CHAPTER IV

INDIVIDUAL PLAYER CONTRIBUTIONS IN EUROPEAN SOCCER

ABSTRACT

This paper looks at applying new techniques to predict match outcomes in professional soccer. To achieve this models are used which measure the individual contributions of soccer players within their team. Using data from the top 25 European soccer leagues, the individual contribution of players is measured using high dimensional fixed effects models. Nine years of results are used to produce player, team and manager estimates. A further year of results is used to check for predictive accuracy. Since this has useful applications in player scouting the paper will also look at how well the models rank players. The findings show an average prediction rate of 45% with all methods showing similar rankings for player productivity. While the model highlights the most productive players there is a bias towards players who produce and prevent goals directly. This results in more attackers and defenders ranking highly than midfield players. There is potential for these techniques to be used against the betting market as most models produce better performance than many betting firms.

1. INTRODUCTION

This paper focuses on producing models with a high prediction rate in professional soccer. The main challenge is to produce accurate estimations of player ability, a problem analogous to research on worker productivity. The study of worker productivity is of great value to businesses in almost every industry. Hiring the best workers will improve the overall performance of business and increase profit. It is of great importance for firms to be able to screen potential employees efficiently to determine their value. Often it is difficult to assess the individual contributions of workers when their productivity is unobserved from previous firms or they are part of a team. For that reason, European soccer is a suitable industry since worker productivity is observed. Twenty-five top flight leagues are considered so that players can be tracked as they move between different teams. High dimensional fixed effect models are used to determine the productivity of individual players.

The models yield on average a 45% prediction rate with the different methods producing very similar player rankings. While wins and losses are predicted well the models struggle with predicting games which end in draws. Compared to betting firms all models perform reasonably with some able to outperform the betting firms for a few leagues. The highest ranked players in the models have often won the most prestigious soccer tournaments and play for the best teams. While the model highlights the most productive players there is a bias towards players who produce and prevent goals directly. This results in more attackers and defenders ranking highly than midfield players.

The paper sets out as follows. Section 2 contains a review of the relevant literature, section 3 presents the theory and section 4 the data. Section 5 shows the predictions and estimations while section 6 contains concluding remarks.

2. LITERATURE REVIEW

Previous work on player estimation ability can be found in team and player efficiency literature. The first methodology to relate team output to team input measures was established by Scully (1974) in his study of US baseball. Some research attempts to estimate production functions with a focus on performance at the game level over one or multiple seasons. Zak et al. (1979) estimate a Cobb-Douglas production function in basketball, identifying specific play variables which contribute towards team output. Scott et al. (1985) use a similar approach but an entire season rather than individual games is used as the unit of observation. Zech (1981) uses the Richmond technique to estimate the potential output of basketball teams. Schofield (1988) estimated production functions for English country cricket to develop strategies on and off the field. Carmichael and Thomas (1995) examine team performance over a season in rugby league by also including team characteristics as well as play variables. Ruggiero et al. (1996) use panel data to estimate the efficiency of baseball teams. Hoeffler and Payne (1997) use a stochastic production frontier model to provide efficiency measures for NBA teams. Carmichael et al (2000) adopt a range of specific play variables and characteristics to estimate a linear production function for the English Premier League. Hadley et al. (2000) use a Poisson regression model to estimate the performance of teams in the NFL.

Other literature looks at also estimating the productivity of team management. Pfeffer and Davis-Blake (1986) look at manager performance and how succession affects subsequent performance. Khan (1993) estimates managerial quality using salary regressions, finding that higher-quality managers lead to higher winning percentages. Dawson et al. (2000) find that coaching performance should be measured in terms of the available playing talent rather than purely on match outcomes.

Frick and Simmons (2008) use a stochastic frontier analysis to estimate coach quality, finding that a team hiring a better coach can reduce technical inefficiency and improve league standing. Gerrard (2005) uses data on the English Premier League to estimate a production function for coaches. Bridgewater et al. (2011) use frontier production functions to estimate managerial ability. Bell et a. (2013) use a fixed effects model with a bootstrapping approach to estimate the performance of English Premier League managers. Del Corral et al. (2015) estimate the efficiency of basketball coaches using a stochastic production function. Muehlheusser et al (2016) investigate the effects of managers on team performance in the German Bundesliga by estimating a manager ability distribution.

High dimensional regression techniques are used for the analysis in this paper. Sparse estimators like the Lasso (Tibshirani 1996) and some extensions (Zou 2006, Meinshausen 2007) are particularly popular because they perform well on high-dimensional data and produce interpretable results. While these methods perform well there is not a consensus on a statistically valid method of calculating standard errors for the lasso predictions. Osborne et al. (2000) derive an estimate for the covariance matrix of lasso estimators. Although these yield positive standard errors for coefficients estimates, the distribution of coefficient estimates will have a concentration at probability zero and may be far from normally distributed. Tibshirani (1996) suggested an alternative method for computing standard errors: the bootstrap. Knight and Fu (2000) argue that the bootstrap has problems estimating the sampling distribution of bridge estimators when parameter values are close to or exactly zero. Kyung et al. (2010) also claim that the bootstrap does not allow valid standard errors to be attached to values of the lasso which are shrunk to zero. In addition they propose a Bayesian Lasso which can be used to produce valid standard errors.

Lockhart et al. (2014) propose a significance test for the lasso based on the fitted values called the covariance test statistic.

Due to the ongoing debate and uncertainty about the validity of high dimensional standard errors the R packages used for the analysis in this paper do not implement standard errors and as such will not be reported in the model results.

3. THEORY

The purpose of this paper is to develop a model which estimates the contributions of individual soccer players. In order to estimate player coefficients we need to use a framework which is flexible for a large number of model parameters, since the dataset contains 33,297 individual players, 1,990 individual managers, and 711 individual teams. The method used will be fixed effects estimators similar to Abowd, Kramaz, and Margolis (1999) which allows for a flexible control of inputs. In professional soccer teams are often rotated within a season and players move to different teams regularly so this condition holds. Using such a large dataset will allow the model to identify how players contribute to team results individually by estimating how their impact on team performance within different lineups and across different leagues. Naturally the problem of collinearity can arise with such a large number of parameters which is why different approaches to estimating the fixed effects model will be included.

Before defining the models it is important to consider other research which estimates performance in sporting contests. Scully (1974) produced an econometric study in Major League Baseball looking at pay versus performance. Tullock (1980) developed a production function where the probability of success is a function of relative resources employed. Szymanski and Smith (1997) adopt a similar approach for English soccer. While the dataset in this paper does not contain financial information, it follows a similar approach to the performance literature in that it relates a variety of match inputs to a measure of performance, in this case goal difference.

The fixed effects model takes the following specification:

$$y_{ij} = \alpha_h + Players_i - Players_j + Manager_i - Manager_j + Team_i - Team_j + League_i - League_j + \varepsilon_{ij}$$

where

- y_{ij} is the goal difference of a match, relative to team *i*. This will be positive when team *i* wins, negative when they lose and equal to zero when the game is a draw.
- α_h corresponds to the advantage acquired by being the home team. This could be a function of referee bias, the bias of home fans towards their team, and may be a function of travelling.
- $Players_i$ are the starting 11 players of $Team$ *i* while $Players_j$ are the starting 11 players of team *j*. Since the results are relative to $Team$ *i* the coefficients for $Players_i$ will take a positive value in the model and likewise the coefficients for $Players_j$ will take a negative value. This is achieved by modelling using contrasts so that we produce only one distinct variable for each player, regardless of which team he plays on.
- $Manager_i$ is the manager of $Team$ *i* while $Manager_j$ is the manager of team *j*. Since the results are relative to $Team$ *i* these coefficients for $Manager_i$ will take a positive value in the model and likewise the coefficients for $Manager_j$ will take a negative value. This is achieved again by modelling using contrasts.
- $Team_i$ is the relative team in the model while $Team_j$ is the opposition team. Since the results are relative to $Team$ *i* these coefficients will take a positive value in the model and

likewise the coefficients for $Team_j$ will take a negative value. This is achieved again by modelling using contrasts.

- $League_i$ corresponds to a league strength coefficient in the model for $Team\ i$ while $League_j$ is the league strength coefficient for *Team j*. Since the results are relative to *Team i* these coefficients for $League_i$ will take a positive value in the model and likewise the coefficients for $League_j$ will take a negative value. This is achieved again by modelling using contrasts. League strength coefficients are only included for the 25 with the highest UEFA associations' club coefficients rankings for 2014/15.[1]
- ε_{ij} is an exponential noise term which accounts for chance factors specific to a soccer contest. This may include weather conditions, errors by the referees or other "luck" based events.

To give the model a relative interpretation, baseline variables are included for the team specific coefficients. Players who have not played at least 35 games over in the data period correspond to the baseline for players. Managers are also treated in a similar fashion. Teams which are not included in the 25 leagues are the baseline for teams. This corresponds to teams in European competition out with these leagues. Finally the baseline variable for leagues corresponds to all other leagues outside of the 25 in the data. This gives a reasonable interpretation for player contributions as being above a "replacement" level player.

Given the large number of fixed effect coefficients in the model sparse matrices will need to be used to improve the computational efficiency. These will be created by using the *Matrix*[2] package

[1] http://www.uefa.com/memberassociations/uefarankings/country/season=2015/index.html
[2] https://cran.r-project.org/web/packages/Matrix/index.html

in R. In order to run a regression using a sparse matrix in R the *glmnet*[3] package will also be used. This presents a series of computationally efficient regularization algorithms that can be used to produce the estimates. These algorithms are standard for research with big data. The three regularizations methods used are LASSO, Ridge Regression and Elastic-net.

The main difference between LASSO and Ridge regression is the specified penalty term. Ridge regression uses a sum of squares penalty to produce proportional shrinking while LASSO produces shrinkage towards zero using an absolute value penalty. This means that LASSO does a sparse selection while Ridge regression does not. For highly correlate variables Ridge regression shrinks the two coefficients towards one another while LASSO generally picks one over the other, setting the other to zero. This means that Ridge regression penalizes the larger coefficients more than the smaller ones whereas LASSO produces a more uniform penalty. Elastic-net is a mix of the two methods, adopting a compromise of the two penalty terms. Ridge regression will be preferable as it does not shrink any coefficients to zero, giving a clear player ranking output, however all three methods be used to produce estimates. A rank order correlation test will then be used to compare the methods. Even though there may be differences in the rank order between methods they should all closely correlate with each other. This will show that we can be happy with the Ridge regression results over the other methods since they are all similar regardless.

4. DATA

This research makes use of a database of player lineups from various European soccer competitions. The database contains 25 top tier leagues which have almost 10 years of lineups, running from 2006/07 to 2015/16. The included leagues represent the 25 with the highest UEFA

[3] https://cran.r-project.org/web/packages/glmnet/index.html

associations' club coefficients rankings for 2014/15. Also included are lineups for the group and knockout stages of The UEFA Champions League and UEFA Europa League (and previous incarnations) over this 10-year period. In total this contributes 133,536 unique lineups over 66,768 individual games. Team managers are included with every lineup along with the game result.

Table 20 presents a breakdown of the database. Listed in the table are the number of unique players, teams and managers appearing in each league. Also listed are the number of unique teams who appear in European competitions. The total number of games and lineups in the data are also listed but this is not fully complete as the source data from footballdatabase.eu is incomplete. The sample sizes are large enough across all leagues to make this small amount of missing data negligible.

The high dimensional analysis relies on the movement of players within teams and leagues to produce accurate estimates. *Table 21* presents information on player movements. It contains how many times a player has transferred, how many unique teams and competitions they appear in, as well as how long they have appeared in the data. We can see that although many players to move between teams on multiple occasions although over half of the players do not. This suggests that there may be some collinearity issues between specific groups of players who stay put on one team. The players who do move should be able to obtain accurate estimates of their ability.

Table 1 - Lineup Breakdown

Competition	Unique Players	Unique Managers	Unique Teams	Intercontinental Teams	Total Games	Total Lineups
Austria	933	65	17	4	1794	3588
Belarus	983	58	25	2	1654	3308
Belgium	1629	91	27	7	2779	5558
Croatia	1243	82	22	3	1850	3700
Cyprus	1398	92	22	5	1546	3092
Czech Rep	1330	64	27	5	2399	4798
Denmark	1008	48	18	6	1978	3956
England	1688	114	37	17	3800	7600
France	1725	103	38	14	3800	7600
Germany	1499	107	33	15	3060	6120
Greece	1944	139	32	8	2504	5008
Israel	1238	77	23	4	2318	4636
Italy	1744	112	36	14	3800	7600
Netherlands	1513	96	25	9	3060	6120
Norway	980	49	24	4	1680	3360

Poland	1584	122	28	3	2567	5134
Portugal	1840	101	30	12	2532	5064
Romania	2010	163	45	9	3021	6042
Russia	1441	112	28	9	2512	5024
Scotland	1267	63	18	3	2280	4560
Spain	1813	128	35	17	3800	7600
Sweden	1141	48	24	4	2103	4206
Switzerland	989	78	16	8	1782	3564
Turkey	1733	113	34	5	3026	6052
Ukraine	1282	64	25	7	2017	4034
European	6827	461	215	N/A	3106	6212

Table 2 - Player movement summary

Count	Players transferring to another team[4]	Players appeared on unique teams[5]	Players appeared in unique competition[6]	Years in the data[7]
1	14186	14186	17752	9015
2	5842	6220	4592	4753
3	3238	3360	2141	3284
4	1670	1596	1155	2522
5	870	683	508	2014
6	364	252	196	1663
7	153	82	49	1187
8	61	26	9	1048
9	20	1	5	921
10	2	1	0	0
11	1	0	0	0
Total	26407	26407	26407	26407

[4] Players moving from a current club to a new club.
[5] The number of unique clubs a player has played for.
[6] The number of leagues and European competitions a player has competed in.
[7] How many of the ten seasons a player has played a game in.

5. RESULTS

5.1 PREDICTIONS

This section focuses on the predictive accuracy of the models tested on the 2015/16 season. Once the validity of the models is tested then we can look in more detail at the coefficients from the estimations. Estimations are based on training data consisting of the 9 seasons from 2006/07 – 2014/15. For making predictions players will be given their coefficient value estimated from the training data. New players appearing only in the final year will be given the player baseline coefficient value. Figure 6 contains 6 graphs presenting a visual representation of predictive accuracy from the perspective of the home team. The top row of plots show predicted vs observed goal difference for the 2015/16 season using each of the three methods. There is not much difference between each method and they are all able to predict more goals when more are observed. There is a high degree of variability observed for each goal difference bin showing the difficulty in capturing the scale of victory for individual games. Since the predicted goal difference is on a continuous scale the likelihood of predicting zero for goal difference is effectively null. To predict draws more accurately a sensitivity parameter is created from the training data which best captures the distribution of results. This parameter is then used for the test data to convert the predicted goal difference into wins, losses and draws. From the perspective of home teams in the training data, 46.2% of games are won, 28.1% of games are lost and 25.7% are drawn. The sensitivity parameter mirrors this distribution for estimated goal difference and by applying it to the test data can create the bins for results. These results are shown in the boxplots in the bottom row. There is little difference between the regularization methods though we find that games predicted as wins have a higher observed mean goal difference. Those predicted a loss have a lower observed mean goal difference however this is close to the observed mean goal difference for games predicted as a draw. The next step is to check the accuracy of the game predictions

Tables 22, 23 and 24 contain contingency matrixes for the predicted results. This will indicate how well the models perform at predicting each specific result. We find that all models predict wins and losses with reasonable accuracy, predicting around 57% and 42% of these cases correctly. The models struggle at predicting draws, where around 26% of cases are predicted correctly. This is not uncommon for any soccer predictive model as draws are more uncommon than wins or losses in the data. We find that the Lasso and Elastic Net models have almost identical predictive accuracy, so the Elastic Net model prefers variable selection over the penalty term from the Ridge regression

Table 25 displays this overall predictive accuracy. This is broken down for individual leagues as well as the complete data set. The predictive accuracy ranges from between 34% to 60% depending on the league but around 45% overall. There appears to be no link between the overall quality of the league and the predictive rate (the correlation between them is 0.28 for Ridge, 0.11 for Lasso and 0.12 for Elastic Net) however it could reflect the balance between teams within the league. Further analysis can be found in the appendix. Overall this prediction rate is consistent with the literature for models which do not update during the season.

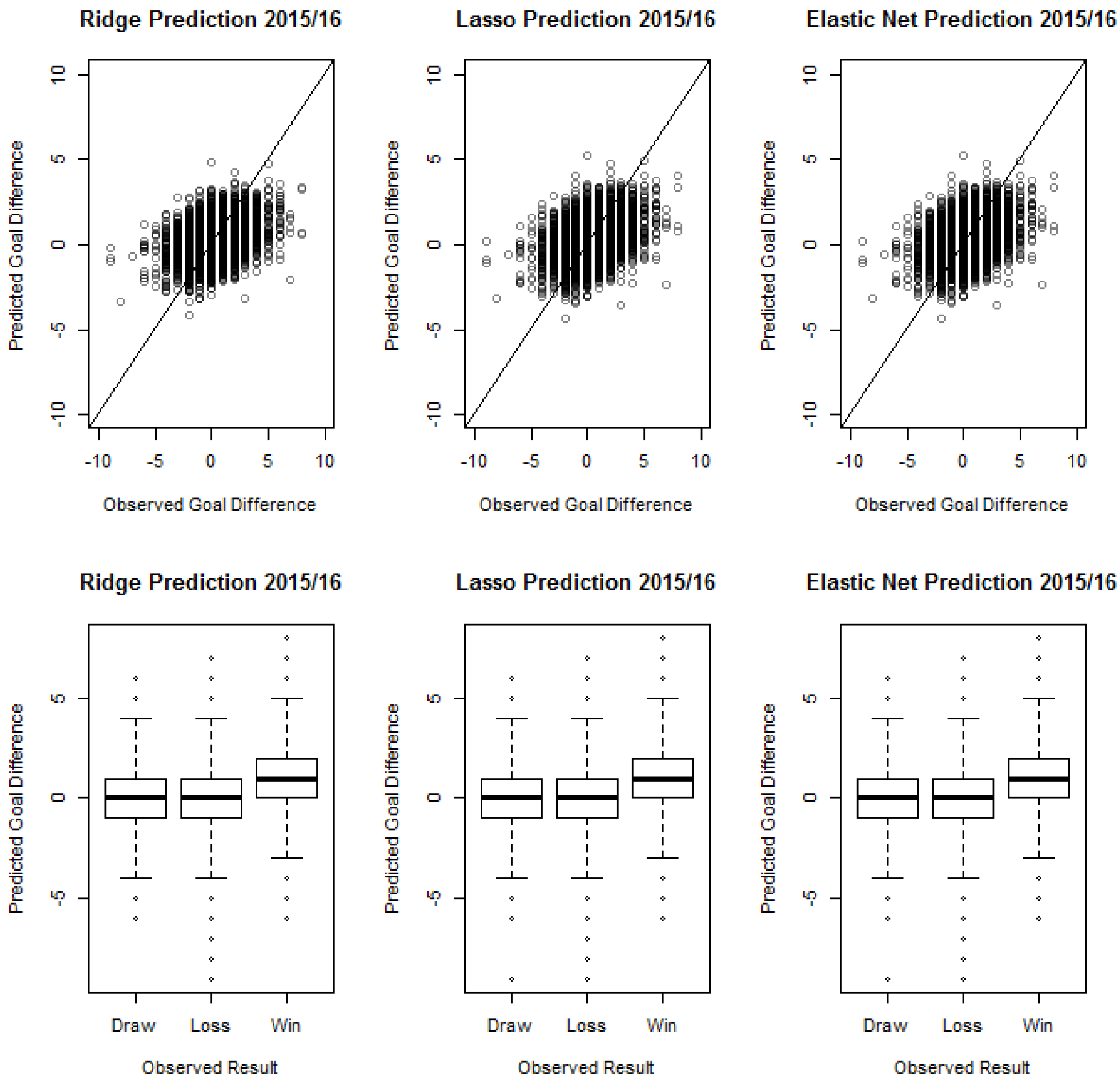

Figure 1 - Observed Goal Difference/Result vs. Predicted Goal Difference

Table 3 - Ridge Regression 2015/16 Contingency Matrix

Ridge	Observed Draw	Observed Loss	Observed Win	Row Total
Predicted Draw	475 26.8%	557 26.9%	752 24.6%	1784

Predicted Loss	540 30.4%	877 42.4%	566 18.5%	1983
Predicted Win	760 42.8%	635 30.7%	1744 57.0%	3139
Column Total	1775 25.7%	2069 30.0%	3062 44.3%	6906

Table 4 - Lasso 2015/16 Contingency Matrix

Lasso	Observed Draw	Observed Loss	Observed Win	Row Total
Predicted Draw	452 25.5%	554 26.8%	692 22.6%	1698
Predicted Loss	553 31.2%	864 41.8%	593 19.4%	2010
Predicted Win	770 43.4%	651 31.5%	1777 58.0%	3198
Column Total	1775 25.7%	2069 30.0%	3062 44.3%	6906

Table 5 - Elastic Net 2015/16 Contingency Matrix

Lasso	Observed Draw	Observed Loss	Observed Win	Row Total
Predicted Draw	451 25.4%	553 26.7%	696 22.7%	1700
Predicted Loss	552 31.1%	866 41.9%	592 19.3%	2010
Predicted Win	772 43.5%	650 31.4%	1774 57.9%	3196
Column Total	1775 25.7%	2069 30.0%	3062 44.3%	6906

Table 6 - Prediction accuracy for 2015/16

Competitions	Ridge	Lasso	Elastic Net
All (mean)	44.8%	44.8%	44.8%
Austria	42.2%	41.7%	41.7%
Belarus	40.7%	42.9%	43.4%
Belgium	45.2%	44.8%	44.1%
Croatia	44.4%	48.9%	48.9%
Cyprus	50.2%	53.9%	53.9%
Czech Republic	45.4%	47.5%	47.9%
Denmark	48.2%	49.7%	49.2%
England	42.1%	45.3%	45.3%
France	44.7%	41.8%	41.1%
Germany	43.8%	41.5%	41.5%
Greece	44.8%	50.2%	50.2%
Israel	44.6%	43.8%	43.8%
Italy	50.5%	47.9%	47.6%
Netherlands	44.1%	44.8%	44.8%
Norway	45.8%	43.3%	43.3%
Poland	35.8%	38.2%	37.8%
Portugal	46.4%	49.3%	49.3%
Romania	35.4%	34.3%	34.3%
Russia	41.7%	42.9%	43.3%
Scotland	37.7%	35.1%	36.0%
Spain	49.7%	47.9%	48.2%
Sweden	42.5%	37.9%	37.9%
Switzerland	46.1%	43.9%	43.9%
Turkey	47.2%	44.3%	43.9%
Ukraine	59.9%	56.4%	57.0%
Europe	46.7%	48.5%	48.5%

5.2 BETTING ANALYSIS

Pre-match betting odds are the best available predictor of match results. We take betting odds supplied by seven firms and calculate the percentage profit and loss achieved by betting £1 on games in 2015/2016 using our predicted results. Tables 26, 27, and 28 present these results. Betting odds are taken from archives online[8]. We find that our models on average perform almost as well as the betting firms with mostly small losses as an outcome. In some cases, the models perform better than the betting firms. The Ridge regression performs well in the Italian and Turkish league while the Lasso and Elastic Net models perform well in England, Greece and Portugal. Since overall predictive performance is almost as good as with the betting firms we can be confident that we can find meaningful conclusions from our estimations on player contributions.

Table 7 - Ridge Regression against betting odds

Ridge	Bet365	Bet&W-in	Interwetten	Ladbrokes	Pinnacle	William Hill	VC Bet
Belgium	-11.7%	-11.2%	-12.0%	-11.3%	-8.7%	-4.2%	-10.9%
England	-6.6%	-9.4%	-10.9%	-8.7%	-6.4%	0.2%	-6.7%
France	-1.0%	-1.4%	-2.6%	-2.5%	0.8%	-7.6%	-0.2%
Germany	-9.3%	-9.9%	-9.8%	-10.4%	-6.9%	-18.5%	-8.0%
Greece	-12.4%	-12.9%	-15.1%	-12.6%	-9.1%	11.6%	-3.5%
Italy	3.1%	2.9%	1.6%	2.0%	5.4%	-5.6%	3.4%
Netherla-nds	-5.4%	-6.3%	-8.1%	-5.0%	-1.6%	-7.4%	-2.9%
Portugal	-2.4%	-2.5%	-4.0%	-2.2%	1.5%	2.9%	-0.5%
Scotland	-15.9%	-19.0%	-18.9%	-15.7%	-14.3%	-22.1%	-12.5%
Spain	-2.4%	-3.3%	-3.9%	-3.1%	0.2%	-8.3%	-0.3%
Turkey	3.7%	3.7%	-0.7%	3.1%	8.7%	-7.8%	5.3%

[8] Betting odds taken from http://www.football-data.co.uk/.

Table 8 - Lasso against betting odds

Lasso	Bet365	Bet&W-in	Interwetten	Ladbrokes	Pinnacle	William Hill	VC Bet
Belgium	-3.5%	-3.4%	-4.6%	-3.4%	0.2%	-4.2%	-2.4%
England	2.0%	-1.3%	-3.2%	-0.5%	2.3%	0.2%	2.2%
France	-7.2%	-7.5%	-8.4%	-8.6%	-5.4%	-7.6%	-6.3%
Germany	-18.0%	-18.6%	-17.8%	-18.9%	-16.1%	-18.5%	-17.0%
Greece	3.4%	2.6%	-0.1%	2.9%	10.0%	11.6%	13.1%
Italy	-5.7%	-5.8%	-6.6%	-6.6%	-3.6%	-5.6%	-5.3%
Netherla-nds	-6.6%	-7.3%	-10.1%	-6.3%	-3.2%	-7.4%	-4.5%
Portugal	3.6%	3.5%	2.0%	3.8%	7.5%	2.9%	5.5%
Scotland	-24.5%	-27.1%	-26.9%	-24.3%	-21.9%	-22.1%	-21.6%
Spain	-8.1%	-8.9%	-9.4%	-8.8%	-5.8%	-8.3%	-6.3%
Turkey	-7.4%	-7.3%	-11.4%	-7.5%	-3.3%	-7.8%	-6.0%

Table 9 - Elastic Net against betting odds

Elastic Net	Bet365	Bet&W-in	Interwetten	Ladbrokes	Pinnacle	William Hill	VC Bet
Belgium	-8.3%	-8.1%	-8.9%	-8.0%	-4.7%	-8.7%	-7.2%
England	2.0%	-1.3%	-3.2%	-0.5%	2.3%	0.2%	2.2%
France	-9.5%	-9.8%	-10.8%	-10.9%	-7.7%	-9.9%	-8.6%
Germany	-18.0%	-18.6%	-17.8%	-18.9%	-16.1%	-18.5%	-17.0%
Greece	3.4%	2.6%	-0.1%	2.9%	10.0%	11.6%	13.1%
Italy	-6.6%	-6.6%	-7.5%	-7.4%	-4.5%	-6.4%	-6.2%
Netherla-nds	-6.6%	-7.3%	-10.1%	-6.3%	-3.2%	-7.4%	-4.5%
Portugal	3.6%	3.5%	2.0%	3.8%	7.5%	2.9%	5.5%
Scotland	-21.3%	-24.2%	-24.0%	-21.2%	-18.6%	-19.1%	-18.3%
Spain	-7.3%	-8.1%	-8.5%	-8.0%	-4.9%	-7.5%	-5.4%
Turkey	-8.3%	-8.2%	-12.2%	-8.4%	-4.2%	-8.6%	-6.9%

5.3 ESTIMATIONS

This section presents player ability estimations for the individual contributions model. Estimations for manager, team and leagues coefficients will be included in the appendix. For this analysis the first 9 years of data are used to produce the estimates. This means that the coefficients are reflective of player contributions going into the 2015/16 season. Two ways of interpreting the coefficients are presented as follows:

$$Coefficient_i = Player_{\hat{\imath}} \qquad (1)$$

$$Coefficient_i = Player_{\hat{\imath}} + Team_{\hat{\imath}} + League_{\hat{\imath}} \qquad (2)$$

The identification of player coefficients can be thought of in layers. As mentioned in the theory section the model controls for player teammates, team, league, manager and home advantage. For specification (1) the player coefficient alone accounts for the extra value unique to a player above and beyond these controls, examined in isolation. This picks up the way a player is outlying within all his typical playing conditions. Specification (2) adds in the coefficients for the player's most recent team and league to level out the playing conditions for players and give an unbiased ranking of player abilities.

Table 29 represents the 25 largest player coefficients from the 11,584 players who have played at least 35 games in this 9 year period using specification (2). This specification produces a weighting that reduces the impact of dominant players in weaker leagues and increase the impact of weaker players in stronger leagues. The rankings list contains many world famous players who have won the UEFA Champions League, many league titles and even international honours. Lionel Messi and Cristiano Rolando, who have won 9 Ballon d'Ors (an award given to the best soccer player in a calendar year) between them rank very highly. Almost every player on the list has played in the top five ranked soccer leagues (see appendix) and the UEFA Champions League. While these models pick out world class players, the order in rankings will not match the perception from soccer fans which results from the variable selection and penalty terms penalizing players with high collinearity. This can often be found in the top sides who amass the best lineups and don't often rotate them. Since the model measures goal difference a high premium is placed on players who both score and prevent goals. The rankings contain many strikers and defenders but not so many midfielders. For example Xavi and Andrés Iniesta have performed very well with Barcelona for the duration of the data set but since Lionel Messi scores most of the goals the models select him to have a higher coefficient when collinearity occurs.

Table 10 - Combined Coefficients Model Results (Players)

Model	Ridge		Lasso		Elastic Net	
Rank	Name	Coef	Name	Coef	Name	Coef
1	Gabriel Paulista	1.494	Gabriel Paulista	2.120	Gabriel Paulista	2.232
2	Iván de la Peña	1.205	Nabil Fekir	2.104	Nabil Fekir	2.085
3	Jon Flanagan	1.197	Frank Lampard	2.001	Frank Lampard	1.982
4	Nabil Fekir	1.182	Cristiano Ronaldo	1.959	Cristiano Ronaldo	1.950
5	Cristiano Ronaldo	1.177	N'Golo Kanté	1.882	N'Golo Kanté	1.866
6	Asier Illarramendi	1.173	Willy Sagnol	1.830	Willy Sagnol	1.818
7	Chechu Dorado	1.160	Chechu Dorado	1.818	Chechu Dorado	1.806
8	Carles Puyol	1.135	Fernandinho	1.789	Fernandinho	1.770
9	Dani Carvajal	1.103	David Albelda	1.774	David Albelda	1.762
10	Rubén de la Red	1.102	Arjen Robben	1.745	Per Mertesacker	1.752
11	Leroy George	1.095	Dani Carvajal	1.739	Arjen Robben	1.734
12	Keylor Navas	1.093	Fernando	1.734	Dani Carvajal	1.730
13	Lionel Messi	1.092	Asier Illarramendi	1.731	Asier Illarramendi	1.725
14	Nicola Pozzi	1.091	Wilfried Bony	1.718	Fernando	1.717
15	Frank Lampard	1.089	Mario Götze	1.711	Mario Götze	1.700
16	David Beckham	1.072	Franck Ribéry	1.708	Wilfried Bony	1.700
17	Wes Morgan	1.067	Iván de la Peña	1.698	Franck Ribéry	1.697
18	Per Mertesacker	1.051	Jô	1.693	Iván de la Peña	1.687

19	Toby Alderweireld	1.044	Sergio Agüero	1.686	Jô	1.675
20	Mario Cotelo	1.043	Martín Demichelis	1.684	Mehdi Benatia	1.671
21	Willy Sagnol	1.041	Sergi Darder	1.682	Sergi Darder	1.671
22	David Albelda	1.030	Mehdi Benatia	1.680	Sergio Agüero	1.670
23	Javier Saviola	1.029	Toby Alderweireld	1.671	Martín Demichelis	1.666
24	Vicente Iborra	1.018	Paco Peña	1.664	Toby Alderweireld	1.658
25	Sergi Darder	1.018	Gaël Clichy	1.662	Keylor Navas	1.650

Table 11 - Player Coefficients Model Results

Model	Ridge		Lasso		Elastic Net	
Rank	Name	Coef	Name	Coef	Name	Coef
1	Shota Arveladze	0.953	Shota Arveladze	1.268	Shota Arveladze	1.266
2	Gabriel Paulista	0.939	Gabriel Paulista	1.247	Gabriel Paulista	1.243
3	Kenneth Omeruo	0.881	Thomas Grogaard	1.184	Thomas Grogaard	1.169
4	Geert-Arend Roorda	0.834	Kenneth Omeruo	1.056	Kenneth Omeruo	1.050
5	Alexander	0.827	Jan Dolezal	1.028	Jan Dolezal	1.026
6	Toni Doblas	0.805	Martin Milec	1.018	Martin Milec	1.01
7	Niko Kovac	0.789	N'Golo Kanté	1.011	N'Golo Kanté	1.007
8	Mateusz Piatkowski	0.787	Toni Doblas	1.000	Toni Doblas	0.997
9	Wim Raymaekers	0.778	Ralf Pedersen	0.999	Ralf Pedersen	0.994
10	Emiliano Dudar	0.773	Kostadin Bashov	0.982	Kostadin Bashov	0.975
11	Joan Tomás	0.768	Niko Kovac	0.976	Niko Kovac	0.974
12	Hezi Dilmoni	0.765	Alexander	0.974	Alexander	0.971
13	Jan Dolezal	0.759	Hezi Dilmoni	0.965	Hezi Dilmoni	0.963
14	Tobias Linderoth	0.740	Emiliano Dudar	0.959	Emiliano Dudar	0.957
15	Mathias Abel	0.738	Wim Raymaekers	0.952	Wim Raymaekers	0.950
16	Ralf Pedersen	0.712	Slobodan Markovic	0.950	Slobodan Markovic	0.947
17	Johan Lind	0.709	Paul McGinn	0.948	Paul McGinn	0.943

18	Terence Kongolo	0.707	Terence Kongolo	0.944	Terence Kongolo	0.942
19	Antonio Rojas	0.701	Antonio Rojas	0.938	Antonio Rojas	0.936
20	Hamza Younes	0.700	Geert-Arend Roorda	0.921	Geert-Arend Roorda	0.920
21	Razak Omotoyossi	0.695	Mads Rieper	0.915	Mads Rieper	0.910
22	Evgeniy Pankov	0.693	Danijel Madjaric	0.900	Danijel Madjaric	0.896
23	Kostadin Bashov	0.690	Tobias Linderoth	0.893	Tobias Linderoth	0.892
24	Danijel Madjaric	0.685	Mateusz Piatkowski	0.887	Mateusz Piatkowski	0.885
25	Nabil Fekir	0.683	Slavko Bralic	0.887	Nabil Fekir	0.884

Table 30 represents the coefficient results from specification (1). Results are presented for each of the 3 regularization methods. Since this specification does not consider team and league ability we should expect to see players who are particularly dominant within their normal playing conditions. The highest ranked player for all methods is Shota Arverladze. He appeared in the data predominantly for Dutch size AZ Alkmaar, winning most games when he was a starting player and finishing high up the table. Most of these games were in the Dutch Eredivisie and the UEFA Cup so will not include the highest quality of opposition. This becomes clear as you look further down the table as many of the players listed performed very well in weaker leagues. That considered, many of the players listed do move onto better teams. For example, N'Golo Kante plays for French side Caen in the training data. He would later win the English Premier League with Leicester City before being transferred to Chelsea.

While many players appear across all 3 regularization methods there are some differences. It is worth noting how large these differences are and whether different methods will result in a notably different set of rankings. Table 31 contains estimates of the Spearman's rank-order correlation test. This determines the strength and direction of association between two ranked variables. Testing between all regularization methods produces correlation values above 0.96 suggesting that player rankings are very close between each of the methods. For that reasons if a complete ranking of players was desired then Ridge Regression would be used since it does not perform variable shrinkage to zero.

Table 12 - Spearman's Rank-Order Correlation

	Ridge	Lasso	Elastic Net
Ridge	1	0.960	0.961
Lasso	0.960	1	0.999
Elastic Net	0.961	0.999	1

6. CONCLUSION

The study of worker productivity is important to businesses in any industry since the best workers will improve the overall performance of business and increase profit. Firms would like to be able to screen potential employees efficiently to determine their potential value. This paper chooses an industry in which worker productivity is observed. The setting is European soccer where twenty-five top flight leagues are considered so that players can be tracked as they move between different teams. High dimensional fixed effect models are used to determine the productivity of individual players.

The models yield on average a 45% prediction rate with the different methods producing very similar player rankings. Some leagues are more easily to predict than others with prediction rates ranging between 35% and 59%. Wins and losses are predicted well though the models struggle predicting games which end in draws. Compared with betting firms the models predict almost as well and in a few leagues, outperform them. The highest ranked players in the models have often won the most prestigious soccer tournaments and play for the best teams. Another specification of player value can determine outlying players within their normally playing conditions which may be of use for player scouting. While the model highlights the most productive players there is a bias towards players who produce and prevent goals directly. This results in more attackers and defenders ranking highly than midfield players. Most of the contribution goes towards players who score goals rather than players who help produce them.

These results have many benefits to teams, fans and business in general. Teams can track players at all levels who can benefit their teams. With such a large dataset, this can help make the scouting process more efficient. Fans will not only be able to gain insight into which players contribute the most towards teams but the prediction accuracy could be of benefit in the betting market. Businesses can use similar approaches to help screen potential new hires as a fixed effects model requires limited information from other firms. Improvements to the model can be made by accounting for team form or by updating the model every week before matches. This would could allow for rolling coefficient values rather than annual updates which may improve overall prediction accuracy.

7. APPENDIX C

In Section 5 a small analysis was presented concerning the correlation between predictive rate and league strength. Figures 7, 8 and 9 present an additional visual element of this relationship. All plots measure the league coefficient (as determined by Ridge regression, Lasso and Elastic Net) against the predictive rate for these models. Historically strong leagues such as the top 5 in Europe sit over on the right of the plots and the weaker leagues towards the left. There appears to be no relationship between league strength and predictive accuracy as most leagues have between a 40%-50% prediction rate regardless of the model.

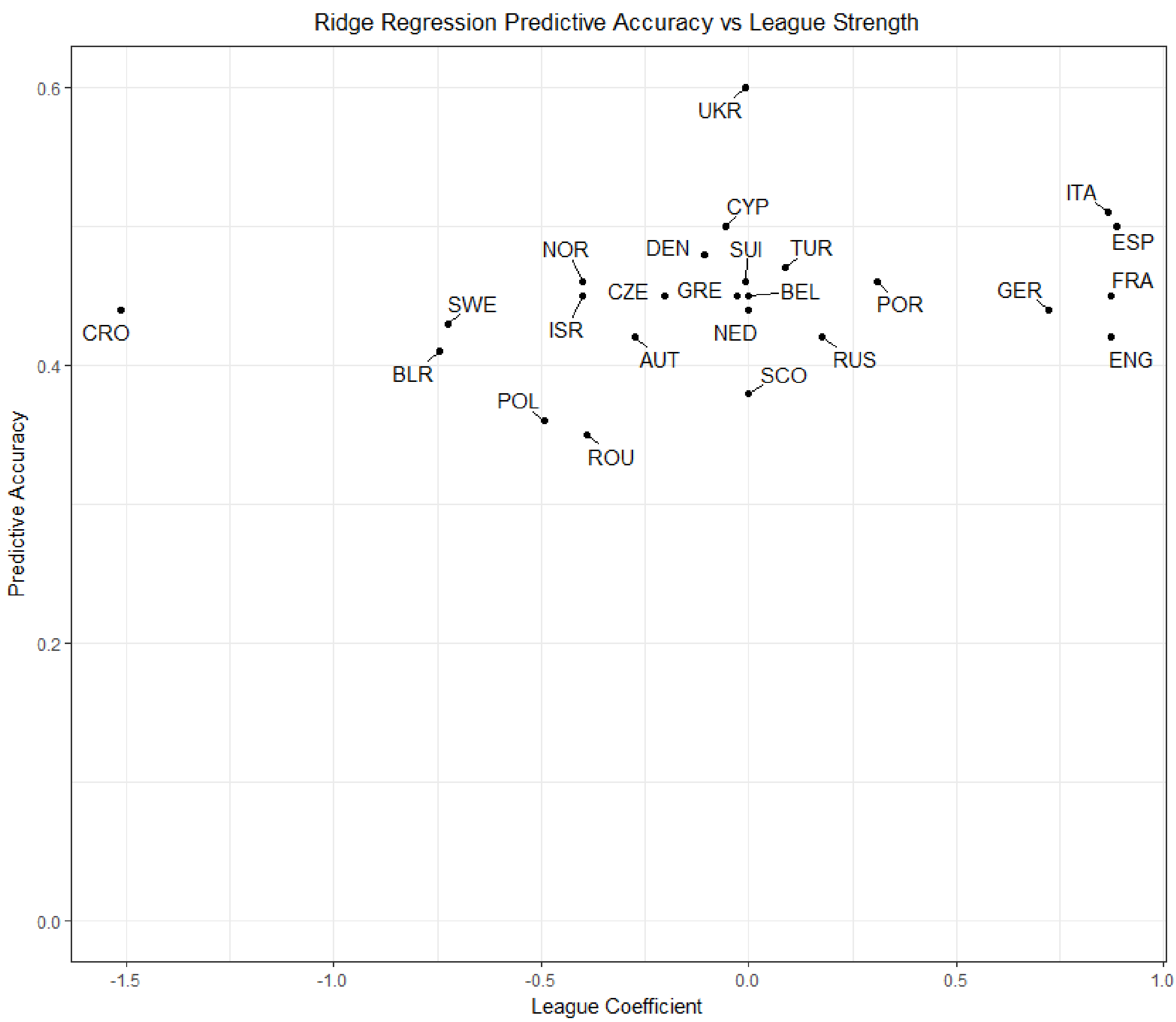

Figure 2 - Ridge Regression

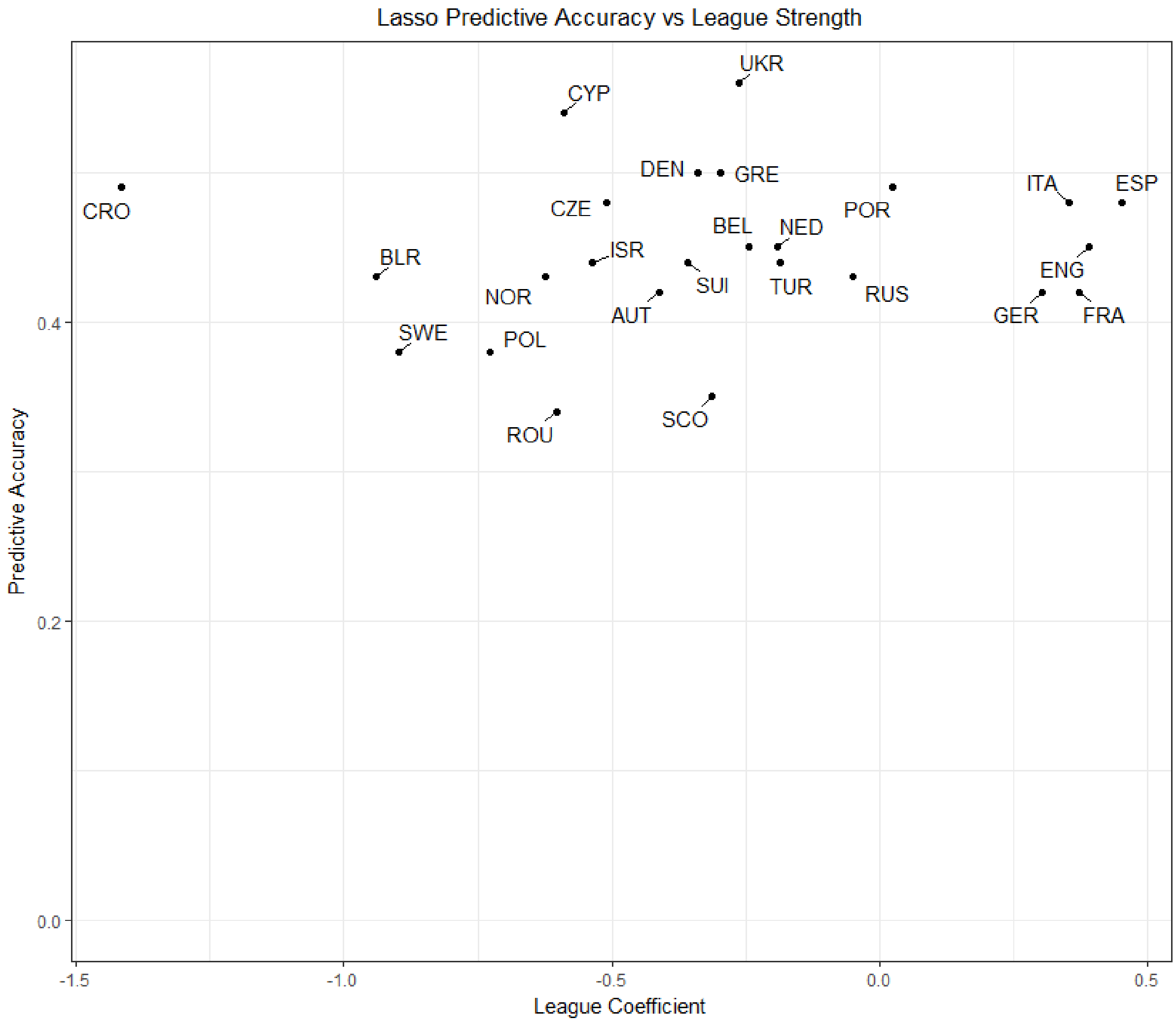

Figure 3 - Lasso

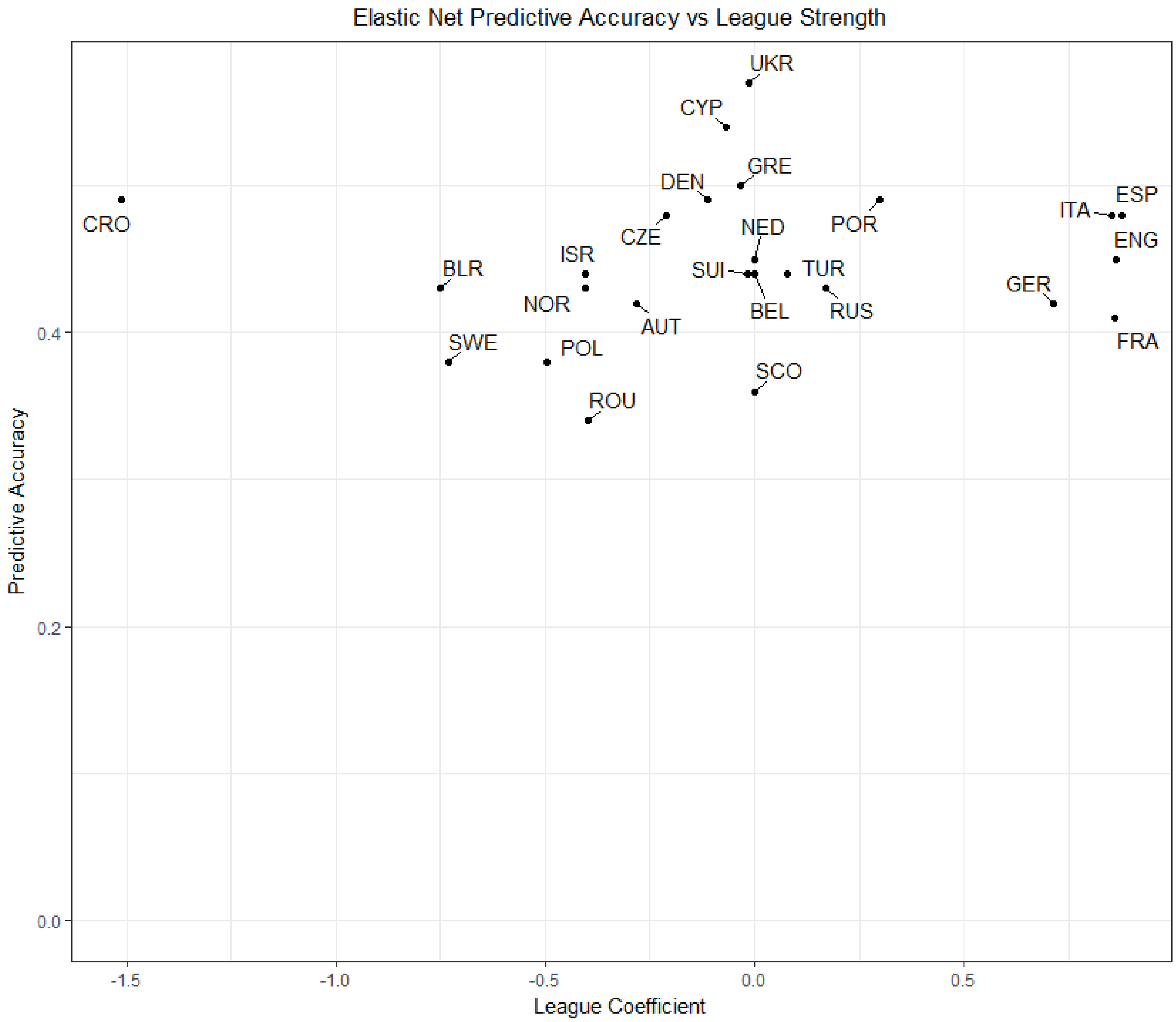

Figure 4 - Elastic Net

While player contributions are the main focus of the paper, manager, team and league coefficients were also produced from the models. Table 32 displays the results for manager coefficients accounting for team and league strength. The list of managers once again contains famous names who have won domestic, intercontinental and international honours. Many of the managers are still active and at top teams to this day.

Table 33 looks at managers who are outlying in their normal managerial conditions. While some famous names exist in the rankings many of the managers there are some who are not so familiar. For example Giorgio Contini who helped Swiss side FC Vaduz to survive in the top flight for the first time with a team record total of points. Again many of the managers do not often face the highest quality of opposition.

Table 34 displays the results for team coefficients accounting for league strength. In the rankings we see many top European sides but also interspersed with some weaker teams. These teams are often teams who have been recently promoted into a top league and performed better than expectations. Many of their players will not have individual coefficient values and so there is a bias towards increasing the club coefficient when they perform better than the baseline coefficient would suggest (which is quite often losing every game).

Table 35 contains just the club coefficients and so should highlight teams who are outlying among teams they normally play. The rankings are mostly filled with teams who historically perform very well within their own league but don't always perform well in European competition. There are some teams who perform well both in domestic and European competition to such a degree that they also appear on this list such as Real Madrid, Bayern Munich and Manchester City.

Table 13 - Combined Coefficients Model Results (Managers)

Model	Ridge		Lasso		Elastic Net	
Rank	Name	Coef	Name	Coef	Name	Coef
1	Luis Enrique	1.522	A. Jonker	1.908	A. Jonker	1.905
2	A. Jonker	1.517	J. Heynckes	1.737	J. Heynckes	1.734
3	R. Schmidt	1.492	R. Schmidt	1.722	R. Schmidt	1.722
4	C. Contra	1.388	J. Guardiola	1.629	J. Guardiola	1.625
5	L. Banide	1.372	Luis Enrique	1.511	C. Ancelotti	1.513
6	J. Lillo	1.368	C. Ancelotti	1.509	Luis Enrique	1.510
7	J. Heynckes	1.366	M. Pellegrini	1.500	M. Pellegrini	1.492
8	C. Ancelotti	1.363	W. Sagnol	1.456	W. Sagnol	1.456
9	W. Sagnol	1.292	C. Contra	1.440	C. Contra	1.437
10	G. Garitano	1.284	L. Banide	1.430	L. Banide	1.428
11	B. Rodgers	1.278	J. Lillo	1.410	J. Lillo	1.407
12	M. Allegri	1.260	J. Lopetegui	1.405	J. Lopetegui	1.401
13	M. Pellegrini	1.254	E. Gerets	1.390	E. Gerets	1.385
14	J. Guardiola	1.253	S. Eriksson	1.388	S. Eriksson	1.380
15	E. Gerets	1.250	M. Allegri	1.323	M. Allegri	1.327
16	M. Sarri	1.248	G. Garitano	1.298	G. Garitano	1.297
17	G. Camolese	1.247	F. Capello	1.258	F. Capello	1.262
18	F. Capello	1.239	J. Muñiz	1.248	J. Muñiz	1.247
19	J. Tigana	1.234	H. Fournier	1.244	O. Hitzfeld	1.245
20	L. Jardim	1.227	L. Jardim	1.238	L. Jardim	1.238
21	A. Wenger	1.205	O. Hitzfeld	1.237	H. Fournier	1.234
22	O. Hitzfeld	1.193	J. Klinsmann	1.227	J. Klinsmann	1.226
23	H. Fournier	1.191	R. Garde	1.220	J. Tigana	1.219
24	P. Chaparro	1.180	M. Gisdol	1.216	M. Gisdol	1.213
25	S. Ferguson	1.179	J. Tigana	1.214	R. Garde	1.213

Table 14 - Manager Coefficients Model Results

Model	Ridge		Lasso		Elastic Net	
Rank	Name	Coef	Name	Coef	Name	Coef
1	Z. Mamic	0.766	G. Contini	1.510	G. Contini	1.492
2	D. Tholot	0.695	D. Canadi	1.039	D. Canadi	1.033
3	A. Bigon	0.626	L. Smerud	1.017	D. Tholot	1.012
4	G. Contini	0.606	D. Tholot	1.017	L. Smerud	0.997
5	R. Schmidt	0.594	M. Kek	0.990	M. Kek	0.976
6	Augusto Inácio	0.588	R. Schmidt	0.956	R. Schmidt	0.946
7	W. Fornalik	0.583	A. Bigon	0.936	A. Bigon	0.933
8	C. Adriaanse	0.557	W. Fornalik	0.929	W. Fornalik	0.923
9	A. Jonker	0.556	Augusto Inácio	0.792	Augusto Inácio	0.787
10	A. Benado	0.551	A. Axén	0.762	A. Axén	0.758
11	M. Jansen	0.548	A. Hütter	0.752	A. Hütter	0.747
12	I. Petev	0.534	A. Benado	0.701	A. Benado	0.698
13	D. Canadi	0.508	A. Jonker	0.681	A. Jonker	0.679
14	A. Hütter	0.498	O. Christensen	0.661	O. Christensen	0.658
15	A. Axén	0.498	Z. Mamic	0.632	Z. Mamic	0.647
16	J. Lillo	0.498	Luis Enrique	0.624	Luis Enrique	0.624
17	C. Contra	0.477	J. Kocian	0.613	J. Kocian	0.608
18	O. Christensen	0.477	J. Boskamp	0.604	J. Boskamp	0.601
19	L. Banide	0.476	E. Rasmussen	0.601	E. Rasmussen	0.599
20	I. Stimac	0.474	H. Hamzaoglu	0.594	H. Hamzaoglu	0.593
21	Luis Enrique	0.474	E. Levy	0.575	E. Levy	0.572
22	J. Lopetegui	0.474	I. Petev	0.568	I. Petev	0.569
23	N. Clausen	0.473	L. Banide	0.559	L. Banide	0.557
24	Y. Sergen	0.468	V. Lavicka	0.555	Z. Barisic	0.553
25	L. Smerud	0.466	Z. Barisic	0.554	C. Contra	0.551

Table 15 - Combined Coefficients Model Results (Teams)

Model	Ridge		Lasso		Elastic Net	
Rank	Name	Coef	Name	Coef	Name	Coef
1	Leicester	1.165	Manchester City	1.388	Manchester City	1.380
2	ESTAC Troyes	1.145	Real Madrid	1.258	Real Madrid	1.262
3	Brescia	1.135	Bayern Munich	1.227	Bayern Munich	1.226
4	Real Madrid	1.131	Lyon	1.220	Lyon	1.213
5	Manchester City	1.108	Qarabag Agdam	1.194	Qarabag Agdam	1.187
6	Novara	1.104	Fiorentina	1.161	Fiorentina	1.162
7	Xerez	1.103	Xerez	1.146	Werder Bremen	1.145
8	Hellas Verona	1.085	Werder Bremen	1.142	Xerez	1.140
9	Eibar	1.067	Lorient	1.136	Lorient	1.137
10	Mallorca	1.058	Juventus	1.092	Juventus	1.096
11	FC Barcelona	1.048	Deportivo La Coruña	1.071	Deportivo La Coruña	1.071
12	Juventus	1.043	FC Porto	1.056	FC Porto	1.049
13	Arsenal	1.035	Hellas Verona	1.031	Hellas Verona	1.031
14	Deportivo La Coruña	1.030	ESTAC Troyes	1.027	ESTAC Troyes	1.028
15	Liverpool	1.014	Parma	1.018	Parma	1.023
16	Fiorentina	1.003	Mallorca	1.009	Mallorca	1.009
17	Manchester United	0.999	Novara	0.995	Arsenal	0.999

18	Lorient	0.998	Brescia	0.994	Brescia	0.995
19	Lyon	0.998	Ludogorets Razgrad	0.990	Novara	0.995
20	Parma	0.997	Villarreal	0.976	Ludogorets Razgrad	0.983
21	Villarreal	0.996	AS Roma	0.941	Villarreal	0.976
22	Elche	0.992	Bordeaux	0.922	AS Roma	0.940
23	Celta Vigo	0.983	Leicester	0.918	Bordeaux	0.928
24	Sevilla FC	0.980	Torino	0.912	Leicester	0.919
25	Sochaux	0.977	Eibar	0.907	Torino	0.913

Table 16 - Team Coefficients Model Results

Model	Ridge		Lasso		Elastic Net	
Rank	Name	Coef	Name	Coef	Name	Coef
1	Qarabag Agdam	0.905	Dinamo Zagreb	1.588	Dinamo Zagreb	1.549
2	Ludogorets Razgrad	0.717	Qarabag Agdam	1.194	Qarabag Agdam	1.187
3	ASA Târgu-Mures	0.517	Grödig	1.003	Grödig	0.994
4	Dinamo Zagreb	0.440	Ludogorets Razgrad	0.990	Ludogorets Razgrad	0.983
5	Istanbul Basaksehir	0.440	Rosenborg	0.946	Rosenborg	0.928
6	Grödig	0.429	PSV Eindhoven	0.897	PSV Eindhoven	0.881
7	MTZ-RIPO Minsk	0.381	Olympiacos	0.849	Olympiacos	0.842
8	Pula	0.345	Panathinaikos	0.766	Panathinaikos	0.754
9	Petrolul Ploiesti	0.331	FC Porto	0.747	FC Porto	0.740
10	Rosenborg	0.313	ASA Târgu-Mures	0.694	ASA Târgu-Mures	0.689
11	Olympiacos	0.295	MTZ-RIPO Minsk	0.645	MTZ-RIPO Minsk	0.642
12	Leicester	0.293	Krasnodar	0.643	Krasnodar	0.635
13	Unirea Alba-Iulia	0.286	Istanbul Basaksehir	0.627	Istanbul Basaksehir	0.629
14	Krasnodar	0.275	Petrolul Ploiesti	0.577	Petrolul Ploiesti	0.574

15	ESTAC Troyes	0.274	La Gantoise	0.575	La Gantoise	0.567
16	Brescia	0.271	Glasgow Rangers	0.559	Glasgow Rangers	0.554
17	Gornik Leczna	0.267	FC Copenhagen	0.550	FC Copenhagen	0.545
18	Panathinaikos	0.264	Manchester City	0.517	Manchester City	0.509
19	Volendam	0.251	Bayern Munich	0.503	Bayern Munich	0.502
20	Real Madrid	0.244	Lokomotiv Moscow	0.480	Lokomotiv Moscow	0.477
21	Lokomotiv Minsk	0.242	APOEL	0.463	APOEL	0.457
22	Olympiakos Volos	0.242	CFR Cluj-Napoca	0.457	CFR Cluj-Napoca	0.451
23	Anorthosis Famagusta	0.242	Sheriff Tiraspol	0.429	Werder Bremen	0.421
24	FC Porto	0.241	Werder Bremen	0.419	Sheriff Tiraspol	0.421
25	Novara	0.240	Red Bull Salzburg	0.401	Red Bull Salzburg	0.402

Section 5 also considered overall predictive accuracy using 9 years of training data and 1 year of testing data. The average predictive accuracy was 44.8% for all models predicting the 2015/2016 season. To further explore predictive accuracy the split between training and testing data is altered by one year. Table 36 shows the predictive accuracy from 8 years of training data and 2 years of testing data. While there are some individual fluctuations within leagues the overall predictive accuracy decreased to around 43.5%. While the accuracy is lower as expected not a large amount of predictive power is lost.

Table 37 contains the results for 7 years of training data and 3 years of testing data. This is almost identical to Table 14 although on average produces a slightly higher prediction rate. Since using smaller training data results in less information about players the results suggest that teams who are expected to perform well stay relatively constant throughout the period. Even with reduced player coefficients there is not much change overall in league results.

Table 17 - Prediction accuracy for 2014/15 - 2015/16

Competitions	Ridge	Lasso	Elastic Net
All (mean)	43.4%	43.5%	43.7%
Austria	41.4%	41.4%	42.2%
Belarus	40.4%	39.5%	40.4%
Belgium	41.7%	41.1%	41.7%
Croatia	48.2%	50.1%	50.1%
Cyprus	45.7%	46.4%	46.2%
Czech Republic	45.4%	43.8%	44.2%
Denmark	41.0%	40.8%	41.5%
England	41.6%	42.5%	42.8%
France	43.8%	42.2%	42.9%
Germany	45.4%	44.6%	44.0%
Greece	42.6%	45.7%	45.7%
Israel	44.4%	43.5%	43.5%
Italy	47.4%	47.4%	47.1%
Netherlands	40.2%	42.2%	42.5%
Norway	39.0%	39.0%	39.4%

Poland	40.7%	39.0%	39.5%
Portugal	44.4%	44.0%	44.3%
Romania	37.6%	40.1%	40.4%
Russia	43.5%	44.4%	44.6%
Scotland	38.8%	38.4%	39.3%
Spain	48.2%	48.2%	48.9%
Sweden	40.2%	40.0%	40.6%
Switzerland	42.8%	41.9%	42.8%
Turkey	41.6%	40.8%	41.4%
Ukraine	52.7%	53.5%	51.3%
Europe	48.0%	48.3%	47.9%

Table 18 - Prediction accuracy for 2013/14 - 2015/16

Competitions	Ridge	Lasso	Elastic Net
All (mean)	43.3%	43.7%	43.7%
Austria	40.2%	41.5%	41.3%
Belarus	43.5%	43.5%	43.5%
Belgium	40.2%	40.7%	40.6%
Croatia	46.0%	50.8%	50.8%
Cyprus	47.4%	48.1%	48.0%
Czech Republic	43.5%	44.2%	44.3%
Denmark	37.3%	38.4%	38.6%
England	43.8%	43.7%	43.9%
France	42.8%	43.2%	43.2%
Germany	43.2%	44.6%	44.6%
Greece	40.0%	43.3%	43.3%
Israel	39.9%	37.9%	38.2%

Italy	46.5%	47.2%	47.1%
Netherlands	45.0%	47.2%	47.3%
Norway	43.6%	42.9%	43.1%
Poland	41.4%	41.4%	41.2%
Portugal	43.5%	44.4%	44.2%
Romania	39.8%	40.5%	40.3%
Russia	42.8%	43.8%	43.9%
Scotland	38.6%	38.6%	38.5%
Spain	47.4%	46.1%	46.1%
Sweden	42.3%	41.2%	41.0%
Switzerland	41.1%	39.4%	39.8%
Turkey	40.5%	40.6%	40.3%
Ukraine	52.3%	51.8%	51.8%
Europe	49.0%	49.4%	49.2%

8. REFERENCES

[1] Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica, 67*(2), 251-333.

[2] Bell, A., Brooks, C., & Markham, T. (2013). The performance of football club managers: Skill or luck? *Economics & Finance Research, 1*(1), 19-30.

[3] Bridgewater, S., Kahn, L. M., & Goodall, A. H. (2011). Substitution and complementarity between managers and subordinates: Evidence from British football. *Labour Economics, 18*(3), 275-286.

[4] Buchanan, J. M., Tollison, R. D., & Tullock, G. (1980). *Toward a theory of the rent-seeking society* Texas A & M Univ Pr.

[5] Carmichael, F., & Thomas, D. (1995). Production and efficiency in team sports: An investigation of rugby league football. *Applied Economics, 27*(9), 859-869.

[6] Carmichael, F., Thomas, D., & Ward, R. (2000). Team performance: The case of english premiership football. *Managerial and Decision Economics,* 31-45.

[7] Dawson, P., Dobson, S., & Gerrard, B. (2000). Estimating coaching efficiency in professional team sports: Evidence from English association football. *Scottish Journal of Political Economy, 47*(4), 399-421.

[8] del Corral, J., Maroto, A., & Gallardo, A. (2015). Are former professional athletes and native better coaches? evidence from Spanish basketball. *Journal of Sports Economics,* 1527002515595266.

[9] Frick, B., & Simmons, R. (2008). The impact of managerial quality on organizational performance: Evidence from German soccer. *Managerial and Decision Economics, 29*(7), 593-600.

[10] Gerrard, B. (2005). A resource-utilization model of organizational efficiency in professional sports teams. *Journal of Sport Management, 19*(2), 143-169.

[11] Hadley, L., Poitras, M., Ruggiero, J., & Knowles, S. (2000). Performance evaluation of national football league teams. *Managerial and Decision Economics,* 63-70.

[12] Hofler, R. A., & Payne, J. E. (1997). Measuring efficiency in the national basketball association. *Economics Letters, 55*(2), 293-299.

[13] Kahn, L. M. (1993). Managerial quality, team success, and individual player performance in major league baseball. *Industrial & Labor Relations Review*, *46*(3), 531-547.

[14] Knight, K., & Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics,* 1356-1378.

[15] Kyung, M., Gill, J., Ghosh, M., & Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis, 5*(2), 369-411.

[16] Lockhart, R., Taylor, J., Tibshirani, R. J., & Tibshirani, R. (2014). A significance test for the lasso. *Annals of Statistics*, *42*(2), 413-468.

[17] Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, *52*(1), 374-393.

[18] Muehlheusser, G., Schneemann, S., Sliwka, D., & Wallmeier, N. (2016). The contribution of managers to organizational success: Evidence from German soccer. *Journal of Sports Economics*, 1527002516674760.

[19] Osborne, M. R., Presnell, B., & Turlach, B. A. (2000). On the lasso and its dual. *Journal of Computational and Graphical Statistics*, *9*(2), 319-337.

[20] Pfeffer, J., & Davis-Blake, A. (1986). Administrative succession and organizational performance: How administrator experience mediates the succession effect. *Academy of Management Journal, 29*(1), 72-83.

[21] Ruggiero, J., Hadley, L., & Gustafson, E. (1996). Technical efficiency in major league baseball. *Baseball Economics: Current Research,* 191-200.

[22] Schofield, J. A. (1988). Production functions in the sports industry: An empirical analysis of professional cricket. *Applied Economics, 20*(2), 177-193.

[23] Scott, 1., 2Frank A, Long, 1., 2James E, & Somppi, K. (1985). Salary vs. marginal revenue product under monopsony and competition: The case of professional basketball. *Atlantic Economic Journal, 13*(3), 50-59.

[24] Scully, G. W. (1974). Pay and performance in major league baseball. *The American Economic Review, 64*(6), 915-930.

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

[26] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society.Series B (Methodological),* , 267-288.

[27] Zak, T. A., Huang, C. J., & Siegfried, J. J. (1979). Production efficiency: The case of professional basketball. *The Journal of Business*, *52*(3), 379-392.

[28] Zech, C. E. (1981). An empirical estimation of a production function: The case of major league baseball. *The American Economist, 25*(2), 19-23.

[29] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418-1429