NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Inna Zaytseva, Daniil Shaposhnikov

MONEYBALL IN OFFENSIVE VS DEFENSIVE ACTIONS IN SOCCER

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 223/EC/2020

Inna Zaytseva[1], Daniil Shaposhnikov[2],

MONEYBALL IN OFFENSIVE VS DEFENSIVE ACTIONS IN SOCCER[3]

There is an established tradition in soccer society, for both soccer fans and club managers to value forwards more than defensive players. However, the soccer rules imply an equally important role of goals scored and goals conceded in a team win. This paper employs these facts to formulate the research hypothesis of undervalued defensive, compared to offensive actions, by professional soccer clubs, known as Moneyball phenomenon in sports economics literature. To test our hypothesis, we use two separate data sets at team and player level (1,224 and 776 observations correspondingly) from two seasons (2017-2019) of the German Bundesliga. We estimate the two groups of models with a dependent variable being, correspondingly, an indicator of win and a market value. We keep the set of controls as similar as possible to make the results of the two groups of models comparable with each other, in terms of a relative contribution of offensive and defensive actions. Offensive actions are measured by shots and key passes, while tackles, interceptions and clearances stay for defensive actions variables. All the key variables are normalized, and the resulting estimates demonstrate both "absolute" and "relative" Moneyball in offense vs defense in different model specifications. This is the first introduction of a term and method of a "relative" Moneyball, to our knowledge. In addition, our results show that there is room for improvement for Bundesliga clubs' cost of win efficiency, in redistributing funds from offense to defense, at least to some extent.

Keywords: Moneyball, soccer, offense, defense, labor market failure, Bundesliga

JEL Classification: Z22, D43, J44

[1] Laboratory for Sport Studies, National Research University Higher School of Economics, izaytseva@hse.ru

[2] Laboratory for Sport Studies, National Research University Higher School of Economics, shaposhnikov.dan@gmail.com

[3] Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged.

Introduction

There is evidence on the soccer transfer market that forwards have been enjoying greater popularity and higher salaries, compared to other players, for a long time. Up to July 2019, only 2 players with a defensive specialization are being included in the list of the top-25 most valuable players[4]. However, nowadays there is a tendency of power rebalancing on the soccer transfer market in favor of defenders. It can be seen in the list of top-25 transfers during the summer of 2019, that defenders have gained 7 positions already[5]. These facts indicate a declining but still prevailing bias on the soccer transfer market potentially resulting in overspending on sporting success. Such a bias can be referred to as the Moneyball phenomenon.

The term “Moneyball”, defined as the undervaluation of specific player skills, was firstly introduced for baseball (Lewis, 2003), and tested in several empirical papers (Berri, Holmes, & Simmons, 2014; Hakes & Sauer, 2006, 2007). In addition to ineffective resource allocation, such a market failure can distort players’ incentives to exert effort, thus deteriorating the spectacle of the game. For the case of soccer, it has already been documented in literature, that the market undervalues productive effort of players in terms of running distance (Weimar, Wicker, 2017). The case of systematic offense vs defense salary bias, is another candidate for revealing Moneyball in soccer, which can be tested by looking at the relative contribution of the offensive vs defensive actions in a team win. An argument in favor of this hypothesis, is the rule of determining the game outcome in soccer by the difference of goals scored and conceded, and treating them with equal importance. It means that the basic economic principle of the marginal productivity theory (Hicks, 1932) can be violated on a soccer transfer market, leaving room for improvement in decision making in professional soccer.

One possible explanation for the consistent defense undervaluation is the difficulty of observing defenders’ efforts as their outcome is in the absence of goal conceded, in contrast to the explicit result of forwards’ efforts - a goal scored. A similar explanation can be applied to running distance undervaluation. It is hard to observe this form of effort for teams on a pitch simultaneously with your own eyes, both on a TV screen and in the stands. Innovative tracking technologies are being introduced in some top championships, which monitor the distance covered by players, however, it is yet not a commonly agreed characteristic for a players’ assessment.

The aim of this study is to test the Moneyball hypothesis for offense vs defense, by analyzing the effect of these two types of action, on a team win probability, and players’ market values, which are commonly used as salary proxies (Franck & Nüesch, 2012). It advances the following main research question: Are offensive and defensive actions paid according to their marginal productivity in the German Bundesliga? The research question is studied using data from two Bundesliga seasons (2017-2019; N=1,224 observations on a team-game day basis and N=772 observations on a player-season basis). We include similar offensive and defensive action measures, and normalize these key variables in order to make the results of the two empirical parts of the research comparable. The regression results allow us to conclude that there is evidence of so called “absolute Moneyball” effect in some of the model specifications, when defensive actions contribute positively to winning probability, but not significantly different from zero to the player’s market value. Other specifications demonstrate the so called “relative Moneyball” effect, when both of the stated contributions are positive, but relative size of winning probability contribution is greater. The research begins with a literature review on sports labor market failures and the role of offense and defense in soccer, followed by the data description and model setup in the empirical part of study. The next section presents the empirical results and discussion and finally we finalize paper with a conclusion.

[4] https://www.transfermarkt.com/spieler-statistik/wertvollstespieler/marktwertetop
[5] https://www.transfermarkt.com/statistik/saisontransfers

Literature Review: Sports labor market failures and the role of offense and defense in soccer

Sports labor market failures have been broadly studied in literature. One of these failures, known as Moneyball, was found in baseball by team manager Billy Beane and described by Lewis (2003). In his book, Lewis hypothesized that certain players' activities are undervalued and therefore some players are underpaid. Hackes and Sauer (2006 & 2007) empirically tested Lewis' hypothesis for Major League Baseball (MLB) and found a significant Moneyball effect for several performance measures (e.g., on-base discipline, hitting for average, hitting for power, and plate discipline) in the seasons preceding the book release (Hackes and Sauer, 2006), however this effect has disappeared, suggesting that the wages were quickly reevaluated (Hackes and Sauer, 2007).

Labor market inefficiencies were also studied in other team sports. In order to establish the Moneyball phenomenon it is necessary to examine the effect of individual players' characteristics, on both team success and player's salary. It has been admitted that identifying the individual contribution to team success is a more challenging task for complex invasion team sports like soccer, ice hockey, or football, than it is for baseball (Gerrard, 2007). Several studies have attempted to overcome the Moneyball challenge for complex invasion team sports. Berri, Brook, and Schmidt (2007) found that the National Basketball Association (NBA) teams overvalued the points scored by a player, in terms of salary and voting points for the All-Rookie team, while other performance characteristics were largely ignored. Staw and Hoang (1995) showed that an increase in the draft number, significantly decreases playing time and shortens career length in the NBA, even after controlling for players' on-court performance, injuries, trade status, and position played. Labor market failure was also observed in the National Football league (NFL) in the sense that high draft picks are overvalued (Massey and Thaler, 2013). These examples provide evidence of the idea that the evaluation of players is mostly based on visually observable player actions (Berri et al., 2006).

Turning to soccer, Wicker, et. al (2013) have managed to separate the individual efforts of soccer players in the Bundesliga, measured as running distance and intensive runs, and showed that players' effort-based performance does not affect their market value. Further Weimar and Wicker (2017) showed that both running distance and intensive runs, significantly increase the team's winning probability. Finally, the authors conclude that players' effort-based performance is undervalued by the soccer transfer market, which is treated as Moneyball. Bryson, Frick, and Simmons (2013) in turn studied the effect of footedness on player salary in European soccer. The authors found that two-footed players enjoy higher salary ceteris paribus, while the additional two-footed player did not increase the winning chances significantly, leading to a conclusion that these players are overvalued by the labor market.

From the literature review on sports labor market failures, it can be seen that a number of cases have already been studied in several sports. This paper contributes to this literature by examining the effect of offensive and defensive actions of soccer players on a team winning probability, and a players' market value. In the case of a greater relative (to offense) contribution of defensive actions to win probability, compared to players' market value, it could be that defensive specialization is undervalued by the soccer labor market. So, we move to the review of literature on the role of offense and defense in team performance, and players' remuneration.

One of the common approaches to study success factors in soccer, is the comparison of characteristics inherent to successful and unsuccessful teams. Lago-Ballesteros and Lago-Peñas (2010) found significant differences across sections of the Spanish Soccer League 2008/2009 table, in goals for, total shots, shots on goal, shots for a goal, assists and ball possession, while no difference was found for defense (crosses against, offsides received, fouls committed, corners against, yellow cards, red cards). Lago-Peñas et al. (2011) examined the games of the UEFA Champions League and found that the shots on goal, crosses, ball possession, venue and quality of opposition, discriminate winning teams from drawing and losing teams, again finding no difference in defensive performance indicators. Other papers using more detailed data on defensive statistics based on video analysis, have managed to identify the specific defensive patterns for successful teams (Vogelbein et al., 2014; Szwarc et al., 2017). For the Bundesliga, it was found that defensive reaction time – team's ability to recover the ball after losing it, is significantly lower for top teams (Vogelbein et al., 2014). At the same time, at national team level, for the last rounds (semi-final, 3rd place game and final) of the World Cup and European Championship from 1990 to 2014, it was found that the players of the winning national soccer teams, showed higher efficiencies in the 1-on-1 duels (Szwarc et al., 2017). Evidence of the importance of the defensive characteristics was also found in World Cup 2014: tackles and aerial advantage positively affect the probability of winning a game, but the latter effect turned to be insignificant in close games (Liu et al., 2015). All in all, there is evidence from different soccer leagues and championships for the significant contribution of defensive actions to success, in soccer.

On the other hand, there is evidence of the undervaluation of defensive skills by the soccer labor market. The Ballon d'Or - an annual award for the most valuable player in European soccer based on the voting of soccer experts, has nominated only one player with a defensive specialization (goalkeeper) in the top 3, during the period 2011-2016. At the same time, according to the plus-minus rating method, five players with defensive specialization (three defenders, one defensive midfielder and one goalkeeper) should have been included in that list (Kharrat et al., 2017). These two facts together, point out that soccer society is likely to overestimate offensive actions, thus, defensive actions are being underrated. This conclusion is consistent with the results of salary research of German soccer from 1996 to 2007 (Frick, 2011). It was indicated that fees for players' skills, ceteris paribus, is lower for defenders and midfielders compared to forwards.

This research contributes to the labor market inefficiencies literature together with the literature studying the role of offensive and defensive actions in soccer, by measuring and comparing the contribution of normalized offensive and defensive action measures, to both team performance and players' remuneration.

Empirical Analysis

Data

For the empirical test of the stated hypothesis, we use data from the 2017/2018 and 2018/2019 seasons of the German Bundesliga. There are two separate datasets. The first database is a team-game level, and includes 612 matches with 2 observations per game. 20 teams took part in the two seasons included in the study. The data originates from four different sources: performance team data were collected on a website of soccer statistics (whoscored.com); distance statistics from the official Bundesliga website (www.bundesliga.com); attendance was taken from the statistics portal (https://fbref.com) and betting odds from a soccer betting website (https://www.football-data.co.uk).

The second database was on a player-season level and includes 772 observations. The data were collected from two different sources: players' market values and individual information were taken from a German transfer market website (www.transfermarkt.de); season performance data were found on the website of the soccer statistics (whoscored.com). All 959 players of German Bundesliga in the given period were considered for the construction of our dataset. Only players with at least one full game in a season were included in the final sample. This exclusion helps to reduce a potential sample selection bias, as fresh players may outperform, compared to the starting 11 players, during a short period at the end of a game. Moreover, the playing time is displayed incorrectly when a player enters the field at the end of the game because the actual overtime is not added to the playing time in the official statistics (Wicker et al., 2013). The players not listed on the transfer market website were excluded from the final sample. In cases when a player was transferred during the season, to a team outside the Bundesliga, his market value at the end of the season was replaced by his market value at the moment of transfer. Goalkeepers are excluded from the sample as their game statistics are not comparable with one of the outfield players. Accordingly, the final sample on a player-season level includes n=772 observations.

Following the established tradition in sports economics literature, we measure players' remuneration with the aggregated expert estimates of players' transfer values, provided by the website www.transfermarkt.de. Although such an estimation procedure seems to be subjective, previous research provides evidence of this measures validity by demonstrating a high correlation of market values and salaries (Franck & Nüesch, 2012; Frick, 2007). In order to avoid reverse causality problems between performance characteristics and market value (Angrist & Pischke, 2008), which is broadly studied in the sports economics field (e.g., Torgler, Schmidt, 2007; Nüesch, 2009), market values are collected at the end of each of the two seasons, while performance variables are averaged within the season, i.e., performance precedes remuneration.

Overview of the variables used in two parts of the research is presented in Tables 1 and 2.

Table 1. Overview of variables in the team level models

Variable	**Description**	**Scale**
Dependent variable (team performance)		
WIN	Team won the game (1 = yes)	Dummy
Offensive statistic		
SHOT	Number of teams' shots in a game	Metric
Defensive statistics		
TACKLE	Number of teams' successful tackles in a game	Metric
INTERCEPT	Number of teams' interceptions in a game	Metric
CLEARANCE	Number of teams' clearances in a game	Metric
Control variables		
AGE	Average age of players in the starting squad	Metric
PASS	Number of teams' total passes in a game	Metric
RUN	Distance run by all players (in km)	Metric
FOUL	Number of teams' fouls commited in a game	Metric
Match characteristics		
HOME	Game is a home game (1 = yes)	Dummy
ATTEND	Number of spectators on match day divided by 10,000	Metric
BET_ODDS	Adjusted winning probability	Metric
Fixed effects		
TEAM 1-21	Observed team (20 teams)	Nominal
TEAM_OPP 1-21	Observed team of opponent in a game (20 teams)	Nominal
GAME_DAY 1-34	Time variable	Ordinal
SEASON	Dummy for the season (2017-2018 season as the baseline)	Dummy

Table 2. Overview of variables in the player level models

Variable	Description	Scale
Market value		
MV_END	Market value of the player in the end of the season, in €	Metric
MV_LOG	Logarithm of market value in the end of the season	Metric
MV_DIFF	Difference between player's market value in the end of the season and in the beginning, in €	Metric
MV_START	Market value of the player in the beginning of the season, in €	Metric
Offensive statistic		
SHOT	Average number of player’s shots in a season per minute	Metric
SHOT&KPASS	Sum of average number of player's shots and key passes in a season per minute	Metric
Defensive statistics		
TACKLE	Average number of player's successful tackles in a season per minute	Metric
INTERCEPT	Average number of player's interceptions in a season per minute	Metric
CLEARANCE	Average number of player's clearances in a season per minute	Metric
Human capital and effort		
AGE	Number of ages in the end of the season (in years)	Metric
AGE2	AGE squared	Metric
POSITION	Dummies for position of the player	Dummy
APPEAR	Number of appearances in season	Metric
FOUL	Average number of player’s fouls in a season per minute	Metric
PASS	Average number of player’s passes in a season per minute	Metric
Fixed effects		
TEAM	Dummies for team of the player	Dummy
SEASON	Dummy for the season (2017-2018 season as the baseline)	Dummy

In line with previous research (e.g., Dewenter & Namini, 2013; Leard & Doyle, 2011; Wicker et al., 2013), the dependent variable WIN in a team level model measures team performance as an indicator of the observed team scoring more goals in relation to the opponent. Other studies on the team performance determinants in soccer have also used ratios and absolute differences in scores as a measure of game outcome (e.g., Clarke & Norman, 1995; Mechtel et al., 2011), as well as ordinal outcomes (3, 1, and 0 points per game) and a scale of win home, draw, and win away (Audas et al., 2002; Bäker et al., 2012; Koning, 2000). However, Wicker et al. (2013) have noted that the points obtained for a victory are more important than a number of goals scored since, in most top-leagues, goals solely are taken into consideration only in cases when teams finish the season with the same number of accumulated points. Moreover, according to Wicker et al. (2013), the ordinal rank order of points is neglected due to violation of the ordinal distribution assumption after the recent adoption of

3- instead of 2-point rule (Stock & Watson, 2015). Thus, an indicator of game victory is perceived to be superior to other performance measures.

We implement the approach of previous studies (e.g., Szwarc, 2008) to measure the offensive (1) and defensive (2) actions: (1) shots and key passes due to they create goal score opportunity and (2) clearances, interceptions, tackles due to they attempt to regain possession of the ball. The official definitions of chosen actions, presented by Opta[6] – the world's leading sports data provider, are shown in a table below. These common measures of offensive and defensive performance are introduced into both parts of the research in order to guarantee a comparability of the results. Moreover, we normalize the defensive and offensive variables using following procedure: $(x - mean(x)/ standard\ deviation(x))$ in order to make the resulting estimated coefficients comparable between the models of team success and plyers' compensation by absolute value.

Table 3. Definitions of the key variables of offense and defense

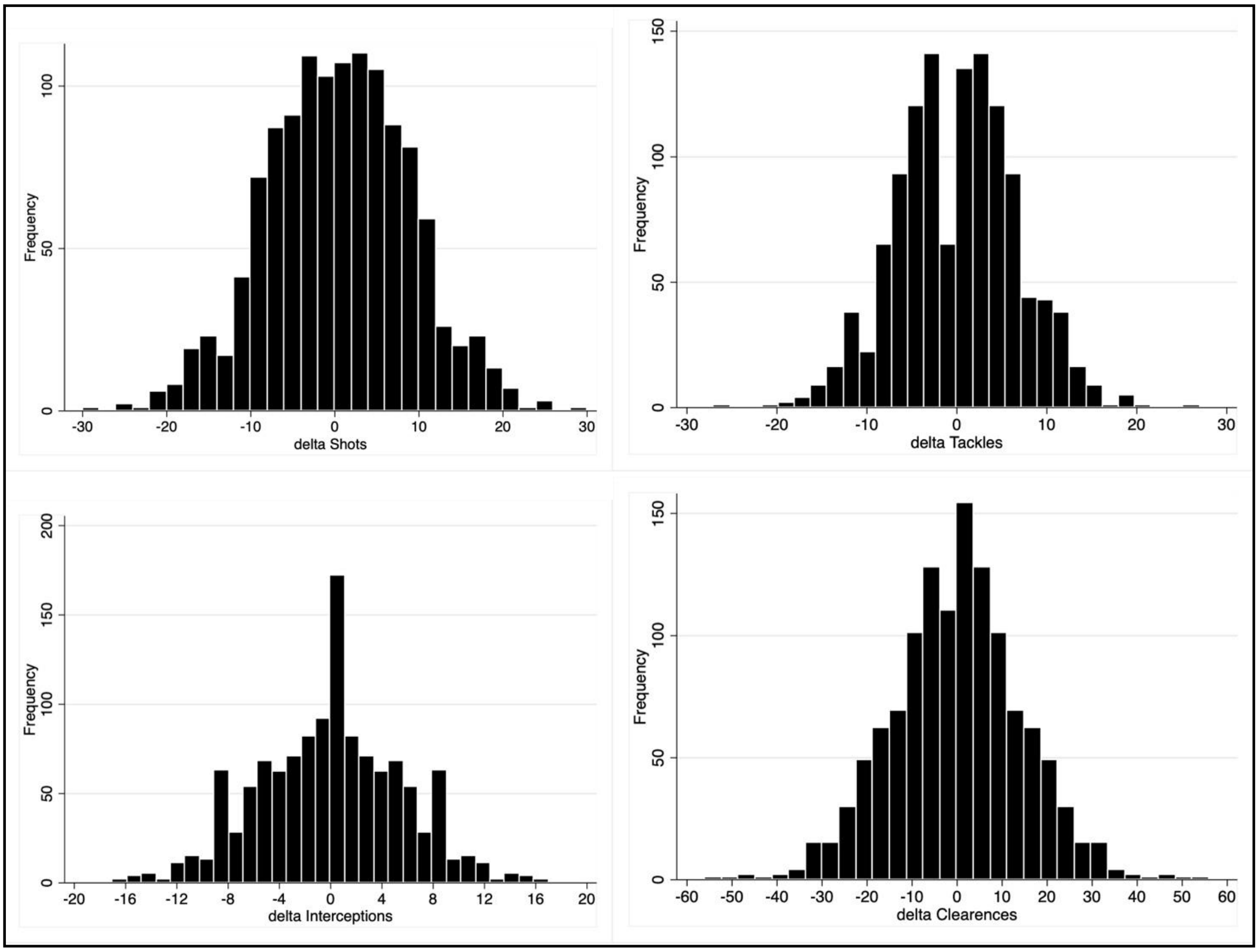
Game event	Definition
Offensive	
SHOT	A shot is defined as any clear attempt to score
KEY PASS	The final pass or pass-cum-shot leading to the recipient of the ball having an attempt at goal without scoring (included only to the player-level model to avoid doubling of attacking actions on a team level)
Defensive	
CLEARANCE	This is a defensive action where a player kicks the ball away from his own goal with no intended recipient
INTERCEPTION	This is where a player reads an opponent's pass and intercepts the ball by moving into the line of the intended pass
TACKLE WON	A tackle is defined as where a player connects with the ball in a ground challenge where he successfully takes the ball away from the player in possession. The tackled player must clearly be in possession of the ball before the tackle is made. A tackle won is deemed to be where the tackler or one of his team-mates regains possession as a result of the challenge, or that the ball goes out of play and is "safe".

These specific measures of offensive and defensive actions in soccer are not only acknowledged by academic society, but also well-known and broadly available to the soccer society, hence, we can expect that clubs' managers and soccer experts take them into account when the wages or transfer market prices are set. The offensive and defensive variables in our study imply both talent and effort as tackles could be won without an attempt, as well as hardly any tackle could be won without a soccer talent. But if we assume that the talent is predetermined and control for the predetermined factors in a model, then we will primarily see the effect of offensive and defensive effort through these variables.

The Figure 1 shows that, all the offensive and defensive variables are characterized by a close to normal distribution with most attacking teams attempting to shot 30 times more than the opponent in a game.

[6] https://www.optasports.com/

Figure 1. Histograms for offense, defense and deltas (N=776)



We use betting odds in our models to control the quality of the squad and several unobservable ex-ante characteristics in our database – physical and emotional form of the team (Coates & Humphreys, 2012). It is important to control betting odds in the studies which attempt to consider efforts due to the reason that in some matches an increased level of player effort will not significantly increase the probability of a positive outcome for the team. For example, it can be the case in matches of weaker vs top clubs like Bayern Munich or Borussia Dortmund, when outsider needs to exert a disproportionally high effort to a get a winning chance. We use averaged odds of 36 bookmakers provided by betting portal Betbrain to construct the variable of pre-game information. The probabilities of a team win are corrected for the bookmakers' margin by dividing the inversed win coefficient by the sum of all three inversed coefficients (win, draw and loss).

As for the determinants of team success in soccer, no traditional production function can be applied here since the performance of team depends on the opponent. Thus, we use absolute difference measures for the team variables (SHOT, TACKLE, INTERCEPT, CLEARANCE, AGE, PASS, RUN, FOUL) according to the tradition set in the research of soccer teams performance factors to control for opponents' actions and playing style (e.g. Leard & Doyle, 2011; Weimar & Wicker, 2017). Passing activity controls for possession (correlation coefficient is significant and is close to 1). It is important to account for this factor as possession is necessary for the opportunity to exert offensive and defensive efforts. We do not have data on possession at players' level, so we use passing activity in both parts of the empirical research instead of possession. Running effort may also be associated with offensive and defensive effort, but the individual tracking data are not available, so we include this variable in the number of control variables only at team level. Other success factors are match characteristics (ATTEND, BET_ODDS), no differences are used for these variables. Attendance allows us to control for the crowd pressure and support during the game.

It also worth mention that the game strategy can be associated with the level of offensive and defensive effort exerted by team. For example, some clubs prefer high pressure in attempt to take the ball away to the opponent's side of field which requires a lot of defensive effort from the players, while other clubs rather prefer a positional defense when a player defends in regards to where the ball and their teammates are, which requires less defensive effort from players. Although the real game tactic is unobservable and complex matter, we can expect to some extend established strategic pattern for each club during a season, so we include TEAM dummies to account for that. We also insert GAME DAY and SEASON dummies to control for some unobservable time characteristics.

In the second part of our empirical analysis the dependent variable is player's logged market value since market value distribution is skewed (Mincer, 1974). In addition, absolute difference in market value between the beginning and the end of the season is calculated to account for the player's market value development factors. Following the traditional signaling theory (Spence, 1978), we assume that players are motivated more by future market value increases rather than by present value, thus we exclude the reverse causality for offensive and defensive performance characteristics. Market value for start of the season is used as a reference (Kahnemann & Tversky, 1979) point since changes in market value tend to decrease for higher market values.

This part of study controls for various covariates including human capital and match characteristics that could potentially affect player's market value. Controlling for these variables is important to isolate the effect of defensive and offensive actions. Human capital is measured by a player's age (AGE & AGE2) in the quadratic form due to the commonly known u-shape profile of earnings during a career, position in the field (POSITION), the number of appearances on the field in the season (APPEAR), which can be associated with defensive and offensive actions. On one hand, the most productive players may appear more frequently, on the other hand, fresh players, who do not appear in each game may have more energy and perform better. We also control for fouls and passes (FOUL, PASS), as may affect the number and efficiency of offensive and defensive efforts. TEAM and SEASON stay for unobservable effects of each team and season.

In the sports labor market, we can observe superstars that have high market values due to a shortage of labor supply and a limited number of positions in the starting squad (Rosen, 1981), what causes a need to account for a potential sample bias due to the superstars effect. We account for it by excluding 16 outliers from our sample (the 1st and 99th percentiles of the sample market value distribution; Vogel & Wagner, 2011) after the full sample estimations. The results have not changed significantly thus our models have passed this robustness check.

Table 4. Summary statistics for team level data set

Variable	Mean	Std. Dev.	Min	Max
WIN	0.372	0.483	0	1
SHOT	13.056	5.108	1	33
TACKLE	18.123	5.475	5	46
CLEARANCE	20.304	8.826	2	61
INTERCEPT	11.547	4.194	0	29
PASS	452.803	122.356	174	1059
RUN	116.488	4.544	98.5	129.7
AGE	25.713	0.946	23.3	27.4
FOUL	12.870	4.108	1	28
ATTEND	44053.56	16623.5	19205	81365
BET_ODDS	0.378	0.182	0.02	0.92

Table 5. Summary statistics for player level data set

Variable	Mean	Std. Dev.	Min	Max
MV_END	10.149	13.692	0	100
MV_DIFF	2.464	8.006	-30	85
MV_LOG	1.613	1.240	-2.303	4.605
SHOT	0.015	0.011	0	0.058
SHOT&KPASS	0.026	0.016	0	0.077
TACKLE	0.021	0.010	0	0.075
CLEARANCE	0.020	0.018	0	0.081
INTERCEPT	0.012	0.008	0	0.053
FOUL	0.015	0.008	0	0.057
PASS	0.446	0.166	0.094	1.041
APPEAR	19.925	8.920	1	34

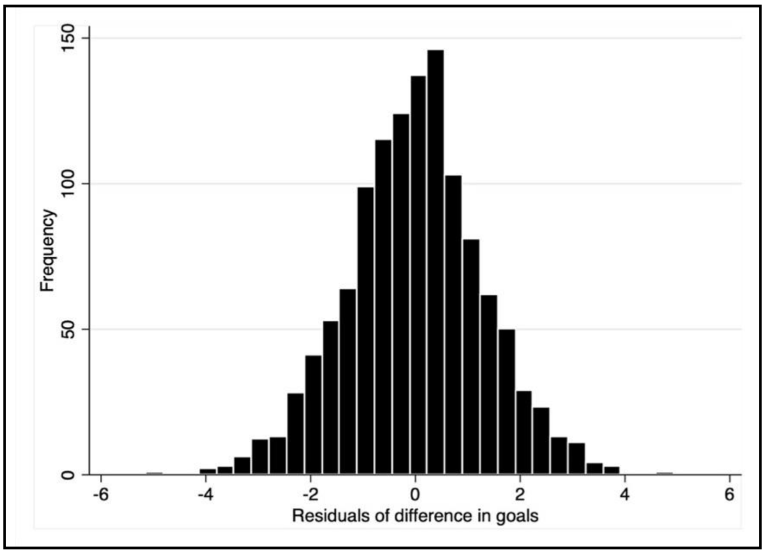
As can be seen from the descriptive statistics in Table 5, the highest value of shots per minute is 0.058, it belongs to a famous polish forward Robert Lewandowski. He is four-time best-scorer of the Bundesliga. Also, almost as high statistics achievement belongs to Dortmund's forward Paco Alcacer (0.049), Andrej Kramaric from Hoffenheim (0.46) and Timo Werner from Leipzig (0.43). If we consider shots plus key passes statistics, James Rodrigues from Bayern Munich had the highest value (0.077). Salif San from Hannover 96 had the maximum number of clearances in a season (0.08). In his case, this achievement together with other factors led to a significant increase in market value from 5 to 18 mln. Benjamin Pavard from Stuttgart also had high value of clearances per minute (0.07), he was one of the major discoveries of the World Cup in Russia 2018. Willi Orban had good results in clearances in both observed seasons had (0.067 and 0.07). During this time his market value increased from 8 to 20 mln. Arturo Vidal - popular defense player from Bayern Munich, had one of the highest numbers of tackles (0,047). The leaders in interceptions statistics are defenders from William Wolfsburg (William - 0,031), Schalke 04 (Benjamin Stambuli - 0.029) and Bayern's (Mats Hummels - 0.025). Thus, we see, that defensive players achieve more in terms of our defensive variables what serves as an evidence of a measure adequacy.

Model Specifications

We begin with regressing the team success and employ linear probability model and probit model to determine the contribution of offensive and defensive actions. We test the hypothesis that error terms are normally rather than logistically distributed (logit model). Following previous research (Weimar & Wicker, 2017) we report the distribution of residuals of difference in goals after regressing it on all the explanatory variables (Wooldridge, 2016) and find its distribution being close to normal (Figure 2).

Before we move to the estimation of models, we analyses Spearman correlation matrix (Table 6) with respect to multicollinearity issue. We observe the significant correlation between the variables of offensive and defensive variables, and betting odds. However, there are no correlation coefficients violating the critical threshold of 0.9. We do not exclude any variables from our models due to multicollinearity factors and after the estimations we get VIF values no greater than 10. Altogether these two tests exclude the multicollinearity issue.

Figure 2. Histogram of the residuals of difference in scored goals (when regressed on all factors, N=1,224)



Altogether, eight models are estimated in this part of our study devoted to team winning probability factors. Firstly, we estimate 4 different specifications with linear probability model (Table 8), and after that we estimate the same specifications with probit model (Table 9). This approach allows us to check for the robustness of results to estimation method. The specification (1)-(3) are based on 3 different subsamples: total, for home and away teams accordingly. It is justified by a concern of a double count for such type of observations for one match and two teams. Moreover, the models in differences can be constructed from two perspectives: $HOME - AWAY$ or $AWAY - HOME$ teams. Thus, we use two subsamples together with full sample. The fourth specification excludes tackles from the list of independent variables as it turns to be insignificant in models (1)-(3), and uses full sample for the estimation.

The second part of our empirical analysis includes 6 different specifications. Ordinary least squares estimation with robust to heteroscedasticity standard errors (White, 1980) is applied to all of the specifications. Absolute transfer market value change stands for the dependent variable in models (1) and (2), while logged market value in the end of season serves as an outcome variable. Models (1), (3) and (5) include SHOT as offensive variable, making this models comparable with the models from the first part of our empirical analysis, while models (2), (4) and (6) use SHOT&KPASS variable as a measure of offensive actions, which allows to get the more precise measure at the individual players' level, as making a key pass may be as equally important as attempting to make a shot for subsequent scoring of a goal. However, it is inconvenient to include the latter variable at the team level as shots and key passes get highly correlated and yield a double counting of attacking actions for the team.

Before the models' estimation we also analyse Spearman correlation matrix (Table 7) with respect to multicollinearity issue. We similar to the first part of empirical research correlation of key variables, so there is also no need to exclude any variable from the models. VIF factors do not raise any concern of multicollinearity either.

Table 6. Spearman correlation matrix for team level data set

	WIN	SHOT	TACKLE	INTERCEPT	CLEARENCE	RUN	FOUL	BET_ODDS	AGE
WIN	1.0000								
SHOT	0.2825*	1.0000							
TACKLE	0.0058	-0.1873*	1.0000						
INTERCEPT	0.0381	-0.2073*	0.0710*	1.0000					
CLEARENCE	0.1567*	-0.4981*	0.2155*	0.2882*	1.0000				
RUN	0.2790*	0.0702*	0.0784*	-0.0210	0.1676*	1.0000			
FOUL	-0.0584*	-0.1411*	0.0927*	-0.0007	0.1134*	0.0455	1.0000		
BET_ODDS	0.3891*	0.6003*	-0.1903*	-0.2489*	-0.3940*	-0.0836*	-0.1639*	1.0000	
AGE	0.0395	-0.0212	0.1672*	0.0814*	-0.0110	-0.2259*	0.0671*	0.0215	1.0000
ATTEND	0.0403	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* p<0.05

Table 7. Spearman correlation matrix for player level data set

	MV_LOG	MV_DIFF	TACKLE	INTERCEPT	CLEARENCE	SHOT	SHOTS&KPASS	APPEAR	FOUL	AGE
MV_LOG	1.0000									
MV_DIFF	0.4933*	1.0000								
TACKLE	-0.1353*	-0.0267	1.0000							
INTERCEPT	-0.0284	0.0292	0.5438*	1.0000						
CLEARENCE	-0.0766*	0.0188	0.2692*	0.5876*	1.0000					
SHOT	0.2247*	0.0807*	-0.3767*	-0.5964*	-0.6514*	1.0000				
SHOT&KPASS	0.2382*	0.0769*	-0.3225*	-0.6011*	-0.7211*	0.9233*	1.0000			
APPEAR	0.5290*	0.4052*	-0.0953*	-0.0112	0.0312	0.1367*	0.1692*	1.0000		
FOUL	-0.1174*	-0.0461	0.1746*	-0.0915*	-0.2023*	0.2270*	0.1855*	-0.0740*	1.0000	
AGE	-0.2300*	-0.3897*	-0.0177	0.0027	0.0861*	-0.0289	0.0163	0.0994*	-0.0337	1.0000
PASS	0.2405*	0.0562	0.3175*	0.5401*	0.4538*	-0.4619*	-0.4183*	0.1028*	-0.2507*	0.1055*

* p<0.05

Results and discussion

We begin with a linear probability model estimates for the win indicator dependent variable (Table 8). All the offense and defense variable except for difference in tackles are highly significant in all the specifications. The relative contribution of offense and defense is about 27 and 73% accordingly (Models (1-4)). It can not be interpreted as a Moneyball until we do not have the corresponding relation for the players' market value so for now we will just keep these numbers in mind and will turn back to them after the second part of the empirical research is accomplished.

In line with the result of Weimer and Wicker (2017) we see a significant contribution of running effort difference to a team win. Age difference has a negative effect which means that physical ability effect turns to be a main driving force for this effect, while experience gets dominated for this sample. We also observe low significance of fouls difference. Match attendance effect is insignificant probably due to a high attendance rate in German Bundesliga, leading to a low variation of match attendance and destroying the effect of crowd pressure and support.

Table 8. Linear probability team level models

VARIABLES	TOTAL (1)	HOME–AWAY (2)	AWAY–HOME (3)	TOTAL (4)
ΔSHOT	0.0794***	0.0793***	0.0721***	0.0786***
	(4.830)	(3.308)	(2.944)	(4.758)
ΔTACKLE	0.0109	0.00988	0.00243	
	(0.895)	(0.552)	(0.136)	
ΔINTERCEPT	0.0354***	0.0630***	0.0124	0.0348***
	(3.009)	(3.654)	(0.734)	(2.965)
ΔCLEARENCE	0.188***	0.203***	0.162***	0.189***
	(13.07)	(9.972)	(7.618)	(13.40)
ΔRUN	0.0580***	0.0645***	0.0521***	0.0584***
	(15.66)	(11.29)	(9.346)	(15.70)
ΔFOUL	-0.00241	0.00127	-0.00623*	-0.00231
	(-1.116)	(0.373)	(-1.883)	(-1.072)
BET_ODDS	0.971***	0.675**	0.915***	0.977***
	(7.201)	(2.269)	(3.105)	(7.268)
ΔAGE	-0.0539**	-0.0380	-0.0548*	-0.0507**
	(-2.500)	(-1.230)	(-1.783)	(-2.361)
ATTEND	-2.28e-07	2.59e-07	-4.72e-06	-2.30e-07
	(-0.239)	(0.0580)	(-1.110)	(-0.240)
SEASON	incl	incl	incl	incl
ROUND 1-34	incl	incl	incl	incl
TEAM DUMMY 1-20	incl	incl	incl	incl
CONSTANT	0.0169	0.170	0.202	0.0150
	(0.151)	(0.608)	(0.812)	(0.135)
OBS	1,224	612	612	1,224
R^2	0.466	0.531	0.450	0.466

Robust t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The next bunch of models for win indicator dependent variable is estimated with probit regression. We report the estimation coefficients rather than marginal effects for the probit models (Table 9), as we are interested in the significance and the relative contribution of the key variables of offense and defense. These coefficients can not be interpreted by their absolute value separately, however, they satisfy our needs to analyse the Moneyball effect, thus we omit the marginal effect from the consideration.

The estimation results are quite similar to the linear probability models which provides an evidence of the robustness to estimation method. Difference in interceptions effect loses its significance for the AWAY-HOME teams subsample, the other key variables' effects stay qualitatively the same. The relative contribution of offense and defense changes slightly to 30 and 70% accordingly (Models (1-4)).

Table 9. Probit team level models

	TOTAL	HOME–AWAY	AWAY–HOME	TOTAL
VARIABLES	(1)	(2)	(3)	(4)
ΔSHOT	0.390***	0.545***	0.384***	0.383***
	(5.165)	(4.695)	(3.058)	(5.031)
ΔTACKLE	0.0618	0.0518	0.0386	
	(1.139)	(0.600)	(0.491)	
ΔINTERCEPT	0.154***	0.360***	0.0334	0.150***
	(3.033)	(4.833)	(0.411)	(2.961)
ΔCLEARENCE	0.812***	1.117***	0.814***	0.816***
	(10.85)	(9.948)	(7.022)	(10.92)
ΔRUN	0.280***	0.395***	0.298***	0.282***
	(13.43)	(11.15)	(8.746)	(13.51)
ΔFOUL	-0.0130	0.0123	-0.0439***	-0.0123
	(-1.332)	(0.743)	(-2.818)	(-1.265)
BET_ODDS	3.830***	2.802**	3.637***	3.840***
	(6.268)	(1.982)	(2.576)	(6.287)
ΔAGE	-0.246**	-0.210	-0.296**	-0.229**
	(-2.528)	(-1.452)	(-1.978)	(-2.350)
ATTEND	-3.91e-06	-2.41e-05	-3.07e-05*	-3.76e-06
	(-0.831)	(-1.304)	(-1.752)	(-0.799)
SEASON	incl	incl	incl	incl
ROUND 1-34	incl	incl	incl	incl
TEAM DUMMY 1-20	incl	incl	incl	incl
CONSTANT	-1.665***	-0.335	-7.293***	-1.672***
	(-3.525)	(-0.278)	(-5.838)	(-3.568)
OBS	1,224	612	612	1,224

Robust z-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Regression results for the second part of empirical research are presented in Table 10. Market value for the start of season has a negative effect on an absolute increase in market value (models (1) and (2)) and a positive effect on the logged market value at the end of season (models (3)-(6)). This is due to a positive and decreasing marginal product of labor. We observe only offense variable being significant in the first to specifications while the variables of defense turn to be insignificant in these models. Combining this fact together with the significance of both offense and defense variables in the first part of our empirical research yields so called "absolute Moneyball" effect similar to the one observed in running efforts by Weimer and Wicker (2017). However, we do observe a significant contribution of defensive actions to market value in the second group of models with logged market value as a dependent variable ((3)-(6)). After excluding tackles variable from models, we compare the relative contribution of offensive and defensive actions to team winning probability and to a player's market value. For model (5) we get 78 and 22% (out of the total contribution of offensive and defensive actions calculated as a sum of estimated coefficients for normalized variables of offense and defense) contribution of offense and defense correspondingly, and 60 to 40% in model (6). Both results compared to corresponding 25 and 75% contribution to a team win provides an evidence of a so called "relative Moneyball" effect, as relative contribution of defensive to offensive actions in more than three or almost two times larger (3.38 and 1.89) in a case of team win (Table 8, model (4)) compared to a contribution to a market value (Table 10, models (5) and (6) correspondingly).

We can also observe a traditional to Mincer type regressions U-shaped relation of remuneration and players age in both types of models. The second group of models ((3)-(6)) yields a standard positive and decreasing effect of age, while the first two specifications for the change in market value as a dependent variable demonstrate the reverse effect of age. It means that the change in market value decreases quadratically with age. The effect of fouls is negative in specifications (1), (2), (5) and (6). This fact can be explained by a negative effect of fouls on a game ceteris paribus, as a team loses the ball after each foul committed. Appearances are associated positively with a market value change. This effect can actually be two-sided: appearances help player to demonstrate his talent and abilities, on one hand, at the same time more talented players appear on the pitch more frequently. As we control for the starting market value, which is a proxy of players' talent, we can expect the former effect to be dominating in this model setup. The effect of tackles is negative in models (1), (3) and (4). This result seems to be counterintuitive, a positive effort yields a negative payoff in terms of player's market value. However, this effect has a logical explanation: the more talented a player is, the better he defends without a ball, that is the need to take the ball away and make a tackle itself may appear due to positional errors committed by a player, while good players force the opponent to lose a ball by themselves. The case of Manchester United defender Aaron Wan-Bissaka can clearly illustrate this thesis. He had one of the highest numbers of tackles per played minute (0,047) in English Premier League (EPL) during the season 2018-2019, but experts admit that this outstanding statistic is driven by the multiple positional errors committed[7].

[7] https://www.telegraph.co.uk/football/2019/03/08/aaron-wan-bissaka-became-effective-defender-europe/, Retrieved 04.02.2020, 13:00

Table 10. OLS player level models

	DEPENDENT VARIABLE					
	TMV_DIFF		TMV_LOG			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
MV_START	-0.162***	-0.159***	0.0421***	0.0419***	0.0426***	0.0426***
	(-3.623)	(-3.469)	(4.996)	(4.893)	(4.869)	(4.770)
TACKLE	-0.453**	/	-0.0859**	-0.0966***	/	/
	(-2.142)		(-2.476)	(-2.924)		
INTERCEPT	0.252	0.0708	0.0890***	0.0961***	0.0577**	0.0593**
	(1.586)	(0.340)	(3.196)	(3.369)	(2.403)	(2.530)
CLEARENCE	0.0429	0.149	0.0139	0.0573	0.0330	0.0799*
	(0.182)	(0.651)	(0.408)	(1.461)	(1.030)	(2.033)
SHOT	0.857**	0.954***	0.186***	/	0.202***	/
	(2.800)	(3.062)	(5.747)		(5.901)	
SHOT&KPASS	/	/	/	0.204***	/	0.212***
				(5.939)		(5.631)
APPEAR	0.352***	0.356***	0.0568***	0.0560***	0.0575***	0.0569***
	(7.164)	(7.173)	(22.55)	(20.31)	(21.49)	(19.65)
FOUL	-62.09**	-72.25**	-4.580	-2.988	-6.786*	-5.407
	(-2.307)	(-2.450)	(-1.042)	(-0.670)	(-1.735)	(-1.344)
AGE	-2.655**	-2.632**	0.382***	0.370***	0.389***	0.378***
	(-2.156)	(-2.129)	(4.971)	(4.889)	(5.010)	(4.943)
AGE2	0.0362	0.0359	-0.00880***	-0.00861***	-0.00892***	-0.00874***
	(1.591)	(1.572)	(-6.416)	(-6.373)	(-6.434)	(-6.403)
PASS	2.080	1.576	0.615***	0.506**	0.539***	0.401**
	(0.644)	(0.482)	(3.020)	(2.638)	(2.979)	(2.454)
TEAM	incl	incl	incl	incl	incl	incl
SEASON	incl	incl	incl	incl	incl	incl
POSITION	incl	incl	incl	incl	incl	incl
CONST	40.82**	40.71**	-3.869***	-3.661***	-3.937***	-3.730***
	(2.446)	(2.420)	(-3.594)	(-3.481)	(-3.606)	(-3.503)
OBS	776	776	775	775	775	775
R^2	0.279	0.277	0.683	0.684	0.678	0.679

Robust clustered by team t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Conclusion

Current situation with relative evaluation of offensive and defensive actions in German Bundesliga is examined in this paper. We first find an evidence of undervaluation of top defensive players compared to forwards by an annual prestigious soccer award – Ballon d'Or (Kharrat et al., 2017) together with an evidence of a significant role of defense in a team performance (e.g. Vogelbein et al., 2014; Szwarc et al., 2017). Moreover, we already have a documented phenomenon of Moneyball in different types of sports including soccer (e.g. Berri et al., 2006; Weimar, Wicker, 2017), so we know that sports labor markets sometimes fail to set efficient wages. Thus, we formulate a research hypothesis that defensive actions are undervalued by transfer market compared to offensive actions.

We examine the stated hypothesis on a Bundesliga 2017-2019 seasons by a consecutive estimation of the two groups of models: firstly, we determine a relative contribution of offensive and defensive actions to a team win, while controlling for playing effort, match characteristics and fixed unobserved team and time effects. After that we move to the estimation of a relative contribution of offensive and defensive actions to a plyer's market value, while controlling for human capital, effort and unobserved team and time effects. We finally get the relative contribution of offense and defense to a team win varying from 25 and 75% to 30 and 70% in two different models correspondingly, while the same contributions change from not significantly different from 0 to 78 and 22% or 60 and 40% in three different sets of market value models. A positive vs zero contribution of defensive actions to a team win vs to a market value is interpreted as an "absolute Moneyball" in contrast to a positive contribution in both cases. For these models we find that relative contribution of defensive actions in from two to three times larger for the team win compared to a contribution to market value. These results are classified as a "relative Moneyball" in our paper. This is a first introduction of this term to a labor market failure, to our knowledge. Thus, this paper contributes to the literature not only by discovering a particular case of a soccer labor market failure, but also suggest an empirical method to detect a "relative Moneyball" effect in any production factor market.

As result, we can expect that the soccer clubs in Bundesliga have a potential to decrease a cost of "producing" a win by redistributing funds between offense and defense. Similar effect is likely to persist in other top soccer leagues as well, as the corresponding labor markets are highly connected to each other. It should be mentioned, that we have investigated the issue of labor market inefficiency by considering sporting success as an ultimate and unique goal of professional soccer club, however, there is a literature which considers the two different club's objectives: financial and sporting success (e.g. Szymanski & Smith, 1997). Thus, taking the financial goal into account may weaken the Moneyball conclusion, as offensive players may have an advantage in indirect incomes from their popularity, however, it is hard to imagine, that these difference between forwards and offensive players may reach 100 or 200% of their market value on the market on average. So, we can expect, that this effect will persist even after the adjustment for non-direct incomes.

The possible extension to this research could be an investigation of players' incentives to exert an extra effort in defense, when it is payed less than marginal productivity. This setting may help to detect a non-financial motivation of players, which may contribute to a sustainability of the Moneyball effect in offense vs defense.

References

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Atkinson, S. E., Stanley, L. R., & Tschirhart, J. (1988). Revenue sharing as an incentive in an agency problem: An example from the National Football League. *The Rand journal of economics*, 27-43.

Audas, R., Dobson, S., & Goddard, J. (2002). The impact of managerial change on team performance in professional sports. *Journal of Economics and Business*, *54*(6), 633-650.

Bäker, A., Mechtel, M., & Vetter, K. (2012). Beating thy neighbor: derby effects in German professional soccer. *Jahrbücher für Nationalökonomie und Statistik*, *232*(3), 224-246.

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

Berri, D. J., Brook, S. L., & Schmidt, M. B. (2007). Does one simply need to score to score. *International Journal of Sport Finance*, *2*(4), 190-205.

Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: Footedness and player remuneration in European soccer. *Journal of Sports Economics*, *14*(6), 606-628.

Carmichael, F., & Thomas, D. (2005). Home-field effect and team performance: evidence from English premiership football. *Journal of sports economics*, *6*(3), 264-281.

Clarke, S. R., & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. Journal of the Royal Statistical Society: Series D (The Statistician), 44(4), 509-521.

Coates, D., & Humphreys, B. R. (2012). Game attendance and outcome uncertainty in the National Hockey League. *Journal of Sports Economics*, *13*(4), 364-377.

Dewenter, R., & Namini, J. E. (2013). How to make soccer more attractive? Rewards for a victory, the teams' offensiveness, and the home bias. *Journal of Sports Economics*, *14*(1), 65-86.

Franck, E., & Nüesch, S. (2012). Talent and/or popularity: what does it take to be a superstar?. *Economic Inquiry*, *50*(1), 202-216.

Frick, B. (2011). Performance, salaries, and contract length: Empirical evidence from German soccer. International Journal of Sport Finance, 6(2), 87.

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. *Journal of Economic Perspectives*, *20*(3), 173-186.

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

Hicks, J. R. (1932). Marginal productivity and the principle of variation. *Economica*, (35), 79-88.

Holmes, P. M., Simmons, R., & Berri, D. J. (2018). Moneyball and the Baseball Players' Labor Market. *International Journal of Sport Finance*, *13*(2).

Kahneman, D. (1979). Tversky A. (1979). Prospect theory: an analysis of decision under risk, 263-292.

Kharrat, T., Pena, J. L., & McHale, I. (2017). Plus-minus player ratings for soccer. *arXiv preprint arXiv:1706.04943*.

Koning, R. H. (2000). Balance in competition in Dutch soccer. Journal of the Royal Statistical Society: Series D (The Statistician), 49(3), 419-431.

Lago-Ballesteros, J., & Lago-Peñas, C. (2010). Performance in team sports: Identifying the keys to success in soccer. *Journal of Human Kinetics*, *25*, 85-91.

Lago-Peñas, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, *27*, 135-146.

Leard, B., & Doyle, J. M. (2011). The effect of home advantage, momentum, and fighting on winning in the National Hockey League. *Journal of Sports Economics*, *12*(5), 538-560.

Lewis, M. (2003). *Moneyball: The art of winning an unfair game.* New York, NY: Norton.

Liu, H., Gomez, M. Á., Lago-Peñas, C., & Sampaio, J. (2015). Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. Journal of sports sciences, 33(12), 1205-1213

Massey, C., & Thaler, R. H. (2013). The loser's curse: Decision making and market efficiency in the National Football League draft. Management Science, 59(7), 1479-1495.

Mechtel, M., Bäker, A., Brändle, T., & Vetter, K. (2011). Red cards: not such bad news for penalized guest teams. *Journal of Sports Economics*, *12*(6), 621-646.

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

Nüesch, S. (2009). Are demographic diversity effects spurious?. *Economic Analysis & Policy*, *39*(3), 379-388.

Rosen, S. (1981). The economics of superstars. The American economic review, 71(5), 845-858.

Spence, M. (1978). Job market signaling. In Uncertainty in economics (pp. 281-306). Academic Press.

Staw, B. M., & Hoang, H. (1995). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. Administrative Science Quarterly, 474-494.

Stock, J. H., & Watson, M. W. (2015). Introduction to econometrics.

Szwarc, A. (2008). The efficiency model of soccer player's actions in cooperation with other team players at the FIFA World cup. *Human Movement*, *9*(1), 56-61.

Szwarc, A., Kromke, K., Radzimiński, Ł., & Jastrzębski, Z. (2017). Efficiency of 1-on-1 play situations for high-level soccer players during the World and European championships in relation to position on the pitch and match time. *International Journal of Sports Science & Coaching*, *12*(4), 495-503.

Szymanski, S., & Smith, R. (1997). The English football industry: profit, performance and industrial structure. *International Review of Applied Economics*, *11*(1), 135-153.

Torgler, B., & Schmidt, S. L. (2007). What shapes player performance in soccer? Empirical findings from a panel analysis. *Applied Economics*, *39*(18), 2355-2369.

Vogel, A., & Wagner, J. (2011). *Robust estimates of exporter productivity premia in German business services enterprises* (No. 207). University of Lüneburg Working Paper Series in Economics

Vogelbein, M., Nopp, S., & Hökelmann, A. (2014). Defensive transition in soccer–are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. Journal of sports sciences, 32(11), 1076-1083.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *econometrica*, *48*(4), 817-838.

Wicker, P., Prinz, J., Weimar, D., Deutscher, C., & Upmann, T. (2013). No pain, no gain? Effort and productivity in professional soccer. International Journal of Sport Finance, 8(2), 124.

Weimar, D., & Wicker, P. (2017). Moneyball revisited: Effort and team performance in professional soccer. *Journal of Sports Economics*, *18*(2), 140-161.

Wooldridge, J. M. (2016). Introductory econometrics: A modern approach. Nelson Education

APPENDIX

There might be a concern of highly correlated defense variables, thus we use principal component analysis to construct an index of defensive actions in team level models in two-way: with all three variables of defense (INTERCEPTION, TACKLE, CLEARANCE) in models (1)-(2) and with excluded tackles (INTERCEPTION, CLEARANCE) in models (3)-(4). The proportion of information included in the first principal component amounts to 46 and 64% correspondingly. The resulting estimates yield the significant relative contribution of offensive and defensive actions to a team win varying from 23,1 and 76,9% correspondingly (Table A1, model (3)) to 32,5 and 67,5% (Table A1, model (4)).

The same principal components are constructed for player level models and the proportion of information absorbed by the first principal component amounts to 60 and 75.5% for PCDEFENSE(3) and PCDEFENSE(2). We observe “absolute Moneyball” in models (1)-(2), that is the effect of defensive actions on market value is not significant but defensive actions significantly increase chances of football club to win a game. The estimates of our Model (3) clearly demonstrate a “relative Moneyball”, that is the relative contribution of defensive actions a team win is relatively larger (67.5-76,9%) compared to the contribution to a market value (about 28%).

Table A1. Linear probability and probit team level models with the principal component of defense variables

VARIABLES	(1) LPM	(2) PROBIT	(3) LPM	(4) PROBIT
ΔSHOT	0.0473***	0.239***	0.0432***	0.213***
	(3.020)	(3.667)	(2.727)	(3.249)
PCΔDEFENSE(3)	0.142***	0.604***		
	(12.48)	(11.14)		
PCΔDEFENSE(2)			0.144***	0.601***
			(12.21)	(10.69)
ΔRUN	0.0611***	0.279***	0.0632***	0.289***
	(16.44)	(13.14)	(17.17)	(13.80)
ΔFOUL	-0.00227	-0.0123	-0.00135	-0.00704
	(-1.026)	(-1.314)	(-0.610)	(-0.747)
BET_ODDS	0.917***	3.577***	0.973***	3.845***
	(6.738)	(6.219)	(7.154)	(6.491)
ΔAGE	-0.0583***	-0.229**	-0.0350	-0.128
	(-2.683)	(-2.492)	(-1.601)	(-1.382)
ATTEND	-2.05e-07	-3.03e-06	-2.28e-07	-2.78e-06
	(-0.211)	(-0.678)	(-0.232)	(-0.615)
SEASON	incl	incl	incl	incl
ROUND 1-34	incl	incl	incl	incl
TEAM DUMMY 1-20	incl	incl	incl	incl
Constant	0.0347	-1.630***	0.0164	-1.742***
	(0.303)	(-3.547)	(0.147)	(-3.891)
Observations	1,224	1,224	1,224	1,224
R-squared	0.434		0.436	

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2. OLS player level models with the principal component of defense variables

	DEPENDENT VARIABLE		
VARIABLES	TMV_DIFF	TMV_LOG	
	(1)	(2)	(3)
MV_START	-0.158***	0.0429***	0.0425***
	(-3.440)	(4.948)	(4.922)
PCDEFENSE(3)	-0.0711	0.0247	
	(-0.283)	(1.048)	
PCDEFENSE(2)			0.0750***
			(3.255)
SHOT	0.827**	0.169***	0.195***
	(2.677)	(4.620)	(5.143)
APPEAR	0.355***	0.0573***	0.0573***
	(7.207)	(20.94)	(21.07)
FOUL	-73.78**	-7.811*	-6.970*
	(-2.654)	(-1.957)	(-1.802)
AGE	-2.632**	0.391***	0.390***
	(-2.133)	(5.110)	(5.117)
AGE2	0.0359	-0.00896***	-0.00893***
	(1.574)	(-6.580)	(-6.594)
PASS	2.283	0.626**	0.482**
	(0.720)	(2.762)	(2.295)
TEAM	incl	incl	incl
SEASON	incl	incl	incl
POSITION	incl	incl	incl
Constant	40.45**	-4.012***	-3.942***
	(2.417)	(-3.713)	(-3.674)
Observations	776	775	775
R-squared	0.277	0.674	0.678
Number of tm	20	20	20

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We test the robustness of our results with respect to overcontrolling of variables. In Table A3 we compare the full linear probability model with the model excluding the BET_ODDS from control variables. The effect of fouls difference on a team win becomes significant and negative. Otherwise, results are qualitatively unchanged and the relative contribution of defensive actions to a team win is still greater than of offensive actions and slightly decreases from 73,8% (model (1)) to 65,7% (model (2)), while the share of explained variance falls.

Table A3. Linear probability team level models with included vs excluded BET_ODDS

VARIABLES	(1)	(2)
ΔSHOT	0.0794***	0.110***
	(4.830)	(6.980)
ΔTACKLE	0.0109	0.0148
	(0.895)	(1.209)
ΔINTERCEPT	0.0354***	0.0295**
	(3.009)	(2.425)
ΔCLEARENCE	0.188***	0.181***
	(13.07)	(12.32)
ΔRUN	0.0580***	0.0589***
	(15.66)	(15.69)
ΔFOUL	-0.00241	-0.00453**
	(-1.116)	(-2.051)
BET_ODDS	0.971***	
	(7.201)	
ΔAGE	-0.0539**	-0.0454**
	(-2.500)	(-2.049)
ATTEND	-2.28e-07	1.72e-07
	(-0.239)	(0.176)
SEASON	incl	incl
ROUND 1-34	incl	incl
TEAM DUMMY 1-20	incl	incl
Constant	0.0169	0.336***
	(0.151)	(3.175)
Observations	1,224	1,224
R-squared	0.466	0.441

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1