Master Thesis

Wages in the Major League Soccer

The correctness of wages

Erasmus School of Economics
Master Program Behavioural Economics 2014-2015
Author: BSc. Tom Verschoor (341808)
Supervisor: Dr Thomas Peeters
Date: 17-06-2015

Table of contents

- **Preface** **3**
- **Abstract** **4**
- **Introduction** **4**
- **Study field** **6**
- **Theoretical background** **7**
  - Moneyball 7
  - Wages 8
  - Heuristics in Decision Making 9
- **Data** **11**
  - Datasets 11
  - Variables 12
- **Methodology** **13**
  - Principal component analysis 13
  - Team performance 14
  - Individual wages 15
- **Results** **16**
  - Principal component analysis 16
  - Team performance 18
  - Individual wage 20
- **Conclusion** **23**
- **Recommendations** **24**
- **References** **26**
- **Appendices** **30**
  - Appendix A: Principal component analysis 30

Preface

When I heard that you cannot become rich as a journalist I decided to shift my focus from sports journalism to economics. My choice to study Economics & Business Economics at the Erasmus University appeared to be a great one. All I learned during the past five years, offers a sound basis for my further professional career. I am glad I am able to finish my career as a student by combining my two passions: sports and economics. This was not able without the help of Dr Thomas Peeters, who has experience in the field and helped me very good during the definition of my research topic. Thanks to Dr Thomas Peeters I took the risk to create a big dataset on my own and to dive in a blank research field. I did not take the easy way, but as my deceased granddad would have said: “I did it my way”. With hard work and dedication I am proud to present my thesis. I hope you will enjoy reading it as much as I have enjoyed writing it.

Abstract

This thesis examines the relationship between team performance, individual wages and individual player skills in MLS. This research shows that MLS teams that spend more on wages than its competitors, have a higher chance of obtaining a good ranking. When attracting high-wage players, teams should take the wage disparity into consideration. Higher wage disparity has a negative impact on team performance. Although goalkeeper shot stopping skills affect the team performance positively, shot stopping skills are undervalued in the wages of players. MLS players with above-average chance creating skills earn the highest wages, which is justified by the contribution these skills deliver to the eventual team performance. The player revenue of chance creating skills is even too low compared to the contribution those skills deliver to the team performance.

Introduction

Soccer is the most popular sport in the world and always gets a lot of attention, even by academics. The latest trend in soccer analytics is that every movement on the pitch is registered and analysed in a statistical way, but what happens in the board room? Lewis (2003) claimed that the valuation of skills in the market for baseball players was grossly inefficient. The publication of his book, Moneyball, triggered a firestorm of criticism from baseball insiders (Lewis, 2004). Furthermore it raised the eyebrows of many economists as well, basic price theory implies a tight correspondence between pay and productivity when markets are competitive and rich in information (Hakes & Sauer, 2006). Hakes and Sauer (2006) showed that the baseball labour market indeed was inefficient at the time Moneyball got published, they also showed that the mispricing diminished thereafter. It appeared that getting *on-base* without hitting the ball was undervalued compared to *slugging,* total bases divided by at-bats. Slugging relates to hitting the ball, while getting on-base can be related to not hitting the ball. In many situations animals and human beings have surprising difficulty noticing and using information provided by the absence or non-occurrence of something (Hearst, 1991). In behavioural economics this phenomenon is described by Tversky and Kahneman (1973) as the availability bias. In soccer, in general, attackers score goals where defenders prevent their opponents from scoring goals. This is obviously a visible and measurable action versus an invisible action which is hard to measure.

Contrary to basketball and American Football matches, in soccer matches goals and goal attempts are scarce. Therefore attackers are being worshipped by soccer fans and coveted by soccer clubs worldwide. Looking at the list of most expensive transfers is looking at the list of the best goal scorers in the history of soccer (Anderson & Sally, 2013). This also applies to the list of winners of the FIFA Ballon d'Or, the award for best soccer player of the year. Since the German defender Franz Beckenbauer won the award in 1976, only two defensive players lifted the FIFA Ballon d'Or or the France Football Ballon d'Or[1]. In the history of the awards, the France Football Ballon d'Or has been awarded from 1956 till 2009, only one goalkeeper won an award. The winner of the FIFA Ballon d'Or is determined by votes from other soccer players, soccer coaches and soccer journalists. From 2009 the FIFA also rewards the best goal of the year with the Puskás Award, there is not a similar award for the best save or other defensive action (FIFA, n.d.[a]). Despite all this attention for attackers, Anderson and Sally (2013) state that scoring the most goals does not necessarily mean that a team will be crowned champion. In twenty seasons of the German Bundesliga, the Italian Serie A, the British Premier League and the Spanish

[1] The FIFA Ballon d'Or and the France Football Ballon d'Or are combined since 2010.

Primera División this was only in 51% of the seasons the case. The team with the least amount of goals against won the championship in 46% of the cases. Between 2001 and 2011 scoring goals and preventing the opponent from scoring goals contribute in the same degree to the amount of wins in the Premier League. In preventing a loss, not suffering a goal against was 33% more valuable than scoring a goal. These numbers imply that defensive players should not be undervalued by the media and the fans, compared to the attackers.

This thesis combines the results of the Major League Soccer (MLS), the player wages in MLS and the individual skill attributes MLS players possess in the years 2008-2014. Because of the availability of the needed data, the wages in MLS are discusses in this thesis. Transfer fees involve noise since player agents try to make money when concluding the agreement and wages of European teams are not publicly available. Furthermore transfer fees depend on contractual specifies like the duration of the contract and break-out clauses.

This thesis provides the answer to the following research questions:

- In which way are wages of MLS players justified by their team performance and individual skill attributes?
  - In which way does the height of wages relate to the eventual team performance?
  - How do the individual skill attributes of players contribute to the eventual team performance?
  - Which individual skill attributes are rewarded in terms of the height of wages?

The corresponding hypotheses are as follows:

- Defensive and less visible individual skill attributes make a contribution to the team performance, but are undervalued by MLS clubs.
  - The more a MLS club spends on wages, the better their ranking will be.
  - Defensive and less visible individual skill attributes make a contribution to the team performance, just as much as attacking and visible skill attributes do.
  - MLS clubs invest more in attacking and visible skill attributes than defensive and less visible skill attributes.
    - Field players earn higher wages than goalkeepers.

To answer these questions and test these hypotheses, multiple regressions and tests have been performed. The regressions, tests and the explanation behind them can be found in the Methodology section.

This thesis is structured in the following way. The next section describes the study field in which research has been done. Hereafter literature on the study field is discussed; this section provides an explanation for the formulated research questions and hypotheses. This section is followed by the data, the variables and the methodology that are used to answer the research questions. Hereafter the results of the tests and regressions are presented with the answers to the research questions. The thesis ends with the conclusion and recommendations for further research.

Study Field

“At the professional level, spectator sports have become a staple of American traditions and culture. Families gather around the television or endure hours of traffic to support their favourite teams, while these athletes are moulded into international celebrities. The process of engaging in athletics in youth and developing those skills with age has fostered the growth of talented professionals that provide athletic entertainment on a national scale. As they learn to walk and talk, kids in America also learn to run, throw and kick” (Santchurn, 2012). Sports also play a major role in high schools and universities all across the country, with schools often providing scholarships to the most talented athletes. Professional sports in the United States are dominated by: American Football (NFL), baseball (MLB), basketball (NBA) and ice hockey (NHL). The Major League Soccer (MLS) is the highest professional soccer league in both the United States and Canada. Ten teams, equally divided among the Eastern and Western Conference, launched MLS in 1996 (MLS, n.d.[a]). MLS reportedly lost $250 million in its first five years, but since the deep run of the United States in the FIFA World Cup of 2002 the popularity of soccer in the United States is rising. At the time of writing, 20 teams take part in MLS and this increase in the amount of teams relates to the increase of popularity of MLS. Increased popularity helped MLS sign a big TV deal with ESPN in 2006, which provided the league’s first media rights fee.

A Major League Soccer club’s first team roster is comprised of up to 30 players. Players occupying roster slots 1-20 count against the club its salary budget. Players occupying roster spots 1-24 will earn at least $44,000 in 2012. Players occupying roster spots 25-30 will earn at least $33,750 in 2012. And any player making $33,750 must be under the age of 25. In addition to the salary budget, each MLS club spends additional funds on player compensation including money from a league-wide allocation pool, the cost of Designated Players outside the salary budget, and money spent on the roster slots 21-30. Allocation Money is money that is available to a club in addition to its salary budget. Each MLS club receives an annual allotment of Allocation Money. A club will also receive Allocation Money if it fails to qualify for the MLS Cup Playoffs, if a player is transferred to a club outside of MLS or if it qualifies for the CONCACAF Champions League. This Allocation Money is tradable. The Designated Player Rule allows clubs to acquire up to three players whose salaries exceed the maximum budget charge, with the club bearing financial responsibility for the amount of compensation above each player’s budget charge. From 2012 Designated Players 23 years or younger carry a lower salary budget charge than the ‘regular’ Designated Players, and Designated Players 20 years or younger carry an even lower salary budget charge. The maximum budget charge for a single player in 2012 was $350,000. Each MLS club has the right to have eight international players on their roster. These spots are tradable among the clubs and there is no limit on the amount of spots per roster. The remaining spots are filled with domestic players. For the MLS clubs based in Canada, Canadian and U.S. players are domestic players (MLS, 2012).

By 2007 MLS was successful enough to attract English superstar David Beckham. The designated player rule that was formed to facilitate the entry of David Beckham has since helped MLS add other popular soccer players like Thierry Henry and Kaká. The salaries of designated players are the only salaries which can exceed cap restrictions, with the extra money landing on the team’s book. Each player’s contract is signed with and paid by MLS and a salary cap ensures control of the player costs. In 2011, average MLS attendance surpassed both the NBA (basketball) and NHL (ice hockey) and it has since grown even further (Smith, 2013). Lawson, Sheehan & Stephenson (2008) estimate that David Beckham doubled ticket sales for his MLS games during 2007, home and away. In 2009 the individual jerseys of designated players were the most popular. The top sellers have been three midfielders and two attackers: David Beckham, Cuauhtémoc Blanco, Freddie Ljungberg, Landon Donovan, and Juan Pablo Ángel (Keh, 2009). These facts indicate that the designated players, who earn the highest wages, are the most popular players among soccer fans.

In the ongoing 2015 season, 20 MLS teams will play 34 games each. Since the league is divided in a Western and an Eastern Conference, the schedule is unbalanced. After 34 games, the top six from each conference qualifies for the MLS Cup Playoffs (MLS, 2015[a]). The winner of the MLS Cup is crowned champion. Within the soccer league, there is no promotion or relegation process. This thesis focuses on the player salaries in MLS from 2008 to 2014 and their relation to the results of clubs and the abilities of players.

Theoretical background

This section covers research on the study field and the research questions in other areas. This section deals with the reasoning behind the chosen research question and hypotheses. This theoretical background starts with a review on the Moneyball hypothesis, which comes from baseball. Thereafter the theory of wages is covered, followed by a background on heuristics in decision making.

Moneyball

Lewis (2003) wrote about the Oakland Athletics baseball team and its general manager Billy Beane. Despite Oakland's disadvantaged financial situation, Beane managed to assemble a very competitive baseball team based on his analytical approach to player statistics. Lewis (2003) claimed that the collected wisdom of baseball insiders over the past century was flawed and subjective and thus the valuation of skills in the market for baseball players was grossly inefficient. Another approach towards statistics like slugging percentage and on-base percentage, gave Beane a competitive advantage over other MLB general managers. Slugging relates to hitting the ball, while getting on-base can be related to not hitting the ball. The essence of the Moneyball hypothesis is that, although it was well known that on base percentage was an important component of offensive productivity, the ability to get on base was seriously undervalued in the baseball labour market. The publication of the book: "Moneyball: The Art of Winning an Unfair Game" triggered a firestorm of criticism from baseball insiders (Lewis, 2004), but it appeared to have an impact on the way certain skills are valued in the market for baseball players afterwards and it impacted other sports as well.

Hakes and Sauer (2006) claim that the baseball labour market indeed was inefficient at the time Moneyball was published, and that the mispricing diminished thereafter. The impact of on-base percentage on winning appeared to be bigger than the impact of slugging percentage on winning. Hakes and Sauer (2006) showed that the impact of on-base percentage on salary was a lot lower during the turn of the twenty-first century than the impact of slugging percentage on salary. During and after the 2003 season the statistical knowledge of the Oakland management diffused across other baseball teams, which made the mispricing disappear. By expanding their research period, Hakes and Sauer (2007) verified that the mispricing has persisted for many years. After 2003 the returns to skill have increasingly matched the impact of skills on winning percentage, the team ability to spend wisely has increased and team-payroll has been a stronger predictor of winning percentage in 2004-2006 than in the seasons before. The Moneyball approach has furthermore been evaluated in soccer (Gerrard, 2007), ice hockey (Mason & Foster, 2007), basketball (Ballard, 2005; Berri, Schmidt & Brook, 2006) and business (Wolfe, Wright & Smart, 2006).

Ballard (2005) noted the proliferation of statistical experts that teams in the NBA have hired to specifically identify opportunities in the labour market; these experts have the goal to identify value that no one else sees. Such evaluation is changing conventional approaches to player recruitment and team planning in basketball. This development took place after the publication of Moneyball. Berri, Schmidt and Brook (2006) offer systematic statistical evidence for the presence of inefficiency in the basketball labour market. Wolfe, Wright and Smart (2006) state that the Moneyball approach has significant overlap with, and lessons for human resource professionals across a variety of industries.

According to Gerrard (2007) there are three barriers to the transferability of the Moneyball approach to other team sports: the technological barrier, the conceptual barrier and the cultural barrier. Over the years, the technological barrier has been largely overcome through the development of video analysis and tracking systems. At the moment of writing, companies like Ortec and Opta collect a lot of soccer statistics. The conceptual barrier is created by the complexity of, for example, soccer. Soccer, as opposed to baseball, consists of highly interdependent team play in which individual player actions are not separable which makes it hard to analyse individual player performance. “The most enduring barrier towards the transferability of the Moneyball approach is the cultural barrier. Moneyball is a very different mind-set in which statistical analysis is integrated into expert judgment-based decision-making systems. It is a different way of doing things that inevitably is dismissed as useless and inappropriate by those wedded to existing methods” (Gerrard, 2007, p.229). Kaplan (2010) dubbed soccer as the least statistical of all major sports, but there are some recent movements towards the Moneyball approach. In March 2015 Dutch soccer club AZ Alkmaar hired Billy Beane as an advisor (Schnitzer, 2015). Rasmus Ankersen uses statistical data only to run Danish club FC Midtjylland. “Players have been signed because algorithms helped identify them as undervalued prospects. Coaches’ evaluations of games are now primarily informed by the mathematical models the club employs. Free kicks have become something of a science project, with monthly set piece meetings between players, coaches, Ankersen, and the occasional outside consultant” (De Hoog, 2015).

Wages

Yellen (1984) considers an economy with identical, perfectly competitive firms. Each firm having a production function of the form $Q = F(e(X)N)$, where N is the number of employees, e is effort per worker, and X is the real wage. A profit-maximizing firm which can hire all the labour it wants at the wage it offers will offer a real wage, w^*. W^* satisfies the condition that the elasticity of effort with respect to the wage is unity. This equation says that labour productivity depends on the real wage paid by firms. If ability and wages are positively correlated, teams with higher wages will attract more able players (Yellen, 1984). Scully (1974) started the literature on wages and performance in sports. Scully (1974) showed that even if MLB players do not receive their full marginal revenue products, in an efficient market the rate of exploitation per unit of talent should be the same. If the rate of exploitation is common across players, team performance should be closely correlated with player salaries. Scully’s (1974) findings find strong support in the NHL (Jones, Nadeau & Walsh, 1999). According to Wiseman and Chatterjee (2003) the greater the MLB team payroll is, the better the on-field performance. According to Coates, Frick and Jewell (2014), the same holds for MLS.

The neoclassical economic theory of human capital states that pay differences between individuals are explained by the stock of skills a person possesses (Mincer, 1974). The demand side of human capital theory states that profit maximizing employers will hire more expensive skilled workers only in jobs where these extra skills increase productivity enough to cover the higher wage (Lang & Dickens, 1988). If the latter is the case in MLS, this research has to show that goalkeepers earn less and contribute less to the team performance than attackers do. Even after adjustments for the required stock of skills a player possesses, wages can differ among different positions. This could be explained by the notion of compensating differentials (Smith, 1979). In this view, the pay of a job consists of wage and the utility experienced while doing the work. This implies that jobs with less hazardous working conditions can be filled with lower wages. If in MLS attackers and older players are more likely to sustain an injury than defenders and younger players are, this would give an explanation for wage differentials. However, according to Morgan and Oberlander (2001) neither player age nor position plays a role in the occurrence or severity of injury.

Since soccer is a team sport and the wages of players differ from each other, the impact of wage disparity on productivity or performance is important for this thesis. Ramaswamy and Rowthorn (1991) stated that workers whose shirking has the greatest impact on the team performance should be paid the greatest; this is known as the damage-potential hypothesis which predicts that wage disparity has a non-negative impact on team performance. Levine (1991) on the contrary developed the team-cohesiveness hypothesis, which predicts that greater wage disparity leads to jealousy among workers in a team and a possible reduction in team performance. Depken (2000) tested these different hypotheses in MLB and found more evidence for the team-cohesiveness hypothesis. Hence, teams with greater wage disparity experienced a reduction in team performance on the field. However, Fuess (1998) showed an opposite relationship in MLB. Coates, Frick and Jewell (2014) provide support for the team-cohesiveness hypothesis in MLS. Berri and Jewell (2004) researched the impact wage disparity has on the team performance of NBA teams. They did not find evidence that wage disparity is a determining factor in team wins. According to Berri and Jewell (2004) players do not consider wage disparity during the course of a game. Since individual players can be substituted when their performance is poor, shirking is not really an option for players. Franck and Nüesch (2011) examined the effect of wage disparity on team performance in professional German soccer. Instead of support for one of the two hypotheses, they found evidence for a U-formed relationship between wage disparity and sporting performance. Teams that have either an egalitarian or a very differential pay structure are more successful on the field than teams with a medium level of wage disparity.

Heuristics in Decision Making

In many situations animals and human beings have surprising difficulty noticing and using information provided by the absence or non-occurrence of something (Hearst, 1991). In behavioural economics this phenomenon is described by Tversky and Kahneman (1973) as the availability heuristic. Tversky and Kahneman (1973) showed that when faced with the difficult task of judging probability or frequency, people employ a limited number of heuristics which reduce these judgments to simpler ones. One of the most well-known and researched heuristics is the availability heuristic. A person is said to employ the availability heuristic whenever he estimates frequency or probability by the ease with which instances or associations could be brought to mind. "It is a common experience that the subjective probability of traffic accidents rises temporarily when one sees a car overturned by the side of the road" (Tversky & Kahneman, 1974). Translating this into the world of soccer, scoring goals is much more salient and easier to bring to mind than preventing opponents from scoring goals is. The reliance on the availability heuristic leads to systematic biases. In the experiments Tversky and Kahneman (1973) conducted, the subjects were not able to recall and count all instances because certain instances do not come to mind easily. Furthermore, Tversky and Kahneman (1973) found that people tend to see an illusory correlation which can be explained by the assessment of availability. MLS managers might see a very strong correlation between scoring goals and winning matches. Although this correlation is obvious, it might be weaker than managers think. Dube-Rioux and Russo (1988) showed that professional managers underestimate 'out of sight, out of mind' probabilities, due to the availability heuristic they employ. This implies that MLS managers undervalue the abilities of players that are not visible; therefore these players will earn lower wages.

Another phenomenon that might relate to the availability heuristic is anchoring. Anchoring is the tendency to rely too heavily on one piece of information (Tversky & Kahneman, 1974). This phenomenon is found by asking subjects to estimate giving them certain different starting points. Tversky and Kahneman (1974) asked subjects for the percentage of African countries in the United Nations. First they gave a value for which the subjects had to indicate if their estimation is higher or lower and then the subjects had to state their estimation. Different groups got different values and these values effected the eventual estimates. Different starting points yield different estimates. Ariely, Loewenstein and Prelec (2003) found that participants' willingness to pay for a bottle of '98 Côtes du Rhône was highly correlated with the price equivalent to the last two digits of their social security number when they were first asked whether they would be willing to pay that particular price. Even if the starting point should clearly be ignored, people are influenced by it. The piece of information people tend to rely too heavily on is usually the first piece of information that is acquired. Relating this to soccer, the first piece of information managers acquire is probably the amount of goals an attacker scored, since this is easily retrievable. The availability heuristic and anchoring implies that this manager will rely too heavily on the goals this attacker made but neglects the chances he missed since he is not a very good finisher for instance. This will possibly lead to a wrong estimation of the value of the attacker, which will be visible in his wage or possible transfer fee.

Important decisions in soccer are made by a small homogeneous group which holds on to existing methods (Gerrard, 2007). Since all MLS managers and technical directors are men who played soccer at a high level, this seems to hold for MLS as well. Janis (1971) sees this homogeneity as a danger because it is a factor that possibly leads to 'groupthink'. "Groupthink refers to a deterioration of mental efficiency, reality testing, and moral judgment that results from in-group pressures" (Janis, 1982, p.9). Since MLS managers compete with each other, this groupthink can better be defined as herd behaviour; doing what everyone else is doing. Especially people who interact with each other regularly tend to think and behave similarly (Shiller, 1995). Thus since MLS managers meet each other regularly during the season it is likely that they make similar decisions. If the majority of the clubs give the highest contracts to attackers and does not invest a lot in goalkeepers, the other clubs will do the same. According to Gerrard (2007), it appeared to be very hard to change old methods or the 'cultural barrier' in soccer. This is probably due to the herd behaviour; it is uncommon to deviate from the others.

Data

In this section all data needed to answer the research questions and test the hypotheses is presented. This section starts with a description of the datasets. Hereafter the variables are presented and explained.

Datasets

Multiple datasets are used in this study. The rosters of the teams and the abilities and general information on the players are from Football Manager 2009 till 2015 (Sports Interactive). The wages of the players are the base salaries from the website of the MLS Players Union. These base salaries exclude signing bonuses, guaranteed bonuses and compensation from any contracts with individual teams or their affiliates (MLS Players Union, n.d.). The final results of the regular seasons are from the website of MLS, as well as the amount of minutes all players played for their team (MLS, n.d.[b]). To find out if players were designated players, the list of designated players from MLS is consulted (MLS, 2015[b]). The panel datasets contain repeated observations over the same units collected over a number of periods.

Football Manager is a series of soccer management simulation games for PC, developed by Sports Interactive. Football Manager has been recognized by real-life football clubs as a source for scouting players. In 2008, English Premiership side Everton signed a deal with Sports Interactive to get exclusive access to the full database of the game. At that time 1000 scouts in 50 countries worked for Football Manager to determine the abilities of 370.000 players (Daily Record, 2008). In 2014 Sports Interactive signed a deal with sports performance analyst Prozone, who uses data from Sports Interactive in their online application Prozone Recruiter. Prozone Recruiter is used by many of the top clubs to scout new players (Bleaney, 2014). Since Sports Interactive uses an extensive network and their database is used by professional soccer teams, the player abilities and information is assumed to be objective and correct.

If players who played in MLS for multiple years are only counted once: the sample contains information on 189 goalkeepers, 401 defenders, 498 midfielders and 328 attackers from 2008-2014. In terms of contracts per year, the sample contains contracts for 338 goalkeepers, 997 defenders, 1246 midfielders and 754 attackers. Below an overview of the average annual wages paid to the players.

Table 1. Average annual wages

	Goalkeepers	**Defenders**	**Midfielders**	**Attackers**	*Average per year*
2008	$61,335.44	$68,011.12	$144,177.28	$129,136.76	$107,855.28
2009	$73,608.89	$76,566.03	$176,927.24	$151,876.80	$126,913.09
2010	$92,282.63	$129,151.10	$168,828.67	$222,506.54	$160.749,91
2011	$100,055.70	$116,229.25	$144,584.98	$209,276.15	$146.415,07
2012	$80,365.20	$134,081.07	$157,590.22	$222,082.37	$157.394,26
2013	$89,208.68	$109,153.04	$130,500.20	$252,264.84	$149,720.14
2014	$102,234.51	$132,740.18	$194,029.66	$326,855.83	$199,476.24
Average per position	$86,537.54	$111,132.81	$158,532.27	$227,597.48	$152,680.30

Variables

In this part the variables are introduced. Team performance is measured in terms of the ranking a team obtained. Since the schedules are unbalanced, the Eastern and Western Conference are not aggregated. Ranking depends on the outcome of at least 30 matches over a seven-month period and therefore is likely to be a good indicator of performance. The wages spend per club are compared to the annual average to see if higher wages are justified by better rankings. Wage dispersion is measured according to Depken (2000). $SALHHI = \sum_{i=1}^{N}(SHARE_i)^2$, where N is the number of paid players on a team and SHARE is the *i*th player its share of a team its total salary expenditure. For the total wages per club, the wages of the players who transferred during the season are assumed to be equally spread among the two teams they played for. For the players their individual shares in salary expenditure, the wages are annualized. Szymanski (2000) states that turnover of players reflects a high level of injuries sustained, and thus is unsettling for team performance. Therefore, the aggregate number of players used in a season relative to the average is included as independent variable.

The database of Football Manager provides information on 32 skill attributes for goalkeepers and 35 skill attributes for field players. Some of these variables are, intuitively, strong related to each other. For example heading and jumping, or reflexes and one-on-ones. This amount of variables requires caution with respect to multicollinearity. Multicollinearity arises when there is a non-negligible relationship between two or more of the explanatory variables (Brooks, 2008). To solve the problem of multicollinearity, the amount of variables is reduced through principal component analysis. A broad explanation on principal component analysis can be found in the Methodology section. The principal components for field players are: goal preventing, chance creating, quickness, goal scoring and mental. The principal components for field players are: shot stopping and sweeper keeper. The principal components scores are calculated for every player and goalkeeper and is used in the team performance regression as well as the individual wage difference regression. For the team performance regression, the components are multiplied with the percentage of total minutes the player played to get individual contributions. These contributions are aggregated per team and compared to the annual average of all teams in the Conference.

The five field player principal components and the log difference between the annual wage of the individual and the average annual wage of all players in the Conference are included in the individual wage difference regression. Furthermore, this regression includes two dummy variables and one control variable. The dummy variable *DES* indicates if a player is signed according to the Designated Player Rule. The dummy variable *home* indicates if a player is from the United States or Canada. Canadian players are only seen as domestic players if their club is Canadian (Montreal Impact, Toronto FC or Whitecaps). The last included variable is *age*, which is equal to the age of the player at the start of the season.

Summary stats on the individual wage variables are reported in table 2.

Table 2. Summary stats field players.

	Defenders		Midfielders		Attackers	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**
Age	25.44	25	24.96	24	24.83	24
Goal preventing	35.43	34.58	27.35	26.11	24.85	23.73
Chance creating	42.41	41.50	49.90	47.87	49.21	46.46
Quickness	22.78	22.14	23.24	22.25	21.70	20.72
Goal scoring	10.96	10.96	10.88	10.28	17.07	16.61
Mental	5.64	5.52	5.01	4.90	6.28	6.12
Observations	997		1246		754	

Methodology

This section covers the regressions and tests based on the datasets and variables. First the principal component analysis is discussed. Thereafter the methodology on team performance is presented and this section ends with the methodology on individual wages.

Principal component analysis

If there are *k* explanatory variables in the regression model, principal component analysis will transform them into *k* uncorrelated new variables called principal components. Principal components are independent linear combinations of the original data. All variables have a certain weight within the principal components, these weights are called loadings. When the loading for a variable in a certain principal component is, for example, 0.5 the eventual principal component is for 50% based on that variable. The principal components are extracted in decreasing order of importance so that the first component accounts for as much of the variation as possible and each successive component accounts for less. Eigenvalues measure the amount of the variation explained by each component, and thus will be largest for the first component. An eigenvalue of 1 indicates that the component accounts for the same variance as one of the original variables does. Therefore the eigenvalue of 1 is commonly used as cut-off point for which principal components are retained (Wold, Esbensen & Geladi, 1987). Cattell and Vogelman (1977) looked at the scree plot to decide how many principal components to include. To get a scree plot, the value of each successive eigenvalue is plotted against the rank order. The smaller eigenvalues tend to lie along a straight line. The point where the first eigenvalues depart from the line distinguishes the trivial components. Points to the left of the straight line segment and the first eigenvalue to the right of this point should be included.

Five field player principal components are included according to the scree method. The underlying loadings of the variables and the principal component analysis for goalkeepers can be found in Appendix A. To examine whether the principal components match the field positions intuitively, the t-test is used. This is a parametric test that tests if the average of two samples are equal. The results of this test should link the player skill attributes from Football Manager to the actual position of players.

Team performance

The fixed effects model to examine the relationship between wages and team performance, follows the work of Szymanski (2000):

$$p_{it} = \alpha_i + \beta_1(w_{it} - \bar{w}_t) + \beta_2(SALHHI_{it} - \overline{SALHHI_t}) + \beta_3(play_{it} - \overline{play_t})$$

The term *p* is position transformed into the log odds of position, which is *ln[p/((n+1)-p)]* with *p* being the eventual ranking and *n* being the amount of teams that take part in the competition. The fixed effects (α_i) account for club-specific attributes that may affect performance. The term $(w_{it} - \bar{w}_t)$ is the log difference of club wage spend to the annual average. The term $(SALHHI_{it} - \overline{SALHHI_t})$ is the log difference between the wage dispersion of club *i* and the average wage dispersion. Since Szymanski (2000) showed that turnover of players is unsettling for team performance, the term $(play_{it} - \overline{play_t})$ is included. It measures the difference between the amount of players club *i* used during the season and the league average. Since there are reasons to believe that the explanatory variables are not completely exogenous or there might be omitted variables in the team performance regression, a problem arises. With panel data, it is possible to exploit the particular nature of the data owing to the availability of repeated observations on the same individuals. This problem is addressed by including individual-specific intercept terms in a fixed effects model, which solves potential omitted variable bias. The fixed effects capture all time-invariant differences across individuals. An F test following a fixed-effects regression indicates that there are significant team-level effects (F-statistic equal to 2.66), which are captured by a fixed effects model. The fixed effects can be eliminated by a first-difference transformation (Verbeek, 2008). Eliminating the fixed effect makes the interpretation of the effect of the independent variables more clear.

To examine the effect of individual skill attributes to the team performance, the minutes each player played are divided by the total amount of minutes during the season. These ratios are multiplied with the principal components to measure the contribution each player has to the total set of skills within a team. Comparing all these skills and including them in the regression, leads to the following regression:

$$p_{it} = \alpha_i + \beta_1(pre_{it} - \overline{pre}_t) + \beta_2(cha_{it} - \overline{cha}_t) + \beta_3(qui_{it} - \overline{qui}_t) + \beta_4(sco_{it} - \overline{sco}_t) + \beta_5(men_{it} - \overline{men}_t) + \beta_6(sho_{it} - \overline{sho}_t) + \beta_7(swe_{it} - \overline{swe}_t)$$

The term *p* is position transformed into the log odds of position. The fixed effects (α_i) account for club-specific attributes that may affect performance. All the other terms are the differences between the volume of a certain principal component a team used and the average volume of that component of all teams in competition. *Pre* stands for the goal preventing component, *cha* for the chance creating component, *qui* for the quickness component, *sco* for the goal coring component and *men* for the mental component of field players. *Sho* stands for the shot stopping component and *swe* for the sweeper keeper component.

To determine if any of the principal components make a contribution to the team ranking which cannot be justified by the wage bill, all principal component differences are separately added to the following equation:

$$p_{it} = \alpha_i + \beta_1(w_{it} - \bar{w}_t) + \beta_2(SALHHI_{it} - \overline{SALHHI_t}) + \beta_3(play_{it} - \overline{play_t})$$

This is the same method Szymanski (2000) used to test discrimination in English soccer leagues. If any of the principal components has a negative significant effect, that principal component is undervalued in the wages.

Individual wages

Since wages are not normally distributed among players (Jarque-Bera tests reject normality), parametric tests to compare medians cannot be used. Because of the central limit theorem, however, it is possible to compare means through parametric tests. The central limit theorem states that the distribution of the average of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution (Rice, 2006). In order to examine if field players earn more than goalkeepers, the unpaired t-test is used. The unpaired t-test is used to compare the means of two samples (Brooks, 2008). The null hypothesis is: the means are equal.

The effect of the individual skill attributes on the wage differences among players is examined through the next model:

$$w_{it} = \beta_1 + \beta_2(pre_{it} - \overline{pre}_t) + \beta_3(cha_{it} - \overline{cha}_t) + \beta_4(qui_{it} - \overline{qui}_t) + \beta_5(sco_{it} - \overline{sco}_t) + \beta_6(men_{it} - \overline{men}_t) + \beta_7 home_{it} + \beta_8 DES_{it} + \beta_9 age_{it}$$

The wage variable w_{it} is equal to the log difference between the earned wage and the average annual wage of all players. The variable *home* takes the value of 1 if the player is from the United States or Canada, if the club is Canadian. The variable *DES* is equal to 1 if the player signed a contract according to the Designated Player Rule. All other variables are mentioned before.

Since the signing of a designated player contract yields an increase of the wage, the chance of signing as a designated player is examined through a probit model. Since the dependent variable is binary (either 0 or 1) the probit model is a binary choice model. These models essentially describe the probability that $y_i = 1$ directly, although they are often derived from an underlying latent variable model (Verbeek, 2008). Therefore, estimated coefficients do not quantify the influence of the independent variables on the probability that the dependent variable takes on the value one. Estimated coefficients are parameters of the latent model. The probit model looks as follows:

$$\Pr(DES)_{it} = \beta_1 + \beta_2(pre_{it} - \overline{pre}_t) + \beta_3(cha_{it} - \overline{cha}_t) + \beta_4(qui_{it} - \overline{qui}_t) + \beta_5(sco_{it} - \overline{sco}_t) + \beta_6(men_{it} - \overline{men}_t) + \beta_7 home_{it} + \beta_8 age_{it}$$

Pr(DES) equals the probability that the binary variable *DES* takes on the value 1.

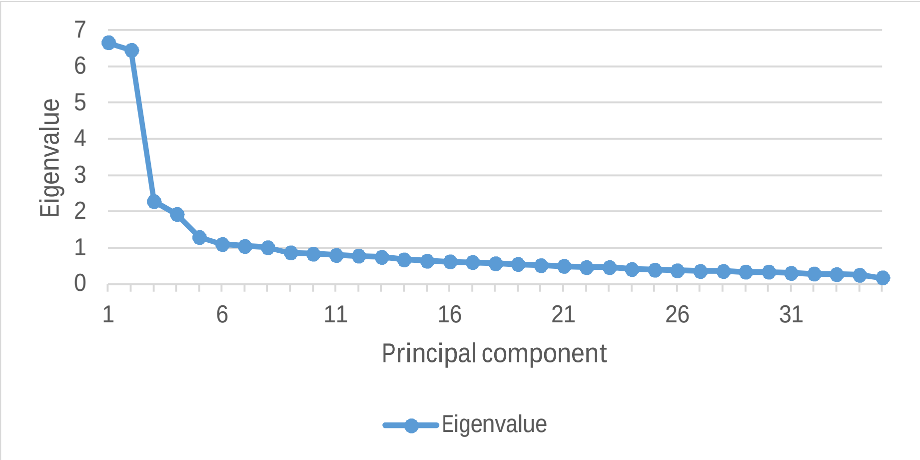
Results

This section covers the results of the regressions and tests according to the Methodology section. First the results of the principal component analysis are presented. Thereafter the result of the team performance regressions are presented and this section ends with the results on individual wages.

Principal component analysis

According to the scree plot in figure 1, five components should be retained for field players.

Figure 1. Scree plot of eigenvalues principal component analysis for field players.



Principal component analysis provides the following ability components for field players: the goal preventing component, the chance creating component, the quickness component, the goal scoring component and the mental component. For goalkeepers, three components remain: the shot stopping component, the sweeper keeper component and the handling component. Since the handling component led to a negative score for a third of the goalkeepers, it is left out of the regressions. Especially the five field player components are appealing to this research since the goal preventing, the quickness and the mental components are mainly based on less visible player attributes where the chance creating and goal scoring component are mainly build on visible attributes. The main parts of the different principal components are reported in table 3. The exact underlying loadings for the mentioned components can be found in Appendix A.

Table 3. Highest loadings of principal components

Goal preventing	**Chance creating**	**Quickness**	**Goal scoring**	**Mental**	**Shot stopping**	**Sweeper keeper**
Marking	Passing	Acceleration	Finishing	Aggression	Determination	Acceleration
Tackling	First touch	Pace	Strength	Anticipation	One on ones	Pace
Positioning	Crossing	Natural fitness	Jumping	Decisions	Strength	Decisions
Bravery	Free kicks	Agility	Heading	Composure	Work rate	Eccentricity
Strength	Technique	Work rate	Bravery	Concentration	Communication	Punching

To determine if the principal components match with the actual positions, t-tests has been performed. The t-test statistics are reported in table 4.

Table 4. Principal component – position analysis, unpaired t-tests.

	Goal preventing	**Chance creating**	**Quickness**	**Goal scoring**	**Mental**
Defender – midfielder	19.44165***	-14.97087***	-1.811639*	0.434681	4.689908***
Defender – attacker	25.35402***	-13.07672***	3.989355***	-24.98857***	-4.093910***
Midfielder - attacker	5.637280***	1.144047	5.446078***	-26.25676***	-8.596030***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

These results suggest that the principal components scores differ significantly per position. From this test results, it can be concluded that the player skill attributes from Football Manager corresponds with the actual positions of players.

Since goalkeepers and field players have other skill attributes, their wages relative to the skills they possess cannot be compared. In order to compare the wages two samples, field players and goalkeepers, have to be evaluated. This is done through the t- test. The test statistic of 2.977422 is significant at the 0.01 level and thus suggests that field players indeed earn more than goalkeepers. As showed above, MLS clubs can improve their ranking by improving their goalkeeper shot stopping component. If the goalkeepers who played in MLS are not that good compared to other players in the world, while the field players are, the wage differences might be fair. But an investment, in terms of wage, in a good shot stopping goalkeeper might yield a performance improvement for MLS teams.

Team performance

The log odds of position are regressed on the wage variables and the player turnover ratio. The empirical results are reported in table 5a.

Table 5. Team performance: coefficient estimates

Regression estimates: first differences.

Dependent variable: log odds of position

Total observations: 101

	a. Wages	**b. Skills**	**c. Skills added separately to wage regression**
Variable	**Coefficient**	**Coefficient**	**Coefficient**
Relative wage bill	-4.981447*** [1.280661]		
Relative wage disparity	2.491424*** [0.931577]		
Player turnover	0.228870*** [0.046320]		
Goal preventing		0.003949 [0.014830]	-0.011835 [0.007748]
Chance creating		-0.021372** [0.010038]	-0.013025** [0.005465]
Quickness		0.005668 [0.018966]	-0.014619 [0.011889]
Goal scoring		-0.17867 [0.023818]	-0.008390 [0.017953]
Mental		-0.004445 [0.024662]	-0.004584 [0.020272]
Shot stopping		-0.065415** [0.031114]	-0.059164** [0.026122]
Sweeper keeper		0.022188 [0.042954]	-0.018747 [0.036439]

All variables are transformed into first-differences. Coefficients represent log odds ratios. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Better league performance leads to a smaller value of the position variable. The three independent variables have the right sign and are highly significant. The expected percent increase in ranking from using one more player, compared to the annual average, is about 25.7% holding other variables constant, since $e^{0.228870} \approx 1.257$. Since player turnover indicates bad luck, this sign was expected. When a club decides to increase their wage bill to 10% above the league average, where it was equal to the league average in the previous year, the club can expect to improve their ranking with about 37.8% holding other variables constant, since $1.1^{-4.981447} \approx 0.622$. When the increase of the wage bill goes hand in hand with an increase in the relative wage disparity, the improvement will be smaller. A 10% increase in the relative wage disparity leads to an expected percent increase in ranking of about 26.8%, since $1.1^{2.491424} \approx 1.268$. Based on this result, it is justified to spend more on wages than the other teams. But, since the relative wage disparity has a negative effect on team performance, the higher wage bill should be divided 'fairly' among the club players. Signing one designated player for a high wage does not necessarily yield a big improvement in ranking if the relative wage disparity increases with the signing. In 2014 Toronto Raptors signed Michael Bradley and Jermaine Defoe both for $6,000,000 per year, it made their wage bill the second largest in the total sample. Although Toronto had the largest wage bill of the 2014 season, they only managed to improve their ranking from 9 to 7. This result shows that there is room for research on the distribution of the wage bills among players.

Team performance is also regressed on the skill components to examine the effect of certain sets of skills on the team ranking. The results of this regression are reported in table 5b. It is remarkable that only the chance creating and the shot stopping component are significant at the 0.05 level. According to the regression of this fixed effects model, only investing in players with a higher score on the chance creating or the shot stopping component will improve team performance significantly. Since $e^{-0.021372} \approx 0.979$, a one unit relative increase in the chance creating components yields an expected ranking improvement of about 2.1% if all other relative components keep constant. Increasing the relative shot stopping component with one unit yields an expected ranking improvement of about 6.3%. According to these results, teams can improve their ranking by improving their shot stopping goalkeeper and their chance creating skills relative to the other teams. Differences in other areas of the game do not significantly influence the team ranking. Since the variables are first-difference transformed, this is only the effect of improving the own score compared to other teams between seasons. The expectation is that team performance is related to all skills, or in any case more than these results suggest. These results therefore raise questions about the correctness of the player skill attributes from Football Manager.

In table 5c the results of the wage regression with an added principal component variable are reported. The negative significant coefficients of the chance creating and shot stopping components imply that a club with a higher than average score on one of these components would expect to achieve a systematically higher league position than the wage bill would appear to justify. This suggests that the return on chance creating and shot stopping talents is lower than the contribution to team performance.

Individual wages

The regression of wage differences on player skill attributes, age, designated player and domestic player yield the results reported in table 6.

Table 6. Player wage: coefficient estimates

Dependent variable: log difference of wages

Total observations: 1227

Variable	Coefficient	One unit increase → expected shift in wage difference [exp(coefficient)-1]
Intercept	-1.243383*** [0.060955]	
Designated player	0.723316*** [0.075050]	106.1%
Domestic player	-0.069385*** [0.016967]	-6.7%
Age	0.037582 *** [0.002519]	3.8%
Goal preventing	0.000294 [0.001061]	
Chance creating	0.007499*** [0.001167]	0.8%
Quickness	-0.007164*** [0.002043]	-0.7%
Goal scoring	-0.000652 [0.001699]	
Mental	0.002800 [0.002849]	
Adjusted R-squared	0.517886	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. White test rejected null hypothesis of homoscedasticity, HAC standard errors in parentheses.

From this regression it is clear that designated players are the top earners in MLS. Recall that their wages do not fully count against the clubs salary budget and clubs bear the financial responsibility for the amount of compensation above their budget charge. Therefore the high significant coefficient for the dummy variable designated player is not unexpected. But it do raise questions about the type of players that are signed as designated players. Instead of the log difference of wages, the dummy variable designated player is the dependent variable to check in which type of players MLS clubs invest. Since the dependent variable is binary, a probit model is used. The results of this regression are reported in table 7.

Table 7. Designated player: coefficient estimates

Dependent variable: probability designated player

Variable	Coefficient	Marginal effect
Intercept	-3.553665*** [0.5234734]	
Domestic player	-0.6178259*** [0.204568]	-0.0410506*** [0.0141051]
Age	0.0684198*** [0.0192216]	0.0045461*** [0.0013211]
Goal preventing	-0.0088198 [0.0089601]	-0.000586 [0.0005966]
Chance creating	0.0329163*** [0.0080868]	0.0021871*** [0.0005543]
Quickness	-0.0475739** [0.0194279]	-0.003161** [0.0013138]
Goal scoring	0.0047266 [0.0157743]	0.0003141 [0.0010485]
Mental	0.0315467 [0.0238919]	0.0020961 [0.0015906]
McFadden R-squared	0.226256	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The significant positive coefficient of the chance creating variable implies that a player whose chance creating score is 1 unit higher than the average, has about 0.2% higher probability of getting signed as a designated player. Although it is easier to sign young designated players since 2012, the probability of signing a designated player contract is about 0.4% higher when getting a year older. A player whose quickness component score is 1 unit above the average has about 0.3% lower probability of getting signed as a designated player. Since this component does not significantly affect team performance, this is an unexpected result. Domestic players have about 4.1% smaller probability of getting signed as designated player. Since Canada has been ranked between 55th and 112th in the years 2008-2014 on the FIFA World Ranking and the United States has been ranked between 8th and 31st (FIFA, n.d.b), it is likely that foreign players in MLS have been better players than domestic players. This can be tested, but is not relevant in this research. Regarding goalkeepers and designated player contract, one out of 129 unique goalkeepers has been signed as a designated player.

The results above show that designated players earn the highest wages in MLS and players who possess more chance creating skills than the average have the highest probability of signing a designated player contract. In the next regression, the designated players have been left out of the sample to see if MLS as organisation acts in the same way as the clubs. The players who are not signed as designated players are actually paid by MLS. The results are reported in table 8.

Table 8. Player wage: coefficient estimates. Designated players excluded.

Dependent variable: log difference of wages

Variable	Coefficient	One unit increase → expected shift in wage difference [exp(coefficient)-1]
Intercept	-1.206667*** [0.061636]	
Domestic player	-0.082315*** [0.016555]	-7.9%
Age	0.036392 *** [0.002554]	3.7%
Goal preventing	0.000177 [0.001053]	
Chance creating	0.007281*** [0.001207]	0.7%
Quickness	-0.007036*** [0.002070]	-0.7%
Goal scoring	-0.000594 [0.001707]	
Mental	0.001618 [0.002835]	
Adjusted R-squared	0.363412	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. White test rejected null hypothesis of homoscedasticity, HAC standard errors in parentheses.

These results corresponds with the earlier results when designated players were included. This implies that the valuation of player skill attributes by the clubs is comparable to the valuation of player skill attributes by MLS.

Conclusion

The publication of *Moneyball* showed misvaluation in the Major League Baseball. Since the world of soccer is seen as a conservative world where statistical analysis is limited, one might expect that the valuation of skills in soccer is not correct. This thesis examined the relationships between team performance, wages and individual skill attributes in MLS. The main findings of this research are reported in table 9.

Table 9. Main findings.

Team performance	**Individual wages**	
Relative wage bill	Chance creating component	**Positively related**
Goalkeeper shot stopping component	Age	
Chance creating component		
Wage disparity	Quickness component	**Negatively related**
Player turnover		
• Not all player skill components affect team performance.	• Goalkeepers earn less than field players. • Chance creating and shot stopping skills are undervalued in terms of wages.	**Remarks**

The found positive correlation between wage bill and team performance supports the findings by Wiseman and Chatterjee (2003) in MLB. Furthermore it supports the theory introduced by Scully (1974); team performance should be closely correlated with player salaries. There are two main theories about wage disparity and its impact on team performance: damage-potential (Ramaswamy & Rowthorn, 1991) and team-cohesiveness (Levine, 1991). This thesis provides support for the team-cohesiveness hypothesis which states that greater wage disparity leads to jealousy among workers in a team and a possible reduction in team performance. The consequences of hiring high-skilled, high-wage players for the intra-team wage disparity should always be taken into account. This finding is in line with the findings of Depken (2000).

Since there has not been done research on the relation between the different skills and eventual team performance in soccer yet, the findings on this relationship cannot be compared with previous research. Anderson and Sally (2013) state that defence and offense are both vital for a good league ranking. Scoring goals and preventing the opponent from scoring goals contribute in the same degree to the amount of wins. In the light of these findings from the Premier League, the finding that chance creating and goalkeeper shot stopping skills are significantly positively related to team performance was expected. The chance creating component is built on skills like crossing and passing, where the goalkeeper shot stopping component is about preventing the opponent from scoring. In preventing a loss and obtaining a draw, not suffering a goal against was 33% more valuable than scoring a goal (Anderson & Sally, 2013). Since a draw yields one point for the ranking, the fact that goalkeeper shot stopping skills contributes more to team performance than chance creating skills is in line with this statement. Since soccer matches consists of a lot of interdependent actions from players, it is intuitively questionable that only the chance creating and the goalkeeper shot stopping component contribute to team performance.

This thesis revealed field players earn higher wages than goalkeepers, and clubs invest most in chance creating skills. The relative wage bills of teams are significantly positively correlated to team performance, which indicates that MLS teams spend their money in the right way although wage disparity has a negative effect on team performance. Despite the expectation that all individual player skills should contribute to the eventual team performance, only chance creating and goalkeeper shot stopping skills significantly affect the team ranking positively. The findings about the contribution to team performance justify the fact that clubs invest in chance creating skills, but the fact that goalkeepers earn less than others should be questioned.

The results of the individual wages regressed on the field player components are in line with the previous results on team performance. MLS teams do reward chance creating skills, which also contribute to the team performance, but not enough according to the results in table 5c. The fact that the valuation of designated players, by the clubs, and other players, by MLS, are the same can be interpreted as herd behaviour by the decision makers in soccer. The found difference in earned wages between goalkeepers and field players does raise some questions. Since the goalkeeper shot stopping component contributes even more to team performance than the chance creating component, the difference in wages cannot be justified by the contribution to team performance. From the results in table 5c, it appeared that the return on shot stopping component skills for individual players is too low compared to the return on team performance. Recall that the shirts of designated midfielders and attackers are the most sold (Keh, 2009) and the transfer fees of attacking players are higher (Anderson & Sally, 2013). For these facts the decisions to pay goalkeepers less than field players cannot be labelled as 'biased'. Because of the economic advantages of field players over goalkeepers it could be completely rational to invest more money in field players than in goalkeepers.

The unexpected result that not all principal components are important for team performance raises questions about the used variables, data and methodology. Since the player skill attributes are from a computer game, similar data has not been used in an academic context yet. The correlation between the skills attributes from Football Manager and the actual performance of a player or a team has not been examined yet as well. Since there is no other public source for this amount of data on players, the use of this data for this thesis is the most obvious choice. But this data needs more extensive research, as will be mentioned in the Recommendations section. In the regressions the team component scores are compared to the average component scores. The differences between the component scores might be negligible since the roster of MLS teams does not differ a lot, which makes the effects of the principal components on team performance insignificant.

MLS clubs which spend more money on wages than its competitors, perform better during regular season. When contracting high-wage players, clubs have to keep an eye on wage disparity. A higher wage disparity among team players leads to a decrease of team performance. Shot stopping goalkeepers and players who create chances improve the ranking of their MLS teams. The chance creating players have a higher probability to get signed as a designated player and earn higher wages than others, which is in line with their contribution to team performance. Shot stopping goalkeepers however, earn less than players who are in front of them on the field. This does not have to imply that the valuation of different players is wrong. Factors like merchandise sales and market value might be higher for field players, which justifies the wage differences.

Recommendations

Since the MLS and soccer team performance is a relative blank field of research, further research is recommended. The individual skill attributes used in this thesis are from Football Manager. Although the database of this computer game is used by professional soccer clubs, it is unknown to what extent the reported skill attributes match with the actual skills of players. The best way to examine this is to estimate the extent to which reported skill attributes by Football Manager can predict the eventual ranking of soccer leagues.

Furthermore it is recommended to enlarge the study field and extend the study period. Since the market for players and the rosters in MLS are regulated, a replication of this study in European soccer competitions might yield different results. The European market for soccer players is barely regulated, there are no salary caps and players sign contracts with their club instead of with the association. The data in this research is from after the publication of *Moneyball*. In MLB the misvaluation of players diminished and eventually disappeared after *Moneyball* got published, this might be the case in MLS as well.

References

Anderson, C. & Sally, D. (2013). *The Numbers Game: Why Everything You Know About Football is Wrong.* Viking.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences, *Quarterly Journal of Economics,* 118, 73-106.

Ballard, C. (2005). Measure of success, *Sports Illustrated*, 103(16), pp. 60-63.

Berri, D. J., & Jewell, R. T. (2004). Wage inequality and firm performance: Professional basketball's natural experiment. *Atlantic Economic Journal*, *32*(2), 130-139.

Berri, D. J., Schmidt, M. B., & Brook, S. L. (2006). *The wages of wins: Taking measure of the many myths in modern sport.* Stanford University Press.

Bleaney, R. (2014). Football Manager computer game to help Premier League clubs buy players. Retrieved from http://www.theguardian.com/football/2014/aug/11/football-manager-computer-game-premier-league-clubs-buy-players.

Brooks, C. (2008). *Introductory econometrics for finance*. Cambridge university press.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research*, *12*(3), 289-325.

Coates, D., Frick, B., & Jewell, T. (2014). Superstar Salaries and Soccer Success The Impact of Designated Players in Major League Soccer. *Journal of Sports Economics*, 1527002514547297.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 187-220.

Daily Record (2008). Everton sign deal to scout for world's best players in Football Manager video game. Retrieved from http://www.dailyrecord.co.uk/news/uk-world-news/everton-sign-deal-to-scout-for-worlds-997789.

De Hoog, M. (2015). How data, not humans, run this Danish football club. Retrieved from https://decorrespondent.nl/2607/How-data-not-humans-run-this-Danish-football-club/230219386155-d2948861.

Depken, C. A. (2000). Wage disparity and team productivity: evidence from major league baseball. *Economics letters*, *67*(1), 87-92.

Dickens, W. T., & Lang, K. (1988). The reemergence of segmented labor market theory. *The American Economic Review*, 129-134.

Dube-Rioux, L., & Russo, J. E. (1988). An availability bias in professional judgment. *Journal of Behavioral Decision Making*, *1*(4), 223-237.

FIFA (n.d.[a]). FIFA Ballon d'Or History. Retrieved from http://www.fifa.com/ballon-dor/history/index.html.

FIFA (n.d.[b]). FIFA/Coca-Cola World Ranking. Retrieved from http://www.fifa.com/fifa-world-ranking/ranking-table/men/concacaf.html.

Franck, E., & Nüesch, S. (2011). The effect of wage dispersion on team outcome and the way team outcome is produced. *Applied Economics*, *43*(23), 3037-3049.

Fuess, S. M. (1998). "Paying Professional Baseball Players: Tournament Pay for Teamwork," mimeo, Presented at the Western Social Science Association meetings.

Gerrard, B. (2007). Is the Moneyball approach transferable to complex invasion team sports?. *International Journal of Sport Finance,* 2(4), 214.

Hakes, J. K., & Sauer, R. D. (2006). An economic evaluation of the Moneyball hypothesis. *The Journal of Economic Perspectives*, 20(3), 173-185.

Hakes, J. K., & Sauer, R. D. (2007). The Moneyball anomaly and payroll efficiency: a further investigation. International Journal of Sport Finance, 2(4), 177-189.

Hearst, E. (1991). Psychology and nothing. *American Scientist*, 432-443.

Janis, I. L. (1971). Groupthink: The desperate drive for consensus at any cost. *Classics of organization theory*, *6*, 185-192.

Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (2nd ed., p. 349). Boston: Houghton Mifflin.

Jones, J. C. H., Nadeau, S., & Walsh, W. D. (1997). The wages of sin: Employment and salary effects of violence in the national hockey league. *Atlantic Economic Journal*, *25*(2), 191-206.

Kaplan, T. (2010). When It Comes to Stats, Soccer Seldom Counts. Retrieved from http://www.nytimes.com/2010/07/09/sports/soccer/09soccerstats.html?_r=0.

Keh, A. (2009). M.L.S. Merchandise Rankings Don't Reflect Performance. Retrieved from http://goal.blogs.nytimes.com/2009/07/23/mls-merchandise-rankings-dont-reflect-performance/?_r=0.

Lawson, R. A., Sheehan, K., & Stephenson, E. F. (2008). Vend It Like Beckham: David Beckham's Effect on MLS Ticket Sales. *International Journal of Sport Finance*, *3*(4), 189.

Levine, D. I. (1991). Cohesiveness, productivity, and wage dispersion. *Journal of Economic Behavior & Organization*, *15*(2), 237-255.

Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. WW Norton & Company.

Lewis, M. (2004). Out of Their Tree. *Sports Illustrated,* March 1.

Mason, D. S., & Foster, W. M. (2007). Putting Moneyball on ice?. *International Journal of Sport Finance,* 2(4), 206.

Mincer, J. (1974). Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.

MLS (n.d.[a]). MLS History. Retrieved from http://www.mlssoccer.com/history.

MLS (n.d.[b]). Statistics. Retrieved from http://www.mlssoccer.com/stats/season.

MLS (2012). 2012 MLS Roster Rules. Retrieved from http://www.mlssoccer.com/2012-mls-roster-rules.

MLS (2015[a]). Competition Rules and Regulations. Retrieved from http://pressbox.mlssoccer.com/content/competition-rules-and-regulations.

MLS (2015[b]). Designated Players. Retrieved from http://pressbox.mlssoccer.com/content/designated-players.

MLS Players Union (n.d.). Player Salary Information. Retrieved from https://www.mlsplayers.org/salary_info.html.

Morgan, B. E., & Oberlander, M. A. (2001). An Examination of Injuries in Major League Soccer The Inaugural Season. *The American Journal of Sports Medicine*, *29*(4), 426-430.

Ramaswamy, R., & Rowthorn, R. E. (1991). Efficiency wages and wage dispersion. *Economica*, 501-514.

Rice, J. (2006). *Mathematical statistics and data analysis*. Cengage Learning.

Santchurn, U. (2012). SOCIETY'S CULTURE: Sports in America. Retrieved from http://www.lemauricien.com/article/societys-culture-sports-america.

Schnitzer, K. (2015). Billy Beane’s Moneyball is heading for Europe. Retrieved from http://nypost.com/2015/03/18/billy-beanes-moneyball-is-heading-for-europe/.

Scully, G. W. (1974). Pay and performance in major league baseball. *The American Economic Review*, 915-930.

Shiller, R. J. (1995). Conversation, information, and herd behavior. *The American Economic Review*, 181-185.

Smith, C. (2013). Major League Soccer's Most Valuable Teams. Retrieved from http://www.forbes.com/sites/chrissmith/2013/11/20/major-league-soccers-most-valuable-teams/.

Smith, R. S. (1979). Compensating wage differentials and public policy: a review. *Industrial and Labor Relations Review*, 339-352.

Sports Interactive. Football Manager 2009-2015.

Szymanski, S. (2000). A market test for discrimination in the English professional soccer leagues. *Journal of Political Economy*, *108*(3), 590-603.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd Ed.). New York: Harper Collins.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Verbeek, M. (2008). *A guide to modern econometrics*. John Wiley & Sons.

Wiseman, F., & Chatterjee, S. (2003). Team payroll and team performance in major league baseball: 1985–2002. *Economics Bulletin*, *1*(2), 1-10.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1), 37-52.

Wolfe, R., Wright, P. M., & Smart, D. L. (2006). Radical HRM innovation and competitive advantage: The moneyball story. *Human Resource Management*, 45(1), 111-145.

Yellen, J. L. (1984). Efficiency wage models of unemployment. *The American Economic Review*, 200-205.

Appendices

Appendix A: Principal component analysis

Table 9. Principal component loadings, field players.

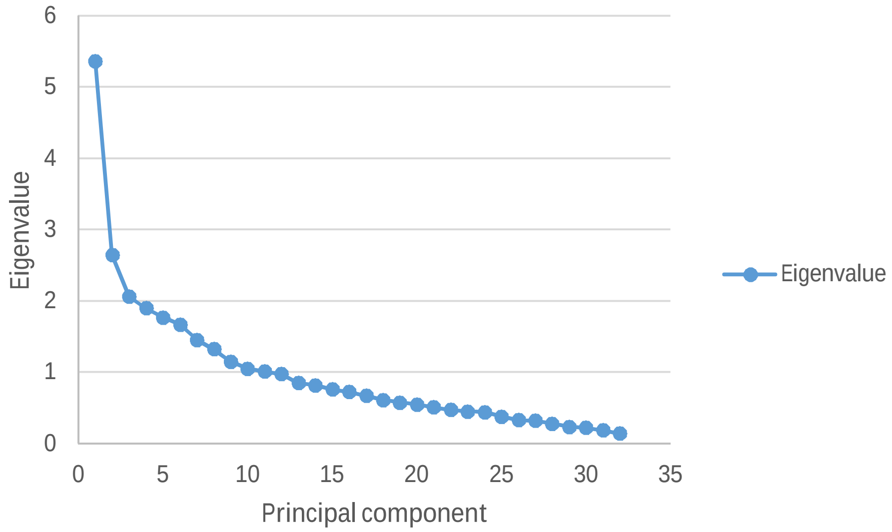
Variable	Goal preventing	Chance creating	Quickness	Goal scoring	Mental
Eigenvalue	6.656997	6.456270	2.277398	1.925606	1.284872
Acceleration	-0.145500	0.008007	0.450629	0.032029	0.175789
Aggression	0.079674	0.061358	0.213691	0.056991	0.406430
Agility	0.078040	0.072650	0.246870	0.094997	- 0.206357
Anticipation	0.160232	0.162990	-0.081703	-0.018583	0.393558
Balance	0.117161	0.098345	0.148685	0.112755	0.126765
Bravery	0.256525	0.087868	0.125933	0.176538	-0.131351
Composure	0.093621	0.189261	-0.125687	-0.048282	0.258568
Concentration	0.192763	0.153704	-0.047934	-0.057338	0.185138
Corners	-0.089651	0.260417	-0.007067	-0.217761	-0.276056
Crossing	-0.076773	0.277271	0.107339	-0.182361	-0.220233
Decisions	0.147334	0.189696	-0.130757	-0.114100	0.347421
Dribbling	-0.179538	0.262402	0.057074	0.051521	1.35E-06
Determination	0.157714	0.118752	0.224090	0.139705	0.045391
Finishing	-0.133142	0.213609	-0.103494	0.378829	-0.064345
First touch	-0.063795	0.284492	-0.063616	-0.054725	0.091130
Flair	-0.221957	0.190200	0.044159	0.094892	0.082112
Free kicks	-0.059601	0.275647	-0.076590	-0.038630	-0.200668
Heading	0.235839	-0.002262	-0.192468	0.321216	-0.019477
Jumping	0.207016	-0.058034	-0.196418	0.337040	-0.057656
Leadership	0.196273	0.084483	-0.043344	-0.033460	-0.006689
Long shots	-0.078247	0.256985	-0.094847	0.111859	-0.148044
Long throws	0.177637	0.074473	0.123567	-0.115258	-0.208273
Marking	0.311323	-0.009384	0.064868	-0.199931	-0.092713
Natural fitness	0.055389	-0.004945	0.286073	0.171157	-0.014870
Off the ball	-0.081596	0.251485	0.031763	0.162532	-0.069332
Pace	-0.124790	-0.015074	0.446431	0.071674	0.141754
Passing	0.024178	0.282666	-0.034205	-0.224244	-0.012473
Penalty taking	0.002990	0.246637	-0.051923	0.158290	-0.015413
Positioning	0.280499	0.044789	-0.022768	-0.175584	-0.016085
Stamina	0.182264	0.112396	0.225210	0.103001	-0.113808
Strength	0.247994	0.047069	-0.102489	0.343726	-0.068278
Tackling	0.308567	-0.024661	0.049494	-0.250925	-0.060158
Teamwork	0.206581	0.088997	0.092257	-0.125530	-0.080889
Technique	-0.047983	0.267707	-0.050474	-0.073066	0.145925
Work rate	0.191887	0.040849	0.235864	0.022112	-0.112144

The most important drivers of the principal components are in red.

Eigenvalues are a set of scalars associated with a linear system of equations. In principal component analysis eigenvalues are used to consolidate the variance. The eigenvalues show how much additional variance is explained by the factors. The factor with the largest eigenvalue explains the most variance, and so on. When the eigenvalue of a factor is smaller than 1, that factor explains less variance than it adds (Tabachnick & Fidell, 1996). The reported factor loadings are the correlation coefficient between the actual variables and the obtained factors. To calculate the scores of players on the principal components, all factor loadings are multiplied with the corresponding skill attributes.

In figure 2 and table 10, the principal component analysis for goalkeepers is reported.

Figure 2. Scree plot of eigenvalues principal component analysis for goalkeepers.



According to the scree plot method, 3 components should be retained. Since the third component yields negative scores for over a third of the sample while it should show the abilities of goalkeepers and not the shortcomings, two components are included in the research. Table 10 shows the eigenvalues and factor loadings of the obtained factors.

Table 10. Principal component loadings, goalkeepers.

Variable	Shot stopping	Sweeper keeper	Handling
Eigenvalue	5.365281	2.650602	2.061370
Acceleration	-0.051690	0.394464	0.111150
Aerial ability	0.150107	-0.117330	0.236837
Aggression	0.080206	0.014583	-0.349671
Agility	0.124469	0.215649	0.259102
Anticipation	0.256268	0.067820	-0.221161
Balance	0.140948	0.008738	0.057517
Bravery	0.209224	0.054355	-0.340677
Command of area	0.214627	-0.034285	0.134300
Communication	0.239534	0.079086	-0.026389
Composure	0.187524	0.207733	-0.302297
Concentration	0.139239	0.050506	0.059714
Decisions	0.061855	0.291224	-0.302726
Determination	0.308078	-0.050401	0.058494
Eccentricity	0.033611	0.242004	0.118875
Flair	0.196170	0.194682	0.080111
Handling	0.199365	0.117390	0.281776
Jumping	0.089616	-0.196622	-0.077393
Kicking	0.020839	0.203621	0.104581
Leadership	0.197292	0.004472	-0.167368
Natural fitness	0.101759	0.037661	-0.129388
Off the ball	0.042777	-0.047466	0.236261
One on ones	0.284053	0.070417	0.120941
Pace	-0.076772	0.309964	0.083967
Positioning	0.193550	0.100979	0.016051
Punching	0.046442	0.240636	-0.104205
Reflexes	0.218280	0.082284	0.265571
Rushing out	0.132983	0.122691	0.092432
Stamina	0.201927	-0.178490	0.043511
Strength	0.257681	-0.198825	-0.144051
Teamwork	0.140412	-0.357487	0.085787
Throwing	0.213246	-0.027054	-0.063342
Work rate	0.252980	-0.225513	0.042380

The most important drivers of the principal components are in red.