Player Productivity and Performance: An Econometric Approach to Team Management in Soccer

by

Guy Benedict Wilkinson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Sport Management)
in the University of Michigan
2017

Doctoral Committee:

Professor Stefan Szymanski, Chair
Professor Rodney D. Fort
Professor Jeffrey A. Smith
Assistant Professor Eric M. Schwartz

Guy B. Wilkinson
gbwilkin@umich.edu
ORCID iD: 0000-0002-3623-1914

DEDICATION

To my father and Uncle Jim for taking me to my first Celtic game, and to Henrik Larsson for inspiring me to go to many more.

ACKNOWLEDGEMENTS

I would like to thank Professor Stefan Szymanski for taking me on as his doctoral student for the past four years. Without his guidance, I would not have been able to complete my research projects with the high quality I strive for.

My thanks also go to my committee members, Professor Rodney Fort, Professor Jeffrey Smith and Assistant Professor Eric Schwartz for their guidance and suggestions throughout the duration of my degree.

Finally, I want to thank my family, friends and cohort for helping me adapt to life in America and supporting me through both the good times and the bad.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES vii

LIST OF FIGURES x

ABSTRACT xi

CHAPTER I – THESIS MOTIVATION 1

1. INTRODUCTION 1
2. REFERENCES 7

CHAPTER II – TESTING THE O-RING THEORY USING DATA FROM THE ENGLISH PREMIER LEAGUE 12

1. INTRODUCTION 13
2. LITERATURE REVIEW 13
   - 2.1 O-RING THEORY 13
   - 2.2 SALARY DISPERSION AND MOTIVATION 15
   - 2.3 WAGE PERFORMANCE RELATIONSHIP 17
3. THEORY 19
   - 3.1 THE AGGREGATE MODEL 19
   - 3.2 THE VALUE MATCHING MODEL 20
4. DATA 23

5. ESTIMATION 27

5.1 EXOGENEITY IN WAGES 27
5.2 DISPERSION ANALYSIS 28
5.3 THE EFFICIENCY OF WAGES AND TPI IN PREDICTING RESULTS 29
5.4 OPTIMALITY 30

6. CONCLUSION 33

7. APPENDIX A 35

7.1 LAGRANGIAN DERIVATION 35
7.2 INDIVIDUAL GAME OPTIMALITY 37
7.3 MODEL ROBUSTNESS 38

8. REFERENCES 41

CHAPTER III – THE IMPACT OF REST PERIODS ON RESULTS IN THE ENGLISH PREMIER LEAGUE 45

1. INTRODUCTION 46

2. LITERATURE REVIEW 47

3. THEORY 49

3.1 PARAMETRIC LINEAR MODELS 51
3.1.1 RELATIVE REST MODEL 52
3.1.2 MIDWEEK EFFECT MODEL 52
3.1.3 WEEKEND EFFECT MODEL 53

4. DATA 53

5. RESULTS 57

5.1 REGRESSION RESULTS 57
5.2 PROPENSITY SCORE MATCHING 65

5.3 ANALYSIS OF ECONOMIC SIGNIFICANCE 68
6. CONCLUSION 69
7. APPENDIX B 71
8. REFERENCES 82
CHAPTER IV – INDIVIDUAL PLAYER CONTRIBUTIONS IN EUROPEAN SOCCER 84
1. INTRODUCTION 85
2. LITERATURE REVIEW 86
3. THEORY 88
4. DATA 91
5. RESULTS 94
5.1 PREDICTIONS 94
5.2 BETTING ANALYSIS 99
5.3 ESTIMATIONS 100
6. CONCLUSION 106
7. APPENDIX C 108
8. REFERENCES 119
CHAPTER V – THESIS CONCLUSION 122
1. CONCLUSION 122
2. REFERENCES 126

LIST OF TABLES

Table 1 - Summary statistics of player share values; the proportion of a team's total spending each ranked player accounts for 26

Table 2 - Difference in the average proportion of resources allocated to player ranks by teams of different stature (standard error in brackets) 27

Table 3 - Exogeneity analysis 28

Table 4 - Dispersion analysis 29

Table 5 - Betting odds comparison 30

Table 6 - The impact of results by the starting eleven players. Top 100 best fitting models are included to highlight the player ranks that contribute most to results 31

Table 7 - Team and optimal TPI allocations over 380 games in 2012/2013 33

Table 8 - The actual and optimal TPI values of the ranked players calculated for the 2007 game played between Watford and Portsmouth 38

Table 9 - The impact on results by the starting eleven players. Different numbers of the best fitting models are included to highlight the player ranks that contribute most to results. (1)

corresponds to the frequency a player rank appears in the top models, (2) corresponds to the player rank coefficient, and (3) corresponds to the t-statistic on the coefficient.....................39
Table 10 - AIC scores average over the number of best fitting models 40
Table 11 - Rest variable summary statistics 56
Table 12 - Relative rest periods on team performance 59
Table 13 - Midweek game effect on weekend team performance 60
Table 14 - Weekend game effect on midweek team performance..63
Table 15 - Logistic Regression of Midweek game probability...66
Table 16 - Matching of exogenous variables...67
Table 17 - Relative rest periods on team performance 73
Table 18 - Midweek game effect on weekend team performance 74
Table 19 - Weekend game effect on midweek team performance..78
Table 20 - Lineup Breakdown 93
Table 21 - Player movement summary 94
Table 22 - Ridge Regression 2015/16 Contingency Matrix 97
Table 23 - Lasso 2015/16 Contingency Matrix 97
Table 24 - Elastic Net 2015/16 Contingency Matrix 97
Table 25 - Prediction accuracy for 2015/16...98
Table 26 - Ridge Regression against betting odds...99
Table 27 - Lasso against betting odds...100
Table 28 - Elastic Net against betting odds 100
Table 29 - Combined Coefficients Model Results (Players) 103
Table 30 - Player Coefficients Model Results 104

Table 31 - Spearman's Rank-Order Correlation 105

Table 32 - Combined Coefficients Model Results (Managers) 112

Table 33 - Manager Coefficients Model Results..113

Table 34 - Combined Coefficients Model Results (Teams) 114

Table 35 - Team Coefficients Model Results...115

Table 36 - Prediction accuracy for 2014/15 - 2015/16..117

Table 37 - Prediction accuracy for 2013/14 - 2015/16..118

LIST OF FIGURES

Figure 1 - Frequency distribution on TPI values for all Premier League players 2013...........25

Figure 2 - The proportion of total team value assigned to each player (1993-2013)...............26

Figure 3 - Unique players in English Premier League starting lineups by year 57

Figure 4 - Histogram of propensity scores..68

Figure 5 - Effect of rest days on goal difference 69

Figure 6 - Observed Goal Difference/Result vs. Predicted Goal Difference...........................96

Figure 7 - Ridge Regression 108

Figure 8 - Lasso 109

Figure 9 - Elastic Net 110

ABSTRACT

The motivation for this thesis concerns worker productivity as estimated from production functions. Identifying worker contributions allows for not just an understanding of economic theory but highlights ways in which business management and strategy can be more efficient. The setting for these analyses is professional soccer, where teams are analogous to businesses and workers are multi-million dollar assets in the form of players. Most of a soccer team's income is tied to its success on the field and so a careful management of staff and players is necessary to their business potential. Sports are the perfect "laboratory" to study economic theory since workers can be observed on a regular basis and there is a large volume of existing data. With almost every game recorded in the modern age this allows for the opportunity to analyse not just worker productivity but also team processes and strategies.

This thesis expands the production function literature using a framework from contest theory literature. Most research in soccer focuses on performance at the aggregate level while this thesis primarily considers performance at the player level. It consists of three papers, each providing a different insight into player productivity. Chapter I presents a brief introduction of the relevant literature and contextualizes the research. Chapter II measures the impact of different workers in a production process depending on their expected productivity, finding support for superstar theories over the O-Ring theory in the English Premier League. Chapter

III looks at the effects of fatigue in professional soccer finding that under current scheduling in the English Premier League and European competition there are no statistically significant effects of receiving different days of rest on team performance. Chapter IV applies high dimensional techniques to European soccer data to predict match outcomes. The models perform almost as well and betting firms and can be used to estimate individual player contributions in the form of rankings. Chapter V concludes.

CHAPTER I

THESIS MOTIVATION

1. INTRODUCTION

The study of productivity has historically been important to economists. Describing how inputs relate to outputs has been researched since the early 1800s. Chambers (1988) chronicles a history of the production function. The first economist to algebraically formulate this relationship is generally believed to be Philip Wicksteed (1894) although Humphrey (1997) presents some evidence Johann von Thünen formulated it first in the 1840’s. The next major theoretical development came from H. L. Moore (1929) who attempted to test marginal productivity theory using statistical techniques. Like von Thünen this work focused on agricultural commodities. Later three more agricultural economists would construct empirical frameworks to help producers make business decisions. Tolley, Black and Ezekiel (1924) attempted to isolate technology in such a way that would allow the application of marginal productivity theory by decision makers. Black along with Cassel (1936) were some of the first economists to use cross-classification tables. Shephard (1970) defines production functions as the relationship between the maximal technically feasible output and the inputs required to produce it. The econometric production functions known today originate from Cobb and Douglas (1928). They originally used macro data to test hypothesis about the marginal productivity theory and the competitiveness of labour markets (Griliches and Mairess 1995) before later refinements shifted to the use of micro data. This study shifted the focus away from supply-and-input-demand relationships and increased the attention on estimating technical relationships. This can be seen from the future work on estimating first-order conditions of a production function after research from Marschak and Andrews (1944). After the estimation of production functions was common practice the next key developments were solving complicated mathematical programming problems. Stigler (1945) solved the diet problem using heuristic methods to find an optimized solution. It is widely believed to be one of the first instances of linear programming. Other studies developed using these methods, such as the study of engineering cost and production functions and efficiency frontier techniques by Farrell (1957). Other programming developments were also found in applied production economics such as Shephard's (1953) first in depth treatment of duality relationships though these had previously been mentioned by Hotelling (1932) and Samuelson (1948). Later, Uzawa (1962) would characterize the class of production functions with constant elasticities of substitution. Heady et al. (1964) used experimental agricultural data to estimate technical production relationships for agricultural products. This was one of the first instances of using experimental data as opposed to market based data using statistical assumptions. Around the same time Mundlak (1963) conducted some empirical modelling of multi-output production relationships, the basis for many multi-output cost and profit function studies. This concludes a brief history on productions functions however more in depth reviews can be found by Fuss and McFadden (1978) and SK Mishra (2007).

The motivation for this thesis concerns the investigation of inputs in the production function: namely worker productivity. This topic has appeal beyond just the field of economics; such as strategy and marketing. Being able to identify worker contributions allows not only an understanding of economic theory but ways in which business management and strategy can be more efficient. The setting for this thesis is professional soccer. Professional soccer can be used to apply this theory, where the top teams are analogous to multi-billion dollar corporations and players are multi-million dollar assets. It is clear to see why the study of worker productivity would be important to teams as nearly all of their income is tied into the success of their players.

Sports are in general a sort of "laboratory" in which to study economic theory. Unlike other markets workers are able to be observed on a regular basis. Productivity is not hidden as each week players are watched by millions who can see their contributions towards game results. Due to the interest generated by fans there is a significant amount of data available which can be analyzed. Thousands of games have their details recorded allowing for the opportunity to explore every aspect of what makes teams successful. This also allows the possibility to not just to analyze individual worker productivity but also to look at team processes and how different strategies can affect output.

There exists a growing research of production functions in sports. Rottenberg (1956) first proposed that the operations of professional sports teams could be modelled using production functions. The first study to estimate a production function for team performance was Scully (1974) who investigated the relationship between wages and marginal revenue product in professional baseball. Zech (1981) used a Cobb-Douglas production function to estimate the major skills involved in professional baseball. Zak et al. (1979) adopt a similar approach using a Cobb-Douglas production function in professional basketball. Similar studies exist for basketball (Scott et al., 1985), American football (Atkinson et al., 1988), and English country cricket (Schofield, 1988). Carmichael and Thomas (1995) formulate a production function for rugby league football using performance influencing variables. Carmichael and Thomas (2001) also estimate a season-based production function for English Premiership soccer teams. Gerrard et al. (2000) estimates a model for manager win ratios in the English Premier League using wages as a measure of available playing talent. Dawson et al. (2000) provide a more comprehensive review of literature on sporting functions as well as using a production function to estimate coaching efficiency in English soccer. More recently Tiedemann et al. (2011) use a Based upon Data Envelopment Analysis to estimate player efficiency in the German Bundesliga. González-Gómez and Picazo-Tadeo (2010) estimate the efficiency of Spanish teams across domestic and European competition in order to measure fan satisfaction. Rimler et al. (2009) perform a Bayesian analysis using Markov Chain Monte Carlo estimation to measure the production efficiency in men's NCAA college basketball.

This thesis aims to expand the applied production function literature using a framework from contest theory literature. Unlike other sports such as baseball the analysis of soccer data, although growing, is in its infancy. This presents a great opportunity to build on the existing research and provide contributions to sports economics. The standard model of a sports league is a contest in which two or more teams compete for a share of success and the productivity of an individual player is his contribution to the success of a team. This framework was first used by El-Hodiri and Quirk (1971) and Quirk and El-Hodiri (1974). This work was later extended by Atkinson et al. (1988), Fort and Quirk (1995), Vrooman (1995), Késenne (2000), Szymanski and Késenne (2004), and Dietl and Lang (2008). Given this underlying theory, Szymanski (2003) claims that outcomes must be determined by relative expenditure unless they are purely random. He summarizes the impact of wage expenditure across several leagues, showing that it is correlated with success although the extent varies by league. Szymanski and Smith (1997) substitute money for talent in the contest success function and show evidence of correlation between aggregate player spending and league performance of teams. The research suggests that on a seasonal basis wages are a reliable measure of productivity.

Most contest theory research in soccer focuses on performance at the aggregate level. This thesis expands on the literature by looking at performance at the player level. It consists of three papers, each providing a different insight into player productivity. Chapter 2 presents the first paper: “Testing the O-Ring theory using data from the English Premier League”. This paper measures the impact of different workers in a production process depending on their expected productivity. The setting is the English Premier league where expected productivity is measured from the transfer fees used to acquire players. In the efficient market for players more productive players should demand a higher transfer fee. The paper examines the impact of these players on the match results. The findings show that the most expensive players tend to have the largest impact on a game whereas the least expensive players have little impact. This is consistent with superstar theories rather than O-Ring theory which states that there is a significant disadvantage to having a weak link in the production process. The optimal spending distribution is found to be more skewed than the observed distribution suggesting a constraint in the market for players.

Chapter 3 presents the second paper: “The Impacts of Rest Periods on results in the English Premier League”. This paper looks at the effects of fatigue in professional soccer using data from the English Premier League. Many managers and players share conflicting views on how much rest is required between games but there is little empirical evidence to support their arguments. The production function is expanded to include information on rest times and distance travelled. Rest can be estimated as the number of days between games played by each player but also as the number of days between games played by the entire team. The spacing between games raise additional factors that can be measured such as team sharpness (the number of consecutive games for players) and team cohesion (familiarity with teammates) as by-products of how much rest and rotation a team experiences. Under current scheduling in domestic and European competition there are no statistically significant effects of receiving different days of rest on team performance. The limited variation in the amount of rest for teams can give concern about the power of the tests used but even if the effects were statistically significant they are found to have a negligible impact on team results.

Chapter 4 presents the final paper in this thesis: "Individual player contributions in European Soccer". This paper applies new techniques to predict match outcomes in professional soccer by estimating player contributions. Using data from the top 25 European soccer leagues, the individual contributions of players is measured using high dimensional fixed effects models. Nine years of data is used to train the model while a further year is used to check for predictive accuracy. The findings show an average prediction rate of 45% with all methods producing similar performance. The model highlights the most productive players but there is some bias towards identifying players who produce and prevent goals directly. This results in attackers and defenders being ranked more highly than midfield players. There is some potential for the models to be used in sports betting as they perform almost as well as betting firms.

2. REFERENCES

[1] 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.

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

[3] Carmichael, F., Thomas, D., & Ward, R. (2001). Production and efficiency in association football. *Journal of Sports Economics*, *2*(3), 228-243.

[4] Cassels, J. M. (1936). On the law of variable proportions. *Explorations in economics: notes and essays contributed in honor of F. W. Taussig*.

[5] Chambers, R. G. (1988). *Applied production analysis: A dual approach* Cambridge University Press.

[6] Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *The American Economic Review, 18*(1), 139-165.

[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] Dietl, H. M., & Lang, M. (2008). The effect of gate revenue sharing on social welfare. *Contemporary Economic Policy, 26*(3), 448-459.

[9] El-Hodiri, M., & Quirk, J. (1971). An economic model of a professional sports league. *Journal of Political Economy, 79*(6), 1302-1319.

[10] Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253-290.

[11] Fort, R., & Quirk, J. (1995). Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature, 33*(3), 1265-1299.

[12] Fuss, M., McFadden, D., & Mundlak, Y. (1978). A survey of functional forms in the economic analysis of production. *History of Economic Thought Chapters, 1*

[13] Gerrard, B. (2000). Football, fans and finance: Understanding the business of professional football. *Edinburgh, Scotland: Mainstream Pubs*

[14] González-Gómez, F., & Picazo-Tadeo, A. J. (2010). Can we be satisfied with our football team? evidence from Spanish professional football. *Journal of Sports Economics, 11*(4), 418-442.

[15] Griliches, Z., & Mairesse, J. (1995). *Production Functions: The Search for Identification,*

[16] Heady, E. O., Jacobson, N., Freeman, A., & Madden, J. P. (1964). Milk production functions incorporating variables for cow characteristics and environment. *Journal of Farm Economics, 46*(1), 1-19.

[17] Hotelling, H. (1932). Edgeworth's taxation paradox and the nature of demand and supply functions. *Journal of Political Economy, 40*(5), 577-616.

[18] Humphrey, T. M. (1997). Algebraic production functions and their uses before Cobb-Douglas.

[19] Késenne, S. (2000). Revenue sharing and competitive balance in professional team sports. *Journal of Sports Economics, 1*(1), 56-65.

[20] Marschak, J., & Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica, Journal of the Econometric Society*, 143-205.

[21] Mishra, S. K. (2007). A brief history of production functions.

[22] Moore, H. L. (1929). Synthetic economics. *Books for college libraries*, Macmillan.

[23] Mundlak, Y. (1968). Elasticities of substitution and the theory of derived demand. *The Review of Economic Studies, 35*(2), 225-236.

[24] Quirk, J. P., & El-Hodiri, M. A. (1973). *The economic theory of a professional sports league*

[25] Rimler, M. S., Song, S., & Yi, D. T. (2010). Estimating production efficiency in men's NCAA college basketball: A Bayesian approach. *Journal of Sports Economics, 11*(3), 287-315.

[26] Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy, 64*(3), 242-258.

[27] Samuelson, P. A. (1948). Foundations of economic analysis.

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

[29] 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.

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

[31] Shepard, R. W. (1953). Cost and production functions. *Princeton: Princeton Univer,*

[32] Shepherd, R. W. (2015). *Theory of cost and production functions* Princeton University Press.

[33] Stigler, G. J. (1961). Economic problems in measuring changes in productivity. *Output, input, and productivity measurement* (pp. 47-78) Princeton University Press.

[34] Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature, 41*(4), 1137-1187.

[35] Szymanski, S., & Késenne, S. (2004). Competitive balance and gate revenue sharing in team sports. *The Journal of Industrial Economics, 52*(1), 165-177.

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

[37] Tiedemann, T., Francksen, T., & Latacz-Lohmann, U. (2011). Assessing the performance of German Bundesliga football players: A non-parametric metafrontier approach. *Central European Journal of Operations Research, 19*(4), 571-587.

[38] Tolley H. R., Black J. D., Ezekiel M. J. B. (1924). Input as related to output in farm organization and cost of production studies, *U.S. Department of Agriculture Bulletin*,1924 1277

[39] Uzawa, H. (1962). Production functions with constant elasticities of substitution. *The Review of Economic Studies, 29*(4), 291-299.

[40] Vrooman, J. (1995). A general theory of professional sports leagues. *Southern Economic Journal,* 971-990.

[41] Wicksteed, P. H., (1965). *An essay on the co-ordination of the laws of distribution.* [London]: [London school of economics and political Science].

[42] 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.

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

CHAPTER II

TESTING THE O-RING THEORY USING DATA FROM THE ENGLISH PREMIER LEAGUE[1]

ABSTRACT

This paper measures the impact of different workers in a production process dependent on their expected productivity. Using the setting of professional soccer, expected productivity is measured from the transfer fees paid to acquire players. It shows that the most expensive players tend to have the largest impact on the game whereas the least expensive players have little impact. The findings support superstar theories rather than O-ring theory. We also find that the optimal spending distribution is more skewed than the observed distribution suggesting some constraint in the market for superstars.

[1] Paper co-authored with Stefan Szymanski and published in the journal *Research in Economics*

1. INTRODUCTION

Most production processes consist of several tasks, often to be completed by different workers. The degree of success in completing an assigned task may be very different depending on the task and the worker. Standard marginal productivity theory usually rests on the assumption that all workers are equally effective. However, some models focus on the productivity of particular workers – e.g. Rosen (1981). By contrast, Kremer's (1993) O-Ring theory suggests that the contribution of the least productive worker may be critical.

In this paper we measure the impact of different workers in a production process dependent on their expected productivity. Our setting is professional soccer. We measure expected productivity from transfer fees paid to acquire players in the English Premier League and we examine the impact of these players on the match results.

We find that the most expensive players tend to have the largest impact on the result of a game, while the least expensive players exert relatively little impact. In that sense our findings are consistent with superstar theories rather than O-Ring theory.

The paper is set out as follows. In the next section we review the relevant literature, Section 3 presents the theory and Section 4 the data. Section 5 presents our estimation methods and results. Section 6 concludes.

2. LITERATURE REVIEW

2.1 O-RING THEORY

The O-Ring theory (Kremer, 1993) proposes a production function which describes a process that is susceptible to mistakes in many of its components. The motivation is that all production tasks must be completed competently in order for any of them to have full value. Kremer defines the production process of the firm as a sequential series of tasks. Workers have varying skill levels represented by a probability that they complete their task. A process employing five workers who each complete their task with 90% probability will find that they produce 59% of the output (0.9^5) of a firm with fully competent workers (100% probability of completion). This implies that there is a significant disadvantage to having a weak link (very low probability of completion) in the process. Four fully competent workers paired with one worker who completes their task only 25% of the time will find their total output at 25% of the firm with perfect workers. In the extreme case, if one worker fails entirely to complete their task then output is worthless.

In the context of professional team sports we can think of each player being matched with an opponent, and their relative capabilities in completing team tasks as determining the outcome of the contest. In football, for example, teams often adopt a “person-marking” strategy – each defender is allocated a specific attacker and their job is to neutralise the attacker′s threat. The success of the team will then depend to a significant degree on the outcome of each of these contests within a contest. While this conceptualisation differs from the conventional O-Ring theory in that production is not sequential and is dependent on relative rather than absolute productivity, the problems are isomorphic.

Anderson and Sally (2013) draw on O-Ring theory to argue that “football is a weakest link game where success is determined by whichever team makes the fewest mistakes”. The implication is that to make a successful team “you need to look less at your strongest links and more at your weakest ones”. Their approach suggests that only the match-up between the two weakest players on each team matters – whoever of these players makes the most mistakes determines the outcome. More generally, it is possible to model all the possible match-ups that might matter. For our analysis we will construct all 2047 combinations of match-ups that could exist among eleven distinct players on each team. These include, for example, only the two weakest players, only the two strongest players, all the players, the five best players on each side and so on.

Kremer's research builds on the analysis of superstars (Rosen, 1981). Rosen developed a model where productivity is multiplicative in order to account for a wage distribution which is more highly skewed than the underlying distribution of abilities. Small differences in ability can lead to very large differences in wages in the multiplicative context, one which is often associated with professional sport. In our data set many workers are valued close to the average, but a small group is valued at a significantly higher rate.

The modelling of the production function in sports can be traced back to Tullock (1980) where the probability of success is a function of the relative share of resources employed. We adapt this function to allow for the contribution of each player on the team. The importance of each player on the team is measured by his market transfer fee value, which we take to be a measure of expected productivity.

2.2 SALARY DISPERSION AND MOTIVATION

In this paper we use market transfer fee values to measure expected productivity. The gap between the most and least valuable player on a team can be very large, a fact which has generated some research in the team sports literature. General theories have provided conflicting views and predictions on the relationship between salary dispersion and firm performance. Akerlof and Yellen (1988) suggested an "effort-wage variance model" that defines a firm's output as a function of the effort by its workers which depends negatively on the variance of wages. They hypothesise that lower salary dispersion between workers results in a more amicable relationship between workers which improves firm performance. This work was extended to produce a model using their "fair wage-effort hypothesis" (Akerlof and Yellen, 1990). In this hypothesis the effort from a worker relates to how his/her received wage differs from what is perceived as a “fair” wage. Those who are paid at or above this “fair” wage contribute their full effort whereas those paid below this wage proportionally reduce their effort. A worker′s conception of a “fair” wage should be motivated by the wages of other workers in the firm and so in turn a greater variability in wages should correspond to lower effort by underpaid workers. Similarly, Levine (1991) formalised cohesiveness among workers in firm production. Levine conjectured that a smaller wage disparity between the low and high skilled workers would create a more harmonious relationship.

Many empirical studies have looked at the effects of salary dispersion on team success across professional sports leagues, most predominantly in MLB. Richards and Guell (1998), Bloom (1999), and Depken (2000) found a negative relationship between salary dispersion and team success by reference to a team′s winning percentage. Molina (2004) also found a negative relationship when fitting a stochastic frontier model. This relationship has also been investigated in other professional sports leagues. Frick et al. (2003) studied the NFL, NHL, NBA, and MLB and while finding no effect from salary dispersion in the NFL or NHL found a negative impact on winning percentages in MLB but a positive relationship in the NBA.

Sometimes contrasting views appear within the same league. For example, in the NHL Sommers (1998) found a negative effect of salary dispersion on team performance while Marchand et al. (2006) also found a negative relationship and Frick et al. (2003) found no relationship. In the NBA, Berri and Jewell (2004) and Katayama and Nuch (2011) find that salary dispersion does not influence team performance whereas Frick et al. (2003) found a positive relationship.

Given the inconclusiveness of this literature, we feel it is safe to assume that these kinds of interactions play a negligible role in determining team performance. Each worker performs to the best of his ability and is rewarded in proportion to this ability. Given that performance on the job is closely observed, many of the moral hazard and adverse selection problems affecting labour markets are absent.

2.3 WAGE PERFORMANCE RELATIONSHIP

The standard model of a sports league is a contest in which two or more teams compete for a share of success. The productivity of a player is his contribution to the success of the team. This framework was first used in the models produced by El-Hodiri and Quirk (1971) and Quirk and El Hodiri (1974), then extended by Atkinson et al. (1988), Fort and Quirk (1995), Vrooman (1995), Késenne (2000), Szymanski and Késenne (2004) and Dietl and Lang (2008). We use both audited wages and player transfer fee values as estimates of productivity for football players in the English Premier League.

Given the underlying contest theory structure, outcomes must be determined by relative team expenditures unless outcomes are purely random. Szymanski (2003) summarises the impact of relative wage spending across a number of leagues and shows that it is indeed correlated with success, although the extent varies by league. Szymanski and Smith (1997) substitute money for talent in the contest success function. They construct a model where clubs maximise a weighted average of profit and success and players aim to maximise their earnings. There is an efficient market for players whose expected productivity is known and observed by consumers (fans) as well as producers. Fans demand success and so more successful clubs generate higher revenues. Given the heterogeneity of capital among clubs there is a dispersion of investment levels at equilibrium, at which spending reflects success. These assumptions seem plausible for the football labour market since there are many buyers and sellers and a lot of public information available about the players.

There is research which shows evidence of a high correlation between aggregate player spending and the league performance of teams (Garcia-del-Barrio and Szymanski, 2009; Forrest and Simmons, 2002; Szymanski, 2003 ; Szymanski and Smith, 1997) suggesting that on a seasonal basis wages are a reliable measure of productivity. Here we examine productivity on a game by game basis. We use data on transfer fees as a measure of productivity. Szymanski (2014) found that transfer fees paid capture the variation of team performance in the English Premier League almost as closely as audited aggregate wage expenditure. The transfer fees used in this paper come from a Transfer Price Index (TPI) constructed by Graeme Riley. The transfer fees are adjusted to allow for the considerable rate of transfer fee inflation. A player's transfer fee value depends on the year the transfer occurred and is inflated by this TPI.

Given the high correlation between player spending and performance, the question of endogeneity naturally arises. Hall et al. (2002) tested for Granger causality and found evidence that the causation runs from wages to success for data from the English Premier League although the power of this test may not be strong. By contrast, Dobson and Goddard (1998) find evidence of causality running from lagged revenue to current performance for Football League clubs. This dependence was larger for smaller than larger clubs suggesting that success is concentrated among a small group of wealthy clubs. Peeters and Szymanski (2013) employ two distinct approaches to deal with unobserved productivity in their data. They firstly add in club-specific fixed effects which imply that productivity is different between teams but remains constant over time. They also use the method of Olley and Pakes (1992) to infer productivity using an instrumental variable based on accounting measures. In this paper we use a slightly different approach – we develop an instrument derived from the charitable donations of football clubs, which are correlated with club wage and transfer spending but unrelated to the success of a club on the pitch.

3. THEORY

When looking at team performance, we measure the outcome on a game by game basis. The possible outcomes for soccer games are the home team winning, the away team winning or both teams playing out a tie. Teams try to win matches by using resources to acquire playing talent and success is determined by the relative share of these resources. We develop a model in which budgeting choices are made at the level of the individual player. First consider a model where only the total budgets of the teams matter. Success is then determined by the ratio of one team's budget relative to the other. For English soccer clubs we can measure the budget ratios in two ways: the flow of wages paid to players and the capital value of the player when they are acquired from another club (transfer value). Our measure of transfer value is based on the TPI, explained more fully in Section 4. Unlike wages, the TPI values are specific to each individual player allowing us an estimate of team budgets for each game rather than by season.

3.1 THE AGGREGATE MODEL

We can model the probability of success for a team in any match as a function of aggregate resources allocated to the team (denoted *bi*) relative to its opponent. Now define a contest success function which relates the resources of each team in the contest to the bins defined below; that is the probability of the home team winning, losing or both teams playing out a tie. The outcome of a game is defined as *yij* which refers to the result of a game between teams *i* and *j*. By adapting the Tullock contest success function to allow for the possibility of a draw we can use the following specification, taking the form of an ordered logit model:

$$y_{ij} = \begin{cases} \text{win } j \Leftrightarrow y^*_{ij} = \dfrac{\alpha_h w_i m_i b_i^{\theta}}{w_j m_j b_j^{\theta}} \exp(\epsilon_{ij}) \le \gamma_1 \\ \text{tie} \Leftrightarrow \gamma_1 < y^*_{ij} = \dfrac{\alpha_h w_i m_i b_i^{\theta}}{w_j m_j b_j^{\theta}} \exp(\epsilon_{ij}) \le \gamma_2 \\ \text{win } i \Leftrightarrow \gamma_2 < y^*_{ij} = \dfrac{\alpha_h w_i m_i b_i^{\theta}}{w_j m_j b_j^{\theta}} \exp(\epsilon_{ij}) \end{cases} \quad (1)$$

where

- y_{ij} is the outcome of the contest, a home win, draw or home loss. We need to map these continuous variables to these three possible outcomes and so introduce y^*_{ij} as a continuous, unobserved variable for this purpose, i.e. a latent index. When y^*_{ij} falls equal or below the cutoff value of γ_1, the away team wins the match. When y^*_{ij} is above the cutoff value of γ_2, the home team wins. When y^*_{ij} lies between both of these threshold values, the contest is a draw.
- b_i and b_j are the budgets for teams i and j respectively. The budget can be measured either by total wage spending or the total capitalized (transfer) value of the team. θ is a parameter which measures the sensitivity of the contest to investments in the teams similar to the Tullock parameter.
- w_i and w_j are team fixed effects, which may reflect underlying differences in productivities among the teams
- m_i and m_j are manager fixed effects
- α_h corresponds to the advantage acquired by being the home team, which might be a function of travelling, referee bias and the bias of home fans towards their team.
- ϵ_{ij} is an exponential noise term which accounts for chance factors specific to a contest. This could include weather conditions, errors by the officials and other 'luck' based events.

Thus our contest success function allows for the quality of the players measured by their wages or transfer values, the productivity of the manager, the club and home advantage.

3.2 THE VALUE MATCHING MODEL

Consider an expansion of the previous ordered logit model now accounting for individual player values. Previously we used the ratio of the *aggregate* TPI values for teams to account for relative spending. We now consider the ratios of each individual player matched against their opponents with the same value rank. In other words, the most expensive players are matched, the second most expensive are matched, and so on, all the way down to the least expensive players. The model now has the following multiplicative specification:

$$y_{ij} = \begin{cases} \text{win } j \Leftrightarrow y_{ij}^* = \dfrac{\alpha_h w_i m_i}{w_j m_j} \prod_{t=1}^{11} \dfrac{x_{it}^{\theta_t}}{x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \leq \gamma_1 \\ \text{draw} \Leftrightarrow \gamma_1 < y_{ij}^* = \dfrac{\alpha_h w_i m_i}{w_j m_j} \prod_{t=1}^{11} \dfrac{x_{it}^{\theta_t}}{x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \leq \gamma_2 \\ \text{win } i \Leftrightarrow \gamma_2 < y_{ij}^* = \dfrac{\alpha_h w_i m_i}{w_j m_j} \prod_{t=1}^{11} \dfrac{x_{it}^{\theta_t}}{x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \end{cases} \quad (2)$$

where

- x_{it} is refers to the payroll (or in this case TPI value) of player *t* on team *i*. *t* ranges from 1 to 11 which includes all starting 11 players for a football match
- The sensitivity parameter θ_t now varies by player. Each of the starting eleven players on a team have a specific sensitivity that we assign to the ratio of players in the contest.

If player salaries in the soccer market are a good indicator of player talent we can infer that the weakest player in a team is the one with the lowest value. Relating this to the O-ring model suggests that only the ratio of the weakest players should be considered. As a theory of team performance this seems extreme. At the other extreme, we might pose a “superstar” theory in which the productivity of the best (most valuable) player determined the success probability. From the point of view of the empirical test, we also want to allow the possibility that the performance of every player on the team has some importance so that the ratios of all eleven players should be included. This specification nests the more extreme cases. In theory, performance could be captured by any combination of the value ratios of the eleven matched players on both teams, that is we have 2^{11}-1=2047 unique model specifications allowing every possible combination of matches. This means that each player would feature in exactly 2014 of the models. We estimate all 2047 of these models and take the average value of coefficients for each value match to determine coefficients and t-values. In this way not only do we test the O-Ring theory but every ranked value ratio and its contribution towards team performance.

Given this structure we need some boundary conditions on the model to analyse the individual contributions of the starting eleven players. The constraints are as follows:

$$x_{i1} + x_{i2} + x_{i3} + \cdots + x_{i11} \leq B \qquad (3)$$

B refers to a team's total budget. This is the total TPI value spent across the starting eleven players. Thus the decision problem of the club is how to allocate a fixed budget across 11 players.

$$x_{i1} \geq x_{i2} \geq x_{i3} \geq \cdots \geq x_{i11} \qquad (4)$$

(4) reflects the fact that there is a strict ordering of player finances by definition of the player ranks.

$$x_{it} \geq R \qquad (5)$$

Finally, (5) says that players have a reservation value, *R*.

After running the model we would like to estimate exactly how much spending should be assigned to each player so as to maximize team performance. To achieve this we maximise equation (2) with respect to the TPI value of any individual player. We have;

$$y_{ij} = \frac{\alpha_h w_i m_i}{w_j m_j} \prod_{t=1}^{11} \frac{x_{it}^{\theta_t}}{x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \qquad (6)$$

Let $\delta_t \frac{\alpha_h w_i m_i}{w_j m_j} \exp(\epsilon_{ij})$ and set up the objective function $f(\vec{x})$ such that

$$y_{ij} = \left(\delta_1 x_{i1}^{\theta_1}\right)\left(\delta_2 x_{i2}^{\theta_2}\right)(\ldots)\left(\delta_{11} x_{i11}^{\theta_{11}}\right) = f(\vec{x}) \qquad (7)$$

where $\vec{x} = (x_1, \ldots, x_{11})$

We set up the constraints as follows;

- $x_1 + \cdots + x_{11} \leq B$
- $x_t \geq R, \qquad t = 1, \ldots, 11$

- $x_t - x_{t+1} \geq 0, \quad t = 1, \ldots, 10$

Which can be used to define the following Lagrangian equation;

$$L(\vec{x}, \lambda, \vec{\mu}, \vec{v}) = f(\vec{x}) + \lambda(x_1 + \cdots + x_{11} - B) + \sum_{t=1}^{11} \mu_t (R - x_t) + \sum_{t=1}^{11} v_t (x_{t+1} - x_t) \qquad (8)$$

where $\vec{\mu} = (\mu_1, \ldots, \mu_{11})$ and $\vec{v} = (v_1, \ldots, v_{11})$

By the theorem of complementary slackness we have that the Lagrangian coefficients are either binding or not binding. Using those constraints we can maximize with respect to the TPI value for individual players. Setting the first order conditions for players 1 and 2 against each other yields.

$$\frac{\theta_1}{x_{i1}} = \frac{\theta_2}{x_{i2}} \qquad (9)$$

$$x_{i2} = \frac{\theta_2 x_{i1}}{\theta_1} \qquad (10)$$

Therefore if a player has a larger θ coefficient then this player contributes more towards results and so should command more in spending relative to his team-mates. When the constraints are binding we have the cases where $x_{it} = R$ and $x_{it} = x_{it+1}$. Adding these constraints gives three conditions for spending on players;

4. DATA

This research makes use of a uniquely constructed database of player transfer values which comprises of 21 English Premier League seasons from 1992/1993 to 2012/2013. The database is supplied by Graeme Riley, author of the annual statistical reference book, *Football in Europe*. His dataset includes information on teams and their players who participated in each individual game over the 21 seasons, giving a total of 8226 individual soccer matches contested by 45 unique Premier League soccer teams. Since each game appears twice (containing the home and away team line-ups) they are given a corresponding match ID used to prevent games from being counted twice in the analysis. Also listed are the managers for both teams and each individual game result. The English Premier League consists of 20 clubs where 3 teams are replaced on a yearly basis due to the promotion and relegation system. Teams will play each other on two occasions, once at home and once at away for a total of 38 games.

The transfer fees values are adjusted to allow for the considerable rate of transfer fee inflation. This is achieved by using his constructed Transfer Price Index (TPI). Each player′s transfer fee value, which depends on the year the transfer occurred, is inflated by this TPI. There is an extensive amount of player trading in the soccer market, both domestically and globally. Since we can observe the productivity of players regularly and player turnover is so high we can be confident that their transfer values reflect market value. This metric excludes players who have never been transferred between clubs and so have no TPI value; namely homegrown players. These players were assigned the median TPI value of their entire team on for that specific year. Of the 90,486 first team places in the starting line-ups that were filled by 2557 players, 25,669 or 28.37% of these places contain players classified as homegrown. As well as players' TPI values we also include information on the wage values for players. Deloitte′s Annual Review of Football Finance Databook contains wage values for each team on a given year. This financial information is taken from the company accounts of every team. These wage values encompass all staff at the club, not just the players but including all backroom and boardroom staff and there is no wage breakdown for each individual player.

Only the starting eleven players in a match are considered and not any of the substitutes. The starting eleven players are on the pitch for most of the match compared to the substitutes. Adding these players into the value matching model may include cases where unused substitutes are matched against an opposition starting player which will add noise within the contest success function.

Figure 1 presents a histogram of the TPI values for all players in the 2012/2013 English Premier League. The crosses on the *x*-axis represent the mean TPI values for players on each of the 20 clubs. The wage distribution is skewed which is common in all labour markets. Table 1 presents summary statistics on team spending. Players are ranked one to eleven from most expensive to least expensive. The proportion that each player makes up of the team′s total TPI value in that game is calculated. The proportion values for each ranked player are calculated over the entire 8226 games in the database. Figure 2 presents a boxplots of the amount of TPI spending allocated to each player. The boxplots representing the more expensive players have larger quartile ranges but also contain a significant number of outliers at large share proportions. Teams are spending significant sums on the best players in their squad but also allocating widely differing amounts of their total resources on them. The least expensive players all receive a similar share of spending.

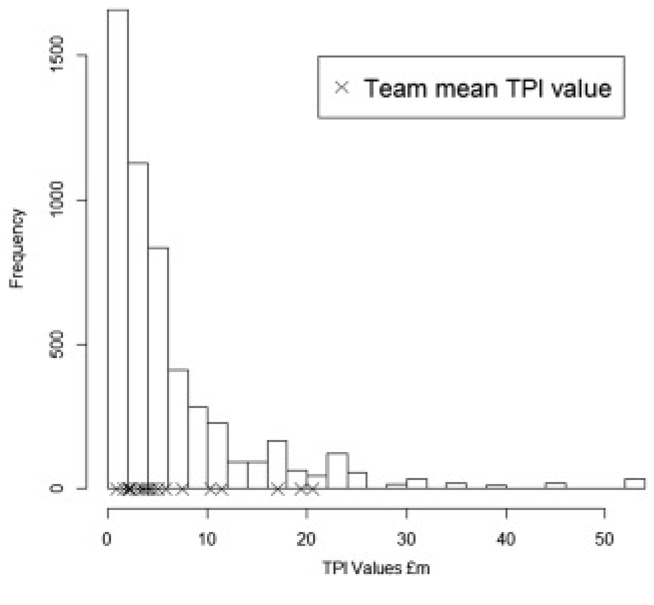

Figure 1 - Frequency distribution on TPI values for all Premier League players 2013

Table 1 - Summary statistics of player share values; the proportion of a team's total spending each ranked player accounts for

Player rank	Mean	Standard deviation	First quartile	Median	Third quartile
One	0.221	0.060	0.180	0.211	0.249
Two	0.155	0.036	0.130	0.151	0.176
Three	0.118	0.026	0.100	0.116	0.134
Four	0.095	0.019	0.083	0.094	0.107
Five	0.082	0.016	0.073	0.082	0.092
Six	0.076	0.015	0.066	0.076	0.085
Seven	0.071	0.015	0.061	0.072	0.081
Eight	0.065	0.016	0.054	0.065	0.076
Nine	0.054	0.017	0.042	0.055	0.067
Ten	0.040	0.018	0.027	0.039	0.052
Eleven	0.023	0.015	0.011	0.019	0.032

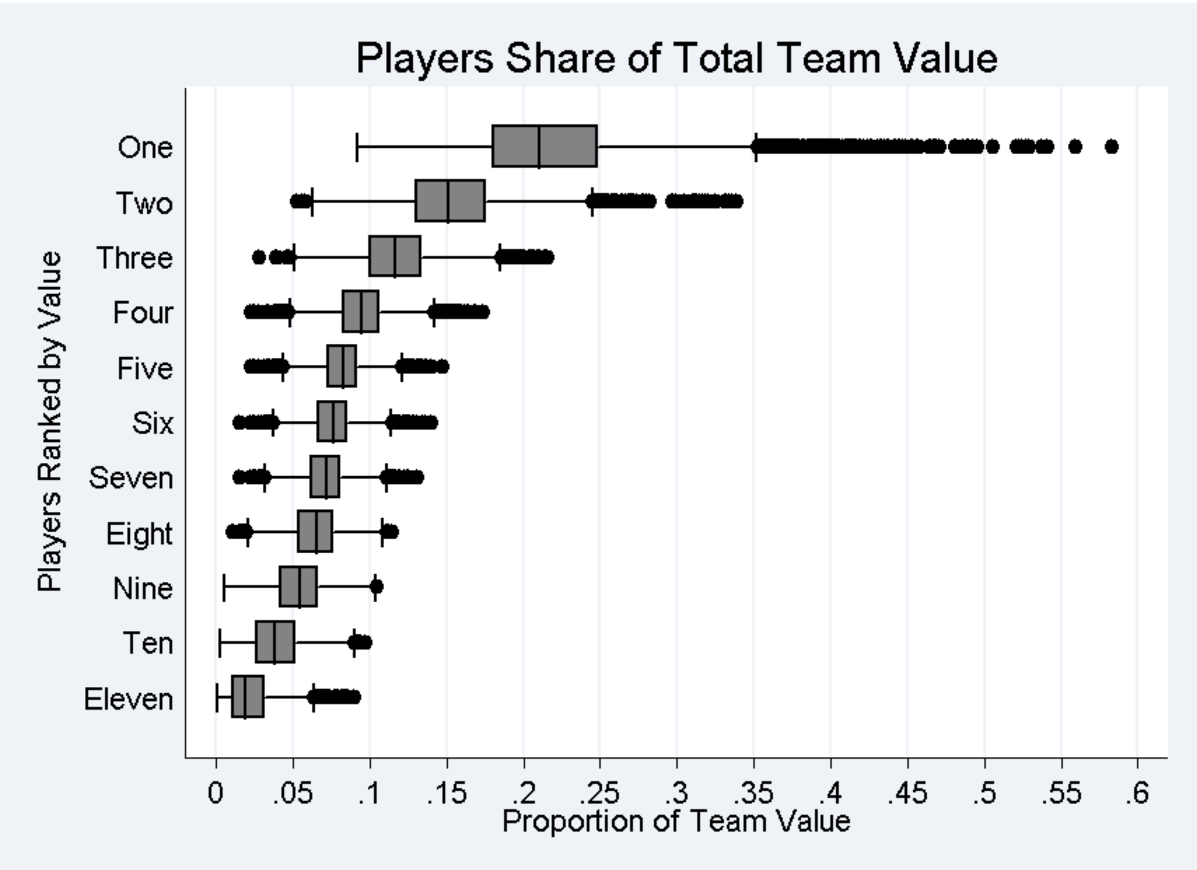

Figure 2 - The proportion of total team value assigned to each player (1993-2013)

Table 2 presents the difference in the proportions of team spending on players relative to the stature of the teams contesting a game. We define “Big Teams” as the five teams who finished top of the league that season. The “Middle Teams” are the next ten highest finishing teams in the league and the “Small Teams” are the lowest finishing five teams. Due to the relationship between spending and success, the “Big Teams” will generally be the teams with the largest financial resources available and likewise the “Small Teams” will have the fewest. The values are averaged over all teams and matches in that season. It is clear that proportional spending on the ranked players does not differ relative to the team's stature and financial size. We have no reason to believe that teams employ different strategies on resource distribution depending on the opponent.

Table 2 - Difference in the average proportion of resources allocated to player ranks by teams of different stature (standard error in brackets)

Player rank	Match-ups by team stature 2012/2013		
	Big vs. small	Big vs. medium	Medium vs. small
One	−0.020 (0.022)	−0.030 (0.000)	0.014 (0.156)
Two	−0.014 (0.009)	0.004 (0.305)	−0.024 (0.000)
Three	−0.003 (0.479)	0.009 (0.004)	−0.016 (0.000)
Four	0.004 (0.215)	0.018 (0.000)	−0.013 (0.000)
Five	0.009 (0.008)	0.006 (0.000)	−0.004 (0.052)
Six	0.009 (0.003)	0.007 (0.000)	0.001 (0.690)
Seven	0.003 (0.207)	−0.001 (0.464)	0.003 (0.176)
Eight	−0.001 (0.855)	−0.006 (0.003)	0.006 (0.004)
Nine	−0.005 (0.106)	−0.009 (0.000)	0.013 (0.000)
Ten	0.006 (0.064)	−0.004 (0.113)	0.014 (0.000)
Eleven	0.003 (0.323)	0.002 (0.497)	0.006 (0.006)

5. ESTIMATION

5.1 EXOGENEITY IN WAGES

We have argued that wages are unbiased predictors of player quality which in turn determines team results and ultimately overall league position. However, reverse causality is also a possibility through means such as team bonuses. This could result in the wage coefficient being overstated in the estimation process. To account for this we introduce an instrumental variable which while correlated with team spending is not correlated with how successful a team is.

This variable is the charitable donations given each season by a team. Richer teams are more likely to donate larger amounts of money to charity.

In their financial accounts clubs must specify their charitable donations. Over the 21 seasons 54% of our observations (where an observation is a team in a given year) included donations to charity. Since this will be used to instrument on the ratio of team wages we take a log transform of the ratio of charitable donations to reduce the effect of large ratio values. We can now follow a two-stage regression as follows:

$$\frac{wage_i}{wage_j} = a + b\log\left(\frac{1+CD_i}{1+CD_j}\right) \qquad (12)$$

where we use an ordinary least squares regression of the wage proportion on the ratio of charitable donations. The residuals from this regression can now be substituted in place of the wage ratio values to remove the overstated effect from the wages when estimating results.

The results from running this regression are shown in Table 3. As expected, the coefficient on spending is lower when using charitable donations as an instrument. The explanatory power of both models is almost identical. This means that even though we use wages to explain results, we can be confident that the overstated effect of wages is not biasing our results.

Table 3 - Exogeneity analysis

Result	Wage ratio	CD ratio
Home advantage	0.963	0.962
	$(0.045)^{***}$	$(0.045)^{***}$
Spending	0.414	0.304
	$(0.052)^{**}$	$(0.046)^{***}$
Pseudo-R^2	0.091	0.090
Observations	8226	8226

5.2 DISPERSION ANALYSIS

Previous studies have examined the effect of wage dispersion on results (see literature review). We examine the effect of dispersion using the TPI values (the wage data is at the team level and cannot be disaggregated). Running the model on the 21 seasons of data will allow us to determine the impact of TPI values on results.

Table 4 displays the results from the analysis of dispersion. We wish to test the reliability of the TPI data relative to the wage values and so the wage model is included. Both the wage and TPI models have a similar fit as demonstrated by their Pseudo-R^2 values. The player wage model has a slightly better fit, but the TPI model is almost identical. In the model that accounts for the variation of spending we see that the coefficient on this variance is not significant. So while we believe that the dispersion of spending matters in team performance, the variance is too coarse a measure and is not included in the full model.

Table 4 - Dispersion analysis

Result	Wage ratio	TPI ratio	TPI & variance ratios
Adv	0.963	0.948	0.948
	$(0.045)^{***}$	$(0.044)^{***}$	$(0.044)^{***}$
Spending	0.414	0.076	0.076
	$(0.052)^{**}$	$(0.018)^{***}$	$(0.018)^{***}$
Variance	–	–	0.000
ratio (TPI)			(0.169)
Pseudo-R^2	0.091	0.089	0.089
Observations	8226	8226	8226

5.3 THE EFFICIENCY OF WAGES AND TPI IN PREDICTING RESULTS

Pre-match betting odds are the best available predictor of match results. We estimate Brier scores to measure the efficiency of wages and TPI in predicting results. Where probabilities of outcomes lie along the continuous interval of 0 to 1, the Brier score measures the absolute difference between these probabilities and the actually occurring event. A lower Brier score reflects better predictive accuracy in the model. For the purposes of this analysis we use data for the 380 games played in the 2012/2013 season.

We take betting odds supplied by two firms, Bet365 and Blue Square who list the historic odds for English Premier League games on their website[2]. Table 5 shows that the two betting firms have the greatest predictive accuracy reflected by the lowest Brier scores which is expected given these odds best reflected the prior information before matches. The models using TPI values and wages not only have very similar predictive accuracy but are almost as accurate as odds produced by the betting firms.

Table 5 - Betting odds comparison

Measure	Brier score
TPI aggregate	0.579
TPI ratios	0.576
Wage spending	0.575
Bet365	0.573
Blue Square	0.572

5.4 OPTIMALITY

Table 6 represents the impact of ranked players on team results. The coefficients reflect the relative importance of the ranked player match-ups relative to the rest of the team. Of the 2047 models we report the top 100 best fitting models and count how many times a player match-up appeared. A robustness check is included in the appendix comparing the results averaged over different numbers of the best fitting models. The coefficients on player match-ups quantify their relative influence of each player value rank. The coefficients were calculated by averaging all 1024 models in which the player appears. A larger coefficient value represents a greater contribution towards team results. Players Eleven and One do not appear in many of these models. The eleventh ranked player appears less frequently than any other player rank. This suggests that the data does not strongly support the O-Ring theory. Player One also appears

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

rarely suggesting that the superstar player in a team does not alone drive team performance. By contrast, Player Two appears in every model suggesting that it is better to have two excellent players in a team rather than one exceptional player who commands most of the budget. Other players which feature often are players Five, Six and Ten implying that a strong team core is important for positive results but not at the expense of having too many weak players.

Table 6 - The impact of results by the starting eleven players. Top 100 best fitting models are included to highlight the player ranks that contribute most to results

Player rank	Average over 2047 models		
	Frequency in top 100 models	Coefficient	*t*-statistic
One	16	1.307	1.857
Two	100	2.751	6.254
Three	37	1.966	2.325
Four	19	1.046	0.974
Five	52	2.283	2.654
Six	44	2.422	1.639
Seven	31	−1.318	−0.816
Eight	31	−1.122	−1.533
Nine	16	0.530	1.131
Ten	59	0.836	5.051
Eleven	15	−0.138	−0.996

Using these results we can now solve equation (11) to obtain an estimate of the optimal spending distribution. Maintaining the strict ranking of players, the optimal spending for a player is given in terms of the ranked player immediately above. For player one there are no higher ranked players and so he retains the value x_{i1}. With a given x_{i1} we can solve for player two. In this case $\theta_2 = 2.751 > 1.307 = \theta_1$ implying investment in player two up to the value of investment in player one and so $x_{i2} = x_{i1}$. For player three we have that $\theta_3 = 1.966 < 2.751 = \theta_2$ meaning that investment is more valuable in player two than three. The investment in player three is a fraction of the investment in player two given by $x_{i3} = \frac{\theta_3 x_{i2}}{\theta_2} = 0.715 x_{i2} = 0.715 x_{i1}$. These rules are applied up to and including player six. For player seven $\theta_7 < 0$

meaning that investment in player seven at the expense of the higher ranked players is suboptimal. The reserve value *R* is assigned to this player. Due to the strict ranking of players, all lower ranked players must now receive the reserve value as their amount of investment. This gives us values for all player investments in terms of either the most expensive player and the reserve value.

$x_{i2} = x_{i1}$

$x_{i3} = 0.715x_{i1}$

$x_{i4} = x_{i5} = x_{i6} = 0.380x_{i1}$

$x_{i7} = \cdots = x_{i11} = R$ (13)

We can substitute these values into equation (3) and solve for x_{i1} given a set *B* and *R*. This yields;

$x_{i1} = x_{i1} = 0.715x_{i1} = 3(0.380)x_{i1} + 5R = B$ (14)

$x_{i1} = \frac{B-5R}{3.856}$ (15)

We can use these results to calculate the optimal spending distribution in every game played. We set the reserve value for 2013 at *R* = £100,000. Reserve values are then calculated for the 1992/93-2012/13 seasons relative to the index of total spending in that given season.

Table 7 illustrates the distribution of expected values, averaged for teams across the 2012/13 season, and compares this with the optimal distribution. For the purposes of illustration we show the clubs with the minimum, median and maximum variance across the season. Generally spending on the most expensive player exceeds spending on the second most expensive player, but for the median club the gap is not large. The major difference between actual and optimal spending is at the bottom of the team. The optimal share of the four worst players is tiny compared to the actual shares. This may reflect the need of clubs to retain better players in case the best players are injured.

Table 7 - Team and optimal TPI allocations over 380 games in 2012/2013

Player rank	Budget allocations for teams 2012/2013			
	Minimum spending variance (%)	Median spending variance (%)	Maximum spending variance (%)	Average Budget (%)
	Everton	Sunderland	Southampton	Optimal[3]
One	18.04	20.44	41.23	25.76
Two	11.14	15.87	14.57	25.76
Three	10.00	10.82	9.10	18.41
Four	9.30	9.60	6.56	9.80
Five	8.69	8.96	5.37	9.80
Six	8.36	8.16	5.21	9.80
Seven	8.25	7.95	5.09	0.14
Eight	8.12	7.69	4.58	0.14
Nine	7.78	6.09	3.88	0.14
Ten	7.15	3.24	3.08	0.14
Eleven	3.18	1.19	1.33	0.14

6. CONCLUSION

In most production settings teams of workers carry out tasks, either simultaneously or in sequence aimed at the production of a final output. When workers are heterogeneous the productivity of particular workers in their given tasks may have a more than proportional impact on the final outcome. Examples include superstar theories where only the most productive workers matter, and the popular O-Ring theory, where the productivity of the least productive worker is crucial.

[3] Distributions for Everton, Sunderland and Southampton are their average spending distributions over 38 games. These teams have the minimum, median and maximum spending variances respectively. The optimal spending distribution is constructed from the average budget across all games in the season.

We use data from professional football to identify the impact on team output of the expected productivity of each worker. We find evidence that the most valuable workers, measured by the transfer fee paid to acquire their services, tend to exert the greatest impact. We find little evidence in support of O-Ring theory.

In a sense our findings are reassuring – the most valuable workers (who also tend to receive the highest wages) ought to make the greatest contribution to team production. However, our results also suggest a puzzle. We find that the optimal distribution of spending is more skewed than the observed distribution. This may reflect some constraint in the market for superstars – a limited number of players at the top end of the distribution such that a few players are paid very high wages but there are not enough players of this quality to go around. Equally it is possible that the observed distribution may reflect equity concerns within the organisation.

7. APPENDIX A

7.1 LAGRANGIAN DERIVATION

We can now maximise with respect to the individual wage values;

$$\frac{d\mathcal{L}}{dx_{i1}} = \frac{\theta_1 f(\vec{x})}{x_{i1}} + \lambda - \mu_1 - \nu_1 \qquad (16)$$

$$\frac{d\mathcal{L}}{dx_{i2}} = \frac{\theta_2 f(\vec{x})}{x_{i2}} + \lambda - \mu_2 - \nu_2 + \nu_1 \qquad (17)$$

$$\vdots$$

$$\frac{d\mathcal{L}}{dx_{i11}} = \frac{\theta_{11} f(\vec{x})}{x_{i11}} + \lambda - \mu_{11} + \nu_{10} \qquad (18)$$

and constraints;

$$\frac{d\mathcal{L}}{d\lambda} = x_{i1} + \cdots + x_{i11} - B \qquad (19)$$

$$\frac{d\mathcal{L}}{d\mu_t} = R - x_{it} \quad t = 1, \ldots, 11 \qquad (20)$$

$$\frac{d\mathcal{L}}{d\nu_t} = x_{it+1} - x_{it} \quad t = 1, \ldots, 10 \qquad (21)$$

By the theorem of complementary slackness we have that either the Lagrangian coefficients are equal to 0 and the constraints are binding or the Lagrangian coefficients are positive and the constraints are not binding. We maximise with respect to the TPI value for player 1 assuming that the constraints are binding. We have from the theorem that either;

$$\lambda \geq 0 \quad \text{or} \quad \sum_{t=1}^{11} x_t - B < 0 \qquad (22)$$

Teams will always utilise their full budgets and so it must be the case that $\lambda=0$ and the constraint is binding

$$x_{it} > R \quad \text{and} \quad \mu_t = 0 \quad \text{or} \quad x_{it} = R \quad \text{and} \quad \mu_t > 0 \qquad (23)$$

So if this constraint is not binding we have that $x_{it}>R$ and $\mu_t=0$

$$x_{it} > x_{it+1} \quad \text{and} \quad \nu_t = 0 \quad \text{or} \quad x_{it} = x_{it+1} \quad \text{and} \quad \nu_t > 0 \qquad (24)$$

When this constraint is not binding we have that $x_{it}>x_{it+1}$ and $\nu_t=0$. Putting this together we can maximise the objective function;

$$\frac{dy_{ij}}{dx_{i1}} = \frac{d}{dx_{i1}} \prod_{t=1}^{11} \frac{\alpha_h w_i m_i x_{it}^{\theta_t}}{w_j m_j x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \qquad (25)$$

$$\frac{dy_{ij}}{dx_{i1}} = \theta_1 x_{i1}^{\theta_1 - 1} \prod_{t=2}^{11} \frac{\alpha_h w_i m_i x_{it}^{\theta_t}}{w_j m_j x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \qquad (26)$$

$$\frac{\theta_1}{x_{i1}} \prod_{t=1}^{11} \frac{\alpha_h w_i m_i x_{it}^{\theta_t}}{w_j m_j x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) = 0 \qquad (27)$$

This is analogous for any of the ten other players. For player 2 we would have;

$$\frac{\theta_2}{x_{i2}} \prod_{t=1}^{11} \frac{\alpha_h w_i m_i x_{it}^{\theta_t}}{w_j m_j x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) = 0 \qquad (28)$$

Next, set equations. (20) & (21) together to determine the relative spending on these two players which maximises team performance.

$$\frac{\theta_1}{x_{i1}} \prod_{t=1}^{11} \frac{\alpha_h w_i m_i x_{it}^{\theta_t}}{w_j m_j x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) = \frac{\theta_2}{x_{i2}} \prod_{t=1}^{11} \frac{\alpha_h w_i m_i x_{it}^{\theta_t}}{w_j m_j x_{jt}^{\theta_t}} \exp(\epsilon_{ij}) \qquad (29)$$

$$\frac{\theta_1}{x_{i1}} = \frac{\theta_2}{x_{i2}} \qquad (30)$$

7.2 INDIVIDUAL GAME OPTIMALITY

We can examine optimality on the individual game level. To see how following the distribution can affect win probability we take as one example the game, played between Watford and Portsmouth in the 2006/2007 season. The match ended 4-2 in favour of Watford. Portsmouth had a TPI budget of $B=£30{,}329{,}936$. Using the full player ratio model on the actual TPI player values, the predicted probabilities for a Portsmouth win, loss and draw are;

$Pr(W)=0.27208$

$Pr(D)=0.30596$

$Pr(L)=0.42196$

We simulate predicted probabilities from replacing the player TPI values with the optimal distribution. The model will now take into account the new match-ups and we should see an increase in the probability of Portsmouth winning the match. Holding the spending for Watford constant, we now have;

$Pr(W)=0.43472$

$Pr(D)=0.30336$

$Pr(L)=0.26188$

The reserve value for 2007 indexed with league spending is $R=£106{,}885$. We can calculate the optimal spending for player one and all other players by using the equations in (13). Table 8 presents the actual TPI values from the match and also the optimal spending distribution using our equations. Note that the optimal and actual spending on the lowest ranked player is not very large.

Table 8 - The actual and optimal TPI values of the ranked players calculated for the 2007 game played between Watford and Portsmouth

Player rank	Actual TPI values	Optimal TPI values
One	4,665,046	7,727,282
Two	4,665,046	7,727,282
Three	4,067,039	5,522,296
Four	3,391,495	2,939,551
Five	2,831,258	2,939,551
Six	2,831,258	2,939,551
Seven	2,831,258	106,885
Eight	2,271,021	106,885
Nine	1,892,517	106,885
Ten	707,196	106,885
Eleven	176,799	106,885

Following the optimal spending distribution suggests that the probability of team success can be greatly increased. Teams do not appear to be following a strategy that matches up with these optimal spending distributions. If the market for buying players is efficient then why are teams adopting inefficient strategies for acquiring playing talent? One reason for this would be that there is not enough talent in the global market to go around. There are only a limited amount of top players and so teams will be unable to find players who match the optimal budget allocations. Also, many players will not want to play for the reserve valuation due to being at prestigious teams. The worst player at a team like Man Utd would still expect to be worth considerably more than the reserve valuation.

7.3 MODEL ROBUSTNESS

The model presented in this paper averages player coefficients over all 2047 models but there are other ways the coefficients can be calculated. Table 9 presents the player coefficients averaged over the top fitting 1000, 500, 100, 50, 10 models, and also the coefficients of the best fitting model top model.

Table 9 - The impact on results by the starting eleven players. Different numbers of the best fitting models are included to highlight the player ranks that contribute most to results. (1) corresponds to the frequency a player rank appears in the top models, (2) corresponds to the player rank coefficient, and (3) corresponds to the t-statistic on the coefficient

Rank	Top 1000			Top 500			Top 100			Top 50			Top 10			Top 1		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
One	494	0.861	2.273	174	0.725	3.404	16	0.676	20.108	5	0.664	24.825	1	0.648	NA	0	NA	NA
Two	881	2.774	6.628	490	2.842	7.664	100	2.950	8.740	50	3.058	11.338	10	3.165	10.838	1	3.182	NA
Three	533	1.506	2.448	214	1.303	3.334	37	1.266	3.970	10	1.269	4.592	1	1.084	NA	0	NA	NA
Four	431	0.733	1.091	183	0.840	1.302	19	1.201	1.885	8	1.074	1.808	0	NA	NA	0	NA	NA
Five	494	2.069	3.797	253	2.162	4.116	52	2.279	4.244	31	2.377	4.876	7	2.495	4.844	1	2.362	NA
Six	496	2.372	1.795	496	2.401	1.772	44	2.787	2.132	21	2.702	1.787	3	3.262	2.546	0	NA	NA
Seven	459	−1.542	−1.099	459	−1.824	−1.460	31	−2.339	−2.523	16	−2.296	−2.421	3	−1.932	−1.491	0	NA	NA
Eight	449	−1.209	−1.884	449	−1.292	−2.091	31	−1.465	−3.060	14	−1.512	−4.066	2	−1.697	−5.632	0	NA	NA
Nine	440	0.480	1.039	440	0.447	0.967	16	0.553	1.242	4	0.355	0.589	0	NA	NA	0	NA	NA
Ten	513	0.818	5.054	513	0.803	4.877	59	0.763	4.610	33	0.734	4.601	5	0.733	4.826	0	NA	NA
Eleven	417	−0.149	−1.089	417	−0.167	−1.232	15	−0.235	−2.460	5	−0.216	−1.764	0	NA	NA	0	NA	NA

Each of the averaged models produced player coefficients with the same sign as the coefficients averaged over 2047 models in the paper. The relative size of the coefficients is also similar. Regardless of the number of models averaged over, similar results are obtained.

Table 10 shows that the AIC scores for each of the average models are almost identical. Regardless of the number of models averaged, the player coefficients and AIC scores are very similar so even though the AIC score is lowest for the best fitting model, we are comfortable using the results averaged over 2047 models as this provides more information overall about each individual player.

Table 10 - AIC scores average over the number of best fitting models

Number of best fitting models	AIC score
Top 1	16,627.95
Top 10	16,628.97
Top 50	16,629.82
Top 100	16,630.39
Top 500	16,632.16
Top 1000	16,633.51
Top 2047	16,637.42

8. REFERENCES

[44] Akerlof, G. A., & Yellen, J. L. (1988). Fairness and unemployment. *The American Economic Review, 78*(2), 44-49.

[45] Akerlof, G. A., & Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics, 105*(2), 255-283.

[46] Anderson, C., & Sally, D. (2013). *The numbers game: Why everything you know about football is wrong* Penguin UK.

[47] 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.

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

[49] Bloom, M. (1999). The performance effects of pay dispersion on individuals and organizations. *Academy of Management Journal, 42*(1), 25-40.

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

[51] Depken, C. A. (2000). Wage disparity and team productivity: Evidence from major league baseball. *Economics Letters, 67*(1), 87-92.

[52] Dietl, H. M., & Lang, M. (2008). The effect of gate revenue sharing on social welfare. *Contemporary Economic Policy, 26*(3), 448-459.

[53] Dobson, S. M., & Goddard, J. A. (1998). Performance and revenue in professional league football: Evidence from granger causality tests. *Applied Economics, 30*(12), 1641-1651.

[54] El-Hodiri, M., & Quirk, J. (1971). An economic model of a professional sports league. *Journal of Political Economy, 79*(6), 1302-1319.

[55] Forrest, D., & Simmons, R. (2002). Team salaries and playing success in sports: A comparative perspective. *Sportökonomie* (pp. 221-238) Springer.

[56] Fort, R., & Quirk, J. (1995). Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature, 33*(3), 1265-1299.

[57] Frick, B., Prinz, J., & Winkelmann, K. (2003). Pay inequalities and team performance: Empirical evidence from the north American major leagues. *International Journal of Manpower, 24*(4), 472-488.

[58] Garcia-del-Barrio, P., & Szymanski, S. (2009). Goal! profit maximization versus win maximization in soccer. *Review of Industrial Organization, 34*(1), 45-68.

[59] Hall, S., Szymanski, S., & Zimbalist, A. S. (2002). Testing causality between team performance and payroll: The cases of major league baseball and English soccer. *Journal of Sports Economics, 3*(2), 149-168.

[60] Katayama, H., & Nuch, H. (2011). A game-level analysis of salary dispersion and team performance in the national basketball association. *Applied Economics, 43*(10), 1193-1207.

[61] Késenne, S. (2000). Revenue sharing and competitive balance in professional team sports. *Journal of Sports Economics, 1*(1), 56-65.

[62] Kremer, M. (1993). The O-ring theory of economic development. *The Quarterly Journal of Economics, 108*(3), 551-575.

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

[64] Marchand, J. T., Smeeding, T. M., & Torrey, B. B. (2006). Salary distribution and performance: Evidence from the national hockey league. *Center for Policy Research,*

[65] Molina, D. J. (2004). Productive efficiency and salary distribution: The case of US major league baseball. *Scottish Journal of Political Economy, 51*(1), 127-142.

[66] Olley, G. S., & Pakes, A. (1992). *The Dynamics of Productivity in the Telecommunications Equipment Industry,*

[67] Peeters, T., & Szymanski, S. (2014). Financial fair play in European football. *Economic Policy, 29*(78), 343-390.

[68] Quirk, J. P., & El-Hodiri, M. A. (1974). *The economic theory of a professional sports league*. In: Noll, R. (Ed.), Government and the Sports Business. Washington, DC: Brookings Institution, 33–90

[69] Richards, D. G., & Guell, R. C. (1998). Baseball success and the structure of salaries. *Applied Economics Letters, 5*(5), 291-296.

[70] Riley, G. *Football in Europe.* Tony Brown, Various years.

[71] Rosen, S. (1981). The economics of superstars. *The American Economic Review, 71*(5), 845-858.

[72] Sommers, P. M. (1998). Work incentives and salary distributions in the national hockey league. *Atlantic Economic Journal, 26*(1), 119-119.

[73] Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature, 41*(4), 1137-1187.

[74] Szymanski, S., & Késenne, S. (2004). Competitive balance and gate revenue sharing in team sports. *The Journal of Industrial Economics, 52*(1), 165-177.

[75] Szymanski, S., Rodríguez, P., Késenne, S., & García, J. (2013). *Wages, transfers and the variation of team performance in the english premier league* Edward Elgar Northampton, MA.

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

[77] Vrooman, J. (1995). A general theory of professional sports leagues. *Southern Economic Journal*, 971-990.

CHAPTER III

THE IMPACT OF REST PERIODS ON RESULTS IN THE ENGLISH PREMIER LEAGUE[4]

ABSTRACT

This paper measures the impact of fatigue in competitive soccer. Using the setting of the English Premier League we set up a production function to estimate results while accounting for the various ways to measure team rest times. The findings show that under current scheduling of domestic and European competition there are no statistically significant effects of receiving different rest days on team performance. Given the limited variation in the amount of rest per team we explore further and find that even if such an effect did exist, it would be negligible.

[4] Paper co-authored with Stefan Szymanski

1. INTRODUCTION

People never seem to tire of debating the role of fatigue in professional soccer. Players' physical condition is an essential factor towards a team's success hence well rested players should be able to contribute more towards overall team performance. While many managers and players share conflicting views on how much rest is appropriate between games, there is little empirical evidence supporting their arguments. This paper focuses on the effects of rest times and the distances travelled between soccer matches for English Premier League teams. Our model takes these factors into account but also includes the length of time between consecutive games. The distance travelled for each game assesses the impact of team travel on performance.

We have attempted to estimate the impact of rest in different ways. Rest periods can be estimated as the number of days between games played by each player, but also as the number of days between games played by the entire team. The spacing between games raise additional factors we can measure such as team sharpness (the number of consecutive games for players) and team cohesion (familiarity with teammates) as by-products of how much rest and rotation a team experiences.

We find that under current scheduling in domestic and European competition there are no statistically significant effects of receiving different days of rest on team performance within the English Premier League. Given the limited variation in the amount of rest per team there might be a concern about the power of our tests. We find that even if the effects were statistically significant, they would have a negligible impact on team results.

The paper sets out as follows. In section 2 we review the relevant literature, section 3 presents the theory and section 4 the data. Results and estimations are shown in section 5 and section 6 contains concluding remarks.

2. LITERATURE REVIEW

Soccer is a sport with both physically and mentally demanding components. Top-class players on average cover a distance of roughly 11km each match though the distance differs highly between players and is partly related to the team position (Bangsbo 1994; Reilly 1997). Players perform around 1,350 individual activities, inclusive of 69 high-speed runs and 109 moderate-speed runs (Mohr et al., 2003). In addition to these runs, other energy-demanding activities include accelerating, dribbling, tackling, jumping and turning (Bangsbo 1994). These studies suggest that the volume of high-intensity exercise in matches can be used as a valid measure of physical performance in soccer. Although it is uncertain to what extent the players undergo fatigue during soccer games, several researchers have observed the total distance covered in the second half being reduced compared with the first (Reilly and Thomas, 1976; Van Gool et al., 1988; Bangsbo et al., 1991). This reduction may reflect the development of fatigue in the second half. It is often speculated whether players are able to recover from this fatigue in time for the next match.

The best teams in the top European professional leagues, such as the English Premier League, La Liga and Bundesliga, can play over 60 matches each season from pre-season friendlies to the end of the domestic league competition. These teams will play in their domestic league and cup competitions but may also play in the Champions League or the Europa League. Often these players will also be called up for additional matches with the national team, increasing their fixture congestion. If teams/players are not well rested then it will have negative consequences on the physical performance of the sport and diminish the performance for fans.

Rest is also important for governing bodies who are responsible for the welfare of players as well as wanting the services of well rested stars for competitions such as the World Cup. In addition, if opposing teams are experiencing different days of rest then the balance of the match could be affected. Scheduling is planned in such a way as to minimize imbalances of rest periods for teams.

The role of fatigue in sports has mostly been researched from outside the perspective of economics but some research exists that looks not just at rest, but other similar factors that may affect performance. Looking at the National Basketball Association (NBA) over a 19-year period, (Ashman, Bowman and Lambrinos 2010) find that the home team performs poorly when playing the second of back-to-back games where the visiting team had 1 or 2 days of rest. This effect is greater when the home team had travelled easterly across time zones between the games. The betting market was not able to account for the home team's fatigue when creating spreads, mispricing the games. (Entine and Small 2008) analyze the relationship between home court advantage and the fewer days of rest between games visiting teams receive, looking at two seasons (2004-2006). The results show that lack of rest for the road team is an important, although not dominant, factor to the home court advantage. (Scoppa 2013) investigated the effect of rest on international soccer teams in the World Cup and European Championship. Scoppa related team performance to the respective days of rest teams had after their previous match, finding that under the current structure of the international tournaments, there are no relevant effects of different days of rest on team performance.

Differences in rest and travelling times between teams can also contribute to home field advantage. (Carmichael and Thomas 2005) find evidence for home field advantage in the English Premier League. (Neville and Holder 1999) provide evidence to suggest that travel factors contribute to part of home advantage provided the journey crosses a number of time zones. In countries where the travel distance is not so large, crowd factors appeared to be the main cause of home advantage rather than travel distance. Some papers analyze the relationship between team performance in soccer and the distance travelled to an away game. Oberhofer, Philippovich and Winner (2010) show that team performance declines the further the distance to the away venue, in particular the travelling team is more likely to concede goals. In the National Football League (Nichols 2012) finds that teams are more likely to lose if they are travelling a longer distance, especially when the crossing at least one time zone from east to west.

(Krumer and Lechner 2016) investigate whether midweek games give an advantage to teams in the German Bundesliga. They find that home advantage disappears for midweek games, since the midweek matches are allocated unevenly among teams. Home teams perform worse for midweek games, due to lower attendances and the perception of these home beings as being less important. This favours teams with fewer midweek home matches.

3. THEORY

To capture the effect of rest times on results we would ideally conduct a randomized controlled trial where we randomly allocate different rest teams between games for all teams and produce some experimental data. Team performance can then be in part attributed to how much rest a team received before the game. However, scheduling is not randomly assigned for professional soccer teams so we must use observational data based over previous seasons. The main reason for the variability of rest times comes from the allocation of midweek games. If we assume that midweek games are allocated randomly to teams then we could compare the results for teams at the weekend by whether they had played a midweek game or not. The difference between the results would be a consistent estimate of the midweek effect. But, scheduling is not always randomly assigned to teams as for example, fixture congestion can result in matches being rescheduled. Successful teams are also more likely to play midweek games through domestic cup and European cup competitions. As a result stronger teams will receive biased scheduling since they are likely to play more games and therefore have a tighter schedule. These deviations must be accounted for in our estimations especially if they are correlated with our outcome variable.

We use a selection-on-observable strategy to identify the effect of midweek games. Since we know the scheduling must take account of certain factors we can capture these characteristics as well as other such as stadium location and team finances which also influence scheduling. Since the effect of midweek games might be different for different teams, and this heterogeneity is unknown we use a flexible propensity score matching approach to control for the various confounding factors. Rubin (1976) presents work on a class of matching methods which are called 'equal percent bias reducing' (EPBR) because they yield the same percentage reduction in bias for all matching variables, and thus for any linear combination of these variables. We use an extension of this work (Rubin 1980) by using Monte Carlo values for the percent reduction in bias where Mahalanobis-metric pair matching is used. This method is implemented in the Stata "psmatch2" package which we will use for our propensity score matching analysis. It is appropriate for this analysis as the metric attempts to find pair matches close on all matching variables.

Rest times are only part of what drives performance in soccer and other research has examined the relationship of pay and performance in sport. Scully (1974) carried out an econometric study on pay versus performance in Major League baseball. Scully estimates a production function which relates team outputs to the win percentage by using several team performance inputs. Szymanski and Smith (1997) adopt a similar approach in English soccer. Money is substituted for talent in the contest success function, finding that on a seasonal basis wages are a reliable measure of productivity (Forrest and Simmons 2002, Szymanski 2003). When considering team performance, we measure results on a game by game basis. There are two different ways we assess a game result. One is by looking at a team's goal difference. A team will have a positive goal difference if they score more goals than they concede and a negative goal difference if they concede more goals than they score. A goal difference of zero corresponds to a tied game. This measure allows us to determine a scale of success beyond simply whether a game is won. The second way results are assessed is by choosing the game outcome. These are either the home team winning, the away team winning, or both teams playing out a tie. Teams try to win matches by using resources to acquire playing talent and success is determined by the relative share of these resources. Using the pay on performance framework we estimate the relationship between player rest and team performance. To estimate this relationship, we look at two different parametric linear models (3.1): the Ordinary Least Squares estimator and the Ordered Logistic model. These models correspond to each of the two ways we express game outcomes. The OLS estimator is used to form a relationship between goal difference and the explanatory variables while the ordered logit model uses game outcome; namely a home win, loss, or draw. For each we develop three frameworks which will define the variables we use to quantify rest (3.1.1 - 3.1.3).

3.1 PARAMETRIC LINEAR MODELS

First consider the model using an OLS estimator specification. Success is then determined by one team's budget relative to the other, in addition to our rest and distance explanatory variables. Team and manager fixed effects are also included. For English football clubs we can measures the budgets in two ways: wages paid to players and the capital value of the player when they are acquired from another club (transfer value). Our measure of transfer value come from a Transfer Price Index (TPI) constructed by Graeme Riley, author of the annual statistical reference book, *Football in Europe*. The transfer fees are adjusted to allow for the considerable rate of transfer fee inflation. A player's transfer fee value depends on the year in which the transfer occurred and is inflated by this TPI index. Unlike wages, the TPI values are specific to each individual player allowing us an estimate of team budgets for each game rather than by season. Results are measured by the goal difference between teams.

The second model uses an ordered logit specification. We consider the second measure of team performance by looking at the game outcome where the possible results are a home team win, loss or a tie. Success is again determined by one team's budget relative to the other, in addition to rest and distance factors. For each of the two estimation methods we adopt the following three frameworks which define the different variables used to quantify team rest.

3.1.1 RELATIVE REST MODEL

The first framework considered is the proportional rest between teams. All English Premier League games are included for this analysis. The regressions alternate between different groupings of rest variables so that we capture the rest effect in different ways. All specifications contain the variables on team finances, home advantage, distance, team sharpness and team cohesion as well as team and manager fixed effects. The rest specifications are set out as follows. The first compares the total rest days for the objective team and the opponent team. The second specification takes a ratio of these rest days and sorts them into bins from where the objective team has the least rest to the most rest compared to their opponent. The final specification takes these bins and further splits them by adding which half of the year the game is being played, to determine whether fatigue plays a role in specific parts of the season.

3.1.2 MIDWEEK EFFECT MODEL

The second framework looks at the effect of playing midweek games on weekend results. Only weekend games from the English Premier League are considered for this model. Regressions alternate between different ways of specifying midweek games. This includes whether teams play a midweek game, the type of midweek game (league, cup, or European game), the rotation of squads due to midweek games and fixture congestion caused by midweek games. All specifications again contain the variables on team finances, home advantage, distance, team sharpness and team cohesion as well as team and manager fixed effects.

3.1.3 WEEKEND EFFECT MODEL

The final model looks at the effect of playing weekend games on the following midweek games. Only midweek games from the English Premier League are considered for this model. Regressions alternate between different ways of specifying weekend games. These are analogous to the Midweek Effect Model where indicators for midweek games are substituted for indicators for weekend games. All specifications again contain the variables on team finances, home advantage, distance, team sharpness and team cohesion as well as team and manager fixed effects.

4. DATA

This research makes use of a uniquely constructed database of player and team rest values which comprises of 21 English Premier League seasons from 1992/93 to 2012/13. The English Premier League consists of 20 clubs where 3 teams are replaced on a yearly basis due to the promotion and relegation system. Teams will play each other on two occasions, once at home and once away for a total of 38 games each, yielding 7980 league games in our database.

In addition to the English Premier League fixtures the data includes the line-ups for every Domestic Cup and European fixture Premier League teams played between the 1992/93 and 2012/13 seasons. Altogether this accounts for 1250 European Cup (Champions League, Europa

League, UEFA Cup and Cup Winners Cup), 1380 League Cup, and 1469 FA Cup fixtures. This allows us to compute the individual rest times for players as well as the distance travelled by teams to those matches in addition to the existing variables.

Variables were created to capture the total rest times for teams. *TRP* corresponds to the Total Rest Period for the objective team measures as the total number of days each starting player on the team has rested before the current game (up to a maximum of 77 days). *TORP* measures the Total Opponent Rest Period. Next, we take the ratio of the *TRP* and *TORP* and order the observations from smallest to largest. The smallest 10% are assigned a dummy variable *R1* where the objective team has the smallest amount of rest relative to the opponent team. Symmetrically the largest 10% are assigned to *R5*. The lowest observations above 10% and up to 30% are assigned to *R2.* Symmetrically the observations between 70% and 90% are assigned to *R4.* The remaining 40% are assigned to *R3* which is used as a baseline for when teams have approximately even rest times. These variables are then further split down depending on the half of the year where *H1* corresponds to games between July-December and *H2* for games between January-June.

The spacing between games allows for additional variables to be measured. The match sharpness variable captures the match fitness of a player; the more often he plays the better he should perform. Sharpness ranges from 1-5 depending on the number the previous 5 games he has played for the team. *SharpProp* is a variable defined as the ratio of match sharpness for all players on a given team against the opponent team. We also define a cohesion variable, which captures how familiar the squad is playing with each other. Cohesive teams are likely to perform better since each player is more familiar with their teammates' playing styles and working to a settled pattern of play. The variable is defined as the number of unique players who played in the last 5 games (ranging between 11 and 55), normalized by dividing through by 11. Thus the variable ranges from 1 to 5 where a lower value corresponds to a more cohesive team. *CohProp* is defined as the ratio between a given team and its opponent. The variables *lnDist* and *lnDistOpp* are the log of distance travelled in kilometers by both teams to the current game. The *lnDistLast* and *lnDistLastOpp* variables are the distance travelled for the previous game.

Additional variables were created to capture the effect of midweek/weekend games and other causes of team rotation. The *Midweek* and *MidweekOpp* variables note if a team has played a midweek game before the current weekend game. *MidEuro, MidCup, MidLeague, OppEuro, OppCup, and OppLeague* reflect whether the midweek game was a European, domestic cup, or league game. *ChangedSquad* and *ChangedOpp* represent how many players have changed from the starting eleven since the last match while *lnTPIChange* and *lnTPIChangeOpp* capture the net TPI value of this change. The number of games played in weekly periods are captured in the *OneWeek*, *TwoWeek*, *ThreeWeek*, and *FourWeek* variables to highlight extended fixture congestion. Finally the *WkndMid* and *WkndMidOpp* variables note if a team has played a weekend game before the current midweek game.

We take the log values of the objective team (*lnTeamXI)* and opponent team *(lnOppXI)* finances to account for the skewed distribution of team budgets due to the richest teams. The *Adv* variable corresponds to the advantage gained by being the home team, which might be a function of traveling, referee bias and the bias of home fans towards their team.

Table 11 presents some summary statistics about the game variables. We find that teams tend to have a similar match sharpness variable but differ somewhat in the cohesion value. The mean value for rest days of 62.58 suggests that teams are almost always fully rested and as such games where one team notably has more rest than another is uncommon.

Table 11 - Rest variable summary statistics

Variable	Mean	Standard Deviation	Median
lnTeamXI	17.77	0.743	17.80
lnOppXI	17.77	0.743	17.80
Adv	0.5	0.5	0.5
SharpProp	1.027	0.293	1
CohProp	1.424	0.178	1.450
lnDist	2.434	2.541	0
lnDistOpp	2.434	2.541	0
lnDistLast	2.445	2.641	0
lnDistLastOpp	2.445	2.641	0
TRP	62.58	15.975	68.00
TORP	62.58	15.975	68.00
R1	0.1002	0.3	0
R2	0.2022	0.402	0
R4	0.2006	0.4	0
R5	0.0983	0.298	0
H1R1	0.0453	0.208	0
H1R2	0.1093	0.312	0
H1R4	0.1088	0.311	0
H1R5	0.0440	0.205	0
H2R1	0.0549	0.228	0
H2R2	0.0929	0.290	0
H2R4	0.0918	0.289	0
H2R5	0.0543	0.227	0

Figure 3 shows the number of unique starting players averaged across English Premier League teams in each year. We can see that this average does not change significantly from year to year so we have no reason to believe squad size or depth is affecting how much rest players receive.

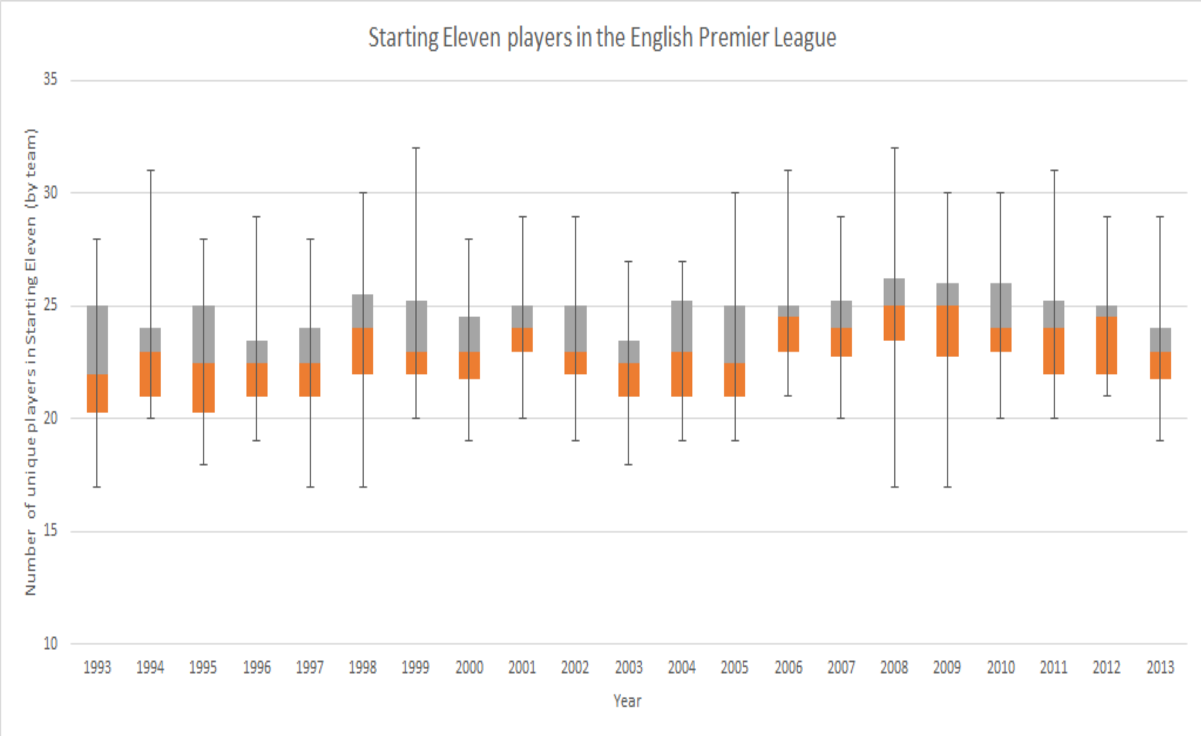

Figure 3 - Unique players in English Premier League starting lineups by year

5. RESULTS

5.1 REGRESSION RESULTS

In this section we conduct an econometric analysis of rest times on team performance while controlling for other variables that could determine team performance. Several specifications are considered using both a parametric linear model and an ordered logit model. For the linear model we use Goal Difference as the dependent variable. The variable Goal Difference is positive if Team A scores more than Team B and is negative if they score less. For the ordered logit model we use Game Result as the dependent variable. Soccer games can have one of three results: a home win, home loss or a draw. The tables below will report only the variables relating to rest. The full analysis can be found in the appendix but in short all variables not related to rest are found to be consistent with the framework built by previous literature.

Table 12 reports the results of the Relative Rest Model estimates using all Premier League games. Columns (1)-(3) contain OLS estimates of the linear model and columns (4)-(6) estimates with the ordered logit model. None of the *R1* to *R5* variables are significant but have varying signs for coefficients. This is also true when including the *H1* and *H2* variables. The *TRP* variable has a positive coefficient while the *TORP* variable has a negative coefficient. Both variables are statistically insignificant.

Table 13 reports the results of the Midweek Effect Model estimates using weekend Premier League games only. Columns (1)-(5) contain OLS estimates of the linear model and columns (6)-(10) estimates with the ordered logit model. The coefficient of the *Midweek* variable is negative and statistically insignificant. By contrast the *MidweekOpp* variable has a positive coefficient. The Midweek variables corresponding to competition types are statistically insignificant. The *ChangedSquad* coefficient is negative and statistically insignificant while the *ChangedOpp* coefficient is positive. The *lnTPIChange* coefficient is negative and statistically insignificant while the *lnTPIChangeOpp* coefficient is positive. The *OneWeek* and *TwoWeek* coefficients are positive and statistically insignificant while the *ThreeWeek* and *FourWeek* coefficients are negative. All interaction coefficients are statistically insignificant.

Table 14 reports the results of the Weekend Effect Model estimates using midweek games only. Columns (1)-(4) contain OLS estimates of the linear model and columns (5)-(8) estimates with the ordered logit model. The coefficient of the *WkndMid* variable is positive and statistically insignificant. By contrast the *WkndMidOpp* variable has a negative coefficient. The *ChangedSquad* coefficient is negative and statistically insignificant while the *ChangedOpp* coefficient is positive. The *lnTPIChange* coefficient is negative and statistically insignificant while the *lnTPIChangeOpp* coefficient is positive. The *OneWeek, TwoWeek, ThreeWeek* and *FourWeek* coefficients are statistically insignificant.

Table 12 - Relative rest periods on team performance

	(1)	(2)	(3)	(4)	(5)	(6)
R1	-0.0054 (0.0628)			-0.0337 (0.0799)		
R2	0.0255 (0.0483)			0.0321 (0.0613)		
R4	-0.0267 (0.0485)			-0.0272 (0.0615)		
R5	0.006 (0.0633)			0.0344 (0.0805)		
H1R1		0.0011 (0.088)			-0.0512 (0.1127)	
H1R2		-0.0172 (0.0603)			-0.021 (0.0763)	
H1R4		0.0231 (0.0605)			0.0148 (0.0765)	
H1R5		0.0005 (0.089)			0.0564 (0.1142)	
H2R1		-0.0059 (0.0804)			-0.0136 (0.1019)	
H2R2		0.081 (0.0646)			0.0881 (0.0822)	
H2R4		-0.08 (0.0649)			-0.0835 (0.0826)	
H2R5		0.0153 (0.0809)			0.02 (0.1024)	
TRP			0.0004 (0.0013)			0.0013 (0.0017)
TORP			-0.0004 (0.0013)			-0.0013 (0.0017)
Observations	8226	8226	8226	8226	8226	8226
AIC	31013.2	31016.5	31011.1	16578.8	16583.8	16576.2

Table 13 - Midweek game effect on weekend team performance

	(1)	(2)	(3)	(4)	(5)
Midweek	-0.0181 (0.0466)	-0.0237 (0.0704)			-0.0172 (0.0466)
MidweekOpp	0.0143 (0.0466)	0.0227 (0.0705)			0.0171 (0.0466)
MidEuro			0.1243 (0.101)		
MidCup			0.0612 (0.0752)		
MidLeague			-0.0549 (0.0692)		
OppEuro			-0.1197 (0.1009)		
OppCup			-0.0626 (0.0752)		
OppLeague			0.0545 (0.0692)		
ChangedSquad		-0.0186 (0.0197)			
ChangedOpp		0.0162 (0.0199)			
Midweek:ChangedSquad		1.2308 (1.0838)			
MidweekOpp:ChangedOpp		0.6608 (0.8744)			
lnTPIChange	-0.0306 (0.0363)				
lnTPIChangeOpp	0.0294 (0.0364)				
Midweek:lnTPIChange	1.216 (1.0787)				
MidweekOpp:lnTPIChangeOpp	0.5993 (0.8736)				

OneWeek				0.0035 (0.0419)	
TwoWeek				0.0108 (0.0434)	
ThreeWeek				-0.0107 (0.0436)	
FourWeek				-0.0009 (0.0287)	
Observations	6397	6397	6397	6397	6397
AIC	24380.1	24379.5	24378.2	24376.0	24375.7

	(6)	**(7)**	**(8)**	**(9)**	**(10)**
Midweek	-0.0507 (0.0587)	-0.0311 (0.0881)			-0.0545 (0.0587)
MidweekOpp	0.0546 (0.0587)	0.023 (0.0882)			0.0544 (0.0587)
MidEuro			0.1593 (0.1314)		
MidCup			0.012 (0.0949)		
MidLeague			-0.0804 (0.087)		
OppEuro			-0.1534 (0.1315)		
OppCup			-0.0152 (0.0948)		
OppLeague			0.0769 (0.0869)		
ChangedSquad		-0.0149 (0.0248)			
ChangedOpp		0.0111 (0.025)			
Midweek:ChangedSquad		-0.0049 (0.0302)			

MidweekOpp:ChangedOpp		0.0088 (0.0303)			
lnTPIChange	-0.0336 (0.0455)				
lnTPIChangeOpp	0.0324 (0.0455)				
Midweek:lnTPIChange	-0.0144 (0.0745)				
MidweekOpp:lnTPIChangeOpp	0.0126 (0.0745)				
OneWeek				-0.0061 (0.0525)	
TwoWeek				0.0159 (0.0542)	
ThreeWeek				-0.0158 (0.0549)	
FourWeek				0.0018 (0.036)	
Observations	6397	6397	6397	6397	6397
AIC	12994.2	12996.7	12996.4	12994.5	12990.8

Table 14 - Weekend game effect on midweek team performance

	(1)	(2)	(3)	(4)
WkndMid	0.1437 (0.1741)	0.1498 (0.1745)	0.1068 (0.2346)	
WkndMidOpp	-0.1356 (0.174)	-0.1483 (0.1745)	-0.0773 (0.2347)	
lnTPIChange		-0.0287 (0.1376)		
lnTPIChangeOpp		0.0325 (0.1373)		
WkndMid:lnTPIChange		0.0846 (0.1625)		
WkndMidOpp:lnTPIChangeOpp		-0.1008 (0.1625)		
ChangedSquad			-0.0341 (0.0613)	
ChangedOpp			0.0431 (0.0614)	
WkndMid:ChangedSquad			0.0183 (0.0718)	
WkndMidOpp:ChangedSquad			-0.0296 (0.0719)	
OneWeek				0.0143 (0.1131)
TwoWeek				-0.0019 (0.1188)
ThreeWeek				-0.0262 (0.1206)
FourWeek				0.0307 (0.0732)
Observations	1829	1829	1829	1829
AIC	3932.7	3936.7	3939.1	3935.5

	(5)	**(6)**	**(7)**	**(8)**
WkndMid	0.1408 (0.1252)	0.1415 (0.1254)	0.1069 (0.1673)	0.1348 (0.1428)
WkndMidOpp	-0.1421 (0.1252)	-0.1407 (0.1254)	-0.105 (0.1675)	-0.1456 (0.1257)
lnTPIChange		0.0086 (0.0994)		
lnTPIChangeOpp		-0.0106 (0.0993)		
WkndMid:lnTPIChange		0.654 (1.9556)		
WkndMidOpp:lnTPIChangeOpp		0.7888 (1.7209)		
ChangedSquad			-0.0216 (0.0429)	
ChangedOpp			0.0228 (0.043)	
WkndMid:ChangedSquad			0.7729 (1.947)	
WkndMidOpp:ChangedSquad				
OneWeek				-0.0206 (0.0952)
TwoWeek				0.0455 (0.0864)
ThreeWeek				-0.0691 (0.087)
FourWeek				0.0442 (0.0529)
Observations	1829	1829	1829	1829
AIC	6952.4	6955.4	6956.2	6956.1

5.2 PROPENSITY SCORE MATCHING

We define the propensity score, which is the probability of a midweek match given the team characteristics. We choose logistic regression where we include variables that could be correlated with team performance. Table 15 presents a logistic regression estimating this probability. The *lnTeamXI* variable is positive and significantly implying that teams with higher squad value are more likely to have played a midweek game previously. Squad rotation variables such as *SharpProp, CohProp, ChangedSquad* are significant implying that teams are likely to change their players after a midweek game. The *MidweekOpp* and *ChangedOpp* variables are also significant suggesting that teams which have played a midweek game are often matched together on the weekend. Finally the *OneWeek* and *TwoWeek* variables are positive and significant confirming that midweek games are more likely to be played during weeks in which a team plays more games.

Overall, no reported variable means change drastically between the matched and unmatched samples although there is a net decrease in the overall bias for the matched sample variables. We can look to see how the matching has affected variables which were significant in the logistic regression. The bias has been reduced for *lnTeamXI* so matches can be found when considering team TPI value. By contrast variables such as *SharpProp* and *CohProp* show a small increase in bias suggesting that matching these variables is more difficult. This is because the exogenous variables are correlated with each other and so matching on one variable to reduce its bias may increase the bias on another correlated variable. The variables which are easier to match will primarily experience the reduction in bias which in this case is the team TPI value.

A 95% confidence interval on the propensity score is (0.293, 0.308) so the average probability for all teams to have played a midweek game is approximately 30%. Using Mahalanobis-metric pair matching with the propensity score metric, for teams who have played midweek games, their goal difference decreases by -0.001 on average with a 95% confidence interval of (-0.003, 0.005). Despite matching the exogenous variables there are no statistically significant effects of playing midweek games on team performance. Table 16 displays the success at matching the exogenous variables.

Figure 4 demonstrates further why this is not a great setting for matching. We find that the propensity score, or the likelihood for teams to play midweek games is very different between teams which do and do not midweek games in our dataset. There are almost no examples of teams which do not play midweek games that have a propensity score of 0.6 or above, demonstrating this imbalance. This reinforces the conclusion that it is not possible to match teams on most of our exogenous variables.

Table 15 - Logistic Regression of Midweek game probability

Midweek	Coef.	Standard Error	*t*-statistic
lnTeamXI	0.265	0.150	0.076
lnOppXI	-0.041	0.155	0.790
Adv	0.170	0.353	0.629
SharpProp	0.393	0.199	0.048
CohProp	0.569	0.286	0.047
lnDist	0.379	0.051	0.461
lnDistOpp	-0.019	0.052	0.722
lnDistLast	0.025	0.015	0.102
lnDistOppLast	0.020	0.015	0.180
MidweekOpp	2.429	0.086	0.000
ChangedSquad	0.146	0.023	0.000
ChangedOpp	-0.154	0.024	0.000
lnTPIChange	0.004	0.051	0.934
lnTIPChangeOpp	-0.059	0.053	0.271
OneWeek	2.826	0.092	0.000
TwoWeek	0.181	0.078	0.021
ThreeWeek	0.037	0.079	0.637
FourWeek	0.036	0.053	0.500

Table 16 - Matching of exogenous variables

	Unmatched Mean		Matched Mean		% bias		t-test Unmatched		t-test Matched	
Variable	Treated	Control	Treated	Control	Unmatched	Matched	t	p >|t|	t	p >|t|
lnTeamXI	17.927	17.706	17.927	17.985	29.5	-7.8	12.36	0.000	-2.72	0.007
lnOppXI	17.762	17.770	17.762	17.699	-1.1	8.6	-0.48	0.635	3.05	0.002
Adv	0.493	0.496	0.493	0.484	-0.6	1.9	-0.26	0.792	0.65	0.513
SharpProp	1.012	1.034	1.012	0.969	-8.3	16.6	-3.40	0.001	6.35	0.000
CohProp	1.027	1.006	1.027	1.058	11.5	-17.6	4.82	0.000	-5.59	0.000
lnDist	2.484	2.452	2.484	2.565	1.2	-3.2	0.52	0.604	-1.12	0.264
lnDistOpp	2.388	2.416	2.388	2.317	-1.1	2.8	-0.46	0.645	0.98	0.325
lnDistLast	2.616	2.379	2.616	2.420	8.8	7.2	3.72	0.000	2.56	0.011
lnDistLastOpp	2.556	2.389	2.556	2.467	6.3	3.4	2.64	0.008	1.19	0.233
MidweekOpp	0.595	0.159	0.595	0.203	100.7	90.5	44.46	0.000	30.69	0.000
ChangedSquad	2.715	1.886	2.715	2.780	41.8	-4.3	18.94	0.000	-1.34	0.180
ChangedOpp	2.273	2.078	2.273	1.755	10.5	27.7	4.50	0.000	9.85	0.000
lnTPIChanged	0.023	-0.010	0.023	0.013	4.7	1.7	1.92	0.055	0.63	0.531
lnTPIChangedOpp	0.013	0.146	0.013	0.245	-0.3	-32.2	-0.12	0.906	-9.16	0.000
OneWeek	1.650	0.752	1.650	1.952	169.7	-57.0	68.11	0.000	-15.80	0.000
TwoWeek	2.836	1.845	2.836	2.992	116.8	-18.3	47.54	0.000	-6.69	0.000
ThreeWeek	3.897	2.932	3879	4.215	85.4	-28.1	34.89	0.000	-10.15	0.000
FourWeek	4.994	4.013	4.994	5.356	69.3	-25.6	28.16	0.000	-9.75	0.000

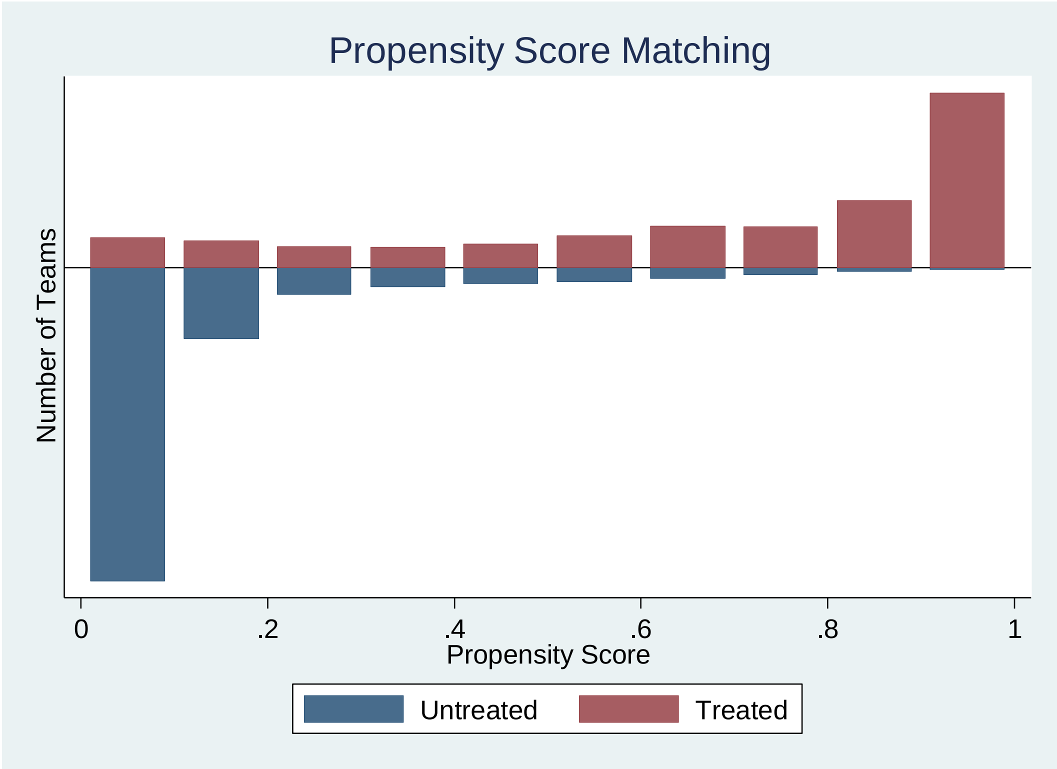

Figure 4 - Histogram of propensity scores

5.3 ANALYSIS OF ECONOMIC SIGNIFICANCE

In the previous section we did not find evidence of a statistically significant effect that a reduction in rest time reduces team performance in the context of English Premier League scheduling, but the coefficients were of the right sign and potentially economically significant. To examine this further, we hold all coefficients constant in the Relative Rest model and allow the total rest days for one team to fluctuate. Figure 5 displays how this fluctuation affects the predicted average goal difference for the team. We find that if this effect were to exist it would only increase the predicted average goal difference by 0.05, or about 1 goal in 20 games when changing from a least rested to fully rested team. This makes us confident that even if we were unable to capture the true effect of rest times in our model, the impact of this effect would be negligible.

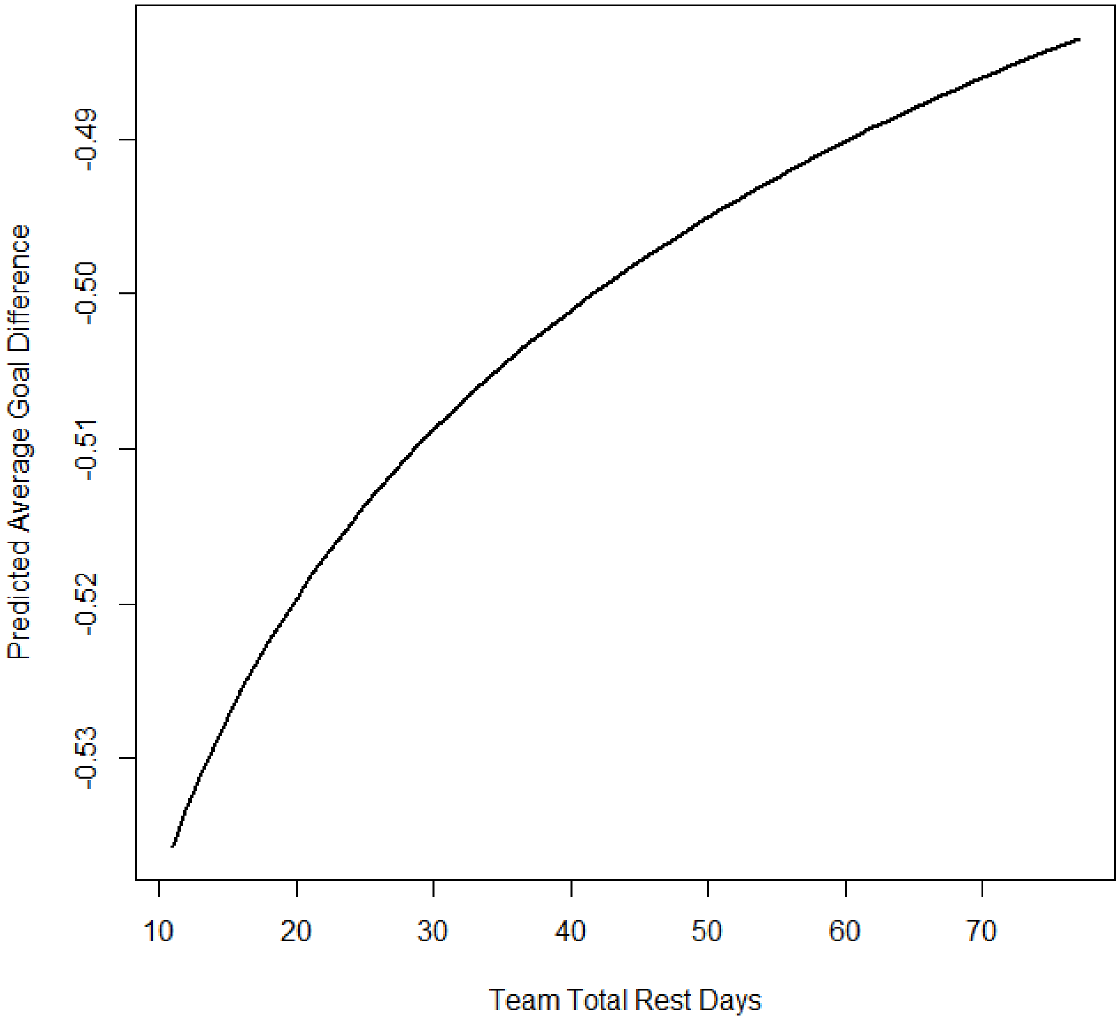

Figure 5 - Effect of rest days on goal difference

6. CONCLUSION

Players' physical condition is an essential factor towards a team's success hence well rested players should be able to contribute more towards overall team performance. We use data from the English Premier League to estimate the effects of rest times and the distances travelled between soccer matches, expanding on previous research that observed the relationship between aggregate player spending and success in soccer.

We can estimate the rest periods of players by calculating the number of days between games but we also want to take account of additional factors. Team sharpness (the number of consecutive games for players) and team cohesion (familiarity with teammates) are by-products of how much rest and rotation a team experiences. The effect of playing midweek games on weekend performance is also explored. The distance travelled for each game assesses the impact of team travel on performance.

We find that under current scheduling in domestic and European competition there are no statistically significant effects of receiving different days of rest on team performance within the English Premier League. A team with increased match sharpness experiences an improvement in results. We show that despite concern in the power of our tests, if a midweek effect did exist it would likely account for 1 goal in 20 games. For policy implications, scheduling is not the problem it is often made out to be by managers and the media. Manager complaints are driven by cognitive biases and convenience. If a team loses on the weekend after playing a midweek game a manager might complain that his players are tired but these excuses rarely appear if the team wins. Premier League teams win 39.6% and lose 34.7% of games played within three days of their last while they win 35.5% and lose 37.1% otherwise.

That is not to say that being tired does not matter. Players train all the time and it may well be that they are just as tired while training. It is often reported that players prefer playing games to training and so the chance to play another game instead of train is would be preferable. Extra games would simply change the type of exercise a player would perform on those given days but there is no statistically significant evidence to suggest that this change is to the detriment of team performance.

7. APPENDIX B

In Section 5 we conduct an econometric analysis of rest times on team performance while controlling for other variables that could determine team performance. Several specifications are considered using both a parametric linear model and an ordered logit model. For the linear model we use Goal Difference as the dependent variable. The variable Goal Difference is positive if Team A scores more than Team B and is negative if they score less. For the ordered logit model we use Game Result as the dependent variable. Soccer games can have one of three results: a home win, home loss or a draw.

Table 17 reports the results of the Relative Rest Model estimates using all Premier League games. Columns (1)-(3) contain OLS estimates of the linear model and columns (4)-(6) estimates with the ordered logit model. The team and opponent financial variables *lnTeamXI* and *lnOppXI* are both statistically significant, opposite and equivalent supporting previous literature on pay and performance. Home advantage is positive and not statistically significant. The distance variables *lnDist* and *lnDistOpp* are both statistically significant, opposite and equivalent. This again supports the literature that home advantage is tied into the distance that teams travel to play games. *SharpProp* is positive and statistically significant showing that teams with better more match sharpness perform better. *CohProp* is negative and not statistically significant. There is no evidence to support that teams that play together more often perform better but the coefficient is of the correct sign. *lnDistLast* and *lnDistLastOpp* are of the correct sign but statistically insignificant suggesting that only distance travelled for the previous game matters and not further lagged games. None of the *R1* to *R5* variables are significant but have varying signs for coefficients. This is also true when including the *H1* and *H2* variables. The *TRP* variable has a positive coefficient while the *TORP* variable has a negative coefficient. Both variables are statistically insignificant.

Table 18 reports the results of the Midweek Effect Model estimates using weekend Premier League games only. Columns (1)-(5) contain OLS estimates of the linear model and columns (6)-(10) estimates with the ordered logit model. The variables not associated with rest are identical in interpretation to those in Table 16. The coefficient of the *Midweek* variable is negative and statistically insignificant. By contrast the *MidweekOpp* variable has a positive coefficient. The Midweek variables corresponding to competition types are statistically insignificant. The *ChangedSquad* coefficient is negative and statistically insignificant while the *ChangedOpp* coefficient is positive. The *lnTPIChange* coefficient is negative and statistically insignificant while the *lnTPIChangeOpp* coefficient is positive. The *OneWeek* and *TwoWeek* coefficients are positive and statistically insignificant while the *ThreeWeek* and *FourWeek* coefficients are negative. All interaction coefficients are statistically insignificant.

Table 19 reports the results of the Weekend Effect Model estimates using midweek games only. Columns (1)-(4) contain OLS estimates of the linear model and columns (5)-(8) estimates with the ordered logit model. The variables not associated with rest are identical in interpretation to those in Table 16 and Table 17. The coefficient of the *WkndMid* variable is positive and statistically insignificant. By contrast the *WkndMidOpp* variable has a negative coefficient. The *ChangedSquad* coefficient is negative and statistically insignificant while the *ChangedOpp* coefficient is positive. The *lnTPIChange* coefficient is negative and statistically insignificant while the *lnTPIChangeOpp* coefficient is positive. The *OneWeek, TwoWeek, ThreeWeek* and *FourWeek* coefficients are statistically insignificant.

Table 17 - Relative rest periods on team performance

	(1)	(2)	(3)	(4)	(5)	(6)
lnTeamXI	0.2638 (0.0744)***	0.2567 (0.0745)***	0.2662 (0.0744)***	0.3379 (0.0951)***	0.3251 (0.0952)***	0.3234 (0.0679)***
lnOppXI	-0.2616 (0.0744)***	-0.2654 (0.0745)***	-0.2624 (0.0744)***	-0.3161 (0.0951)***	-0.3254 (0.0952)***	-0.3313 (0.0676)***
Adv	0.1208 (0.1689)	0.1225 (0.1689)	0.1236 (0.1689)	0.1068 (0.2134)	0.1078 (0.2134)	0.1124 (0.213)
SharpProp	0.1751 (0.0765)*	0.1815 (0.0772)*	0.1789 (0.0767)*	0.3284 (0.118)**	0.3285 (0.1176)**	0.327 (0.1171)**
CohProp	-0.212 (0.127)	-0.2075 (0.1273)	-0.2162 (0.127)	-0.2238 (0.1698)	-0.2256 (0.1697)	-0.235 (0.1685)
lnDist	-0.0733 (0.0247)**	-0.0724 (0.0247)**	-0.072 (0.0247)**	-0.0902 (0.0312)**	-0.0925 (0.0313)**	-0.0925 (0.0311)**
lnDistOpp	0.0715 (0.0247)**	0.0721 (0.0247)**	0.0722 (0.0247)**	0.0937 (0.0312)**	0.0915 (0.0312)**	0.0904 (0.0312)**
lnDistLast	-0.0004 (0.0073)	-0.0003 (0.0073)	-0.0006 (0.0073)	-0.0034 (0.0093)	-0.0027 (0.0093)	-0.0034 (0.0093)
lnDistLast-Opp	0.0004 (0.0073)	0.0002 (0.0073)	0.0001 (0.0073)	0.0031 (0.0093)	0.0035 (0.0093)	0.0031 (0.0093)
R1	-0.0054 (0.0628)			-0.0337 (0.0799)		
R2	0.0255 (0.0483)			0.0321 (0.0613)		
R4	-0.0267 (0.0485)			-0.0272 (0.0615)		
R5	0.006 (0.0633)			0.0344 (0.0805)		
H1R1		0.0011 (0.088)			-0.0512 (0.1127)	

H1R2		-0.0172 (0.0603)			-0.021 (0.0763)	
H1R4		0.0231 (0.0605)			0.0148 (0.0765)	
H1R5		0.0005 (0.089)			0.0564 (0.1142)	
H2R1		-0.0059 (0.0804)			-0.0136 (0.1019)	
H2R2		0.081 (0.0646)			0.0881 (0.0822)	
H2R4		-0.08 (0.0649)			-0.0835 (0.0826)	
H2R5		0.0153 (0.0809)			0.02 (0.1024)	
TRP			0.0004 (0.0013)			0.0013 (0.0017)
TORP			-0.0004 (0.0013)			-0.0013 (0.0017)
Observati-ons	8226	8226	8226	8226	8226	8226
AIC	31013.2	31016.5	31011.1	16578.8	16583.8	16576.2

Table 18 - Midweek game effect on weekend team performance

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
lnTeamXI	0.2996 (0.0874)***	0.2792 (0.0859)***	0.2807 (0.0858)**	0.2775 (0.0858)**	0.2805 (0.0859)***
lnOppXI	-0.3002 (0.0875)***	-0.2765 (0.0858)***	-0.2767 (0.086)**	-0.2829 (0.0858)***	-0.2811 (0.0858)***
Adv	0.2136 (0.1991)	0.2069 (0.1991)	0.2128 (0.1991)	0.2122 (0.199)	0.2092 (0.199)

SharpProp	0.3291 (0.1009)**	0.2964 (0.1038)***	0.3368 (0.1012)***	0.3223 (0.1011)**	0.3263 (0.101)**
CohProp	-0.0223 (0.152)	0.0175 (0.1554)	-0.0281 (0.1521)	-0.0331 (0.152)	-0.0238 (0.152)
lnDist	-0.0643 (0.0291)*	-0.0655 (0.029)*	-0.0648 (0.0291)*	-0.0666 (0.0291)*	-0.0648 (0.0291)*
lnDistOpp	0.0646 (0.0291)*	0.064 (0.0292)*	0.0641 (0.0291)*	0.0623 (0.0291)*	0.0646 (0.0291)*
lnDistLast	0.0005 (0.0083)	0.0012 (0.0083)	0 (0.0083)	0.0007 (0.0083)	-0.0002 (0.0083)
lnDistLastOpp	-0.0004 (0.0083)	-0.0003 (0.0083)	0.0003 (0.0083)	0.0003 (0.0083)	-0.0008 (0.0083)
Midweek	-0.0181 (0.0466)	-0.0237 (0.0704)			-0.0172 (0.0466)
MidweekOpp	0.0143 (0.0466)	0.0227 (0.0705)			0.0171 (0.0466)
MidEuro			0.1243 (0.101)		
MidCup			0.0612 (0.0752)		
MidLeague			-0.0549 (0.0692)		
OppEuro			-0.1197 (0.1009)		
OppCup			-0.0626 (0.0752)		
OppLeague			0.0545 (0.0692)		
ChangedSquad		-0.0186 (0.0197)			
ChangedOpp		0.0162 (0.0199)			
Midweek:ChangedSquad		1.2308 (1.0838)			

MidweekOpp:Changed-Opp		0.6608 (0.8744)			
lnTPIChange	-0.0306 (0.0363)				
lnTPIChangeOpp	0.0294 (0.0364)				
Midweek:lnTPIChange	1.216 (1.0787)				
MidweekOpp:lnTPICha-ngeOpp	0.5993 (0.8736)				
OneWeek				0.0035 (0.0419)	
TwoWeek				0.0108 (0.0434)	
ThreeWeek				-0.0107 (0.0436)	
FourWeek				-0.0009 (0.0287)	
Observations	6397	6397	6397	6397	6397
AIC	24380.1	24379.5	24378.2	24376.0	24375.7

	(6)	**(7)**	**(8)**	**(9)**	**(10)**
lnTeamXI	0.4037 (0.1108)***	0.374 (0.1086)***	0.3708 (0.1085)***	0.3776 (0.1085)***	0.3808 (0.1086)***
lnOppXI	-0.3938 (0.1107)***	-0.3698 (0.1085)***	-0.3808 (0.1087)***	-0.3774 (0.1085)***	-0.3726 (0.1084)***
Adv	0.3025 (0.2512)	0.2979 (0.2513)	0.3002 (0.2513)	0.2996 (0.2511)	0.3016 (0.2512)
SharpProp	0.4211 (0.1355)**	0.373 (0.1381)**	0.4297 (0.137)**	0.4033 (0.1347)**	0.4088 (0.1349)**

CohProp	-0.1421 (0.1941)	-0.1205 (0.1973)	-0.1568 (0.1947)	-0.1718 (0.1937)	-0.1583 (0.1937)
lnDist	-0.0747 (0.0366)*	-0.073 (0.0367)*	-0.0756 (0.0367)*	-0.0741 (0.0367)*	-0.0726 (0.0367)*
lnDistOpp	0.0738 (0.0367)*	0.0752 (0.0367)*	0.0734 (0.0367)*	0.0745 (0.0367)*	0.0756 (0.0367)*
lnDistLast	-0.0026 (0.0104)	-0.0011 (0.0105)	-0.003 (0.0105)	-0.0018 (0.0105)	-0.0022 (0.0104)
lnDistLastOpp	0.0009 (0.0104)	0.0009 (0.0105)	0.0022 (0.0105)	0.0023 (0.0105)	0.0014 (0.0104)
Midweek	-0.0507 (0.0587)	-0.0311 (0.0881)			-0.0545 (0.0587)
MidweekOpp	0.0546 (0.0587)	0.023 (0.0882)			0.0544 (0.0587)
MidEuro			0.1593 (0.1314)		
MidCup			0.012 (0.0949)		
MidLeague			-0.0804 (0.087)		
OppEuro			-0.1534 (0.1315)		
OppCup			-0.0152 (0.0948)		
OppLeague			0.0769 (0.0869)		
ChangedSquad		-0.0149 (0.0248)			
ChangedOpp		0.0111 (0.025)			
Midweek:ChangedSq-uad		-0.0049 (0.0302)			
MidweekOpp:Changed Opp		0.0088 (0.0303)			

lnTPIChange	-0.0336 (0.0455)				
lnTPIChangeOpp	0.0324 (0.0455)				
Midweek:lnTPIChange	-0.0144 (0.0745)				
MidweekOpp:lnTPIC-hangeOpp	0.0126 (0.0745)				
OneWeek				-0.0061 (0.0525)	
TwoWeek				0.0159 (0.0542)	
ThreeWeek				-0.0158 (0.0549)	
FourWeek				0.0018 (0.036)	
Observations	6397	6397	6397	6397	6397
AIC	12994.2	12996.7	12996.4	12994.5	12990.8

Table 19 - Weekend game effect on midweek team performance

	(1)	**(2)**	**(3)**	**(4)**
lnTeamXI	0.1462 (0.2322)	0.1003 (0.2372)	0.1397 (0.233)	0.1499 (0.2323)
lnOppXI	-0.1362 (0.2324)	-0.1287 (0.2374)	-0.1267 (0.2328)	-0.1273 (0.2326)
Adv	-0.5103 (0.4716)	-0.4884 (0.4728)	-0.5031 (0.4721)	-0.5185 (0.4714)
SharpProp	0.1777 (0.2569)	0.164 (0.2572)	0.1236 (0.2671)	0.1725 (0.2594)
CohProp	-0.5389 (0.4121)	-0.569 (0.4126)	-0.4611 (0.4282)	-0.5565 (0.4145)
lnDist	-0.1663 (0.0715)*	-0.1601 (0.0716)*	-0.1625 (0.0714)*	-0.1646 (0.0713)*

lnDistOpp	0.1652 (0.0713)*	0.1664 (0.0716)*	0.1662 (0.0716)*	0.1681 (0.0711)*
lnDistLast	-0.0098 (0.0251)	-0.0097 (0.0251)	-0.0095 (0.0251)	-0.009 (0.0251)
lnDistLastOpp	0.0076 (0.0251)	0.0084 (0.0251)	0.0065 (0.0251)	0.0085 (0.0251)
WkndMid	0.1437 (0.1741)	0.1498 (0.1745)	0.1068 (0.2346)	
WkndMidOpp	-0.1356 (0.174)	-0.1483 (0.1745)	-0.0773 (0.2347)	
lnTPIChange		-0.0287 (0.1376)		
lnTPIChangeOpp		0.0325 (0.1373)		
WkndMid:lnTPIChange		0.0846 (0.1625)		
WkndMidOpp:lnTPIChangeOpp		-0.1008 (0.1625)		
ChangedSquad			-0.0341 (0.0613)	
ChangedOpp			0.0431 (0.0614)	
WkndMid:ChangedSquad			0.0183 (0.0718)	
WkndMidOpp:ChangedSquad			-0.0296 (0.0719)	
OneWeek				0.0143 (0.1131)
TwoWeek				-0.0019 (0.1188)
ThreeWeek				-0.0262 (0.1206)
FourWeek				0.0307 (0.0732)
Observations	1829	1829	1829	1829

AIC	3932.7	3936.7	3939.1	3935.5

	(5)	**(6)**	**(7)**	**(8)**
lnTeamXI	0.2119 (0.1641)	0.183 (0.1679)	0.2106 (0.1646)	0.1922 (0.1647)
lnOppXI	-0.2036 (0.1643)	-0.1831 (0.1676)	-0.2025 (0.1644)	-0.2184 (0.1643)
Adv	-0.1745 (0.3455)	-0.155 (0.3462)	-0.1701 (0.346)	-0.1657 (0.3461)
SharpProp	0.0304 (0.1574)	0.0367 (0.1494)	0.0088 (0.1654)	0.0245 (0.1547)
CohProp	-0.607 (0.2842)*	-0.5903 (0.2812)*	-0.5688 (0.2966)	-0.6139 (0.2834)*
lnDist	-0.1008 (0.0517)	-0.0973 (0.0518)	-0.0982 (0.0517)	-0.0996 (0.0517)
lnDistOpp	0.0993 (0.0517)	0.0994 (0.0518)	0.1009 (0.0519)	0.0982 (0.0519)
lnDistLast	-0.0048 (0.0179)	-0.0041 (0.018)	-0.0044 (0.018)	-0.0036 (0.018)
lnDistLastOpp	0.0034 (0.0179)	0.0049 (0.018)	0.0036 (0.018)	0.004 (0.018)
WkndMid	0.1408 (0.1252)	0.1415 (0.1254)	0.1069 (0.1673)	0.1348 (0.1428)
WkndMidOpp	-0.1421 (0.1252)	-0.1407 (0.1254)	-0.105 (0.1675)	-0.1456 (0.1257)
lnTPIChange		0.0086 (0.0994)		
lnTPIChangeOpp		-0.0106 (0.0993)		
WkndMid:lnTPIChange		0.654 (1.9556)		

WkndMidOpp:lnTPIChangeOpp		0.7888 (1.7209)		
ChangedSquad			-0.0216 (0.0429)	
ChangedOpp			0.0228 (0.043)	
WkndMid:ChangedSquad			0.7729 (1.947)	
WkndMidOpp:ChangedSquad				
OneWeek				-0.0206 (0.0952)
TwoWeek				0.0455 (0.0864)
ThreeWeek				-0.0691 (0.087)
FourWeek				0.0442 (0.0529)
Observations	1829	1829	1829	1829
AIC	6952.4	6955.4	6956.2	6956.1

8. REFERENCES

[1] Ashman, T., Bowman, R. A., & Lambrinos, J. (2010). The role of fatigue in NBA wagering markets: The surprising ‘‘Home disadvantage situation’’. *Journal of Sports Economics, 11*(6), 602-613.

[2] Bangsbo, J. (1994). The physiology of soccer--with special reference to intense intermittent exercise. *Acta Physiologica Scandinavica.Supplementum, 619,* 1-155.

[3] Bangsbo, J., Norregaard, L., & Thorso, F. (1991). Activity profile of competition soccer. *Canadian Journal of Sport Sciences = Journal Canadien Des Sciences Du Sport, 16*(2), 110-116.

[4] Carmichael, F., & Thomas, D. (2005). Home-field effect and team performance: Evidence from English premiership football. *Journal of Sports Economics, 6*(3), 264-281.

[5] Entine, O. A., & Small, D. S. (2008). The role of rest in the NBA home-court advantage. *Journal of Quantitative Analysis in Sports, 4*(2), 1-9.

[6] Forrest, D., & Simmons, R. (2002). Team salaries and playing success in sports: A comparative perspective. *Sportökonomie* (pp. 221-238) Springer.

[7] Krumer, A., & Lechner, M. (2016). *Midweek Effect on Performance: Evidence from the German Soccer Bundesliga,*

[8] Mohr, M., Krustrup, P., & Bangsbo, J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences, 21*(7), 519-528.

[9] Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport. *Sports Medicine, 28*(4), 221-236.

[10] Nichols, M. W. (2014). The impact of visiting team travel on game outcome and biases in NFL betting markets. *Journal of Sports Economics, 15*(1), 78-96.

[11] Oberhofer, H., Philippovich, T., & Winner, H. (2010). Distance matters in away games: Evidence from the German football league. *Journal of Economic Psychology, 31*(2), 200-211.

[12] Reilly, T. (1997). Energetics of high-intensity exercise (soccer) with particular reference to fatigue. *Journal of Sports Sciences, 15*(3), 257-263.

[13] Reilly, T., & Thomas, V. (1976). A motion analysis of work-rate in different positional roles in professional football match-play. *Journal of Human Movement Studies, 2*(2), 87-97.

[14] Riley, Graeme. Football in Europe. Tony Brown, Various years.

[15] Rubin, D. B. (1976). Multivariate matching methods that are equal percent bias reducing, II: Maximums on bias reduction for fixed sample sizes. *Biometrics,* 121-132.

[16] Rubin, D. B. (1980). Bias reduction using mahalanobis-metric matching. *Biometrics,* 293-298.

[17] Scoppa, V. (2015). Fatigue and team performance in soccer: Evidence from the FIFA world cup and the UEFA European championship. *Journal of Sports Economics, 16*(5), 482-507.

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

[19] Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature, 41*(4), 1137-1187.

[20] Szymanski, S., & Smith, R. (2010). The English football industry: Profit, performance and industrial structure. *Football economics and policy* (pp. 1-26) Springer.

[21] Van Gool, D., Van Gerven, D., & Boutmans, J. (1988). The physiological load imposed on soccer players during real match-play. *Science and Football, 1,* 51-59.

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 in the betting market as most models almost as well as 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 al. (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, Kramarz, 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.[5]

- 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*[6] package in R. In order to run a regression using a sparse matrix in R the *glmnet*[7] 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

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

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 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 on one team. The players who do move should be able to obtain accurate estimates of their ability.

Table 20 - 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 21 - Player movement summary

Count	Players transferring to another team[8]	Players appeared on unique teams[9]	Players appeared in unique competition[10]	Years in the data[11]
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

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

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

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.

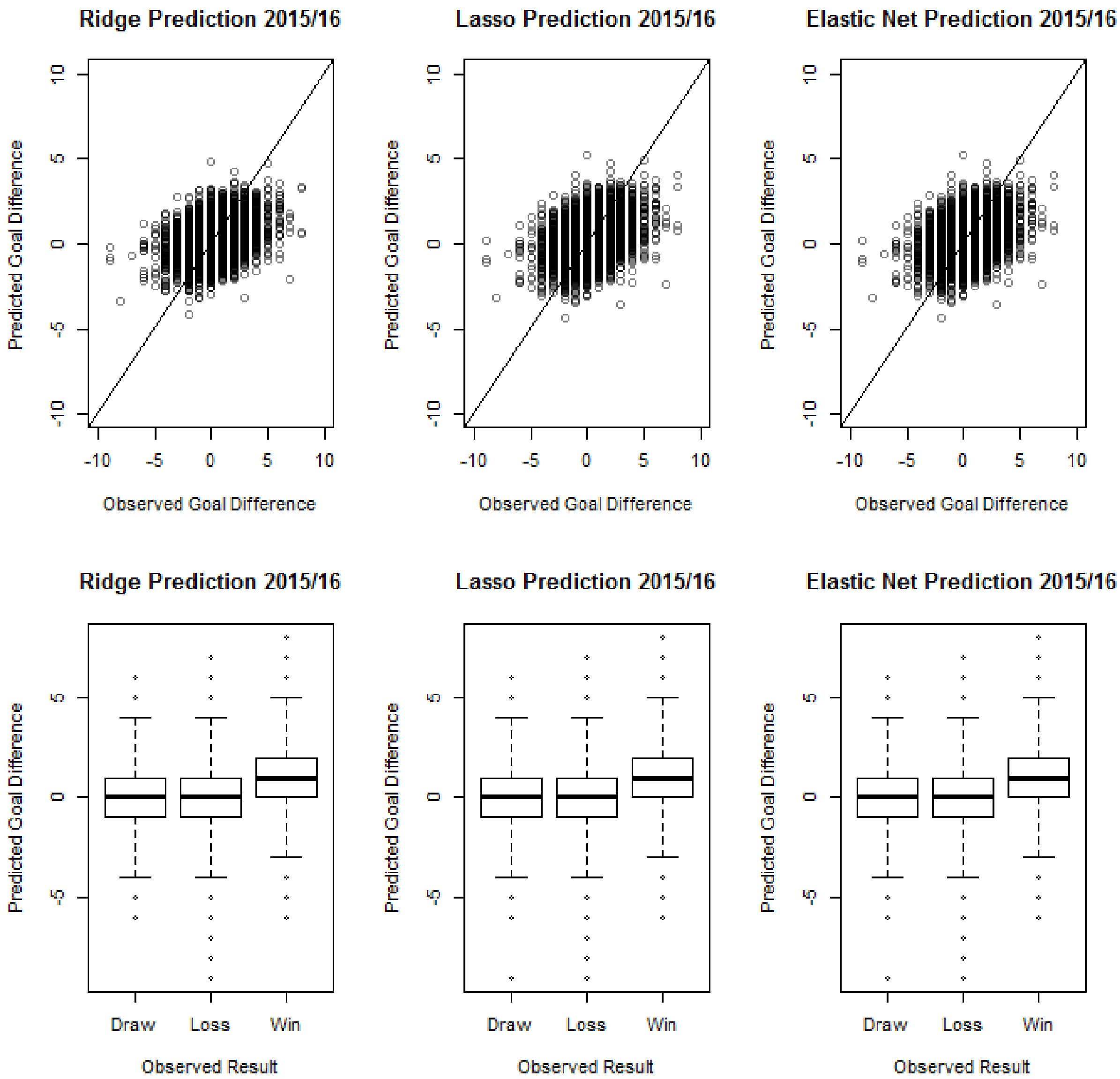

Figure 6 - Observed Goal Difference/Result vs. Predicted Goal Difference

Table 22 - 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 23 - 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 24 - 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 25 - 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[12]. We find that our models on average perform almost as well as the betting firms with mostly small losses occurring. In some leagues, 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 26 - 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%

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

Table 27 - 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 28 - 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 a player's teammates, team, league, manager and team 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, since the variable selection and penalty terms penalize 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 29 - 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 30 - 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 31 - 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.

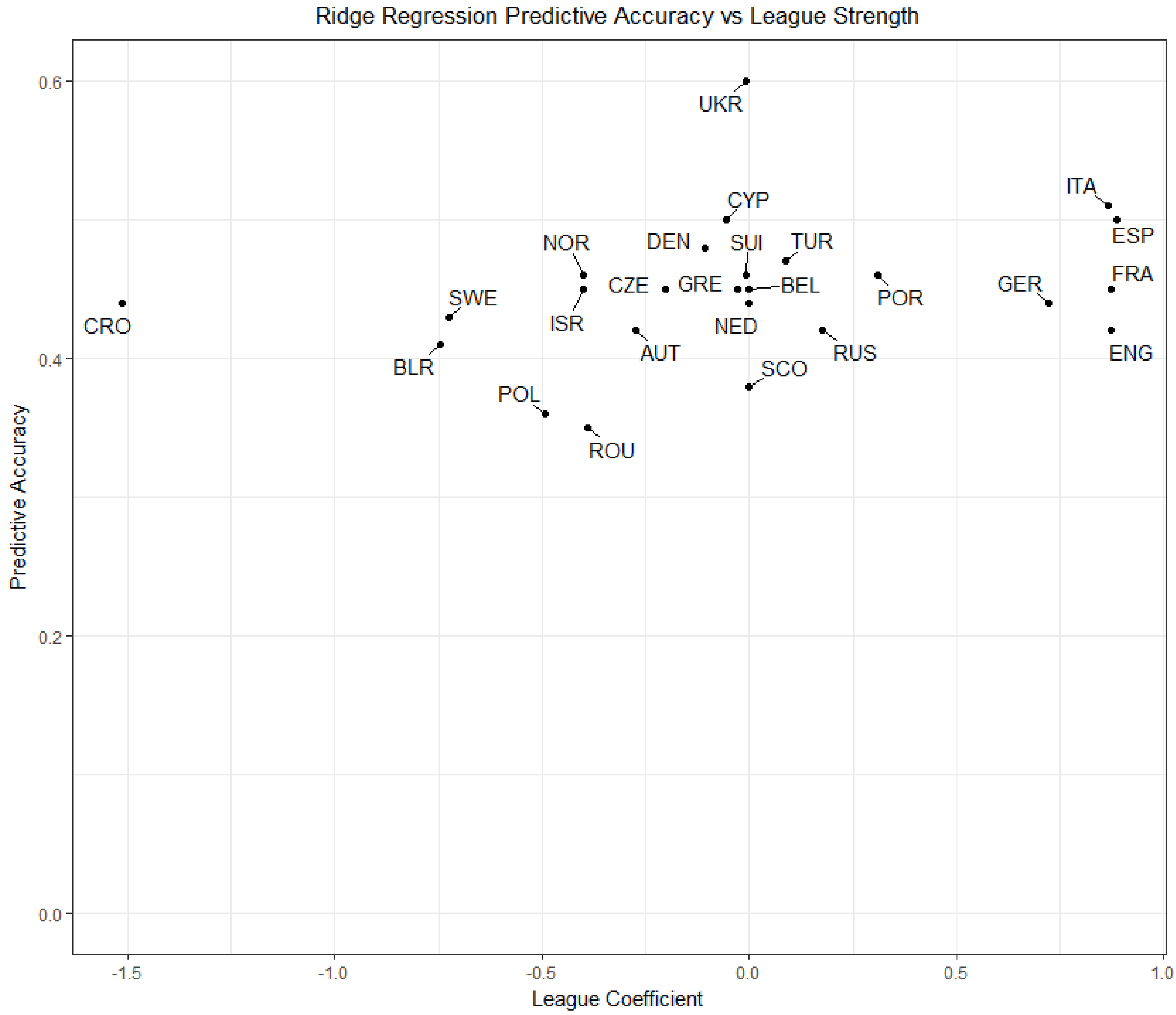

Figure 7 - Ridge Regression

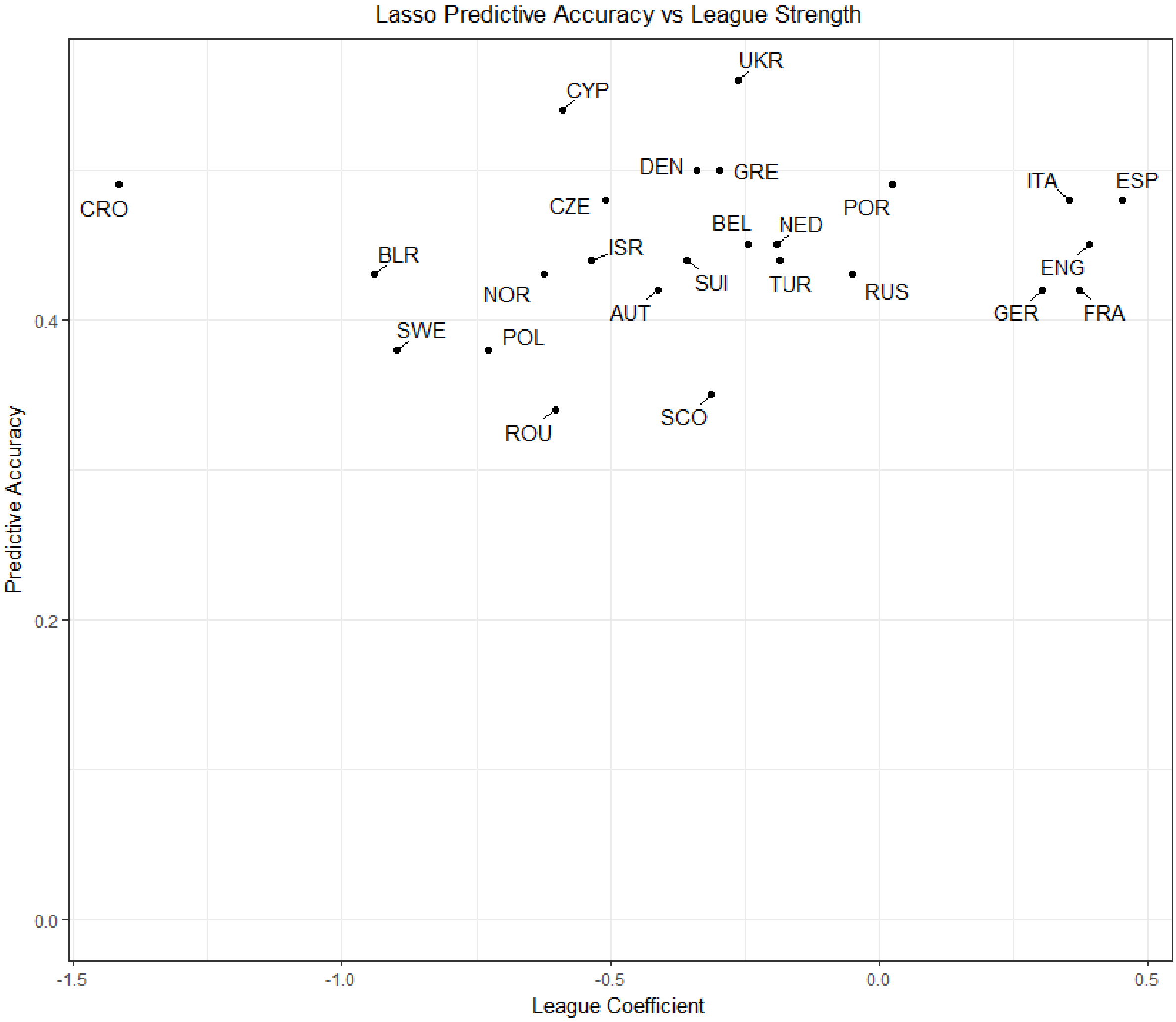

Figure 8 - Lasso

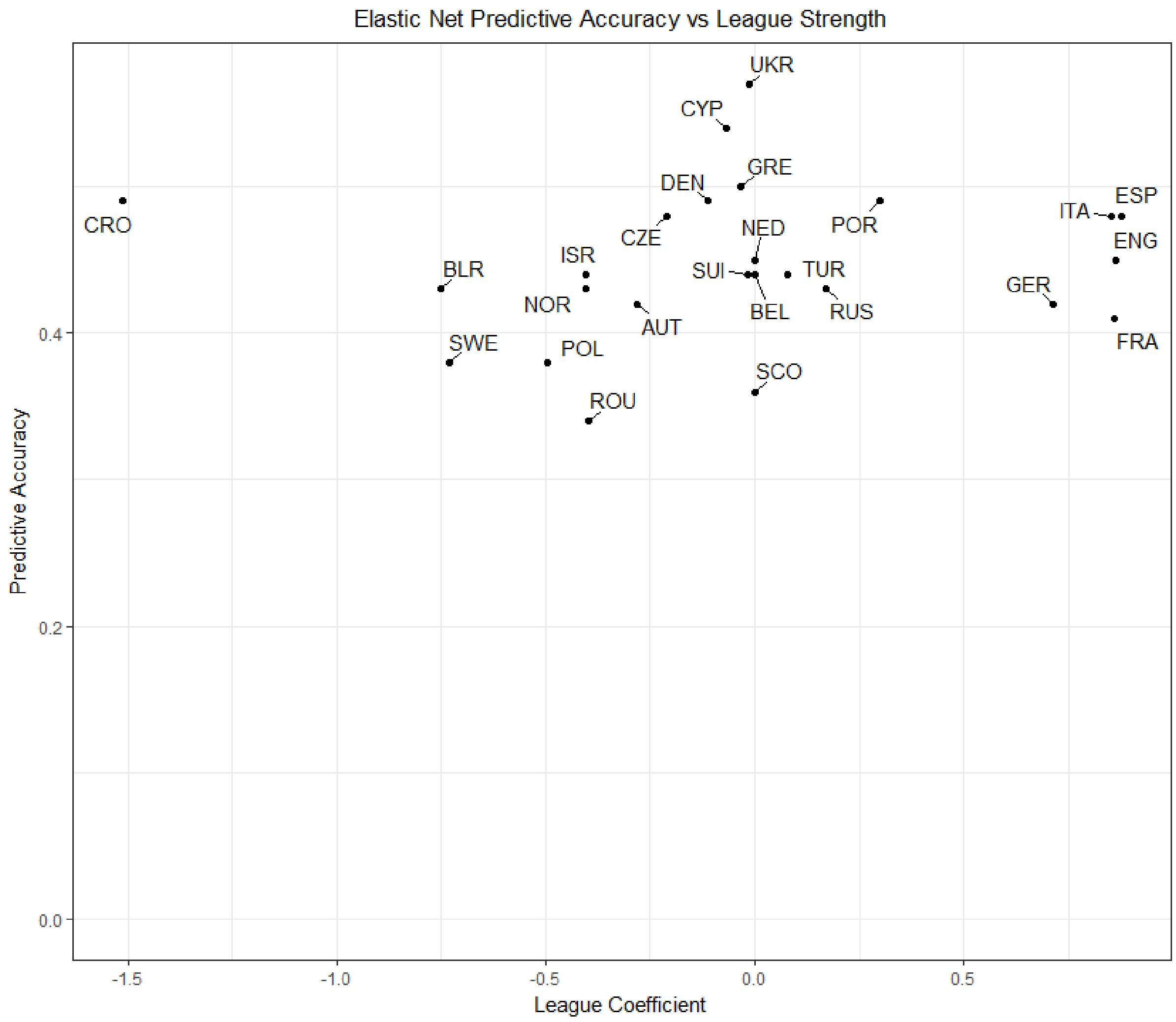

Figure 9 - Elastic Net

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 32 - 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 33 - 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 34 - 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 35 - 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 36 - 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 37 - 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.

CHAPTER V

THESIS CONCLUSION

1. CONCLUSION

This thesis examined production functions in sport using the setting of contest theory literature. It focuses on the inputs of the production function, namely player productivity. The setting for the analyses were in European soccer with a focus on the English Premier League. This research is important to soccer teams (and all sport organizations) where the top teams are multi-billion dollar corporations and players are multi-million dollar assets. Better performance by players will allow teams to experience increased financial success and global prestige. Beyond that, the methods and techniques developed in this thesis can be used generally in the fields of economics and management by substituting performance variables and output appropriately. Sports are almost unique in that unlike other markets workers can be observed on a regular basis. Productivity is easily observed as each week players can be seen playing and their contributions to games are observed.

Previous research has almost exclusively considered at performance on the aggregate level, looking at teams' inputs and relating them to league performance, often at the season level. Originating with Rottenberg (1956) and Scully (1974), many techniques and approaches have been used to model the process relating to inputs to outputs in a variety of sports. This was first adopted into a contest success function by El-Hodiri and Quirk (1971) and Quirk and El-Hodiri (1974). Szymanski and Smith (1997) would eventually substitute money for talent in the contest success function show it as a driving factor behind team results. This thesis expanded on the literature with a focus in breaking down the contest success function from an aggregate level. Finances and other input variables are therefore able to be considered on a player level by using high dimensional regression techniques. The main body of the thesis consists of three papers presented in Chapters 2-4, each looking at different inputs at the player level and how the relate to team performance.

Chapter 2 presented the first paper: “Testing the O-Ring theory using data from the English Premier League”. This paper measures the impact of different workers in a production process depending on their expected productivity. The setting is the English Premier league where expected productivity is measured from the transfer fees used to acquire players. The findings show that the most expensive players tend to exert the largest impact on games whereas the least expensive players have relatively little impact. This is consistent with superstar theories rather than O-Ring theory. The optimal spending distribution is found to be more skewed than the observed distribution suggesting a constraint in the market for players. This paper adds to the literature labour market theory, particularly in those with very few high ability workers. It shows in particular that you should aim to improve on your higher ability workers rather than spend the money to make modest improvements on your lower ability workers.

Chapter 3 presented the second paper: “The Impacts of Rest Periods on results in the English Premier League”. This paper looks at the effects of fatigue in professional soccer using data again from the English Premier League. Many managers and players share conflicting views on how much rest is required between games but there is little empirical evidence to support their arguments. The production function is expanded to include information on rest times and distance travelled. Under current scheduling in domestic and European competition there are no statistically significant effects of receiving different days of rest on team performance. The limited variation in the amount of rest for teams can give concern about the power of the tests used but even if the effects were statistically significant they are found to have a negligible impact on team results. This paper while expanding the literature on contest success theory in soccer also targets a popular argument in the professional game. Managers frequently complain about the amount of rest time their teams experience due to playing more European or domestic cup games. There is a cognitive bias that when results go poorly after these games that the managers can use the extra game as an excuse. The paper results show that such an effect does not exist and that teams are generally unaffected by the fatigue additional games bring during a season.

Chapter 4 presents the final paper in this thesis: "Individual player contributions in European Soccer". This paper applies new techniques to predict match outcomes in professional soccer by estimating player contributions. Using data from the top 25 European soccer leagues, the individual contributions of players is measured using high dimensional fixed effects models. Nine years of data is used to train the model while a further year is used to check for predictive accuracy. The findings show an average prediction rate of 45% with all methods producing similar performance. The model highlights the most productive players but there is some bias towards identifying players who produce and prevent goals directly. This results in attackers and defenders being ranked more highly than midfield players. There is some potential for the models to be used in sports betting as they predict almost as accurately as betting firms. This paper is more ambitious in scope in that it directly tries to measure the abilities of a very large number of players. It uses high dimensional modelling techniques not often found in the literature in order to rank players, managers, teams and leagues. The results suggest potential for such techniques to be used by teams for future player scouting in a variety of sports.

This thesis set out to add an applied contribution to production function and contest success theory. It contributes by breaking down inputs in the production function at the worker level and by adopting some high dimensional regression techniques rarely applied to sport data. This empirical research can be replicated not just in sports but in the fields of economics, management and strategy.

Future research can extend these results to better understand production functions in sports, focussing on the player level, and adopting a variety of advanced high dimensional techniques. Paper 1 could be extended by adopting some budget constraints which were not considered in the model, tightening up the optimization on player spending. This in particular refers to being unable to allocate players the reserve value in reality. Other approaches to play matching could be considered, such as taking into account the player position. Paper 2 covered a lot of ground when looking at the different ways to consider rest. This could be expanded by looking across other leagues in Europe, particularly when there were cases for such an effect existing in Germany found in the literature. Paper 3 has the potential for a lot of future work. We are very much in the early stages of such large amounts of data being modelled in soccer. In the future entire careers of players could be captured as opposed to just a ten year snapshot. On a smaller scale of just considering individual leagues, particular player attributes could be included into the model to see what drives some of the fixed effect player contributions. Predictive models could also be improved by adding in form functions and having updated rolling coefficients. On the whole, these suggestions would be a great asset in understanding the motivations and productivity of individual workers in a production process.

2. REFERENCES

[1] El-Hodiri, M., & Quirk, J. (1971). An economic model of a professional sports league. *Journal of Political Economy, 79*(6), 1302-1319.

[2] Quirk, J. P., & El-Hodiri, M. A. (1973). *The economic theory of a professional sports league*

[3] Rottenberg, S. (1956). The baseball players' labor market. *Journal of Political Economy, 64*(3), 242-258.

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

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