect, because $\hat{P}(z^*|\hat{p}) = P(z^*)$ for $\hat{p} \geq p[1|z^*]$.

Proposition 5 *The proportion of novices, the hiring threshold, and total surplus are increasing in the maximum penalty.*

The proof is in the Appendix. Intuitively, as a binding cap is relaxed, hiring novices becomes more attractive and some clubs switch into novice-hiring. As a direct consequence of more novices being hired, the talent levels at veteran-hiring clubs increase, at the cost of lower talent at formerly poaching clubs. This trade-off is illustrated in Figure 1 as the middle case. By varying the level of the maximum penalty, the market outcome—fraction of novices, total surplus, profits, wages, and prices of each level of talent—varies continuously between the extreme cases of complete abolition of transfer fees and unrestricted transfer fees.

5 The Bosman Ruling

Jean-Marc Bosman was a Belgian 2nd division player, whose club had prevented him from moving to a French club at the end of his contract as the two clubs could not agree on a transfer fee. The ensuing case in the European Court of Justice ended with a ruling in favor of Bosman in 1995 that had two effects: first, clubs were prevented from collecting a transfer fee at the end of the contract, and second, leagues were prevented from imposing quotas on foreign players from other countries within the European Economic Area.

Despite complaints by the industry at the time, there are good reasons to believe that the long-run effect of free transfers at the end of the contract should have been benign or insignificant. Inasmuch as clubs may have been unable to commit to promises of letting their players go free at the end of the contract, the pre-Bosman system actually limited the set of available contracts. Such limitations on parties' ability to contract tend to reduce welfare, although caveats due to externalities or asymmetric information apply in some settings.[20] The crucial feature of the Bosman ruling was that it left the parties with the ability to agree on the length of the contract, during which the current club is able to demand a transfer fee. If it were inefficient, other things equal, to let players go free at the end of a contract of given length, then it would be in the parties' interest to sign longer contracts to begin with—and there is indeed evidence that contracts became longer post-Bosman. We believe that the economic factors behind the determination of contract lengths are quite

[20]See, e.g., Aghion and Bolton (1987) and Aghion and Hermalin (1990).

different from those considered in our model, and have been well analyzed by previous literature.[21]

The other and arguably more sweeping effect of the Bosman ruling was the integration of the European labor market for football players. The European Commission had declared as early as 1987 that foreign quotas for EU nationals were in breach of the free movement of labor guaranteed by the Treaty of Rome.[22] Traditionally, national associations had been free to set their own restrictions on the number of foreign players (at the time the quota was two players in all member countries, except three in Belgium). Finally, in 1992, the governing body of European football associations, UEFA, responded by setting the quota at three players, although the member associations were allowed to choose more liberal quotas. The Bosman ruling opened the floodgates. In the English Premier League (EPL), the biggest league by revenue, there were over 6 foreign players per club in 1999, and by 2005 the average was above 12. A famous culmination point of this integration came in 1999 when, for the first time, an EPL club fielded an all-foreign starting line-up.[23]

As with any removal of trade barriers, an increase in efficiency is an immediate consequence. Also, it is quite obvious that the integration of labor markets will lead to movements of high-quality players from smaller footballing nations to the bigger leagues. What is less clear is how trade will affect the aggregate level of experimentation, and who are the resulting winners and (possible) losers.

To study the effects of labor market integration under a system of unrestricted transfer fees we need to make some simplifying assumptions.[24] In particular, we assume that there are initially two completely isolated markets, "small" and "large," with the same population distribution of talent. The smallest clubs in both markets are of the same size, but the biggest clubs are bigger in the large market. This stylized description of market size seems reasonable: the biggest clubs in small or mid-size markets like Finland or Norway would only pass as relatively small or mid-range clubs in a big market like England or Spain; yet even the largest markets have a whole range of clubs down to the lower leagues. We also assume that there is a unit measure of clubs in both markets. This is not an essential assumption, but neither is it unreasonable, as the integration took place between a few big countries and many smaller countries.

---

[21] See Feess and Muehlheusser (2003b) and Feess, Frick, and Muehlheusser (2004), and also Antonioni and Cubbin (2000).

[22] European Comission Press Release IP/87/261.

[23] Chelsea on December 26. On February 14, 2005, Arsenal became the first team to not include any domestic players even on the bench.

[24] As shown in Section 3, in the absence of long-term contracts the level of experimentation is unaffected by the club size distribution and therefore would not react to the integration.

Pre-Integration Denote the profile of club size in the small and large market respectively by $b[i]$ and $b[i]\rho(i)$, where $\rho(0) = 1$ and $\rho'(i) > 0$ for $i \in (0, 1]$. So we assume that the size profile of clubs begins at the same level for both, but is everywhere steeper in the large market. This definition of "larger" implies first order stochastic dominance, but not vice versa.

Proposition 6 *In autarky, the large market has a higher fraction of novice players then the small market.*

Proof. We need to prove that $z_L^* > z_S^*$. Recall the equilibrium condition (13) and use the formula for the average transfer fee in (11) to define the equilibrium z in the large market:

$$\left(a^*(z) - \overline{\theta}\right) b[z]\rho\left(z\right) = \frac{1}{z} \int_z^1 \left(1 - i\right) a'[i|z]b[i]\rho\left(i\right) \mathrm{d}i. \tag{17}$$

Now divide both sides by $\rho\left(z\right)$, so that the LHS is exactly the same as in the definition of z_S^*, but the integrand on the RHS, which is positive, is multiplied by $\rho\left(i\right)/\rho\left(z\right)$, which is greater than one for all $i \in (z, 1]$. Thus, evaluated at z_S^*, the LHS is less than the RHS. The LHS is increasing in z, and the RHS decreasing, by the same arguments that showed the uniqueness of the equilibrium z in (13). Thus, for LHS to equal RHS, it must be the case that $z_L^* > z_S^*$.

It is harder for players to make the grade in the larger market. The intuition is that the value of finding high talents is increasing in the size of the biggest clubs, while the opportunity cost of experimentation depends on the size of the marginal novice-hiring club. Loosely, an increase in the ratio of the sizes of big clubs to small clubs will increase the economic value of finding high talents relative to the opportunity cost of experimentation at a given z; hence z must increase for equilibrium to hold. By contrast, a constant ratio, $\rho\left(i\right) \equiv \bar{\rho}$, would just cancel out in (17).

Post-Integration After integration, there will be a common distribution of club size, and the distribution of player talent is determined by the fraction of novices in the integrated market. To keep the proofs of our last two propositions manageable, we now make a further distributional assumption.[25]

Assumption. *All distributions are uniform. The minimum club size is unity in both markets, but the maximum club size is larger in the large market.*

[25] Unfortunately we don't have a proof for general distributions. We believe that power law distributions would be the most realistic here, and indeed the same results hold there but with much messier calculations (they are available upon request).

Proposition 7 *The integrated market has a higher fraction of novice players than the small market. It has a higher (lower) fraction of novice players than the large market if the biggest clubs in the large market are (in)sufficiently large compared to the biggest clubs in the small market.*

The proof is in the Appendix. The comparison between the integrated market and the small market is qualitatively the same as the comparison of the large and small market in Proposition 6. However, the comparison between the large and the integrated market is fundamentally different. Both the minimum and maximum club size are the same in both markets, but in between the size profile of the integrated market must be below that of the large market. (The profiles are illustrated in Figure 2.) The idea of the proof is to construct a ratio of integrated-to-large clubs sizes, $\gamma(i)$; this is decreasing at the lower quantiles and increasing at the upper quantiles. If the increasing part of $\gamma(i)$ begins early enough then the argument in the proof of Proposition 6 works, as the weighting $\gamma(i)/\gamma(z)$ is above one for all $i > z$. If $\gamma(i)$ begins its ascent sufficiently late—which happens if the biggest clubs in the two markets are not very different by size—than the weights will be mostly below one and the opposite is true.

Again, what matters is the size of the marginal seller (novice-hiring) club relative to the buyer clubs. The optimal level of experimentation in any market is, very roughly, increasing in the relative heterogeneity of clubs by size. Unless the club size distributions in the two markets are very similar, then the clubs in the integrated market are more heterogeneous in this sense than those in the large market, resulting in a higher optimal level of experimentation.

Proposition 8 *Integration benefits the smallest clubs in the small market and the biggest clubs in the large market, and hurts the smallest clubs in the large market.*

The proof is in the Appendix. The smallest clubs in both the small and large market hire novices both before and after integration. The revenue they generate from output is unchanged but the expected transfer fee is increasing in the size of veteran-hiring clubs. Loosely speaking, there is a higher fraction of big clubs in the large market than in the integrated market, so the expected transfer fee is higher in the autarkic large market than in the integrated market, where in turn it is higher than in the autarkic small market. Therefore integration is good news for novice-hiring clubs in the small market, but bad news for those in the large market. The biggest clubs in the large market are sure to benefit, because they face relatively fewer close competitors in the integrated market.

The intuition about winners and losers is in the spirit of trade theory a la Hecksher-Ohlin-Samuelson: opening up trade causes the income to a country's relatively abundant factor to increase. To be sure, the setup here is quite different: production takes place in a matching setup and there is only one final good, which is jointly produced with its own intermediate good (information about talent). But the point is that the small market is more abundant in small clubs, for whom integration results in an increase in demand for the talent they discover; similarly, the big clubs in the large market will face a relatively more abundant supply of high talent. For small clubs in the large market, integration brings more in way of competition in the seller side of the transfer market than in terms of demand. The biggest clubs in the small market may be better or worse off than before integration, depending on how large the gains from trade are.

The above exercise is related to two trade theory papers. In Grossman and Maggi (2000), two tradable goods are produced with different matching technologies, but the factors cannot migrate across borders. They show how differences between countries in the levels of dispersion of ability give rise to gains from trade. In Antràs, Garicano and Rossi-Hansberg (2005), one good is produced by matching managers with workers, one country ("North") has a better distribution of ability, and the decision of becoming a manager or a workers is endogenous. They find that after it becomes possible for managers to match with foreign workers, the best workers in North become managers, and the worst managers in South become workers. As a result, the lowest ability workers in one country must lose out from trade, but in which country depends on a parameter of the production function ("communication costs").

6 Discussion

We have analyzed the effects of transfer fee regulations with a model that emphasizes the role of on-the-job learning and the complementarity of club size with ex-ante unknown player ability. In the model, individual clubs face a decision of whether to hire novices or experienced players. The industry as a whole faces the decision of how much to use scarce playing time in developing new talent, and how to allocate players of different levels of talent and experience between the clubs. In this setup the market price of talent, as expressed in transfer fees, conveys crucial information about the marginal opportunity costs of experimentation and the marginal value for discovering more talent.

We found that the ending of long-term wage contracts—which amounts to an outright abolition of transfer fees—causes an across-the-board increase in the price of talent, i.e., the wages of all player types increase by more than what had been the corresponding transfer fee cost. Therefore not just the clubs previously selling talent, but also the net buyers would be worse off as a result. Scarce playing time is reallocated towards reduced experimentation: some positions that should be used to try out new players with upside potential will instead be given to older players with better expected performance, eventually reducing the amount of high talent available and increasing the average career length of players.

We also argued that penalties for early contract termination would result in the same allocation as unrestricted transfer fees if the penalties could be enforced regardless of their level. By varying the level of the maximum enforceable penalty, it is possible to achieve the whole range of outcomes (total surplus, wage and profit levels) between the levels achieved with unrestricted fees and without long-term contracts. Since the highest levels of transfer fees have tended to cause uproar, the level of enforceable penalties is in danger of becoming relatively low. The distribution of transfer fees has been right-skewed, so, for example, a penalty cap close to the average (unrestricted) fee would wipe out most of the fee income and might not be significantly better for efficiency than a complete abolition.

Some might argue that caps on penalties or transfer fees won't affect small clubs on the grounds that the highest fees are almost always paid between big clubs, for players that have already been transferred before. However, any restrictions on fees would most likely trickle down to the small clubs, as the price that a seller can get depends on the buyer's expected transfer fee revenue from selling the same player in the future.

To facilitate analysis at the level of an industry (as opposed to a worker-firm pair) we assumed away many features that we don't believe are quite as crucial for understanding the effects of transfer market regulations. There was no asymmetric information or effort cost and thus no moral hazard, adverse selection or other incentive problems in our model. Neither were there any kinds of frictions or firm-specific learning. Assuming one player per club left out complicated real-world complementarities and substitutabilities between different types of players and positions within a club.

Most significantly, with the assumption of one-shot learning and two-period careers, our model did not give rise to realistic career dynamics. Long-term commitment meant locking into a single wage for the whole career. In reality, player careers consist of series of overlapping contracts: players are typically first traded while still on their initial contract, at which point the new contract is extended beyond the duration of the original. If the player is moving up, then the wage is typically revised upwards immediately, while decaying or disappointing players may move down on free transfers.[26] A more sophisticated learning process would be needed for any empirical study of these dynamics. Such a model should include gradual learning (e.g., as in Jovanovic 1979), a trend to life-time ability (an initially increasing and eventually decreasing "fitness"), and perhaps also signals about ability that

[26]A multi-period setup brings up the further issue that it can be costly for clubs to part with disappointing players that had showed promise earlier, because their contract wages are above the reservation level.

are more precise when generated in leagues with more able opponents and co-players.

Despite discussing efficiency we did not explicitly model consumer surplus. We find it plausible that a decrease in surplus in our model corresponds to a decrease in welfare, because the driving factors behind the changes in surplus are changes in the quality of players and in their matching with clubs, and these do not affect the social costs of production. We have also bypassed a recurring theme of the public debate on the player market, namely its impact on competitive balance. The simple reason is that any regime considered will result in the players moving to wherever they have the highest marginal product. This idea goes back to Rottenberg (1956), who noted, in the spirit of Coase, that the equilibrium allocation of talent between clubs is independent of who owns the contractual rights.[27] Furthermore, the institutional setup in football—country-level league hierarchies with merit-based promotion/relegation—guarantees a dynamic sorting of clubs into various leagues, by which the clubs that meet in sporting competitions are much more equal than they are in the industry and in the labor market as a whole.

The last 15 years have been a time of tremendous growth for the football industry.[28] The driving forces have been the technological progress in the distribution of football matches, the changes in the structure of the sporting competition, and the integration of the labor markets. The proliferation of satellite television, pay-per-view, and cable channels, have increased the role of non-local revenue and resulted in stronger "superstar economics" (Rosen 1981)—where both top players and top clubs are the stars.[29] The introduction and increasing preeminence of the UEFA Champions League has further enhanced the revenue potential of talent at the biggest clubs. These developments are shifting some of small market consumers' attention into foreign clubs, where they can now also follow the performance of their most talented compatriots. The structure of demand itself may be changing in favor of the biggest clubs. In terms of our model, the relative size of the biggest clubs is increasing. We believe that the gains from trade between different leagues are increasing, at the same time when increasing regulatory pressure is restricting the level of compensation for this trade. The net effect of these two changes on the industry remains to be seen.

[27] For a survey of competitive balance and related issues in the economics of sports, see Fort and Quirk (1995).

[28] The combined revenue of the "Big Five" European leagues almost tripled (to 5.6 billion Euros) in the 7 years since 1995.

[29] Of EPL revenue in 2002/03, 44% came from broadcasting, 29% from matchday income, and 27% from commercial income. Only ten years earlier, most revenue was still earned at the gate.

7 Conclusion

Like other economists who have discussed the system, we found that transfer fees serve an important allocational purpose. However, we argued that the standard defense that transfer fees provide compensation for the cost of training is insufficient. Far from being just rents to talent or compensation for training, transfer fees are needed to efficiently allocate scarce competitive playing time among players of various levels of actual and potential ability, and to allocate the players between clubs with various opportunity costs of experimenting with new talent.

The football industry is currently in a situation where the regulators are seriously hampering the functioning of the transfer fee system (since Monti 2003, not since Bosman 1995), an institution that solves an externality problem in a decentralized manner. We believe that it would be impossible to approximate the efficiency benefits of the transfer fee system even with the most intricate training cost formula, or any other training-related regulations. Our modest proposal for deterring future regulatory moves by the EU is an immediate self-imposed ban on the term "transfer fee" in the industry. Friends of the sport should argue for the right of clubs and players, on mutual consent, to agree on contracts with "early termination penalties" with as much freedom as possible to deviate from the EU-approved default parameters. This rhetorical reform, by which players would no longer be "bought and sold" but merely fined on broken promises, would be a fitting response to the arguments laid out against the transfer system.

Appendix

Proof that the equilibrium defined by condition (13) is unique. We need to show that $\frac{\partial}{\partial z}\left(\frac{P(z)}{z}\right) < 0$. This proof will be easier by using talent θ rather than the quantile i as the variable of integration in (6).[30] Using $F(\theta)$ to denote the distribution function, and inverting equation (7), we get the relation $z(a^*) = \frac{1}{2-F(a^*)}$. The price of talent can then be written as

$$p(a|a^*) = \int_{a^*}^{a} b[1-\frac{1-F(\theta)}{2-F(a^*)}]\mathrm{d}\theta \text{ if } a > a^* \text{ and zero otherwise.} \tag{18}$$

Note that the expected transfer revenue to sellers is the expected value $\frac{P(z)}{z} = E[p(a|a^*)] \equiv P^E(a^*)$, where $p(a|a^*)$ is the price of talent a when there are $z(a^*)$ sellers. Since a^* is increasing in z, it suffices to show that $P^E(a^*)$ is decreasing in a^*. Denote the support of θ

[30] For change of variable, apply $F(\cdot)$ to both sides of $a = \theta[1-(1-i)/z]$, and solve for $i = 1-(1-F(a))z$. Note also that $\theta'[j] = 1/f(a)|_{F(a)=j}$.

by $[a_{\min}, a_{\max}]$. The expected transfer fee is

$$P^E(a^*) = \int_{a_{\min}}^{a_{\max}} p(a|a^*) f(a) \mathrm{d}a = \int_{a^*}^{a_{\max}} \left(\int_{a^*}^{a} b[1 - \frac{1-F(\theta)}{2-F(a^*)}] \mathrm{d}\theta \right) f(a) \mathrm{d}a, \tag{19}$$

where f is the density function of θ. We need to show that this is decreasing in z. Applying Leibniz's rule yields

$$\begin{aligned} \frac{\partial}{\partial z} P^E(z) &= -0 - \int_{a^*}^{a_{\max}} b[1 - \frac{1-F(a^*)}{2-F(a^*)}] f(a) \mathrm{d}a \\ &\quad - \int_{a^*}^{a_{\max}} \left(\int_{a^*}^{a} b'[1 - z(1-F(\theta))] \frac{1-F(\theta)}{(2-F(a^*))^2} f(a^*) \mathrm{d}\theta \right) f(a) \mathrm{d}a < 0. \end{aligned} \tag{20}$$

All terms behind the minus signs are positive.

Proof of Proposition 4. Denote the relation of talent level and talent price by $p(\theta|z^0)$ and $p(\theta|z^*)$ before and after the abolition. First, for $\theta \in (\bar{\theta}, a^*(z^*)]$ we see that $p(\theta|z^0) > 0$ and $p(\theta|z^*) = 0$. Next, use $b(a|z)$ to denote the club size b to be matched with talent $a \geq a^*(z)$, conditional on z, and change the variable of integration from j to a in (6). This yields $p(\theta|z) - p(a|z) = \int_a^\theta b(a|z) \mathrm{d}a$. Since $p(a^*(z)|z^0) > p(a^*(z)|z^*) = 0$, for $p(\theta|z^0) > p(\theta|z^*)$ to hold at all $\theta > a^*(z)$, it is sufficient to show that $\frac{\partial}{\partial \theta} p(\theta|z^0) > \frac{\partial}{\partial \theta} p(\theta|z^*)$. This is equivalent to $b(a|z^0) > b(a|z^*)$, which follows from Proposition 1. The decrease in profits is obvious for clubs in $[0, z^0]$, i.e. clubs that hire novices before and after, because the only change is the loss of transfer fees. For clubs in $(z^1, 1]$, i.e. clubs that hire veterans before and after, profits are reduced because each of them (except $i = 1$) now has to match with a lower talent than before and because the price of every level of talent is higher than before, as shown above. Finally, consider the clubs in $(z^0, z^1]$ that switched from hiring novices to hiring mediocre veterans. They get more revenue from output than before, because they hire better talents then before (in expectation), but they also lose the transfer fee income. The loss must be larger than the gain, because the types that they hire now are below $a[z^*|z^0] < a^*(z^*)$, and the type $a^*(z^*)$ would have been available at zero fee before.

Proof of Proposition 5 in Section 4. I.e., that $\hat{z}, a^*$ and Y increase in $\hat{p}$ for $\hat{p} \in (0, p[1|z^*])$. First note that the left hand side of (16) is increasing in z, while the right hand side is increasing in $\hat{p}$. Thus for the total differential $\frac{\mathrm{d}z}{\mathrm{d}\hat{p}}$ to be positive it is sufficient that the right hand side be decreasing in z. Since $z(a^*)$ is increasing in a^*, this is equivalent to showing that the right hand side is decreasing in a^*. To show this we will again use θ as the variable of integration, so total talent rents are then again given by (19). However, a cap on payments $\hat{p}$ will be binding above some talent level $\hat{a}$, at which

$$p(\hat{a}|a^*) = \hat{p}. \tag{21}$$

This defines a function $\hat{a}(a^*|\hat{p})$. The expected compensation of novice-hiring clubs, i.e. the right hand side of (16), is then

$$\begin{aligned}\hat{P}^E(a^*|\hat{p}) &= \int_{a^*}^{\hat{a}(a^*|\hat{p})} p(a|a^*)f(a)\mathrm{d}a + (1 - F(\hat{a}(a^*|\hat{p})))\,\hat{p}. \\ &= \int_{a^*}^{\hat{a}(a^*|\hat{p})} \left(\int_{a^*}^{a} b[1 - \frac{1 - F(\theta)}{2 - F(a^*)}]\mathrm{d}\theta \right) f(a)\mathrm{d}a + (1 - F(\hat{a}(a^*|\hat{p})))\,\hat{p}. \qquad (22)\end{aligned}$$

Taking the derivative yields

$$\begin{aligned}\frac{\partial}{\partial a^*}\hat{P}^E(a^*|\hat{p}) &= -0 - \int_{a^*}^{\hat{a}(a^*|\hat{p})} b[1 - \frac{1 - F(a^*)}{2 - F(a^*)}]f(a)\mathrm{d}a \\ &\quad - \int_{a^*}^{\hat{a}(a^*|\hat{p})} \left(\int_{a^*}^{a} b'[1 - \frac{1 - F(\theta)}{2 - F(a^*)}]\frac{1 - F(\theta)}{(2 - F(a^*))^2} f(a^*)\mathrm{d}\theta \right) f(a)\mathrm{d}a \qquad (23) \\ &\quad + \left\{ \left(\int_{a^*}^{\hat{a}(a^*|\hat{p})} b[1 - \frac{1 - F(\theta)}{2 - F(a^*)}]\mathrm{d}\theta \right) f(\hat{a}(a^*|\hat{p})) - f(\hat{a}(a^*|\hat{p}))\hat{p} \right\} \frac{\partial \hat{a}(a^*|\hat{p})}{\partial a^*}.\end{aligned}$$

The first two lines are clearly negative. The integral inside the brackets in the third line is $p(\hat{a}(a^*|\hat{p})|a^*)$, which is equal to $\hat{p}$ by definition (21), so the whole line cancels out to zero. Hence $\frac{\partial}{\partial a^*}\hat{P}^E(a^*|\hat{p}) < 0$ and $\frac{\mathrm{d}z}{\mathrm{d}\hat{p}} > 0$. The rehiring threshold is increasing in $\hat{p}$ because $a^*(z)$ is an increasing function, and total surplus is increasing in $\hat{p}$ because it moves z closer to the unrestricted equilibrium value at which total surplus is maximized.

Proof of Proposition 7. Holding the maximum size in the small market fixed at $\beta > 1$, the club size profile is $b[i|\delta] = 1 + (\beta + \delta - 1)i$, where $\delta > 0$ in the large and $\delta = 0$ in the small market. The corresponding distribution function is $G(b|\delta) = (b - 1)/(\beta + \delta - 1)$ in the support $[1, \beta + \delta]$. The distribution function of the integrated market is thus $G_I(b|\delta) = \frac{1}{2}(G(b|0) + G(b|\delta))$, which can be inverted to yield the profile for the integrated market:

$$B[i|\delta] = \begin{cases} 1 + \phi(\delta)\, i, & i \in [0, \hat{\imath}(\delta)] \\ 2 - \beta - \delta + 2(\beta + \delta - 1)\, i, & i \in [\hat{\imath}(\delta), 1] \end{cases} \qquad (24)$$

Here $\hat{\imath}(\delta) \equiv G_I(\beta|\delta) = (2(\beta - 1) + \delta) / (2(\beta + \delta - 1))$ is the fraction of firms in the overlapping part of the supports and $\phi(\delta) = (2(\beta - 1)(\beta + \delta - 1)) / (2(\beta - 1) + \delta)$ denotes the slope in that part.

Use $z_L(\delta)$ and $z_I(\delta)$ to denote the implicit functions defined by the equilibrium condition (13) for size profiles $b[i|\delta]$ and $B[i|\delta]$ respectively. They are by definition equal at $\delta = 0$, and by Proposition 6, increasing functions, because the slopes of both profiles are everywhere increasing in δ. We also know that $1 > z = 1/(2 - F(a^*)) \geq 1/\left(2 - F(\bar{\theta})\right) > 1/2$, because at least some veterans are hired, and below-mean talents are always discarded. The inequality $z_L(\delta) > \hat{\imath}(\delta)$ holds for sufficiently large δ, because $\lim_{\delta\to\infty} \hat{\imath}(\delta) = 1/2$. Now consider $\gamma(i) \equiv B[i|\delta]/b[i|\delta]$ for $i \in (\hat{\imath}(\delta), 1]$:

$$\gamma'(i) = \frac{B'[i|\delta]b[i|\delta] - B[i|\delta]b'[i|\delta]}{b[i|\delta]^2} = \frac{1}{b[i|\delta]^2}(\beta+\delta-1)\left(2b[i|\delta] - B[i|\delta]\right). \tag{25}$$

Clearly $\gamma'(i) > 0$ for $i \in (\hat{\imath}(\delta), 1]$, because $b[i|\delta] > B[i|\delta]$ for all $i \in (0,1)$. Thus, when δ is sufficiently large, then the argument used in the proof of Proposition 6, with $\gamma(i)$ replacing $\rho(i)$, shows that $z_I(\delta) > z_L(\delta)$.

Next we prove that $z_L(\delta) > z_I(\delta)$ for sufficiently small δ by showing that $z_L(\delta)$ has a strictly higher slope than $z_I(\delta)$ at $\delta = 0$. Assuming that talent is distributed uniformly in $[0,\alpha]$, we have $\bar{\theta} = \alpha/2$, $a^*(z) = \alpha(2-1/z)$ and $a[i|z] = \alpha(1-(1-i)/z)$ and $a'[i|z] = \alpha/z$ for $i \in [z,1]$, with $z > 2/3$ required by feasibility. First consider the large market. The LHS of the equilibrium condition (13) is

$$\begin{aligned}\left(a^*(z) - \bar{\theta}\right) b[z|\delta] &= \alpha\left(2 - 1/z - 1/2\right)\left(1 + (\beta+\delta-1)z\right) \\ &= \alpha\left(\frac{3}{2} - \frac{1}{z}\right)\left(1 + (\beta+\delta-1)z\right).\end{aligned} \tag{26}$$

Using (11), the RHS of (13) becomes

$$\begin{aligned}&\frac{\alpha}{z^2}\int_z^1 (1-i)\left(1+(\beta+\delta-1)i\right)\mathrm{d}i \\ =\;& \frac{\alpha}{6z^2}(1-z)^2\left(2+\beta+\delta+2(\beta+\delta-1)z\right).\end{aligned} \tag{27}$$

Applying the implicit function theorem to (26) = (27), we get (after some simplifications)

$$\frac{dz_L}{d\delta} = \frac{1+(3-7z)z^2}{3z\left(6(3-\beta-\delta)+7z(\beta+\delta-1)\right)} \tag{28}$$

Similarly, consider the integrated market. Since $\hat{\imath}(0) = 1 > z$, we have $B[z|\delta] = 1 + \phi(\delta)z$ for δ near zero, and the LHS of (13) is

$$\left(a^*(z) - \bar{\theta}\right) B[z|\delta] = \alpha\left(\frac{3}{2} - \frac{1}{z}\right)\left(1+\phi(\delta)z\right). \tag{29}$$

The RHS is

$$\frac{\alpha}{z^2}\left(\int_z^{\hat{\imath}(\delta)} (1-i)\left(1+\phi(\delta)i\right)\mathrm{d}i + \int_{\hat{\imath}(\delta)}^1 \left(2-\beta-\delta+2(\beta+\delta-1)i\right)\mathrm{d}i\right). \tag{30}$$

The implicit function theorem applied to (29) = (30) yields the slope

$$\begin{aligned}\frac{dz_I}{d\delta} \;=\;& \left\{\left(24z^2 - 56z^3\right)(\beta-1)^2(\beta+\delta-1)^2 + \left(2(\beta-1)+\delta\right)^2\left(2\left((\beta-1)(\beta+\delta-1)\right)+\delta^2\right)\right\} \div \\ & \left\{24z^2(\beta+\delta-1)^2\left(2(\beta-1)+\delta\right)\left(8\beta+4\delta-6+7(\beta-1)(\beta+\delta-1)z - 2\beta(\beta+\delta)\right)\right\}\end{aligned}$$

Evaluating the slopes at $\delta = 0$, we find (after considerable simplifications) that

$$\left.\frac{dz_I}{d\delta}\right|_{\delta=0} = \frac{1 + (3 - 7z)z^2}{6z\left(6 + 7z\left(\beta - 1\right) - 2\beta\right)} > \frac{1}{2}\left.\frac{dz_L}{d\delta}\right|_{\delta=0}. \tag{31}$$

Thus, for sufficiently small δ, $z_L(\delta) > z_I(\delta)$.

Proof of Proposition 8. First the case when $z_I > z_L$ (large δ), which implies $a_I^* > a_L^*$. Changing the variable of integration from i to θ as in (18), and denoting $j(\theta|a^*) = 1 - (1 - F(\theta)) / (2 - F(a^*))$, the prices of retained talent of level a in the large and integrated market are

$$p_L(a) = \int_{a_L^*}^{a} b[j(\theta|a_L^*)|\delta]\mathrm{d}\theta \text{ and } p_I(a) = \int_{a_I^*}^{a} B[j(\theta|a_I^*)|\delta]\mathrm{d}\theta. \tag{32}$$

Since $b[i|\delta] > B[i|\delta]$ for all $i \in (0, 1)$, and because $a_I^* > a_L^*$, we have $b[j(\theta|a_L^*)|\delta] > B[j(\theta|a_I^*)|\delta]$ for all $\theta \in [a_L^*, \alpha)$. Therefore $z_I > z_L$ is sufficient for $p_I(a) < p_L(a)$ to hold at all $a \in (a_L^*, \alpha]$. As the price for every level of talent is lower in the integrated market than in the large market, the expected transfer fee is also lower.

Now consider the case when $z_L > z_I$ (small δ), which implies $a_L^* > a_I^*$. Denote the expected transfer fee in the large and integrated market respectively by $P_L^E(\delta)$ and $P_I^E(\delta)$. In equilibrium, expected transfer fees, the right side of (13), are equal to the left side, so

$$P_L^E(\delta) = \left(a_L^* - \bar{\theta}\right) b[z_L(\delta)|\delta] \text{ and } P_I^E(\delta) = \left(a_I^* - \bar{\theta}\right) B[z_I(\delta)|\delta]. \tag{33}$$

To show that $P_L^E(\delta) > P_I^E(\delta)$ it is sufficient to show that $b[z_L(\delta)|\delta] > B[z_I(\delta)|\delta]$. As these are equal at $\delta = 0$, all that remains to show is that, starting from $\delta = 0$, the former increases faster in δ. So we need to compare

$$\frac{\partial}{\partial\delta}\left(b[z_L(\delta)|\delta]\right) = b'[z_L(\delta)|\delta]\frac{\partial z_L}{\partial\delta} + z_L(\delta) \text{ and} \tag{34}$$

$$\frac{\partial}{\partial\delta}\left(B[z_I(\delta)|\delta]\right) = B'[z_I(\delta)|\delta]\frac{\partial z_I}{\partial\delta} + z_I(\delta)\frac{\partial}{\partial\delta}\phi(\delta). \tag{35}$$

We know that $\partial z_L/\partial\delta > \partial z_I/\partial\delta$ for small (but nonzero) δ, and $b'[z|\delta] = (\beta + \delta)$ and $B'[z|\delta] = \phi(\delta)$. Each term in (34) is then greater than the corresponding term in (35) because

$$\phi(\delta) = \frac{2\left(\beta - 1\right)\left(\beta + \delta - 1\right)}{2\left(\beta - 1\right) + \delta} < \beta + \delta \text{ and} \tag{36}$$

$$\frac{\partial}{\partial\delta}\phi(\delta) = \frac{2\left(\beta - 1\right)^2}{\left(2\left(\beta - 1\right) + \delta\right)^2} < 1 \tag{37}$$

hold for all $\beta > 1$, $\delta > 0$. Since talent below the rehiring threshold has zero price, the talents between the two rehiring thresholds are more expensive in the integrated market: $p_I(a) > 0 = p_L(a_L^*)$ for $a \in (a_I^*, a_L^*]$. Because the expected price of talent has decreased, while the price of the lowest talent has increased, the price of some of the higher talents must have decreased. In fact, it is straightforward to show (we omit the proof) that the difference $p_L(a) - p_I(a)$ is convex in (a_L^*, α) and maximized at $a = \alpha$. Therefore a segment of the very highest talents are cheaper in the integrated market. Finally, consider the biggest club: it must certainly be better off after integration, because it is matching with the same type ($\theta = \alpha$) before and after, but is paying a lower transfer fee after the integration.

References

AGHION, PHILIPPE AND PATRICK BOLTON (1987). "Contracts as a Barrier to Entry," *American Economic Review*, 77, 388–401.

AGHION, PHILIPPE AND BENJAMIN HERMALIN (1990). "Legal Restrictions on Private Contracts Can Enhance Efficiency," *Journal of Law, Economics, and Organization*, 6, 381–409.

ANTONIONI, PETER AND JOHN CUBBIN (2000). "The Bosman Ruling and the Emergence of a Single Market in Soccer Talent." *European Journal of Law and Economics,* 9, 157–173.

ANTRÀS, POL, LUIS GARICANO AND ESTEBAN ROSSI-HANSBERG (2006). "Offshoring in a Knowledge Economy." *Quarterly Journal of Economics*, forthcoming.

CARMICHAEL, FIONA, DAVID FORREST AND ROBERT SIMMONS (1999). "The Labour Market in Association Football: Who Gets Transferred and For How Much?" *Bulletin of Economic Research*, 51, 125–150.

CARMICHAEL, FIONA AND DENNIS THOMAS (1993). "Bargaining in the Transfer Market: Theory and Evidence." *Applied Economics*, 25, 1467–1476.

CAVES, RICHARD E (2000). *Creative Industries: Contracts Between Arts and Commerce.* Harvard University Press.

FALCONIERI, SONIA, JÓZSEF SÁKOVICS AND FRÉDÉRIC PALOMINO (2004). "Collective Versus Individual Sale of Television Rights in League Sports." *Journal of the European Economic Association*, 2, 833–862.

FEESS, EBERHARD, BERND FRICK AND GERD MUEHLHEUSSER (2004). "Legal Restrictions on Buyout Fees: Evidence from German Soccer." *IZA discussion paper 1180.*

FEESS, EBERHARD AND GERD MUEHLHEUSSER (2003a). "Transfer Fee Regulations in

European Football." *European Economic Review*, 47, 645–668.

FEESS, EBERHARD AND GERD MUEHLHEUSSER (2003b). "The Impact of Transfer Fees on Professional Sports: An Analysis of the New Transfer System for European Football." *Scandinavian Journal of Economics*, 105, 139–154.

FORT, RODNEY AND JAMES QUIRK (1992). *Pay Dirt: The Business of Professional Team Sports*. Princeton University Press.

FORT, RODNEY AND JAMES QUIRK (1995). "Cross-subsidization, Incentives, and Outcomes in Professional Team Sports Leagues." *Journal of Economic Literature*, 33, 1265–1299.

FRANK, ROBERT H. AND PHILIP J. COOK (1995). *The Winner-Take-All Society*. The Free Press.

GROSSMAN, GENE M. AND GIOVANNI MAGGI (2000). "Diversity and Trade." *American Economic Review*, 90, 1255–1275.

ROSEN, SHERWIN (1981). "The Economics of Superstars." *American Economic Review*, 71, 845-858.

ROSEN, SHERWIN AND ALLEN SANDERSON (2001). "Labour Markets in Professional Sports." *Economic Journal*, 111, F47–F67.

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

SANDERSON, ALLEN AND JOHN SIEGFRIED (1997). "The Implications of Athlete Freedom to Contract: Lessons from North America." *Economic Affairs*, 17(3), 7–12.

SATTINGER, MICHAEL (1993). "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature*, 31, 831–880.

SZYMANSKI, STEFAN (2001). "Income Inequality, Competitive Balance and the Attractiveness of Team Sports: Some Evidence and a Natural Experiment from English Soccer." *Economic Journal*, 111, F69–F84.

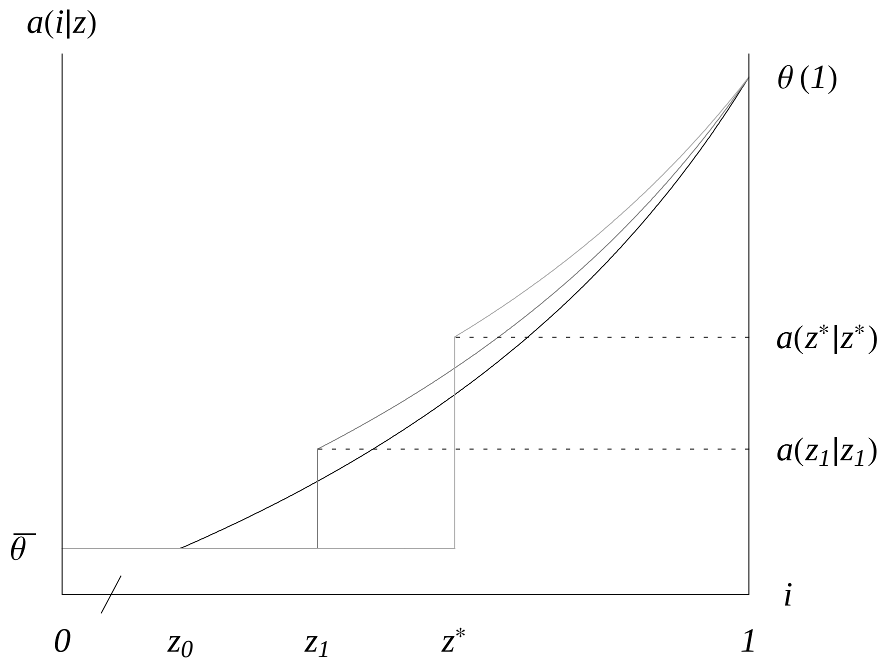

Figure 1. Three profiles of talent. The profile that jumps at z^* corresponds the equilibrium under transfer fees. The lowest profile, which kinks at z_0, depicts the situation without transfer fees. The middle case corresponds to a maximum termination penalty p such that $z(p)=z_1$.

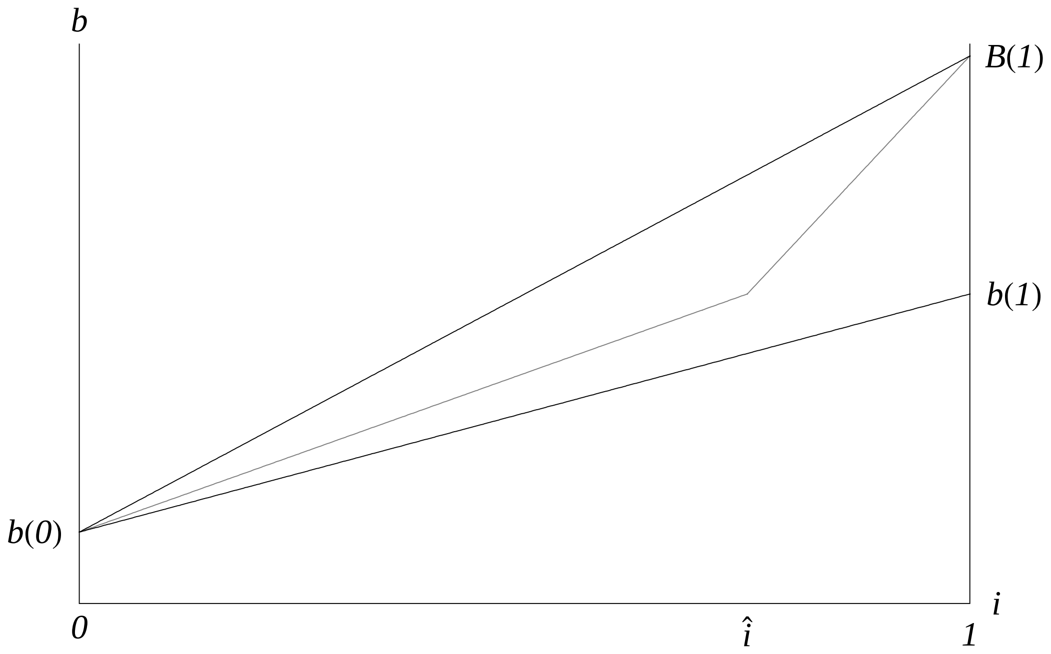

Figure 2. Integration of two profiles of club size.