The Football Team Composition Problem: A Stochastic Programming Approach

G. Pantuso
Department of Mathematical Sciences
University of Copenhagen
2100 Copenhagen, Denmark

Abstract

Most professional European football clubs are well-structured businesses. Therefore, the financial performance of investments in players becomes crucial. In this paper, after the problem is discussed and formalized, an optimization model with the objective of maximizing the expected value of the team is presented. The model ensures that the team has the required mix of skills, that competition regulations are met, and that budget limits are respected. The model explicitly takes into account the uncertainty in the career development of football players. A case study based on the English Premier League is presented. Our results show that the model has significant potential to improve current decisions ensuring a steady growth of the value of the team. The team value growth reported is particularly driven by investments in young prospects.

Keywords: soccer,team composition,sports management,stochastic programming

1 Introduction

European Football (or *soccer* in U.S. parlance), besides being one of the most practiced leisure activities, has become a flourishing industry. Hundreds of professional football clubs all over the world are well-structured businesses, and complex entertainment systems are built around football competitions. Deloitte (2016) estimates the cumulative revenue of the top 20 football clubs to be €6.6 billion for season 2014/15, and expects further growth towards €8 billion for season 2016/17. This corresponds to doubling of revenue in six years. The study ascribes the main earning potential of football clubs to: (a) match-day revenue through stadium attendance, (b) broadcasting rights (including the distributions for participating, and possibly advancing, in national and international competitions), and (c) commercial income through sponsorship (see Bouchet, Doellman, Troilo, and Walkup (2015)), merchandising, and social media visibility.

In this context, one of the key decisions for football club managers, in order to stimulate earnings, is the hiring of professional players to compete for the club. In fact, Dobson and Goddard (2001, Ch.2) state that the main component of a football club's cost is expenditure on players through wages and transfer fees. Furthermore, combining the data in Forbes (2015b) and Forbes (2015a), it emerges that 17 of the 20 highest-salaried players play for the top 8 most valuable clubs. This data suggests that valuable players stimulate all the above-mentioned sources of earnings. In fact, skilled players can reinforce and trigger the interest of supporters in the club, which in turn is translated into stadium attendance, merchandise sale, social media visibility, and sponsorship deals. As an example, Forbes (2015b) shows that top-paid football players can generate significant social media interest. Bouchet et al. (2015) register a positive effect on the stock return after an enterprise sponsors an international match featuring popular clubs. As far as broadcasting rights are concerned, at least part of them are conditional on qualification and thus on successes on the field of play. However, as Dobson and Goddard (2001, Ch.2) point out, sport successes also depend on the club's capacity to strengthen the team by purchasing and retaining the best players. Finally, for clubs with limited spending potential, a crucial source of sustainment is arguably the sale of players. Such clubs, in fact, can make significant profits by scouting and purchasing young talents and selling them once they reach a high market value. Thus, the composition of the team is a crucial driver for both the competitive and financial performance of a club.

In this paper we consider a football club's strategic problem of investing an available budget to purchase football players. The objective is to maximize the expected value of the team - represented by the sum of the transfer market appreciation of the players in the team. The need to obtain players with certain skills is taken into account as well as competition rules (which may as well influence the mix of players). In addition, we take into consideration the uncertainty in the future market evaluation of the players which, besides being driven by performance, is influenced by a number of unpredictable elements such as injuries, motivation, and luck. We refer to this problem as the *Football Team Composition Problem* (FTCP). Thus, the contribution of this paper is a novel stochastic optimization model for assisting football club managers when investing in football players while explicitly dealing with uncertainty in the career development of players. To our knowledge, this problem has not been considered before. The model is tested on a case study based on the English Premier League. Decisions to the model are compared to those of the corresponding real-life clubs in order to assess whether the model might help decision makers in reality. The problem is formalized in Section 2, while in Section 3 we connect it to the existing literature. In Section 4 we provide a mathematical model for the FTCP and in Section 5 we present a computational study. Finally, conclusions are drawn in Section 6.

2 The Football Team Composition Problem

Football clubs engage a number of football players in order to participate in national and international competitions. The *International Federation of Football Associations* (FIFA) dictates that clubs register new players only during *registration periods* (FIFA, 2015), or transfer market windows (TMWs) in football jargon. Particularly, each national football association must arrange two TMWs, the first between two consecutive seasons for at most twelve weeks, and the second in the middle of the season for at most four weeks. In order to be registered for competitions, professional football players are engaged to clubs by means of contracts.

If a club wants to hire a player currently engaged to another club, an agreement between the clubs must be reached. Such agreement typically involves a remuneration for the club who currently has a contractual agreement with the player. The remuneration reflects the value of the player, which is influenced by market logic and measures the increase in the buying team's performance and club revenue (Dobson and Gerrard, 1999). A new contract between the purchasing club and the player is then signed. *Free-agent* players, that is, players who are currently free from contracts, can also be engaged. In what follows we say that a club *owns* a player if there is an ongoing contractual agreement between the two. Similarly, we say that club A *buys* a player when they find an agreement with club B, who currently owns the player, in order for the player to play for club A. In that case we also say that club B *sells* the player. Players may also be borrowed/lent by one team to another (FIFA, 2015). In this case we say that the player is on a loan. A loan fee is typically paid to the club which owns and lends the player and the salary is typically paid by the club which borrows the player. The most common duration of a loan is one season.

When planning the team, clubs must take a number of restrictions into account. Managers of football clubs are typically given a budget to spend in the ongoing TMW. The budget is typically decided by the financial management or by the owners of the club and may be limited by competition regulations, such as the *UEFA Financial Fair Play* (UEFA, 2015b). However, in addition to the given budget, clubs may reinvest earnings from selling and lending players. The number of players which can participate in a given competition is also limited by the organizers of the competition. As an example, for the period 2015-2018, a club cannot register more than 25 players for the UEFA Champions League (UEFA, 2015a). Both owned and borrowed players count against the 25-player limit. Furthermore, an upper bound on the number of players owned by clubs may be imposed by national football associations or by a club's internal policies.

Players are characterized by a set of skills. The set of skills indicates, at the very least, the position(s) on the field a player can occupy, e.g., midfield, right-wing and central attack. However, clubs may take into account several other abilities such as speed, stamina, and technique, as well as personality traits such as motivation and determination to succeed. Clubs typically need to compose a team with a mix of different and compatible skills. Most often the mix of characteristics is determined with input from the coach of the team who needs players with skills compatible with the team's organization of play. In addition, competition rules might also influence the mix of players. As an example, the Italian Football Federation requires that clubs register at least four players who grew up in the club's young selections and four players who grew up in an Italian club's young selection (FIGC, 2014). We refer to a skill set as a *role*.

Clubs consider a number of *target players* to purchase or borrow. A list of target players is often the result of extensive scouting and analysis. Such analysis might rely on statistics which describe and quantify the abilities of the players. Example statistics include goals scored, pass accuracy, speed, and air-duels won. However, the analysis also heavily relies on the specific expertise of the scouts who observe the players in action and can also assess skills which are difficult to translate numerically, such as leadership. In any case, target players are to a large extent the result of screening and filtering scouting lists. For each target player, clubs know whether the player can be purchased or borrowed (or both), their price, loan fee, and current salary. With respect to the sale of players, the club knows which players can be sold or lent (i.e., for which players a potential buyer or borrower exists) as well as the selling prices and loan fees. In addition, the club may not want to sell or lend some crucial players, regardless of the price and loan fee. Finally, clubs must take into account that players retire at an age which is typically in their late thirties (Dobson and Goddard, 2001, Ch.4), and that the retirement age varies from player to player.

The value of a football player is influenced by several factors (Dobson and Gerrard, 1999), including skills, past performances, age, and personality. When deciding with which players to negotiate, clubs know the current value of each player. However, the future value depends on a number of unpredictable and intangible elements such as fitness, injuries, successes, and luck. Therefore, the future value of football players (i.e., their value at future TMWs) is uncertain, and such uncertainty is transferred onto future purchase and selling prices, loan fees, and salaries. Thus, decisions about the team composition have to be made under uncertainty.

Therefore, the FTCP can be summarized as the problem of deciding which players to buy, sell, borrow and lend, and when to do so, in order to maximize the expected value of the team over the planning horizon, while satisfying budget limits, competition restrictions, and coach requirements. Particularly, notice that maximizing the expected market value of the team implicitly corresponds to maximizing a measure of the on-the-field performance of the team. The performance of players is generally measured by a number of role-specific statistics such as goals scored, pass accuracy, and distance run. However, accounting for these measures simultaneously in the objective function would introduce a number of non-trivial challenges such as summing and weighing numerically different quantities, and dealing with performances not easily associated with statistics such as team spirit and motivation. Instead, the market value of a player, although influenced by market dynamics, is an implicit indicator of performance which solves many of these challenges. In fact, it ensures a unique metric and measurement unit across different roles, it captures abilities of the players which are difficult to express with other statistics and, finally, it enables decision makers to consider the financial and on-the-field performance simultaneously.

3 Connections to the existing literature

The FTCP shares features with several well-studied problems from the research literature. However, all these problems possess significant differences with the FTCP.

Similar to portfolio optimization problems (Cariño, Kent, Myers, Stacy, Sylvanus, Turner, Watanabe, and Ziemba, 1994; Mulvey and Shetty, 2004; Mansini, Ogryczak, and Speranza, 2015), the FTCP can be seen as the problem of investing in a number of assets (football players) with stochastic returns in order to maximize some utility function (in this case the value of the team). As in cardinality-constrained portfolio optimization problems (Chang, Meade, Beasley, and Sharaiha, 2000), the FTCP requires a fixed number of assets (players). However, in portfolio optimization problems, fractions of a wealth are typically allocated to different assets (e.g., stocks), while in the FTCP, decisions are of a Boolean type (i.e., a football player can be either bought or not). As an example, while the FTCP requires a fixed number of players for each role, cardinality-constrained portfolio optimization problems may impose a minimum and maximum proportion of wealth to be allocated to a certain class of assets. Furthermore, the FTCP includes the possibility of borrowing and lending players which does not directly correspond to features of portfolio optimization problems.

In capacity renewal problems (Rajagopalan and Soteriou, 1994; Chand, McClurg, and Ward, 2000), machinery must be replaced over time due to a variety of factors such as obsolescence and technological breakthroughs (Hopp and Nair, 1994; Adkins and Paxson, 2013). Similarly, in the FTCP, players are replaced over time due to several factors such as ageing. However, in capacity replacement problems, machinery are not seen as assets used to maximize some utility function, rather as items needed to satisfy a given demand in the most efficient way. Furthermore, in capacity replacement problems, individual machines of a given type are often indistinguishable from each other (see Mørch, Fagerholt, Pantuso, and Rakke (2017) for the case of transportation fleets) while in the FTCP each player has specific distinguishing characteristics.

Knapsack problems also have similarities with the FTCP. A capacity (budget) has to be allocated to items characterized by a weight (cost) and reward (value). As an example, Kirshner (2011) models the problem of signing free-agent NBA players as a knapsack problem where rewards are measures of the ability of the players, weights are salaries, and the capacity is the budget of the team. Gibson, Ohlmann, and Fry (2010) use a version of the of the stochastic knapsack problem to address draft-phase team composition decisions. Essentially, at the beginning of the seasons, teams can select players to hire according to a pick order. Therefore, players become stochastically unavailable over time depending on the hiring decisions of other teams. Thus, the authors model the problem as a stochastic knapsack problem where the future availability of items is stochastic. However, significant differences can be found between the FTCP and knapsack problems. The main difference is that the FTCP is concerned not only with adding (buying, borrowing) items to a knapsack, but also with removing (selling, lending) them. Particularly, removal decisions not only free knapsack space for more valuable items, but also yield a reward (selling price, lending fee) which stochastically changes over time. Furthermore, in the dynamic stochastic knapsack problem, items generally arrive at random times (Kleywegt and Papastavrou, 1998), while in the FTCP the current players in the team and the focal players are available at the current and future TMWs.

Similar to the FTCP, staffing problems seek to match the supply and demand of personnel of different categories or with different skills (Komarudin, Guerry, Feyter, and Vanden Berghe, 2013; Bruecker, den Bergh, Belien, and Demeulemeester, 2015). However, significant differences exist. First, in staffing problems there is generally no distinction between workers in the same category (e.g., two individual nurses are homogeneous) while in the FTCP players are heterogeneous. In fact, even if two players have the same skills, their future career developments (and market values) are different. Second, the FTCP also considers selling or lending players while staffing problems may consider dismissal of personnel. Finally, the FTCP sees players as assets which increase the value of the team, and not just as items which contribute to satisfy a demand.

As far as the football literature is considered, Tavana, Azizi, Azizi, and Behzadian (2013) propose a two-phase method for player selection. In the first phase, the method evaluates and ranks a number of players in order to find the best performers. In the second phase, alternative combinations of the best performers are evaluated and the best combinations is selected for the team. This method does not consider the financial impact of the investment in the players suggested. A similar problem is that of selecting the best lineup. In Boon and Sierksma (2003) and Sierksma (2006), the starting 11 are selected in such a way that on-field quality is maximized. Quality is measured as the fitness of a player to the role assigned by the coach. Similarly, Ozceylan (2016) proposes a combination of the analytic hierarchical process and integer programming for selecting the best 11 from within the team. In this case, for each position on the field, players are evaluated based on a number of criteria extracted from a popular computer game. Finally, a related problem for hockey teams has been studied by Chan, Cho, and Novati (2012) who use a clustering technique to identify distinct hockey player types, and then, by means of regression, quantify the relationship between player types and team performance. With respect to the above mentioned approaches, the FTCP comprises decisions at a higher strategic level and considers a longer planning horizon. In addition, in the FTCP the scope is not only to select players with the desired skills and expected performances, but also to take into account the financial soundness of the investments. The above mentioned approaches are thus complementary with the optimization model proposed in this paper.

Finally, a similar problem has also been studied in the fantasy football literature. Analytic methods for assisting draft-phase decisions have been proposed by Fry, Lundberg, and Ohlmann (2007) and Becker and Sun (2016). While similarities with the FTCP exist (e.g., in both cases a fixed number of players with different roles have to be selected), significant differences can also be found. For instance, in fantasy football competitions a new team is built from scratch every season (or every time the game is started) while in the FTCP, managers modify an existing team periodically. Furthermore, the pick order makes draft phase decisions significantly different from those made in the FTCP, where club managers choose from a list of target players that can be bought or borrowed.

The contribution of this paper to the existing literature consists of a novel optimization model for assisting the management of a football club in the composition of a team of players. Particularly, the model is designed to explicitly take into account uncertainty in the career development of football players. To our knowledge, this problem has not been considered before in the OR literature.

4 Mathematical Model

In this section, we present a mathematical model for the FTCP. In Section 4.1 we discuss our modeling assumptions, and then we present the notation and the formulation of the model. In Section 4.2 we comment on how uncertain future players values may be described.

4.1 Stochastic Programming Formulation

The mathematical model is based on the following assumptions:

- decision times correspond to TMWs,
- the current value of the players, their purchase and selling price, salary, and loan fees are known,
- the future value of the players, their purchase and selling price, salary, and loan fees are stochastic but their probability distribution is known; Section 4.2 discusses how such distribution may be obtained,
- an ongoing contract exists between the team and the players owned, thus decisions regarding the length and renewal of contracts are addressed in a separate decision problem,
- loans last until the next TMW.

Notice that current purchase and selling prices, loan fees and salaries, are in reality the result of negotiations. However, in general, market operators have an understanding of closing prices. Alternatively, an instance of the FTCP can be solved at any step in the negotiation phase to assess the option of buying (selling, borrowing, lending) a player after new bids.

We model the FTCP as a multistage stochastic program. Multistage stochastic programming is a framework for modeling problems involving a sequence of decisions under uncertainty, conditional on the realization of random events, see e.g., the textbooks Kall and Wallace (1994), Birge and Louveaux (1997) and King and Wallace (2012). The FTCP involves a sequence of team composition decisions under uncertainty. Figure 1 sketches the interplay between decisions and market information. At every TMW t, the market status $\xi_t(\omega_t)$ – which will be formally introduced shortly – is determined by the realization of a random event ω_t, and the club is to make team composition decisions. Thus, at TMW t the club knows the market status up to $\xi_t(\omega_t)$ and the future market status only in a probabilistic sense. Let $\mathscr{P}$ be the set of players

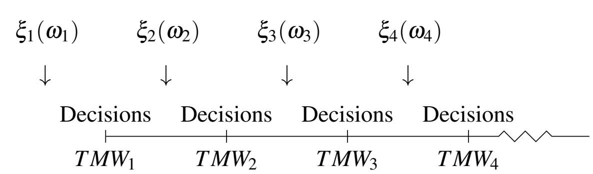

Figure 1: Multistage Stochastic Program Decision Timing

upon which decisions must be made. $\mathscr{P}$ includes both the players already in the team (which can thus be sold or lent) and the target players. Let $\mathscr{T} = \{1, 2, \dots\}$ be the set of TMWs (i.e., stages). Let $\mathscr{R}$ be the set of roles. A role is a well-defined skill set which typically includes, but is not limited to, the position on the field of play (see Section 2). Finally, let $\mathscr{P}_r \subseteq \mathscr{P}$ the set of player having role r. Let $\omega_1, \ldots, \omega_{|\mathscr{T}|}$ be a random process defined on some probability space $(\Omega, \mathbb{F}, \mathbb{P})$, where Ω is the set of elementary outcomes, $\mathbb{F}$ is the set of possible events, and $\mathbb{P}$ is a measure of the likelyhood of the events. As an example, ω_t may represent the joint career development of the players in $\mathscr{P}$ until TMW t.

The FTCP relies upon the following uncertain parameters. For player p at time t and outcome ω_t, define $V_{pt}(\omega_t)$ as the player value, $S_{pt}(\omega_t)$ as the price to receive for selling the player, $P_{pt}(\omega_t)$ as the price to pay to buy the player, $I_{pt}(\omega_t)$ as the fee to pay borrow the player, $O_{pt}(\omega_t)$ as the fee to receive to lend the player, and $W_{pt}(\omega_t)$ the player salary. Notice that, for a given player and TWM, the purchase and selling price are assumed, without loss of generality, different. This is done to account for different transaction costs born by buyers and sellers. Similarly, the fee to pay for borrowing a player is generally different from the fee to receive for lending the same player. The *player vector* for player p at TMW t is then $\xi_{pt}(\omega_t) = (V_{pt}(\omega_t), S_{pt}(\omega_t), P_{pt}(\omega_t), I_{pt}(\omega_t), O_{pt}(\omega_t), W_{pt}(\omega_t))$. Let the *market status* at TMW t be $\xi_t(\omega_t) = (\xi_{pt})_{p \in \mathscr{P}}$. Knowledge of $(\Omega, \mathbb{F}, \mathbb{P})$ is not necessary, but we assume the probability distribution of $\xi = (\xi_1, \ldots, \xi_{|\mathscr{T}|}(\omega_{|\mathscr{T}|}))$ is known and that ξ_1 is deterministic.

The FTCP relies upon the following known parameters. Let $\bar{N}$ be the maximum number of players the club can own. Let N be the number of player which can participate in competitions. Let N_r be the number of players required (typically by the coach) in role r. Let K^O_{pt} be a binary parameter equal to 1 if the club is available to lend player p at TMW t, 0 otherwise. Let K^S_{pt} be a binary parameter equal to 1 if the club is available to sell player p at TWM t, 0 otherwise. Let K^I_{pt} be a binary parameter equal to 1 if the club can borrow player p from the owning club at TMW t, 0 otherwise. Let K^P_{pt} be a binary parameter equal to 1 if the club can buy player p from the owning club at TMW t, 0 otherwise. Note that parameters K^S_{pt} and K^O_{pt} are set by the decision-maker club, i.e., the club which is solving a FTCP. Parameters K^I_{pt} and K^P_{pt} depend instead on the availability of the club which currently own player p, to lend or sell, respectively, the player to the club which is solving a FTCP. Let $\bar{A}_p$ the retirement age for player p and A_{pt} their age at time t. Let Y_p be equal to 1 if player p belongs to the team at the beginning of the planning horizon, 0 otherwise. Let B_t be the budget of the club for time t and ρ be the discount rate. Notice that we consider the budget to be a deterministic parameter set by the club's ownership or management independent of the random outcomes.

Finally, the club's decisions are represented by the following decision variables. Variable $y_{pt}(\omega_t)$ is equal to 1 if player p belongs to the club at the end of TMW t given outcome ω_t, 0 otherwise. Variable $y^P_{pt}(\omega_t)$ is equal to 1 if player p is purchased during TMW t given outcome ω_t. Variable $y^S_{pt}(\omega_t)$ is equal to 1 if player p is sold during TMW t given outcome ω_t. Variable $l^I_{pt}(\omega_t)$ is equal to 1 if player p is borrowed from another club during TMW t given outcome ω_t. Variable $l^O_{pt}(\omega_t)$ is equal to 1 if player p is lent to another club during TMW t given outcome ω_t. Variable $v(\omega_{|\mathscr{T}|})$ represents the *sunset value* of the team given outcome $\omega_{|\mathscr{T}|}$, i.e., the value of the team after the end of the planning horizon. The FTCP is hence

$$\max \ \mathbb{E}_\xi \left\{ \sum_{t\in\mathscr{T}} \sum_{p\in\mathscr{P}} \left[\frac{1}{(1+\rho)^{(t-1)}} V_{pt}(\omega_t) y_{pt}(\omega_t) \right. \right. \tag{1a}$$

$$-P_{pt}(\omega_t) y^P_{pt}(\omega_t) + S_{pt}(\omega_t) y^S_{pt}(\omega_t) \tag{1b}$$

$$+O_{pt}(\omega_t) l^O_{pt}(\omega_t) - I_{pt}(\omega_t) l^I_{pt}(\omega_t) \tag{1c}$$

$$\left. \left. -W_{pt}(\omega_t) \left(y_{pt}(\omega_t) + l^I_{pt}(\omega_t) - l^O_{pt}(\omega_t) \right) \right] + \frac{1}{(1+\rho)^{|\mathscr{T}|}} v(\omega_{|\mathscr{T}|}) \right\} \tag{1d}$$

subject to

$$y_{p1}(\omega_1) = Y_p + y^P_{p1}(\omega_1) - y^S_{p1}(\omega_1) \qquad p \in \mathscr{P}, \tag{1e}$$

$$y_{pt}(\omega_t) = y_{p,t-1}(\omega_t) + y^P_{pt}(\omega_t) - y^S_{pt}(\omega_t) \qquad p \in \mathscr{P}, t \in \mathscr{T} \setminus \{1\}, \tag{1f}$$

$$\sum_{p\in\mathscr{P}} y_{pt}(\omega_t) \le \bar{N} \qquad t \in \mathscr{T}, \tag{1g}$$

$$\sum_{p\in\mathscr{P}} \left[y_{pt}(\omega_t) + l^I_{pt}(\omega_t) - l^O_{pt}(\omega_t) \right] = N \qquad t \in \mathscr{T}, \tag{1h}$$

$$\sum_{p\in\mathscr{P}_r} \left[y_{pt}(\omega_t) + l^I_{pt}(\omega_t) - l^O_{pt}(\omega_t) \right] \ge N_r \qquad r \in \mathscr{R}, t \in \mathscr{T}, \tag{1i}$$

$$y_{pt}(\omega_t) - l^O_{pt}(\omega_t) \ge 0 \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1j}$$

$$y_{pt}(\omega_t) + l^I_{pt}(\omega_t) \le 1 \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1k}$$

$$l^O_{pt}(\omega_t) \le K^O_{pt} \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1l}$$

$$y^S_{pt}(\omega_t) \le K^S_{pt} \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1m}$$

$$l^I_{pt}(\omega_t) \le K^I_{pt} \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1n}$$

$$y^P_{pt}(\omega_t) \le K^P_{pt} \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1o}$$

$$A_{pt}(y_{pt}(\omega_t) + l^I_{pt}(\omega_t)) \le \bar{A}_p \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1p}$$

$$\sum_{p\in\mathscr{P}} \left[P_{pt}(\omega_t) y^P_{pt}(\omega_t) + I_{pt}(\omega_t) l^I_{pt}(\omega_t) \right.$$

$$\left. -S_{pt}(\omega_t) y^S_{pt}(\omega_t) - O_{pt}(\omega_t) l^O_{pt}(\omega_t) \right] \le B_t \qquad t \in \mathscr{T}, \tag{1q}$$

$$v(\omega_{|\mathscr{T}|}) = \sum_{p\in\mathscr{P}} V_{p|\mathscr{T}|}(\omega_{|\mathscr{T}|}) y_{p|\mathscr{T}|}(\omega_{|\mathscr{T}|}), \tag{1r}$$

$$y_{pt}(\omega_t), y^P_{pt}(\omega_t), y^S_{pt}(\omega_t), l^I_{pt}(\omega_t), l^O_{pt}(\omega_t) \in \{0,1\} \qquad p \in \mathscr{P}, t \in \mathscr{T}, \tag{1s}$$

$$v(\omega_{|\mathscr{T}|}) \geq 0. \tag{1t}$$

Objective function (1a)-(1d) consists of maximizing the expected net present value of the team, which includes the value of the players owned minus the money spent for buying and borrowing players and paying salaries, plus the income generated by selling and lending players. Notice that salaries are not paid for players lent to other clubs. The FTCP has, in principle, an infinite planning horizon which we approximate by $|\mathscr{T}|$ TMWs. This might generate an end-of-horizon effect which causes decision variables in the last stages to behave as if the corresponding decisions had no future consequences. To mitigate this effect we include the sunset value in the objective function. This is equivalent to stating that after $|\mathscr{T}|$ TMWs the decision maker wants to have a team of as high value as possible. Constraints (1e)-(1f) ensure the balance between players joining and leaving the club, for the first and the following TMWs, respectively. Constraints (1g) and (1h) ensure that the club does not own more than $\bar{N}$ players and that exactly N players are registered for competitions, respectively. Constraints (1i) ensure that the squad has at least N_r players in each role r. Constraints (1j) and (1k) ensure that players are lent only if owned, and borrowed only if not owned, respectively. Furthermore, constraints (1l), (1m), (1n), and (1o) state whether a player can be lent, sold, borrowed and purchased, respectively. Constraints (1p) ensure that players are not considered after their retirement. Constraints (1q) ensure that the net spending can be covered by the available budget. Constraint (1r) sets the sunset value as the value of the players owned at the end of the planning horizon. Finally, constraints (1s)-(1t) set the domain for the decision variables.

Finally, notice that the dependence of the decision variables on ω_t indicates that decisions are only made once the outcome of ω_t is known. That is, decisions at time t are only based on information available at t and on probabilistic information about $\xi_{t+1},\dots,\xi_{|\mathscr{T}|}$, i.e., they are *nonanticipative*. More formally, it can be said that decisions at time t are $\mathscr{F}_t(\xi_1,\dots,\xi_t)$-*measurable*, where $\mathscr{F}_t(\xi_1,\dots,\xi_t)$ is the sigma-algebra generated by $\xi_1,\dots,\xi_t$, with $\mathscr{F}_t(\xi_1,\dots,\xi_t) \subseteq \mathscr{F}_{t+1}(\xi_1,\dots,\xi_{t+1}) \subseteq \cdots \subseteq \mathscr{F}$.

4.2 Possible Stochastic Models for the Uncertain Parameters

Model (1) requires a probability distribution of ξ, the collection player vectors (i.e., the vectors containing future values, prices, loan fees, and salaries). Below we point out possible ways to estimate the probability distribution.

- **Distribution of forecasting errors**. It can be argued that football clubs rely on some projection of the player vector for the future when evaluating the acquisition or sale of players. This projection might be stated more or less formally and may or not be based on some analytic support. A probability distribution of future player vectors might be obtained by analyzing how accurate the club's projections have been in the past. That is, future player vectors may be represented as a function of a deterministic term representing the club's projection, and a random term representing the projection error. Then a probability distribution of past projection errors would suffice.

- **Property Matching**. If complete probability distributions cannot be achieved, it may be possible to estimate statistical properties of the player vectors. As an example, data might show that forwards who, by the age of 25, have scored less than 10 goals per season, are worth on average €15 million, with a standard deviation of €5 million. Property matching (see Høyland and Wallace (2001) and Høyland, Kaut, and Wallace (2003)) can be used to generate scenarios matching the statistical information available.

- **Handpicked scenarios**. If none of the above methods is readily applicable (e.g., when historical data is not available) a simple alternative is to rely on the club's expert opinion on possible future career developments for the players. That is, a club's manager may envisage alternative career scenarios and value developments for a player and the corresponding likelihoods. It can be argued that such endeavor is preferable to completely relying on a unique projection (King and Wallace, 2012, Ch.4).

Finally, if data is available, probability distributions may be estimated through regression analysis as we illustrate in Section 5.

5 Computational Study

In this section, we present a computational study based on the English Premier League (EPL). The scope of this section is: (a) to illustrate the expected team value development attainable with model (1), (b) to assess whether the model might help to improve current decisions and, finally, (c) to understand the drivers of the value development. When possible and meaningful, we compare the decisions and results of our model to those of the corresponding clubs in the EPL.

5.1 Instances

Instances are based on the case of the EPL 2013/14 using data available on TransferMarkt (see `www.transfermarkt.com`). Season 2013/14 is the most recent season which allows us to assess the model on a sufficiently long planning horizon. Our case study considers the 20 teams in the competition, see Table 1. Each team becomes in turn the *focal team*, that is, the team a FTCP is solved for while dealing with the summer 2014 TMW (TMW14 -

in preparation to season 2014/15). For each focal team, the set $\mathscr{P}$ is made of the players in the team during season 2013/14 and a number of target players. In general, target players are sensitive information known only to the club managers, thus not publicly available. Therefore, we generated target players based on the actual transactions made by the clubs. Particularly, we adopted the following procedure:

1. We sorted the 20 teams in decreasing order of their value during season 2013/14.
2. We divided the teams into groups A through E of similar value, see Table 1.
3. For each group we defined a set of target players as the set of all players purchased or borrowed by the teams in the group during TMW14.
4. Teams in the same group will have the same target players.

We believe establishing the set of target players in this manner reasonably approximates the players the focal club might consider. In fact, a subset of the target players are exactly the players the club bought or borrowed, and thus were certainly target players. In addition, it contains target players which were bought by teams with comparable purchase power and appeal. The set of target players is valid for the entire planning horizon. When solving the FTCP for different focal teams, it is possible for the purchase or rental of the same player to be recommended by the optimal solutions of multiple focal teams. Thus, the results we present should be interpreted as the outcome for the focal team had they implemented the solution to the FTCP.

Table 1: Teams in the English Premier League 2013/14

Group	Team	Group	Team
A	Chelsea FC (CH)	C	West Ham United (WH)
A	Manchester United (MU)	C	Stoke City (ST)
A	Manchester City (MC)	D	Swansea City (SW)
A	Arsenal FC (AR)	D	Aston Villa (AV)
B	Tottenham Hotspur (TO)	D	Southampton FC (SO)
B	Liverpool FC (LI)	D	Norwich City (NO)
B	Everton FC (EV)	E	Cardiff City (CC)
B	Newcastle United (NU)	E	West Bromwich Albion (WB)
C	Sunderland AFC (SU)	E	Hull City (HU)
C	Fulham FC (FU)	E	Crystal Palace (CP)

For the deterministic parameters of the problem we set the following default values. For each player, the role, age, and current market value are known. Roles correspond to positions on the field of play. All the roles considered by TransferMarkt have been included. The minimum number of players for each role (N_r) corresponds to the least number of players for the role among all the teams considered. Consistently with EPL regulations, $N = 25$ players are used for competitions while the maximum number of players is set to $\bar{N} = 48$ (the maximum team size among the 20 clubs). For the entire planning horizon, a player can be lent (K^O_{pt}), sold (K^S_{pt}), bought (K^P_{pt}), or borrowed (K^I_{pt}) if they were lent, sold, bought, or borrowed, respectively, during TMW14. The retirement age is set to 42 years for all players consistently with the age of the oldest player in the EPL 2013/14. All players reaching the age limit during the planning horizon are automatically considered sellable. The discount rate is set to 7%. The budget of the team B_t is set equal to the net spending of the actual club during TMW14. That is, we set $B_t = \max\{E - I, 0\}$, for each $t \in \mathscr{T}$, where E is the amount of the actual expenses of the club in TMW14 (actual player purchases and rentals) and I is the amount of the actual income of the club in TMW14 (actual player sales and loans). The same budget is valid for every TMW in the planning horizon. This ensures that our model cannot spend more than the actual club though avoiding representing a position of debt. We assess our budget assumptions in Section 5.6. The planning horizon $|\mathscr{T}|$ is three summer TMWs and every TMW is considered a new stage, that is, new information is obtained at each TMW. We believe this is a fair length for the planning horizon as it is arguably difficult to foresee the number and type of new players emerging from young selections, and thus updates to target lists. However, the impact of the length of the planning horizon and of the number of stages on the results is assessed in Section 5.6.

5.2 Modeling the Uncertainty

In this section we describe how we model the uncertain parameters for our case study. We remind the reader that uncertain parameters, summarized by the player vector $\xi_{pt}(\omega_t)$, for every $p \in \mathscr{P}$ and $t \in \mathscr{T}$, are the value of the player, the selling price, the purchase price, the fees for borrowing and lending, and the salary. First, we illustrate how we model the player value, and then we describe how we model the remaining parameters.

We use a regression model to forecast the future value of a player given current available information. While thorough regression analysis can itself be the subject of a much deeper investigation, for the scope of this paper we look for a regression model with sufficient forecasting ability. We use as explanatory variables the current value of the player, their age, and their role. The response variable is the value of the player during the next TMW. Particularly, we seek a regression model showing the following properties:

- sufficiently high proportion of the variability of the dependent variable predictable from the explanatory variables, measured by the coefficient of determination, R^2,
- sufficiently high statistical significance of the explanatory variables, measured by the p-value,
- non-negative response and no evidence of heteroscedasticity.

Regression analysis was performed using historical EPL data from season 2011/12 to season 2016/17 obtained from `www.transfermarkt.com`. Each of the 2095 records shows the value of a player, their current age and role, and value of the same player during the following season. After testing several regression models we decided upon the regression model described by equation (2):

$$\sqrt[4]{V_{pt}} = \alpha \sqrt[4]{V_{p,t-1}} + \beta A_{p,t-1} + \sum_{r \in \mathscr{R}} \gamma_r \delta(p, \mathscr{P}_r) + \varepsilon, \qquad p \in \mathscr{P}, t \in \mathscr{T} \setminus \{1\} \tag{2}$$

Particularly, the value of the coefficients of model (2) is summarized Table 2, while $\delta(p, \mathscr{P}_r)$ is the indicator function, i.e., $\delta(p, \mathscr{P}_r) = 1$ if $p \in \mathscr{P}_r$, 0 otherwise. The model has $R^2 = 0.9874$, the p-value is smaller than 2.2^{-16}, and the residuals-fitted plots do not show evidence of heteroscedasticity. Furthermore, the error term can be safely assumed normally distributed with $\varepsilon \approx N(0, 0.1791^2)$. Notice, in model (2), that the future player value, significantly depends on the current value and that, as intuition suggests, the value of the player decreases with age. Notice also that the model does not include an intercept term. This is mainly due to the difference of value between players in different roles. However, by means of the term in $\delta(p, \mathscr{P}_r)$, the model creates a role-specific intercept γ_r. Therefore, in what follows, for every TMW and for every player, the market value during the following TMW is a random variable described by equation (2).

As far as the other random parameters are concerned, the available 143 purchase records and 195 sales records for TMW14 show that the average purchase price to average value ratio is 1.22, while the average selling price to average value ratio 0.97. That is, it appears that, on average, the transaction fee is heavier on the buyer than on the seller. Note that the records include several transactions with clubs not in the EPL, thus a sale record does not necessarily correspond to a purchase record. Not many records of loan fees were available. The 17 available records suggest a loaning fee of approximately 15% of the current value. Therefore, in our computational study, we set the purchase and selling price of a player 1.22 and 0.97 times their current value, respectively, and the borrowing and lending fee 0.15 times their current value. Finally, we set the player salary as 10% of the current market value. Sensitivity analysis with respect to this parameter is reported in Section 5.6.

Coefficient	Value
α	0.860081
β	-0.024199
γ_{keeper}	0.883168
$\gamma_{attackingmidfielder}$	0.884657
$\gamma_{centralmidfielder}$	0.864462
$\gamma_{defensivemidfielder}$	0.874801
$\gamma_{leftwing}$	0.888214
$\gamma_{rightbacks}$	0.879782
$\gamma_{secondarystriker}$	0.870941
$\gamma_{centreback}$	0.867006
$\gamma_{leftback}$	0.866930
$\gamma_{rightmidfielder}$	0.754924
$\gamma_{centreforward}$	0.873991
$\gamma_{leftmidfielder}$	0.810373
$\gamma_{rightwing}$	0.856120

Table 2: Value of the coefficients of regression model (2)

5.3 Scenario Generation

Equation (2) generates a continuous probability distribution for $\xi = (\xi_t)_{t=1}^{|\mathscr{T}|}$, which in turn makes model (1) intractable. Therefore, in order to solve model (1), the continuous distribution is approximated by means of a *scenario tree* where each scenario represents a complete realization of ξ for the entire planning horizon. An example three-stage scenario tree in depicted in Figure 2. The example scenario tree contains two possible conditional realizations of ξ_t of per stage (i.e., four scenarios $S1,\ldots,S4$).

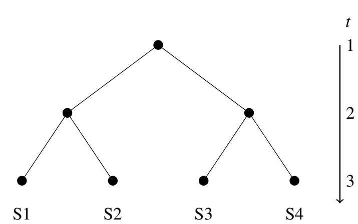

Figure 2: Example three-stage scenario tree

(A) 24 year-old center forward (B) 33 year-old center back

Figure 3: Example scenario trees for two players

We obtain scenarios by sampling, for each stage and for each player, realizations of ε from the underlying $N(0, 0.1791^2)$ distribution and thus generating the corresponding ξ_{pt} realizations. Particularly, we use 30 samples for two-stage problems, and 18 samples per stage (324 scenarios) for three-stage problems. These sample sizes were chosen by performing the in-sample stability test described by Kaut and Wallace (2007). Particularly, the in-sample stability test ensures that the results discussed in Section 5.5 are not biased by the specific set of scenarios used, i.e., the optimal objective value is approximately the same for any sample of the same size. Therefore, using more scenarios would not lead to stability improvements which would justify the increased complexity. Figure 3 reports two example scenario trees. It can be noticed that for the 33-year old player, most scenarios envisage a value decline by the end of year 3.

5.4 Size of the Problem and Complexity

We implemented the FTCP using a node formulation. That is, we created one copy of the decision variables for each node in the scenario tree (see, e.g., Pantuso, Fagerholt, and Wallace (2015) for an example node formulation). Thus, for instance, variables $y_{pt}(\omega_t)$ are replaced by variables y_{pn}, where n is a node in the scenario tree. As an example, the scenario tree in Figure 2 contains seven nodes, thus we would have variables y_{pn} for $p \in \mathscr{P}$ and $n \in \{1, \ldots, 7\}$. Similarly, the constraints of the FTCP must hold for every node in the scenario tree. In general, node formulations have significantly fewer decision variables and constraints than scenario formulations. In scenario formulations, a copy of the decision variables is created for each stage and scenario. As an example, $y_{pt}(\omega_t)$ would be replaced by y_{pts} where s is a scenario, and the constraints of the FTCP must hold for every scenario. However, given a number of scenarios S, the number of nodes in a scenario tree is (significantly) smaller than $|\mathscr{T}| \times S$. The node formulation corresponding to problem (1) is provided in Appendix A.

The problems we solved for the results in Section 5.5 have 105,000 variables on average (minimum 82,000 and maximum 126,000) and 173,000 constraints on average (minimum 137,000 and maximum 208,000). In the sensitivity analysis of Section 5.6, the biggest problem we solved (i.e., the case with four periods and three stages) had on average 205,000 variables and 339,000 constraints.

Model (1) was implemented and solved using Cplex 12.6.2 Java callable libraries. All tests have been run on a Linux machine with 24.7 GB RAM and Six-Core AMD Opteron 2.4 GHz CPU. We could solve the base case to an optimality gap of 0.5% in less than two-hours in the worst case.

5.5 Results

We present the results obtained by solving an instance of the FTCP for each team in the EPL 2013/14 while dealing with TMW14. Figure 4 illustrates the three-year team value development in the solution of model (1) and in real life by the corresponding teams. In addition, error bars show the domain of the team value distribution for the second and third stage (we remind the reader that future team values are scenario-dependent). A number of elements emerge from Figure 4. The first observation is that the expected team value provided by model (1) is almost always higher than the actual team value realized by the clubs (except for the cases of Crystal Palace, Southampton FC and West Ham United upon which we comment later). The second observation can be made by looking at the error bars in Figure 4. The error bars show, for seasons 2015/16 and 2016/17, the range of values of the team value attainable by implementing the solutions to model (1). It can be noticed that the actual realized value for a team often falls in the lowest part of the team value distribution, or even below. This corresponds to saying that the solutions to the FTCP exhibit a high probability of performing better than actual solutions. In some cases, solutions to the FTCP perform better than actual realizations even in the worst-case realization of the player values (see e.g., the cases of Aston Villa, Chelsea FC and Fulham FC). Thus, Figure 4 suggests that model (1) has potential to help decision makers improve current decisions, provided that the assumptions made reflect the actual preferences and operating scenario of the actual clubs.

Figure 4: Expected team value progression in the solutions to the FTCP versus actual (realized) team values. Error bars indicate the support of the team value distribution for each season.

It can also be noticed that cases exist where actual decisions made by the clubs outperform expected results from model (1). Cases like these are completely normal. Stochastic programs provide solutions bearing the scenario-wise distribution of the results (team value in this case) with the highest expected value. Given any individual realized scenario, there may be a different course of action that performs better than the solution that maximizes the expected value. However, the expectation of such solution is not higher than the one provided by the solution to the stochastic program. With this in mind, we can understand the results of Crystal Palace, Southampton FC and West Ham United in Figure 4. That is, the decisions made by the clubs turned out to be better than the solution to the FTCP for the particular scenario which materialized. However, beforehand, such solutions would have had a lower expected value than the solutions to the FTCP.

As an example, we tested the effects of forcing model (1) to make, for the case of Southampton FC, the same purchases and sales made by the actual club. For this case, we obtained the expected team value is lower than the optimal by €10.75 million in season 2014/15, €12.45 million in season 2015/16, and €14.18 million in season 2016/17.

Table 3: Compound Annual Growth Rate in real life and in the model

Team	Actual CAGR[%]	Model Expected CAGR [%]
Arsenal FC	6.65	9.84
Aston Villa	-7.58	25.07
Cardiff City	-16.48	30.24
Chelsea FC	-0.24	6.50
Crystal Palace	21.64	33.99
Everton FC	11.83	10.62
Fulham FC	8.93	21.17
Hull City	-11.46	24.31
Liverpool FC	10.03	11.52
Manchester City	3.12	1.03
Manchester United	1.62	0.82
Newcastle United	-3.07	17.16
Norwich City	-1.75	21.12
Southampton FC	10.87	22.00
Stoke City	9.94	14.35
Sunderland AFC	-4.78	16.58
Swansea City	1.23	16.59
Tottenham Hotspur	7.93	8.42
West B. Albion	3.20	19.39
West Ham United	25.99	17.17

Finally, we can observe that the expected growth of the value of the team tends to be steady, with the final expected value of the team being higher than the initial value in almost all cases, and significantly higher in some cases. Even if the solution to the FTCP is not the best in the short run (see e.g., the cases of Hull City and West Bromwich Albion), the team value growth is maintained throughout the planning horizon and the solution turns out better in the long run. This is confirmed also by the expected *Compound Annual Growth Rate* (CAGR) from season 2014/15 to season 2016/17 reported in Table 3. The CAGR represents the average annual growth rate of an investment for a given period of time. It is calculated as $(\mathscr{F}/\mathscr{I})^{(1/|\mathscr{T}|)} - 1$, where $\mathscr{F}$ represents the final value of the team and $\mathscr{I}$ represents the initial value of the team, i.e., the value of the team before TWM14. It can be noticed that the expected CAGR provided by the FTCP is always positive and in most cases higher than the actual CAGR.

Figure 5: Number of purchases and sales and their average age in the first-stage solution to the FTCP and in real life.

Patterns in the solutions to the FTCP help explain the drivers of the expected team value growth. Figure 5 reports the first-stage solution to model (1) in terms of number of players bought and sold as well as average age of players bought and sold. In addition, it reports the same information for the actual decisions made by the corresponding clubs. Model (1) suggests, in general, a comparable but smaller number of purchases and sales, thus fewer transactions, with respect to actual decisions. Transactions are performed only when these lead to a higher expected team value. In addition, the model tends to suggest buying at an age slightly above 20 and selling at an age close to 30. In most cases, the model buys at an age younger than actual decisions, and sells at an age older than actual decisions. Thus, a key for expected team value growth is to be found in fewer but targeted investments in cheaper, younger players with high growth potential.

5.6 Sensitivity of the Results

We illustrate how the results are influenced by the available budget, by the salary of the players, by the length of the planning horizon and by the number of stages.

Figure 6 illustrates how the expected team value changes as a function of the available budget. Particularly, it reports the progression of the expected team value when the base budget (see Section 5.1) is incremented and decremented by €10 and €20 million. As intuition suggests, higher budgets foster higher values of the team due to the increased capability of hiring valuable and high potential players. However, it can also be observed that, for certain teams with already high budgets (e.g., Liverpool FC and Manchester United), higher budgets help increasing the here-and-now value of the team but do not necessarily ensure a steeper growth and a higher final value. This has to do with the fact that, in our instances, as in real life, the number of talents with high growth potential is not unlimited. On the other hand, teams with smaller budget obtain, by means of extra budget amounts, slightly steeper growth rates and, in general, higher final values (see, e.g., Fulham FC and Cardiff City). Finally, notice that for certain teams with an already low base-case budget, the value progression corresponding to a budget decrease is not shown. This is due to the fact that model (1) becomes infeasible, i.e., the budget is not sufficient to ensure a team composition with the desired characteristics (see, e.g., Norwich City and Stoke City). In real-life, the team would have to revisit the requirements of the coach or consider additional target players of lower current value. Finally, Figure 6 sheds light on how model (1) could be used by actual clubs to support budget allocation decisions, by assessing the expected team value progression as a function of the budget allocated.

Figure 6: Expected team value for the base case and with a budget increase of €10 and €20 million, and a budget decrease of €10 and €20 million.

As far as salaries are concerned, the base case assumes that the salary of a player is 10% of their value (see Section 5.1). We also run the tests with salaries set to 5%, 15%, and 20% of the value of the player. However, the results illustrated for the base case are to a very large extent confirmed also with the other salary levels and no significant changes in the team value or in the solutions could be observed.

Finally, we tested the sensitivity of the results to the number of stages and periods. That is, we test the effects of considering a longer planning horizon, as well as considering fewer stages than periods, i.e., assuming that from some TMW on the decision maker obtains perfect knowledge about the future. Particularly, considering fewer stages than periods is a simplification which reduces the size of the problem but approximates the original problem. We remind the reader that the base case has three periods and three stages, that is, new information is obtained at every TMW. Particularly, we assessed the following cases: three periods and two stages, four periods and two stages and four periods and three stages. It emerges that the two-stage problems, both in the case with three and four periods, significantly overestimate the future value of the team. Particularly, for the two-stage three-period case, the expected team value in season 2016/17 is on average 13.7% (approximately €35 million) higher than that reported by three-stage three-period models. This is due to the fact that, in two-stage problems, the future after the first TMW is deterministic. Thus, the problems can adapt their solutions to the specific scenario in the best possible way, since there is no more uncertainty to deal with. However, such situation does not correspond to what happens in real-life decision problems, where the decision maker faces uncertainty at every TMW. This also illustrates the benefit from modeling the interplay between decisions and uncertainty in a more strict way.

6 Conclusions and Future Research

In this article we formally described the decision problem, faced by football clubs, of investing in professional football players. We introduced a novel stochastic programming model for composing a football team with the desired mix of skills such that the expected value of the team is maximized. In addition, the model takes into account competition regulations and budget limits, as well as the uncertainty in the future market value of the players.

The model has been tested on a case study based on the English Premier League 2013/2014. Our results show that the model has a significant potential to improve real-life decisions in actual football clubs. Solutions to the model ensure a steady growth of the value of the team generated primarily by investments in young prospects. The expected growth is in most cases higher than the actual growth of the corresponding real-life teams. Results also show that the budget is an important driver of growth, and that reducing the number of stages leads to overestimated expected team values. In addition, a number of future research avenues can be pointed out.

The FTCP introduced in this paper assumes that ongoing contracts exist between the club and the players owned. However, contract length and contract renewal decisions also impact team composition decisions. As an example, it is arguable that clubs prefer to evenly spread the expiration of ongoing contracts as this would prevent replacing too many players between two seasons. In addition, the duration of the contracts or, more precisely, the time-to-expiration, has an impact on selling prices. In fact, potential buyers gain market power due to the possibility of signing the player as a free agent after the expiration of the contract. Therefore, a possible extension of the FTCP might include decisions regarding the length and renewal of contracts.

In this paper we used the market value as an indicator of the financial impact the players. A possible extension of this work might look at ways to explicitly consider on-the-field performance. Performance is typically quantified by means of a set of role-specific and numerically heterogeneous statistics. Additional research is needed to shed light on: (1) how to sum and weigh numerically different statistics (e.g., goals scored and pass accuracy) in an attempt to obtain an overall measure of team performance, (2) how to quantify intangible performances such as "making the correct movement" and "occupying the correct position", and (3) how to translate individual performance into competitiveness of the team. In the model we presented it is possible to enforce some level of (expected) performance by requiring a certain number of players with a sufficient value of given statistics through constraints (1i).

Recently, a new type of transaction is emerging where a club purchases a player but leaves them to the selling club (as a loan) for one or more seasons. This is especially done by top clubs who purchase young prospects but leave them to smaller clubs until they reach maturity. Modeling the joint purchases-and-loan might improve the realism of the model.

Finally, one implicit assumption in this paper is that the probability distributions of the player values are independent of decisions. However, one might argue for cases in which the distribution of the value of a player is influenced by the team they play for or for a positive correlation between the value of players in the same team. To model such situations one requires multistage stochastic programs with endogenous uncertainty. To the best of our knowledge, methods have been proposed for the cases where decisions influence the timing of new information (Apap and Grossmann, 2016) and where decisions influence the probability measure on a fixed sample space in a two-stage problem (Laumanns, Prestwich, and Kawas, 2014; Peeta, Salman, Gunnec, and Viswanath, 2010). However, no tractable method for multistage stochastic programs where decisions change the probability distribution at every stage is available. Thus, in order to assess the impact of decisions influencing the stochastic values of the players, a methodological advancement of the field beyond the scope of this paper is necessary.

References

R Adkins and D Paxson. Deterministic models for premature and postponed replacement. *Omega*, 41: 1008–1019, 2013.

R M Apap and I E Grossmann. Models and computational strategies for multistage stochastic programming under endogenous and exogenous uncertainties. *Computers & Chemical Engineering*, 103:233–274, 2016. ISSN 0098-1354.

A Becker and X A Sun. An analytical approach for fantasy football draft and lineup management. *Journal of Quantitative Analysis in Sports*, 12:17–30, 2016.

J R Birge and F Louveaux. *Introduction to Stochastic Programming*. Springer, New York, 1997.

B H Boon and G Sierksma. Team formation: matching quality supply and quality demand. *European Journal of Operational Research*, 148:277–292, 2003.

A Bouchet, T W Doellman, M Troilo, and B R Walkup. The impact of international football matches on primary sponsors and shareholder wealth. *Journal of Sport Management*, 29:200–210, 2015.

P. D. Bruecker, J Van den Bergh, J Belien, and E Demeulemeester. Workforce planning incorporating skills: State of the art. *European Journal of Operational Research*, 243:1–16, 2015.

David R Cariño, Terry Kent, David H Myers, Celine Stacy, Mike Sylvanus, Andrew L Turner, Kouji Watanabe, and William T Ziemba. The Russell-Yasuda Kasai model: an asset/liability model for a Japanese insurance company using multistage stochastic programming. *Interfaces*, 24:29–49, 1994.

Timothy C Y Chan, Justin A. Cho, and David C. Novati. Quantifying the contribution of NHL player types to team performance. *Interfaces*, 42(2):131–145, 2012. ISSN 00922102.

S Chand, T McClurg, and J Ward. A model for parallel machine replacement with capacity expansion. *European Journal of Operational Research*, 121:519–531, 2000.

T.-J. Chang, N. Meade, J.E. Beasley, and Y.M. Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302, 2000. doi: 10.1016/S0305-0548(99)00074-X.

Deloitte. Football money league: top of the table. Report, Deloitte, 2016.

S Dobson and B Gerrard. The determination of player transfer fees in english professional soccer. *Journal of Sport Management*, 13:259, 1999.

S Dobson and J A Goddard. *The Economics of Football*. Cambridge University Press Cambridge, 2001.

FIFA. *Regulations on the status and transfer of players (October 2015)*. FIFA, Zurich, 2015.

FIGC. *Comunicato ufficiale 84/A*. FIGC, Rome, 2014.

Forbes. Real Madrid tops ranking of the World's most valuable soccer teams. Online, 2015a. URL `http://www.forbes.com/sites/mikeozanian/2015/05/06/real-madrid-tops-ranking-of-the-worlds-most-valuable-soccer-teams/{\#}75c570e43b36`.

Forbes. Cristiano Ronaldo edges Lionel Messi as the World's highest-paid soccer player. Online, 2015b. URL `http://www.forbes.com/sites/christinasettimi/2015/05/06/cristiano-ronaldo-edges-lionel-messi-as-the-worlds-highest-paid-soccer-player/{\#}5ade9ada56f0`.

Michael J. Fry, A W Lundberg, and J W Ohlmann. A player selection heuristic for a sports league draft. *Journal of Quantitative Analysis in Sports*, 3:1–35, 2007.

Matthew R. Gibson, Jeffrey W. Ohlmann, and Michael J. Fry. An agent-based stochastic ruler approach for a stochastic knapsack problem with sequential competition. *Computers and Operations Research*, 37(3): 598–609, 2010. doi: 10.1016/j.cor.2009.02.028.

W Hopp and S K Nair. Markovian deterioration and technological change. *IIE Transactions*, 26:74–82, 1994.

K Høyland and Stein W. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

K Høyland, M Kaut, and Stein W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24:169–185, 2003.

P Kall and Stein W. Wallace. *Stochastic Programming*. John Wiley and Sons Ltd, Chichester, 1994.

M Kaut and Stein W. Wallace. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3:257–271, 2007.

A J King and Stein W. Wallace. *Modeling with Stochastic Programming*. Springer New York, New York, 2012.

S Kirshner. Optimizing an NBA team's approach to free agency using a multiple-choice knapsack model. In *MIT Sloan Sports Analytics Conference 2011*, 2011.

Anton J. Kleywegt and Jason D. Papastavrou. The dynamic and stochastic knapsack problem. *Operations Research*, 46(1):17–35, feb 1998. doi: 10.1287/opre.46.1.17.

Komarudin, M.-A. Guerry, T De Feyter, and G Vanden Berghe. The roster quality staffing problem – A methodology for improving the roster quality by modifying the personnel structure. *European Journal of Operational Research*, 230:551–562, 2013.

Marco Laumanns, Steven Prestwich, and Ban Kawas. Distribution shaping and scenario bundling for stochastic programs with endogenous uncertainty. In J L Higle, W Römisch, and S Sen, editors, *Stochastic Programming E-Print Series*, Stochastic Programming E-Print Series. Institut für Mathematik, 2014.

R Mansini, W Ogryczak, and M G Speranza. *Linear and mixed Integer programming for portfolio optimization*. Springer International Publishing, 2015.

Ove Mørch, K Fagerholt, Giovanni Pantuso, and Jørgen Rakke. Maximizing the rate of return on the capital employed in shipping capacity renewal. *Omega*, 67:42–53, 2017.

John M Mulvey and Bala Shetty. Financial planning via multi-stage stochastic optimization. *Computers & Operations Research*, 31:1–20, 2004.

E Ozceylan. A mathematical model using AHP priorities for soccer player selection: a case study. *South African Journal of Industrial Engineering*, 27:190–205, 2016.

Giovanni Pantuso, K Fagerholt, and Stein W. Wallace. Solving hierarchical stochastic programs: application to the maritime fleet renewal problem. *INFORMS Journal of Computing*, 27:89–102, 2015.

S Peeta, F S Salman, D Gunnec, and K Viswanath. Pre-disaster investment decisions for strengthening a highway network. *Computers & Operations Research*, 37:1708–1719, 2010.

S Rajagopalan and A C Soteriou. Capacity acquisition and disposal with discrete facility sizes. *Management Science*, 40:903–917, 1994.

G Sierksma. Computer support for coaching and scouting in football. In *The Engineering of Sport 6*, pages 215–219. Springer, 2006.

Madjid Tavana, Farshad Azizi, Farzad Azizi, and Majid Behzadian. A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review*, 16(1): 97–110, 2013. ISSN 14413523.

UEFA. *Regulations of the UEFA Champions League 2015 – 18 Cycle*. UEFA, Geneva, 2015a.

UEFA. *UEFA Club licensing and financial fair play regulations editions 2015*. UEFA, Geneva, 2015b.

Appendix A A Node Formulation for the FTCP

In order to obtain a node formulation of model (1) the notation presented in Section 4 must be adjusted and complemented. Given a scenario tree which represents the uncertain parameters (see Section 5.3), let $\mathscr{N}$ be the set of nodes in the scenario tree, where 0 is the root node, and $\mathscr{N}^L \subset \mathscr{N}$ is the set of leaf nodes, that is the nodes at the last stage. Let $a(n)$ be the parent node of node n, Π_n the probability of node n, and $t(n) \in \mathscr{T}$ the stage (TMW) corresponding to node n. A copy of the decision variables is created for each node in the scenario tree. Particularly, variables $y_{pt}(\omega_t)$ are replaced by variables y_{pn}. Notice, therefore, that the combination stage-realization, that is $t-\omega_t$ is fully captured by the node n. Thus, a node $n \in \mathscr{N}$ represents a possible state at the corresponding TWM, and captures a joint realization of the random variables. Similarly, variables $y^P_{pt}(\omega_t)$, $y^S_{pt}(\omega_t)$, $l^I_{pt}(\omega_t)$, $l^O_{pt}(\omega_t)$, and $v(\omega_t)$ are replaced by y^P_{pn}, y^S_{pn}, l^I_{pn}, l^O_{pn} and v_n, respectively. In a similar fashion, all random variables in model (1) are replaced by the corresponding node realizations. As an example, the player value $V_{pt}(\omega_t)$ is replaced by a number of realizations V_{pn}, with $t(n)=t$. Instead, the deterministic parameters remain unchanged. The node formulation corresponding to model (1) is thus:

$$\max \sum_{n\in\mathscr{N}} \sum_{p\in\mathscr{P}} \Pi_n \Bigg[\frac{1}{(1+\rho)^{(t(n)-1)}} V_{pn} y_{pn} \tag{3a}$$

$$-P_{pn} y^P_{pn} + S_{pn} y^S_{pn} \tag{3b}$$

$$+O_{pn} l^O_{pn} - I_{pn} l^I_{pn} \tag{3c}$$

$$-W_{pn}\left(y_{pn} + l^I_{pn} - l^O_{pn}\right)\Bigg] + \tag{3d}$$

$$\sum_{n\in\mathscr{N}^L} \Pi_n \frac{1}{(1+\rho)^{|\mathscr{T}|}} v_n \tag{3e}$$

subject to

$$y_{p0} = Y_p + y^P_{p0} - y^S_{p0} \qquad p\in\mathscr{P}, \tag{3f}$$

$$y_{pn} = y_{p,a(n)} + y^P_{pn} - y^S_{pn} \qquad p\in\mathscr{P}, n\in\mathscr{N}\setminus\{0\}, \tag{3g}$$

$$\sum_{p\in\mathscr{P}} y_{pn} \le \bar{N} \qquad n\in\mathscr{N}, \tag{3h}$$

$$\sum_{p\in\mathscr{P}} \left[y_{pn} + l^I_{pn} - l^O_{pn}\right] = N \qquad n\in\mathscr{N}, \tag{3i}$$

$$\sum_{p\in\mathscr{P}_r} \left[y_{pn} + l^I_{pn} - l^O_{pn}\right] \ge N_r \qquad r\in\mathscr{R}, n\in\mathscr{N}, \tag{3j}$$

$$y_{pn} - l^O_{pn} \ge 0 \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3k}$$

$$y_{pn} + l^I_{pn} \le 1 \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3l}$$

$$l^O_{pn} \le K^O_{p,t(n)} \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3m}$$

$$y^S_{pn} \le K^S_{p,t(n)} \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3n}$$

$$l^I_{pn} \le K^I_{p,t(n)} \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3o}$$

$$y^P_{pn} \le K^P_{p,t(n)} \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3p}$$

$$A_{p,t(n)}(y_{pn} + l^I_{pn}) \le \bar{A}_p \qquad p\in\mathscr{P}, n\in\mathscr{N}, \tag{3q}$$

$$\sum_{p \in \mathscr{P}} \left[P_{pn} y^P_{pn} + I_{pn} l^I_{pn} \right.$$

$$\left. - S_{pn} y^S_{pn} - O_{pn} l^O_{pn} \right] \leq B_{t(n)} \qquad n \in \mathscr{N}, \qquad (3r)$$

$$v_n = \sum_{p \in \mathscr{P}} y_{pn}, \qquad n \in \mathscr{N}^L, \qquad (3s)$$

$$y_{pn}, y^P_{pn}, y^S_{pn}, l^I_{pn}, l^O_{pn} \in \{0,1\} \qquad p \in \mathscr{P}, n \in \mathscr{N}, \qquad (3t)$$

$$v_n \geq 0 \qquad n \in \mathscr{N}^L. \qquad (3u)$$