Investigating the level of interdependency between the performance(s) of direct opponent(s) in professional football: a study on teams, positional units and individual players competing in the German Bundesliga

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/21109](https://dspace.lboro.ac.uk/2134/21109)

Publisher: © Mikael Jamil

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
Investigating the Level of Interdependency Between the Performance(s) of Direct Opponent(s) in Professional Football: A Study on Teams, Positional Units and Individual Players Competing in the German Bundesliga

By
Mikael Jamil

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

December 2015

©Mikael Jamil, 2015
Table of Contents

List Of Figures ................................................................................................................................... v
List Of Tables ...................................................................................................................................... vii
Acknowledgements ............................................................................................................................ ix
Abstract ............................................................................................................................................... x

Chapter One - Introduction ................................................................................................................ 1
  1.0 Overview .................................................................................................................................... 1
  1.1 Aims and Objectives .................................................................................................................. 2
  1.2 Why Football? ........................................................................................................................... 4
  1.3 Why the German Bundesliga? ................................................................................................... 12
  1.4 Contributions ............................................................................................................................ 14
  1.5 Thesis Structure ......................................................................................................................... 16

Chapter Two - Production in Economics ........................................................................................ 18
  2.1 Introduction ............................................................................................................................... 18
  2.2 Alternative Market Structures ................................................................................................. 19
    2.2.1 Perfect Competition ............................................................................................................ 19
    2.2.2 The Isoquant/Isocont Approach to Production .................................................................... 24
  2.3 Monopoly, Externalities and Market Failure ............................................................................. 29
    2.3.1 Monopoly ........................................................................................................................... 29
    2.3.2 Oligopoly ............................................................................................................................ 31
    2.3.3 Externalities ......................................................................................................................... 32
  2.4 Wage Determination and Monopsony ...................................................................................... 37
  2.5 Individual and Team Productivity ............................................................................................ 41
    2.5.1 Individual Production Function .......................................................................................... 41
    2.5.2 Human Capital ..................................................................................................................... 42
    2.5.3 Moral Hazard and Incentive Contracts ............................................................................. 45
    2.5.4 Contract Design and Incentive Contracts ........................................................................... 46
    2.5.5 Tournament Contracts ....................................................................................................... 48
    2.5.6 Effort and Efficiency Wages ............................................................................................... 51
    2.5.7 Team Production ................................................................................................................ 53
    2.5.8 Summary ............................................................................................................................ 56
  2.6 Economics of Professional Sport ............................................................................................... 57
    2.6.1 The Role of Sports Clubs .................................................................................................... 59
    2.6.2 Sports Competitions as Economic Contests ...................................................................... 61
    2.6.3 Peculiar Economics of Sport .............................................................................................. 63
    2.6.4 Market Structure in Sport .................................................................................................. 64
    2.6.5 Cross Subsidisation in Professional Team Sports Leagues .............................................. 69
    2.6.6 Key Concepts .................................................................................................................... 72
  2.7 Conclusion ................................................................................................................................ 74

Chapter Three - Production in Team Sports ................................................................................... 76
  3.1 Introduction ............................................................................................................................... 76
  3.2 Tournament Theory and Contest Design .................................................................................. 78
  3.3 Uncertainty of Outcome .......................................................................................................... 82
    3.3.1 Uncertainty of Outcome in North American Sports .......................................................... 82
    3.3.2 Uncertainty of Outcome in European Football ............................................................... 85
6.5. Results: Second Stage Results ................................................................. 216
6.6 Discussion ............................................................................................ 223
6.7 Conclusion ........................................................................................... 228

Chapter Seven - Investigating Interdependency Between the Performances of Positional Units and their Opposing Team(s) in the German Bundesliga

7.1 Introduction ............................................................................................ 230
7.2 Data and Variables ................................................................................ 232
7.3 Analysis Strategy .................................................................................. 241
  7.3.1. Impact of the away team's attacking unit upon the collective performance of the home team ................................................................. 241
  7.3.2. Impact of the away team's defending unit upon the collective performance of the home team .................................................................. 243
  7.3.3. Impact of the home team's attacking unit upon the collective performance of the away team ................................................................. 245
  7.3.4. Impact of the home team's defensive unit upon the collective performance of the away team ................................................................. 246
7.4. First Stage Regression Results: Impact of Positional Units Upon the Performance of the Opposing Team ................................................................. 247
7.5. Second Stage Regression Results: Individual Impact of Team Units Upon the Performance of the Opposing Team ................................................................. 255
7.6. OLS Estimation Results: Individual Impact of Positional Units Upon the Collective Performance of the Opposing Team ................................................................. 263
7.7 Discussion ............................................................................................ 267
7.8 Conclusion ........................................................................................... 269

Chapter Eight - Conclusion ....................................................................... 271
8.1 Introduction ............................................................................................ 271
8.2 Review of the Thesis ............................................................................ 272
8.3 Research Study Contributions, Limitations and Recommendations for Future Research ................................................................................................. 275
8.4 Concluding Remark ............................................................................ 278

Bibliography ............................................................................................. 279

APPENDIX ONE .......................................................................................... 307
APPENDIX TWO .......................................................................................... 311
APPENDIX THREE ........................................................................................ 315
List Of Figures

Figure 1.1a Revenue Thresholds for Positions in the Deloitte Money League .......................................................... 6
Figure 1.1b – Proportional breakdown of Revenues Received For the Top Twenty Highest Earning Football Clubs in Season 2004/05................................. 8
Figure 1.1c – Proportional Breakdown of Revenues Received For the Top Twenty Highest Earning Football Clubs in Season 2013/14......................... 8
Figure 2.a - Industry supply and demand in perfect competition................................................................. 20
Figure 2.b - Costs and Revenues of an individual firm in perfect competition.................................................... 21
Figure 2.c – An Isoquant............................................................... 24
Figure 2.d - An Isoquant Map......................................................... 25
Figure 2.e - A Graphical Representation of an Isocost................................................................. 26
Figure 2.f - The least cost method of production.................................................. 27
Figure 2.g - Maximising output for a given total cost....................................................................................... 28
Figure 2.h - Monopoly supply of Output....................................................................................... 30
Figure 2.i - Monopoly Vs. Perfect Competition....................................................................................... 31
Figure 2.j - External Costs in production....................................................................................... 33
Figure 2.k - External Benefit in production....................................................................................... 34
Figure 2.l - External cost in consumption....................................................................................... 35
Figure 2.m - External benefit in consumption....................................................................................... 35
Figure 2.n - Wages and employment under monopsony and perfect competition....................................................................................... 39
Figure 2.p - Long Term Incentives....................................................................................... 47
Figure 6.2.1 - Team X and Team Z Lining up Against Each Other in a Typical 4-4-2 Team Formation....................................................................................... 196
Figure 6.2.2 - Typical 4-4-2 VS 4-4-2 Team Formation Match
Up........................................................................................................197

Figure 7.2.1 – Positional unit match-ups in fixture n.............................234

Figure 7.2.2 – Positional unit match-ups in fixture n+1..........................235

Figure 7.2.3 – Positional unit match-ups in fixture n+2.........................236
List Of Tables

Table 5.2.1: Table listing all teams represented throughout the four season sample period and how often each team was observed..........................158/159
Table 5.2.2: List of all variables used throughout the analysis and their definitions.................................................................160/161
Table 5.4.1: First Stage Regression Estimations Obtained When Assessing the Impact of the Performance of the Away Team on the Performance of the Home Team..............................................................172
Table 5.4.2: First Stage Regression Estimations Obtained When Assessing the Impact of the Performance of the Home Team on the Performance of the Away Team..............................................................177
Table 5.5.1: Second Stage Regression Results.................................................183
Table 5.6.2 Tests of Over-identifying Restrictions............................187
Table 5.6.3 Tests of Endogeneity..........................................................188
Table 6.2.1 List of variables used and definitions...............................236/237
Table 6.4.1 First stage regression estimation results (DEFTP)...............209
Table 6.4.2 First Stage Regression Results (ATTTP)...............................212
Table 6.4.3 Explanatory power of Instrumental Variables.......................216
Table 6.5.1. Second Stage Regression Results........................................218
Table 6.5.2 Tests of over-identifying restrictions.............................222
Table 6.5.3. Tests of endogeneity.......................................................223
Table 7.2.1. List of all variables used throughout the analysis and their definitions.................................................................238/239/240
Table 7.4.1 First Stage Regression Estimates Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of the Opposing Team..............................................................249/250/251
Table 7.4.2. Statistics Examining the Explanatory Power Of the Instrumental Variables.................................................................254
Table 7.5.1 Second Stage Regression Estimates Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of the Opposing Team..............................................................256/257
Table 7.5.2. Tests of over-identifying restrictions.............................261
Table 7.5.3. Tests of Endogeneity.............................................................262
Table 7.6.1 OLS Estimations Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of their Opposing Team.................................................................265
Table 7.6.2. Summary Statistics for OLS Estimation.................................266
Acknowledgements

I would like to take this opportunity to thank a number of people and organisations. First and foremost, I am most thankful to my supervisor Professor Paul Downward. His excellent guidance, constructive criticisms, total support, encouragement and commitment have enabled me to complete this project successfully. Paul’s expertise has been invaluable throughout the last four years and he has readily made himself available to guide me when I have required his support. I would also like to thank Loughborough University for allowing me to continue in higher education and research this area of great personal interest. As a PhD researcher I have been able to meet and learn from leading academics, build a network of contacts, learn to overcome obstacles and work alongside fellow PhD colleagues I now consider dear friends.

I would also like to take this opportunity to thank the team at Opta, who were kind enough to provide me with ample data of high commercial value free of charge. Without their support or belief in my ability I would not have been able to conduct this research or indeed continue in higher education.

Finally, I would like to acknowledge the encouragement and support of my friends and family. My father, mother and both of my brothers have continually supported me throughout my time in higher education and their unwavering faith in my ability has enabled me to complete this research project. Without their support, the last few years would have undoubtedly been much more difficult. Last, but by no means least, I would like to thank my partner. Her energy, positivity, continuous encouragement and support throughout this journey motivated and inspired me to pursue my dreams and specialise in this field.
Abstract

The purpose of this thesis is to investigate the levels of interdependency (simultaneity) between the performances of two direct opponents in professional football. More specifically, interdependency between performances is investigated at three different levels. Firstly, empirical analyses are conducted in order to assess the levels of interdependency between the performances of two teams in direct competition using team-game observations. Secondly, data on team formations and player starting positions within these formations is utilised in order to uniquely match individual players to a sole opponent on the field of play. Further empirical analyses are then conducted in order to investigate the levels of interdependency at this more isolated individual player level using player-game observations. Finally, an empirical investigation into the levels of interdependency between the performances of a positional unit (defence or attack) and their opposing team (as a collective) is conducted using positional unit-game observations.

An exclusive and detailed data set ranging from the 2007-08 season to the 2010-11 season is utilised in order to estimate several production functions for teams, individual players and positional units competing in the German Bundesliga. The results in all empirical analyses confirm that the performance of the opponent is significant. At a team and positional unit level, no evidence is found to suggest that the performances of two direct opponents(s) are interdependent, however the results reveal that the recent past performances of the opponent(s) have a significant linear impact upon the performance of the subject. In particular, relative team form going into a match is revealed to have a significant impact upon the performance of their opponent. At an individual player level, evidence is found to confirm that the performances of players in direct competition are interdependent thus supporting the sports economics theory of joint production. Specifically, the results reveal that the performances of defenders have a significant and negative impact upon the performances of their opposing attackers.

Keywords: Production functions; Team productivity; Positional unit productivity; Player productivity; German Bundesliga football; Joint production in sports; Sports economics; 2SLS IV estimation
Chapter One

Introduction

1.0 Overview

This chapter provides the rationale for this research as well as providing an outline of the thesis. Section 1.1 outlines the aims and objectives of this thesis and states the research question. Section 1.2 illustrates the growing economic significance of football and presents an overview of the financial growth football has experienced in recent years. Section 1.3 offers several reasons for the selection of the German Bundesliga football league as a case study for the empirical analyses conducted in subsequent chapters. Section 1.4 emphasises the originality of the research and summarises the main contributions to knowledge. Finally section 1.5 details the structure of the thesis.
1.1 Aims and Objectives

Rottenberg (1956), Neale (1964) and Sloane (1971) are widely credited with having made three key contributions in the economics of professional team sports. These seminal contributions are centred on the joint nature of ‘production’ in professional team sports through highlighting the mutual interdependence between two competitors in the production of a sporting contest. This thesis is built upon the theory of joint production and contributes to the burgeoning literature on sports economics by conducting a series of empirical analyses in order to explore the joint productivity of competitors in German Bundesliga football.

This study aimed to answer the following research question:

“To what extent is playing performance affected by the performance of the opponent in professional football?”

This thesis proposed to apply econometric analysis, motivated by economic theory, to examine if, and to what extent, player productivity in football is affected, by the performance of the opponent. As presented in the literature review in chapter 3, previous studies have paid minimal attention to the performance of the opposition and the assumed simultaneity between direct competitors. This thesis attempts to partially fill this gap by utilising an exclusive data set that allows the estimation of team, positional unit and individual player production functions whilst specifically taking into account the relative performance of the direct opponent(s), at a disaggregated match specific level.

In order to answer the research question the following research objectives were set:

1) To review the relevant economic, theoretical and empirical literature, to establish what factors have been identified as relevant to the determination of team and player productivity.

2) To extend the existing empirical literature by:

A. Organising and developing several data sub-sets that can be used to test for the impact of opponents.
B. Applying an appropriate econometric strategy to perform the tests.

C. Providing an accurate assessment of disaggregated team, positional unit and individual player performance in the German Bundesliga, over the course of four seasons, whilst specifically accounting for the relative performance of the opponent(s) at each of the following levels.

   i) Team Vs. Team
   ii) Positional Unit Vs. Team
   iii) Player Vs. Player

The construction of several exclusive data sets, in conjunction with the methodologically rigorous approach adopted in this study enabled three areas of investigation, each of which attempted to fill identified gaps in previous literature and advance knowledge. To begin with the mutually interdependent performances of home and away teams (in direct competition) was analysed at a disaggregated match level. This was followed up by the construction of an additional data set that enabled the analysis of individual player performance whilst up against a sole direct opponent on the field of play. Finally, the development of additional data sets enabled a match level investigation on the performances of specific positional units against opposing teams (as a collective).
1.2 Why Football?

Since the early 1990s, professional football in many countries has experienced an astonishing transformation. Player salaries have risen exponentially, television contracts yield revenues on a scale unimaginable only a few years ago, many football stadia have been completely rebuilt, and the importance of commercial sponsorship and merchandising has increased beyond measure. (Dobson and Goddard 2011, preface)

Football¹ (soccer) continues to grow in cultural, social and economic importance (Frick 2007; Dobson and Goddard 2011). The recently published Deloitte Football Money League report 2015, reveals that the aggregate annual revenue of the top twenty highest earning football clubs in the world for the 2013/2014 season exceeded €6 billion for the first time. A further affirmation of football’s growing economic significance is offered in the Forbes publication of July 2014 listing the most valuable sports teams around the world. The top three positions within this Forbes ranking were occupied by the three football teams Real Madrid C.F, FC Barcelona and Manchester United FC, with the New York Yankees (baseball) and Dallas Cowboys (American football) making up the top five as the only representatives of other sports. The value of Real Madrid C.F was estimated to be $3.44 billion, almost an entire billion higher than the estimated value of the New York Yankees ($2.5 billion) whom ranked number 4 and represented the highest valued team from a sport other than football.

Since the mid 1990’s the football industry has experienced a surge in revenues and costs. Substantial revenue increases from broadcasting rights began concurrently with the development of the Premier League². These developments instigated changes in the players’ labour market such as the Bosman Rule³, which ultimately gave players greater bargaining power and led to a substantial rise in player salaries (Simmons 1997). Figure 1.1a below illustrates the rate at which economic growth in football has continued in recent years. Figure 1.1a displays the steady revenue increases of the 1st,

---
¹ A form of football played by two teams of eleven players with a round ball that cannot be handled except by goalkeepers. Also known as Soccer and Association football. The objective of the game is to score goals in the opponent’s goal by kicking or heading. The game originated in England and is played according to rules established by the football association.
² The top division of professional Soccer in England and Wales.
³ A reference to a European Court ruling which obliges professional football or other sports clubs to allow players over the age of 25 to move freely between clubs once their contract have expired. Named after Jean Marc Bosman, a Belgian footballer who brought a legal case that resulted in the ruling.
5th, 10th and 20th ranked highest earning clubs within football over the ten year period between the 2003/04 season to the 2013/14 seasons.
Figure 1.1a Revenue Thresholds for Positions in the Deloitte Money League

Figure 1.1a shows that revenues have steadily risen over the last decade. This has been due largely to increases in three main sources of revenue, match day income, the sale of broadcasting rights and commercial revenue. In the 2004/2005 season, aggregated club revenue from the top twenty of the world’s highest earning clubs was just over €3 billion. In the nine full seasons since (up to the end of the 2013/14 season), the total aggregated revenue of the top twenty highest earning football clubs was recorded at over €6 billion meaning the level of revenue in football has doubled in just nine seasons (Deloitte 2006, 2015).

Although all three of the main revenue streams have risen in recent years and contributed to the €6 billion aggregated revenue recorded for the season 2013/14, the majority of this increased revenue can be attributed in particular, to substantial increases in commercial and broadcasting revenues. Figure 1.1b below presents the proportions of each of the three main revenue streams that accounted for the aggregated revenue of the twenty highest earning football clubs for the 2004/05 season. Figure 1.1b shows the largest contribution to the recorded €3.13 billion (2dp) aggregated revenue of the top twenty highest earning clubs in the season 2004/05 came from broadcasting revenue, which accounted for 41.03% (€1.28 billion, 2dp). Commercial revenue was the second largest source of revenue accounting for 32.01% (€1 billion, 2dp) followed by match day revenue, which accounted for just fewer than 27% (€0.8 billion, 2dp) of total aggregated revenue.

In comparison with aggregated revenue for the top twenty highest earning clubs recorded in season 2013/14 (Figure 1.1c), broadcasting revenues and commercial revenues accounted for a combined 80% of the total €6.2 billion aggregate revenue. Commercial revenue alone accounted for 40.45% (€2.5 billion, 2dp), representing an increase of €1.5 billion in nine seasons, whereas broadcasting revenue was responsible for 40% (€2.48 billion, 2dp) representing a €1.2 billion increase. In season 2013/14 match day revenue only accounted for 19.89% (€1.2 billion, 2dp), representing a comparatively small €0.4 billion increase throughout the same period.
Figure 1.1b – Proportional breakdown of Revenues Received For the Top Twenty Highest Earning Football Clubs in Season 2004/05

Source: Deloitte 2006

Figure 1.1c – Proportional Breakdown of Revenues Received For the Top Twenty Highest Earning Football Clubs in Season 2013/14

Source: Deloitte 2015
Figures 1.1b and 1.1c above highlight the changing composition of revenues in football and emphasise the requirement for clubs to be competitive off the pitch (as well as on the pitch). Figures 1.1b and 1.1c above emphasise the growing importance of commercial revenue. Throughout the ten season period between the 2004/05 season and 2013/14 season, commercial revenue has surpassed broadcasting revenue as the largest source of income even though broadcasting revenues have increased by an enormous €1.2 billion throughout the same period.

Although the three main sources of revenues are increasing in football, they are not increasing evenly across teams, which can present potential problems. Szynanski (2001), states that the injection of greater TV and commercial revenue into football is causing income inequalities between clubs to grow, consequently causing widespread concern that competitive imbalance will grow as a result.

According to Dobson and Goddard (2011), team performance and financial performance in football are inextricably linked. At the micro level, a club’s capacity to generate revenue depends on its team’s success on the field of play. Furthermore, a club’s capacity to strengthen its team by purchasing better players in the transfer market and by offering remuneration at high levels that can attract and retain the best players depends on the strength of their finances, highlighting the two-way relationship between team performances and team finances in football (Dobson and Goddard 2001).

Studies on American sports and in particular baseball have revealed that teams with strong financial performances have not always been the most successful. According to Bauer and Sauer (2005), the four teams with the highest estimated brand value in American sports in 2001 were the Dallas Cowboys (NFL), Washington Redskins (NFL), New York Yankees (MLB) and New York Knicks (NBA). Of these four teams, only the Yankees had reached the championship series within the three preceding years, while the other teams in the two preceding seasons were not even able to make it to the playoffs (Bauer and Sauer 2005).

Furthermore, the best-selling book authored by Michael Lewis entitled, Moneyball: The Art of Winning an Unfair game (2003) tells the story of how the Oakland Athletics in Major League Baseball (MLB) had achieved a sustained competitive advantage despite being one of the lowest wage spenders (Gerrard 2007).
Moneyball story revolved around the Oakland Athletic’s general manager Billy Beane and his systematic use of player performance data in order to guide player recruitment, player valuation and field tactics.

The statistical analysis of baseball data is more commonly known as sabermetrics (Gerrard 2007). Moneyball, told the story of Billy Beane and his utilisation of sabermetrics in order to convert the Oakland Athletics from a low budget, low achievement team to a low budget, high achievement team. The focus of the book was the 2001 and 2002 seasons where the Oakland A’s had enjoyed win ratios of .63 and .636 in a traditional 162 game season, meaning the Oakland A’s enjoyed over 100 wins each season, which ensured their qualification for the post season playoffs, even though they recorded one of the lowest payroll expenditures for both seasons (Gerrard 2007).

Over eight seasons between 1998 and 2006, the New York Yankees (generally considered MLB’s top brand), averaged only 3.9% more wins than Oakland even though they outspent Oakland by a factor of 3.22. In other words, a 216.7% pay premium yielded the New York Yankees only 6 extra wins per season (Gerrard 2007). Furthermore, in seasons 2000 and 2002, Oakland recorded a higher win ratio than the New York Yankees.

Moneyball attracted a global audience and inevitably raised questions on the extent to which the approach adopted by the Oakland Athletics could be replicated in other team sports such as football. According to Gerrard (2007) there is significant value to be obtained from the systematic (statistical) analysis of player and team performance in football. In his benchmarking analysis for the English Premier League, (Gerrard 2007) concluded that teams with fewer financial resources adopted a knowledge-based strategy to overcome their financial disadvantage.

Statistical analysis has become a key input into the knowledge-based strategy that allows low budget teams to compete effectively with the market giants (Gerrard 2007). A successful knowledge based strategy however is not achieved by simply acquiring player data, but rather analysing and interpreting player data in the correct manner. Gerrard (2007) emphasises this point by stating the strategic resource in Oakland Athletics success (described in Moneyball) was not the availability of player performance data, (a commodity available to all MLB teams) but rather the ability of
the Oakland management to create valuable information through the analysis, interpretation and application of the data. The book Moneyball focussed upon the science of Sabermetrics (a statistical analysis of player performance) however as the title of the book implies, the application of this science is an art. Winning requires risks, which can never be a matter of pure calculation as knowing which risks to take require judgement.

Organisational effectiveness is both an art and a science; ignoring one in favour of the other is likely to diminish rather than enhance the organisations effectiveness. Successful performance management requires both “hard” data analysis and “soft” expert judgement (Gerrard 2007, p230)
1.3 Why the German Bundesliga?

In order to fulfil the aims and objectives of this thesis, the top tier national league in Germany, the German Bundesliga, was adopted as a case study. There were three main justifications for the selection of the German Bundesliga. Firstly, Bundesliga data has been commonly utilised in recent literature for investigations into various aspects of football. For example, German Bundesliga data has been frequently utilised for investigations into various dimensions of the players’ labour market such as, remuneration, transfers, contracts and mobility (Feess et al 2003; Ashworth and Heyndels 2007; Frick 2007, 2011; Lehmann and Schulze 2008). German Bundesliga data has also been utilised for investigations into different aspects of team and individual player performance (Kern and Süssmuth 2005; Torgler 2004; Torgler et al 2006; Torgler and Schmidt 2007; Kuss et al 2007; Frick and Simmons 2008; Tiedemann et al 2011; Haas and Nüesch 2012; Bryson et al 2013; Vogelbein et al 2014). Furthermore, studies analysing issues such as home advantages and head coach performances have also utilised data from the German Bundesliga (Hautsch et al 2001; Sutter and Kocher 2004; Barros et al 2009; Frick et al 2010).

Another reason the German Bundesliga was selected for the purposes of this study is because of the existence of a statute known commonly as the ‘50+1 rule’. Since 1998, professional football clubs in Bundesliga 1 and 2 have been able to operate as incorporated businesses (Dobson and Goddard 2011). For the few clubs that have become incorporated, the governance system of regular corporations does not apply. In accordance with German football regulations the Verein must hold 50 per cent plus one vote of the corporation (hence the 50+1 moniker) in floated clubs. This restriction effectively means that the Verein will always retain 51% ownership, meaning any outside investor can never own more than 49% of a German clubs shares, preventing a single entity taking control (Merkel 2013). This statute therefore offers partial protection from the impact of an exogenous shock (a billionaire takeover for example) upon team and player performances.

---

*Traditionally German football clubs were established as registered associations (eingetragener Verein). The eingetragener Verein is a non-profit member organisation, managed by representatives democratically elected by the members.*
An exogenous shock such as a billionaire takeover can have a huge bearing on a team’s league ranking (Barros and Leach 2006). In June 2003 the Russian billionaire Roman Abramovic purchased Chelsea FC and injected £760 million into the club in his first five years as owner up until the end of the 2008 season. At the end of the same season Chelsea’s salary expenditure was £172 million, 42% higher than the next highest spending EPL team Manchester United on £121 million (Dobson and Goddard 2011). At the end of the 2014/15 season, Chelsea claimed their fourth Premier League Championship victory with the previous three having been won in 2005/06, 2006/07 and 2009/10. Similarly, the billionaire Sheik Mansour bin Zayed Al Nahyan of Abu Dhabi and his acquisition of Manchester City in September 2008 has also heralded success in the form of two Premier League championship triumphs (their first ever and second) in the seasons 2011-12 and 2013-14. Due to the presence of the 50+1 ruling in German Bundesliga football billionaire takeovers are not possible and hence the impact of exogenous shocks upon performance can be limited, offering further justification for the selection of German Bundesliga data.

The final reason for the selection of Bundesliga data was due to the availability of supplementary data. Information on physical characteristics of players such as their height, age and preferred foot are relatively easy and inexpensive to acquire from published magazines such as Kicker and Transfermarkt as well as the website www.transfermarkt.de. Each of these sources has excellent coverage of players ranging from high profile to low profile players for every team competing in Germany (Bryson et al 2013). Furthermore the information contained within these sources has been verified by a stable team of experts (Bryson et al 2013) and hence have been used as a source of data in previous studies (Torgler 2006; Frick 2006; Torgler and Schmidt 2007; Bryson et al 2013; Scelles et al 2014).

---

5 See www.chelseafc.com for a full list of honours
6 See www.mcfc.co.uk for a full list of honours
1.4 Contributions

This is an innovative topic largely due to the difficulty of obtaining detailed performance data on football. Despite the proliferation of analysis of production in sports, previous research tends to be predominantly based on American sports. This is due to two main reasons. Firstly, there is a dearth of match play statistics on other more interactive European sports (Carmichael and Thomas 2014). This is due largely to detailed player performance data being confidential and usually provided to teams on a commercial basis, hence making it either unavailable in the public domain or prohibitively expensive for academic research (Gerrard 2007). Secondly, it has been argued that attempting to determine the levels of productivity of the team as a collective or individual players within the team in football is problematic due to the interdependent team play nature of football making it difficult to isolate individual player contributions (Weimar and Wicker 2014). Football is not like an atomistic striking and fielding sport such as baseball or cricket that are said to be:

… more conducive to the systematic exploitation of player performance data due to the high degree of separability of individual playing contributions. (Gerrard 2007, p214)

Furthermore, several measurement problems, technological limitations, conceptual and cultural barriers limit the examination of performance in complex invasion team sports such as football (Gerrard 2007; Weimar and Wicker 2014). In spite of these purported limitations there is significant value to be obtained from the systematic (statistical) analysis of player and team performance (Gerrard 2007).

This research is based on a secondary data set exclusively obtained from Opta statistics, one of the world’s leading sports statistics companies. The data consists of over 200 player performance measures for outfield players and over 80 player performance measures for goalkeepers, per match, across four seasons between 2007/08 and 2010/11. Crucially, this data set reports the team formations utilised by each team in each fixture and hence identifies the specific positions in which each player performed. Theoretically, a player performing in a specific position could be matched up to an opposing player who was performing in the same area of the playing field (see chapter 6) enabling investigation into productivity at a more isolated
A methodologically rigorous approach is devised in order to investigate the joint contribution of two competitors to output in German Bundesliga Football. The model constructed accounts for endogeneity and the assumed simultaneity present in sports contests at a team level and a more isolated player level.

Advancements to knowledge on team production are made as the relative performance of the opponent is measured and hence their contribution to output is estimated. The use of relative performance measures means endogeneity and simultaneity between the performances of two teams in direct competition is accounted for at a match level.

Knowledge on player productivity is advanced as production functions for individual players performing in specific positions are estimated. The relative performance of the opponent is measured and endogeneity and simultaneity between the performances of two players in direct competition is accounted for at an individual player level.

Knowledge on positional unit production is advanced as the impacts of specific positional units upon the performances of their opposing team (as a collective) are estimated. Relative performance measures allow endogeneity and simultaneity between opponents’ performances to be accounted for.
1.5 Thesis Structure

This research project proceeds as follows. Chapter two outlines some basic economic models before moving on to explain key elements of the economic theory of production. This provides a platform for the analysis of production in football. The concept of perfect competition is outlined, as this is the core theoretical framework that underpins the ideal state of production and markets within economic analysis. Market failures and the monopoly and oligopoly market structures that are often compared to this economic ideal are also detailed. Individual and team productivity is outlined and factors such as, the acquirement of human capital and the subsequent enhancement of cognitive and non-cognitive skills, the concept of moral hazard and the design of incentive and tournament contracts, are all explained. Chapter two also outlines the ‘peculiar economics of sports’ (Neale 1964) and describes the unique characteristics of sports that produce peculiar conclusions from an economic perspective. One of these characteristics, the joint production of sporting events is taken forward and expanded upon in chapter 3.

Chapter 3 proceeds by exploring the process of production in team sports. In particular, the role of contest organisers in the production of a contest, the uncertainty of outcome hypothesis (UOH) and the use of production functions and their associated production frontiers in the analysis of team performance in sports are detailed. Chapter three also highlights the general lack of attention awarded to the impact of the opponent and explains how limitations in previous research will be overcome in the empirical analyses conducted in subsequent chapters.

Chapter 4 presents the methodology adopted in order to achieve the research aims and objectives stated above. This chapter outlines the methodological and philosophical foundations upon which this study is built and reviews the key theoretical approaches that affect the research design. The data used to conduct this study is described and the selected econometric process is outlined. Key performance indicators (KPI) for playing positions are identified and the justification for the inclusion of these KPI’s in subsequent empirical chapters is offered.

Chapter 5 presents the first of three empirical chapters and investigates the level of simultaneity between the performances of two teams in direct competition. Chapter 6
investigates the level of simultaneity between the performances of two players in
direct competition and hence analyses the level of interdependency at a more isolated
individual player level. The final empirical chapter 7 investigates the level of
simultaneity between the performances of positional units and their opposing team(s)
collectively. Each of these chapters presents a detailed description of the sample, the
chosen performance and instrumental variables, the analysis strategy and the model
specification. The results obtained from all empirical analyses conducted in each
chapter are presented and a discussion of the results obtained and their implications
are also presented in each chapter.

Finally, chapter 8 discusses the conclusions of the study informed by the results
obtained from the empirical analyses in previous chapters. Limitations of the study
are discussed and suggestions for future research are presented. The empirical
analyses conducted in this study revealed that the impact of an opponent’s
performance is significant at a team level, positional unit level and an individual
player level. At a team and positional unit level the opponent’s past performance
(relative form going into a match) was revealed to have a significant impact upon the
performance of the subject, but no evidence was found to confirm that performances
between the opponent and the subject (on matchday) are interdependent. At an
individual player level however, performances between an observed player and their
direct opponent were revealed to be interdependent, thus confirming the existence of
simultaneity between the performances of individual players in direct competition in
matches.
Chapter Two

Production in Economics

2.1 Introduction

In this chapter the key elements of the economic theory of production will be outlined in order to provide a platform for the analysis of production in football. Section 2.2 will begin by briefly highlighting the concept of perfect competition, as this is the core theoretical framework that underpins the ideal state of production and markets within economic analysis. Section 2.3 will briefly highlight the concept of market failures and outline the monopoly and oligopoly market structures that are often compared to this economic ideal.

Section 2.4 shifts the focus towards individual and team productivity. More specifically, this section explores how individual workers contribute to output at both an individual and team level. This section discusses how various factors such as, the acquirement of human capital and the subsequent enhancement of cognitive and non-cognitive skills, the concept of moral hazard and the design of incentive and tournament contracts, can all potentially impact the levels of individual productivity.

Section 2.5 details the ‘peculiar economics of sports’ (Neale 1964) and outlines some unique characteristics of sports that produce peculiar conclusions from an economic perspective. Market structures in sports, market failures in sports and in particular market failures caused by externalities in sports are all explored. This section outlines tournament organisation with reference to contest theory and then highlights the theoretical importance of the concept of competitive balance in sports because of the uncertainty of outcome hypothesis as argued by Rottenberg (1956). Finally, this chapter concludes with section 2.6 that summarizes the peculiar characteristics of professional sports that lend themselves to economic analysis. One of these characteristics, the joint production of sporting events is taken forward and expanded upon in chapter 3 and is examined empirically in chapters 5, 6 and 7.
2.2 Alternative Market Structures

In the textbook theory of the firm, resource allocation through output prices and input prices are examined with reference to alternative market structures assuming that the firms in the markets seek to maximize profits (Stiglitz and Driffill 2000). What is of relevance to firms is the level of profit they can make, the level of output that can be produced, their efficiency in production (use of resources) and the prices they can charge consumers and pay for inputs.

The answer to all of these questions depends theoretically upon the level of competition the firm faces. A firm in a highly competitive environment will behave in a completely different manner to a firm facing little or no competition. A firm facing much competition from rival firms will be forced to follow competitors in their prices and be as efficient as possible thus reducing costs to their lowest level. Conversely, a firm facing no competition is likely to have considerable power over prices, meaning consumers will pay more as a result. This section and the subsequent section 2.3 focuses on the three main extremes: perfect competition (where many firms are competing), monopoly (where there is only one firm in the industry) and oligopoly, (where a few firms share market power) and more specifically production within each market structure.

2.2.1 Perfect Competition

In perfectly competitive markets, firms are assumed to take the price of outputs and inputs as given and make decisions only about the volume of outputs and the quantities of inputs utilised. Furthermore, all participants in competitive markets are assumed to be profit maximisers that are fully informed about prices and are price takers. This is because there are so many firms in the industry that each one produces an insignificantly small portion of the total industry supply and they therefore cannot individually affect prices. Products are not differentiated in any way (homogenous) and the level of production by one firm is not directly dependent on the behaviour of other firms. This allows firms to expand or contract their production levels and to

---

7 Readers should note that alternative objectives for firms have also been examined such as Revenue maximisation and Sales maximisation (Geetika et al 2011).
enter or leave a particular industry freely. In order to fully understand how this competitive theory works readers should be made aware of a firm’s demand and supply decisions. Figure 2.a below illustrates the demand and supply function of a firm in a perfectly competitive industry.

**Figure 2.a - Industry supply and demand in perfect competition**

The demand curve above reflects that as prices rise, consumer demand for the product falls and as prices fall, consumer demand rises. The position of the curve is conditional on other factors. In textbook economic analysis this is assumed to be the price of other goods, consumer incomes and preferences (Eaton et al 2002). The curve represents potential purchases of the product at a variety of different price levels. The supply curve can be understood as the aggregate supply of the individual firms, derived from their costs.

In this price taking market, the interaction of market demand and supply determines the market price (P) and quantity (q) (Downward et al 2009). If demand and supply in the market set the price of the product, individual firms must supply the product at this price (Downward and Dawson 2000). This implies that firms face a horizontal demand curve as only one price can prevail in the market (hence, firms are price takers). Consequently, a firm’s marginal revenue (MR), or revenue received from the
sale of the last unit, is equal to average revenue (AR), (total revenue divided by the total sales volume) or market price (P) (Downward and Dawson 2000).

Figure 2.b below illustrates the economic costs and revenues of an individual firm within the competitive market. As all rivals in a perfectly competitive model are assumed to have homogenous products, the costs and revenues represented in figure 2.b will be the same for each firm. The supply of the firm is determined by their marginal costs (MC), which are the additional costs a firm will incur should they produce an additional unit of output (Downward et al 2009). Marginal costs are calculated by dividing the change in total costs by the change in output (Eaton et al 2002).

**Figure 2.b - Costs and Revenues of an individual firm in perfect competition**

The upward slope of the marginal cost curve can be explained with reference to both the productivity of inputs and also their monetary costs. Underlying production in this model is the concept of a production function. The production function refers to the process by which resources are transformed, that is, that output is a function of the inputs. Inputs are described as factors of production and often comprise of land, capital and labour services (Estrin and Laidler 1995). It is assumed that the production
function captures the efficient combinations of resources in production. Specifically, the maximum set of outputs that are obtainable from the minimum set of inputs\(^8\).

A simple production function could be expressed as in equation 2.1 below, where the output \(Y\) is a result of the three inputs land (LD), labour (L) and capital (C).

\[
Y = Y(LD, L, C) \tag{2.1}
\]

This function represents how a dependent variable \(Y\), could be affected by a related set of independent variables LD, L and C. Changes to output following from changes in inputs can then be understood with reference to marginal productivity which, according to Downward et al (2009, p11/12) is:

The additional output following the employment of an additional unit of a variable factor of production, or resource, combined with other factors of production, or resource, being fixed.

It is assumed that marginal productivity diminishes. This means that for each increment in inputs to production, output increases but at a diminishing rate. As well as marginal productivity, average productivity could be defined as “total output divided by the total amount of input used” (Eaton et al 2002, p669).

These relationships mean that for any given (monetary) cost of an input or factor, for example determined by the competitive market for resources, marginal costs are as described in equation 2.2 which will rise in accordance to diminishing marginal productivity.

\[
MC = \frac{FC}{MPF} \tag{2.2}
\]

In equation 2.2 above MC, represents marginal cost, FC represents factor cost and MPF represents the marginal productivity of the factor. It is this that generates the upward sloping marginal cost curve in figure 2.b. Figures 2.a and 2.b indicate that the market sets the price of output and because of the lack of power of the perfectly competitive firms this is the price they have to take from the market.

As demonstrated in figures 2.a and 2.b above \(P=AR=MR\). Maximum profits for the firm will thus occur where \(MR = MC\) as illustrated in figure 2.b. This identifies the

\(^8\) Profit maximisation, and in this case, perfect competition, implies that production takes place in this way as it keeps costs as low as possible as implied in this section.
level of supply required in order to maximise profits referred to as the short run
supply function (represented by point Q in figure 2.b). Any further production beyond
this point and the marginal revenue obtained from the sale of this additional unit will
no longer be sufficient to cover marginal costs (as \( P = MC > MR \)). Similarly, any
level of production below this point would result in profit being sacrificed (as \( P = MR > MC \)) and therefore profit maximisation would not be achieved (Downward et al
2009).

Average costs (AC) refer to fixed and variable costs and therefore describe the total
cost per unit of product. Fixed costs are those that must be paid regardless of the
levels of production. Variable costs refer to those costs that can be varied as firms
commit resources to production and therefore only need to be paid when a firm varies
their level of output (Downward et al 2009). The firm’s profit is determined by the
difference between AR and AC at point Q. In the perfectly competitive market in the
long run \( P=MR=AR=AC \), as competition will prevent returns above those required to
keep capital in that activity. In the short run if \( P > AC \) then this will encourage entry
to the market and price will be competed down as the supply curve shifts to the right
in figure 2.a. If \( P < AC \), then losses would cause firms to leave the market and price
would rise back to AC as the supply curve shifts to the left in figure 2.b\(^9\).

The major contribution of the perfect competition economic theory is that it formally
demonstrates the concept of efficiency and efficient allocation of resources. The
production function outlined above indicates what is meant by productive efficiency.
However, it is the role of prices in allocating resources that demonstrate the allocative
efficiency of the competitive market. At the market equilibrium in figure 2.a prices
reflect what fully informed consumers are prepared to pay for output, which is
produced at its most efficient possible. Allocative efficiency is otherwise known as
Pareto – optimality. This results in the competitive equilibrium allocation being
Pareto- optimal or efficient, as no party can be better off while leaving the other party
at least as well off (Eaton et al 2002).

---

\(^9\) The firms AC curve in Figure 2.b comprises average variable costs (AVC) and average fixed costs (AFC). In the long run as the
firm adjusts its capacity AVC and AC are effectively the same. In the short run analytically it is assumed that labour inputs can
vary but not capital stock and land. In this case Average Variable Costs are driven by the productivity of labour. As average fixed
costs have to be met regardless of the level of production, the firm will cease trading if \( P < AVC \).
2.2.2 The Isoquant/Isocost Approach to Production

The subsection above is essentially based on a consideration that one factor of production can be varied in production in the short run. Analytically, in the long run, firms must choose the optimal combination of factors of production. This can be illustrated through the isoquant/isocost approach. A firm’s choice of the optimum combination of factors can be explained graphically. Figure 2.c below reveals several estimated levels of production labelled a, b, c, d and e for an output of a hypothetical firm using just two variable factors, land (L) and capital (K).

**Figure 2.c - An Isoquant**

As can be seen from figure 2.c above, production technique (a) is a capital intensive technique. Moving towards technique (e) production becomes more labour intensive as greater units of labour (workers) are used and less units of capital. The isoquant displayed in figure 2.c above, presents 5 alternate production techniques for a given level of output and demonstrates the range of alternate ways the production of a given level of output can be achieved.

For every given level out output, additional isoquants can be drawn. If the given level of output is to be increased, the isoquant will be further out to the right. If the given level of output is decreased, the isoquant will be further to the left. This is illustrated in figure 2.d below where the original level of production presented in figure 2.c above is now labelled i₁. As can be seen from figure 2.d below, i₂ and i₃ lie to the
right of $i_1$ meaning these isoquants represent a higher level of output. The Isoquant $i_0$ lies to the left of $i_1$ indicating a lower level of output.

The slope of the isoquant is explained by the diminishing marginal rate of technical substitution (MRTS), which refers to the rate at which one factor can be substituted for another whilst maintaining a constant level of output. In other words, the marginal rate of technical substitution is the amount of one factor (e.g. K) that can be replaced by one additional unit of the other factor (e.g. L) in cases where the output is held constant.

The shape and slope of the isoquant demonstrates that the production function exhibits diminishing marginal productivity associated with each input to the production function (Downward et al 2009). Increasing the amount of L relative to K means that the marginal productivity of L falls relative to the marginal productivity of K. Consequently, reorganising the inputs in order to maintain production at the same level means that more L is lost relative to additional K. The marginal rate of technical substitution is determined by the ratio of marginal productivities of L and K, as illustrated in equation 2.3 below.

$$MRTS = \frac{MP(L)}{MP(K)}$$  \hspace{1cm} (2.3)

**Figure 2.d - An Isoquant Map**
Figures 2.c and 2.d above reveal ways in which factors can combine in order to produce a given level of output. A firm however must determine the level of output, which involves taking into account costs. Assuming factor prices are fixed, a firm can evaluate various factor combinations for a fixed cost via the construction of an isocost line. If $C$ is the fixed level of cost, the isocost line represented in equation 2.4 below contains all input combinations $(L, K)$ that cost $C$.

$$ C = wL + rK \quad (2.4) $$

Equation 2.4 can be rearranged in terms of $K$ as in equation 2.5 below.

$$ K = \left( \frac{C}{r} \right) - \left( \frac{w}{r} \right) L \quad (2.5) $$

This straight line equation (2.5) can be presented graphically as in figure 2.e below.

**Figure 2.e - A Graphical Representation of an Isocost**

Several isocosts can also be plotted on the same graph each one representing different costs. The higher the cost, the further out to the right the isocost will be, the lower the cost, the further to the left the isocost will be. This indicates the least costly way of producing a particular level of output and the highest level of output that can be achieved for a given cost of production.
The isoquant curve $i_1$ in figure 2.f below (taken originally from figure 2.c) represents the level of desired output for a firm. The three isocosts drawn represent the different levels of total cost. The least cost combination of labour and capital is shown at point $r$. This is the point at which the isoquant is tangent to the lowest possible isocost.

**Figure 2.f - The least cost method of production**

Isoquants and isocosts must logically also represent the highest output possible given a total cost of production. In this case, a series of isoquants can be drawn each representing different levels of output as in figure 2.g below. The point at which the isocost is at a tangent to the highest isoquant (point $s$) reveals the factor combination yielding the highest output for that particular level of cost. Here the MRTS = $\frac{MPL}{MPK} = \frac{w}{r}$. Factors of production are paid their marginal productivity.
When firms are not operating in perfect competition, it is argued that market failure has occurred. These failures, to varying extents relax the assumptions of the perfectly competitive model and suggest more realism (Downward et al 2009). Economic theory of the firm suggests that these failures are likely to follow from monopoly power and the existence of externalities.
2.3 Monopoly, Externalities and Market Failure

Sometimes prices paid by consumers or received by firms do not reflect the true costs and benefits involved to society (Downward et al 2009). In other words, markets fail to achieve efficiency. Economists have previously examined the conditions under which efficiency is not achieved. One such example would be a firm that operates in a monopoly market. Imperfect markets such as monopoly markets highlight the limitations of production as firms’ with market power will use less factors of production and hence produce less output than the pareto-optimum (Eaton et al 2002). Another reason efficiency is not always achieved is due to production and consumption externalities. These inefficiencies are now discussed in turn.

2.3.1 Monopoly

The monopoly model relaxes the assumption about the number of suppliers in a market. Theoretically it occurs where there “is a single firm that serves an entire market” (Nicholson 2002, p495). The fundamental difference between monopoly firms and competitive firms is that a monopolist can choose to produce or to set its price at any point on the market demand curve as a “price maker” (Eaton et al 2002). Consequently, a monopolist does not face the horizontal demand curve (as in perfect competition) meaning the relationship between MR and AR will be different for a monopolist than for a competitive firm. In order to increase sales a monopolist will offer additional units of product at a lower price than before. The MR earned by the monopolist on this additional produce will be lower than the AR earned from previous sales. Therefore, in general MR < AR for a monopoly or any other form of market that is not perfectly competitive and where firms influence market price (Downward et al 2009).

A monopoly market could be referred to as the opposite of perfect competition (Downward et al 2009). Theoretically then, the supply curve in a monopoly market, will represent the MC of the monopoly and the sum of the ACs will represent the monopoly’s AC. Similarly the market demand curve will represent the monopolist’s AR (Downward et al 2009). These different assumptions mean that in the monopoly model, marginal cost is equal to marginal revenue in a profit maximising case but
MR=MC is less than price or AR. Figure 2.h below illustrates the situation of a monopoly.

**Figure 2.h - Monopoly supply of Output**

As can be seen from figure 2.h, MR is less than AR for a monopoly. The profit maximising point is at Q* where the MR intersects the MC and price is represented by P* as this is the price the firm are able to set, as this represents what consumers are willing to pay for the product. This market is inefficient as the firm is able to earn a mark up of P*- MC, through a restriction of output and higher price than competition (Q* < Qc and P* > Pc) as illustrated in figure 2.i below.

The allocative inefficiency that follows is due to three main sources. First, the profitability implied by P > AC across all output represents profit that is a redistribution from the consumer to the producer of consumer surplus. This is the value accruing to consumers because their willingness to pay as described by the demand curve is above the market price paid for a given volume of output on the market up to the equilibrium. However, as the market is also smaller than the competitive case, there are ‘deadweight’ losses of consumer surplus of XWY and also deadweight losses of producer surplus ZWY. The latter represents the contribution to profit to firms by P > MC for any given output up to the equilibrium. Strictly defined, producer surplus refers to the amount producers receive above that what is required to induce them to supply goods in a market.
2.3.2 Oligopoly

In the absence of pure monopoly, oligopoly markets exist where a few firms that have market power use various strategies in order to outmanoeuvre each other. Oligopoly could be referred to as a middle ground between perfect competition and monopoly (Eaton et al 2002).

Oligopoly models try to capture elements of realism through their description of markets, while retaining the assumptions that rational economic agents both have perfect information and optimize their decisions. (Downward et al 2009, p136).

Another key difference of oligopoly markets is that they require analysis of the strategic interactions of firms using game theory. The reason this is important is because the strategy of any one firm in an oligopoly will consequently affect another firm’s total payoff (Eaton et al 2002). The profit maximisation levels of one particular firm will therefore be partially dependent on the strategies and actions of other firms in the industry. One strategy is when firm’s maximise their own profits, given the strategies, that is output choice, of other firms. This is called a Nash- Cournot

---

10 Game theory consists of various different strategies such as co-operative/non- cooperative strategies, dominant strategies, Nash equilibrium, repeated games and sequential games. Co-operative strategies refer to binding contracts being arranged in order to solve conflicting objectives (essentially bargaining over consumer surplus and profit). With non co-operative strategies this is not possible. Dominant strategies refer to economic agents adopting the optimal strategy regardless of what other agents choose to do. Nash equilibrium refers to when agents adopt the best possible strategy, given the actions of other agents. Many transactions are sometimes repeated; a repeated game could therefore be modelled, referring to the repeated games strategy. Sequential games refers to the way in which games are different if agents make decisions and act simultaneously rather than act in sequence.
equilibrium strategy. Profits in the Cournot model will be greater than under perfect competition as price will remain greater than MC, however profits will be less than in a monopoly or cartel (colluding firms) as price is lower than the monopoly price (Eaton et al 2002).

Another oligopoly model is the Bertrand model. The Bertrand model assumes that each firm in the market will choose price instead of level of output (as their strategy) and that each firm will take their rival firms price as a given. Competition suggests that if one firm undercut the other in pricing, then they become the monopoly in the market and if firms set equal prices then they will split the market evenly. Assuming costs are constant for all firms, the best strategy for each firm would therefore be to undercut their rivals by the smallest possible unit until the marginal cost level is reached (however, setting prices at the marginal cost level will result in zero profit).

Because of uncertainty over rival responses, oligopoly provides a context for collusion to share the market. Collusive profit maximisation levels may be higher than the individual equilibrium levels from the Cournot and Bertrand models. This implies that on occasion both the Cournot and Bertrand models are individually rational but collusively irrational. Although collusion could ultimately lead to a monopolistic outcome, this does not mean collaboration is always best. A firm’s profit maximisation incentive can always provide enough temptation for a firm to renege on previous agreements and therefore deceive fellow collaborators in the market. This implies that collusion is only collectively rational when an agreement between firms is self-enforcing. If the arrangement is not self-enforcing then collaboration models are collectively rational but individually irrational (Eaton et al 2002).

2.3.3 Externalities

Market failure can also occur due to the existence of externalities, which can be regarded as production or consumption side effects. The existence of externalities, even in an otherwise perfect market, means efficiency will never be achieved. When society (people other than the producer or consumer) experience benefits from production or consumption, there is said to be external benefits. Conversely, when society is affected in a negative manner, there is said to be external costs. Thus, the full social cost (cost to society) of the production of any good is the private cost faced
by firms’ plus any externalities of production. Similarly, the full social benefit (benefit to society) from the consumption of any good is the private benefit enjoyed by the consumer plus any externalities of consumption.

One type of externality is the external cost of production. This externality arises in cases where a community bears additional costs to those borne by the firm. In these cases, the marginal social cost (MSC) of production will exceed the private cost (MC), therefore assuming the market is in all other respects perfect, the MSC curve lies above the MC curve as shown in figure 2.j below. The socially optimum output would be at the point $Q_2$, where $P=MSC$, however the firm’s output is actually at point $Q_1$, meaning their output is greater than the optimum. From society’s point of view external costs lead to overproduction.

**Figure 2.j - External Costs in production**

![External Costs in production](image)

Alternatively, there can be external benefits of production. In such cases the marginal social cost of production will be less than the marginal private cost, meaning the MSC curve will drop below the MC curve as illustrated in figure 2.k below. In this case, a firm’s level of output will be at the point $Q_1$ where $P=MC$, which is a lower level than the social optimum at point $Q_2$ where $P=MSC$. 
As there can be external costs and benefits of production there can also be external costs and benefits of consumption. External costs of consumption arise in cases where the marginal social benefit (MSB) of consuming a product is less than the marginal private benefit (MPB). Figure 2.1 below shows the marginal private benefit and price to a consumer of using (consuming) a product. The user's consumption of the product is represented by point $Q_1$ where $MPB = P$. The social optimum however would be less than this at point $Q_2$ where $MSB = P$.

Finally, figure 2.2 below illustrates the external benefits of consumption. This occurs in cases where the marginal social benefit is greater than the marginal private benefit and hence the $MSB$ curve will be above the $MB$ curve. In this case the actual level of consumption ($Q_1$) will be less than the social optimum ($Q_2$).
Figure 2.1 - External cost in consumption

Figure 2.m - External benefit in consumption
To summarise, whenever there are external benefits, there will either be too little being produced or too little being consumed. Whenever there are external costs, there will either be too much being produced or too much being consumed. Although, imperfect market structures and externalities are major reasons for the lack of efficiency, market failure can also occur due to a number of other reasons such as, ignorance and uncertainty on a firm’s part or the immobility of factors and their associated time lags (Eaton et al 2002). Under the assumptions of perfect competition, firms’ and factor suppliers have perfect knowledge of costs and benefits. In the real world however, there is a great deal of ignorance and uncertainty and thus people are not always able to equate marginal benefits with marginal costs. Similarly, market failure could occur as factors can be slow to respond to changes in demand or supply, which then has a knock on effect on prices, meaning the economy is in a constant state of disequilibrium and the long run never arrives (Eaton et al 2002).

So far the key aspects of three economic models have been discussed and their differences have been explained. These differences primarily exist because of differences in the structure of markets and they ultimately impact the levels of production. In the following section 2.4 below, the effect of production on wages is discussed and the concept of monopsony is introduced and detailed.
2.4 Wage Determination and Monopsony

The discussion above provides a basis under which wage determination can now be understood in perfect competition and other market structures. For a profit maximising firm in a perfectly competitive market, MC will equal MR, which itself equals and therefore represent Price (P). Equation 2.2 presented above represented the calculation for the determination of the marginal costs of an input. If the input is labour, marginal costs would be determined as in equation 2.6 below:

\[ MC = \frac{W}{MP} \quad (2.6) \]

Where MC is marginal costs, W is wages and MP is marginal productivity. Equation 2.6 can be rewritten as:

\[ P = \frac{W}{MP} \quad (2.7) \]

Where P is price, W is wages and MP is marginal productivity. Re-arranging equation 2.7 in order to determine wages gives:

\[ P \times MP = W \quad (2.8) \]

Which is the same as:

\[ MR \times MP = W \quad (2.9) \]

This could be expressed otherwise as:

\[ MRP = W \quad (2.10) \]

Workers therefore are paid according to their contribution to the production of output, which is then sold on the market. As with perfect competition, implications for the labour market can also be derived from the monopoly market structure. If the monopolist has no power in the market for inputs then wages should be set in this market as described above, recognising that price is not equal to MR. If the monopoly firm has power in the input market a monopsony position can apply.

In theory a monopsonist is a single buyer of goods in the market (Nicholson 2002). Monopsonists will therefore have significant control over the price they pay for
inputs. Furthermore, a monopsonist, as the sole buyer in the labour market will face the entire market supply curve of labour. In order for the monopsonist to increase hiring its labour by a single unit it will have to increase wages for marginal workers as well as the workers already employed. The marginal cost of hiring the extra unit of labour therefore, exceeds the firms wage rate (Nicholson 2002). This means that in figure 2.n (presented below) the monopsonist faces higher input costs per unit of output than in a competitive market.

The monopsony model therefore differs from the competitive model primarily because the MC of hiring additional employees comprises of the wage received by the additional employee and the increased wages paid to those already employed (Downward et al 2009). This ultimately means the marginal revenue product of labour (MRPL) will not equal and therefore represent wage rates paid. The higher wages needed to attract additional workers will also be paid to all other employees (Downward et al 2009). Figure 2.n below presents a comparison between competitive and monopsonistic labour conditions.

The line A1, A2 represents the perfectly competitive demand for labour (MRPL) defined over hours of work, while the line A1, A3 represents the monopsonists MRPL. The line B1, B2 is the supply of hours to competitive firms in the labour market. The line B1, B3 is the monopsonists cost of labour (MCL).

The perfectly competitive firm will maximise profit at the point V (intersection between B1, B2 and A1, A2), so this point V therefore represents the perfectly competitive equilibrium wage-hours combination. The monopsonists profit maximisation point is represented by point J (intersection between A1, A3 and MCL), but the monopsonists wage hours equilibrium is at point D (Where W2 and E2 meet). As can be seen in figure 2.n, the monopsonists MRPL (line A1, A3) lies above point D, indicating the exploitation of labour (Downward et al 2009).
As illustrated in figure 2.n above, wages, hours worked and the number of people employed will be lower in cases where firms have the power to determine prices. Naturally, as this monopsony power declines, wages and the number of people employed will increase. Figure 2.n above also illustrates the difference between the actual wages paid and an employee’s reservation wage. An employee’s reservation wage is the sufficient wage amount required in order to persuade an employee to work that particular hour (Downward et al 2009). It can be seen at point V in figure 2.n that the competitive wage rate at equilibrium is W0. Figure 2.n also reveals that many of these hours can be obtained at lower wage rates. If W0 is reduced the marginal hour will no longer be obtained but the intra-marginal hours will still be supplied, as the now lowered wage rate will still exceed an intra-marginal workers reservation wage (Downward et al 2009).
Different levels of wages will inevitably cause different levels of earnings. An employee’s reservation earnings level is also referred to as their transfer earnings. This is the minimum level of earnings for which an employee would work for any particular firm. If a worker receives less than their transfer earnings then the worker will likely quit. Any earnings above their transfer earnings (or in excess of their reservation wage limit) provide their recipients with “economic rent” which similarly to consumer surplus is a pure welfare gain (Downward et al 2009).
2.5 Individual and Team Productivity

In this section of the chapter, the focus shifts from firm behaviour to individual behaviour and contributions to understanding a worker’s productivity will be examined more exclusively (irrespective of which market structure they may work in). This section outlines and then discusses factors that can potentially influence the exertion of effort and hence the productivity of an individual. Lazear (1998) argues that many factors such as incentives, team work, work and life balance, benefits and authority all play a large role in determining a worker’s productivity level. Some of these factors are discussed in greater detail in the subsections below. However to begin with, an individual production function will be defined.

2.5.1 Individual Production Function

Conceptually individual production functions are different than the generalised production function discussed above. Generally in economics, the individual production function refers to how an individual allocates their time and their effort, which are both considered inputs, in order to produce outputs that vary as workloads differ across personnel and types of job (Coviello et al 2011). An individual’s production function is particularly important in analysing how much a worker is worth and can be of great economic significance to a firm (Lazear 1998). The economic models detailed above abstract from this difficulty. In an obvious sense, knowing how good workers are would allow firms to hire and build a workforce full of efficient workers and avoid a workforce with low productivity. This would therefore aid companies in their attempts to maximise profits.

Determining an individual’s productivity can be expensive and hence difficult. This then begs the question when does attempting to determine individual productivity justify the expenses incurred in doing so? The answer to the question above depends on whether the information gained on an individual can be kept confidential (Lazear 1998). If certifying a worker as having ability, in turn results in all competitors to the firm being informed of the workers value, then the wage the firm will have to pay the worker will rise due to market pressure. The firm may not be any better off by having learned their workers ability.
Obtaining information on an individual’s productivity can be beneficial to a firm as it can allow firms to allocate existing workers in divisions to which they are suited and use the findings of their current workforce as a benchmark of comparison to which they will then be able to hire new workers who can be “screened”. The concept of screening refers to a recruitment procedure that many firms incorporate in order to determine whether potential new employees have the required skill levels to sustain a position of employment with the firm and ultimately increase the firm’s productivity (Lazear 1998). Many of the big firms in a particular industry will often outsource this screening responsibility to a recruitment firm as this is more often than not a much more cost effective method to filter out the “weaker” candidates (Lazear 1998).

Sometimes effort alone is not the sole contributor to performance and productivity as luck can occasionally play a role (Barney 1986; Lazear 1991; Bertrand and Mullainathan 2001; Garvey and Milbourn 2006; Gopalan et al 2010). As luck becomes relatively more important, effort has a smaller effect on the chance of an individual winning promotion at their firm. Therefore the returns to effort decline, and a larger spread in salary is required to induce effort. Another factor that can contribute to the levels of individual productivity is the amount of human capital an individual acquires and this is detailed further in the subsection below.

2.5.2 Human Capital

Human capital refers to the acquisition of skills and knowledge, which are principally a product of deliberate investment (Schultz 1961). Human Capital theory is based on the premise that individuals can acquire human capital by investing time and money (including deferred earnings) in education, experience, training and other qualities that ultimately increase their productive capacity and hence their worth to a potential employer.

There are many ways for individuals to invest in human capital including, schooling, on the job training, medical care and acquiring information about the economic system (Becker 1962). Each of these differ in their relative effects on earnings and consumption, in the amount of resources invested, in the size of returns and in extent to which the connection between investment and return is perceived. All of these
however can improve the physical and mental abilities of people and therefore raise income prospects (Becker 1962).

According to Lazear (1998), the two main investments an individual can make in human capital come in the form of education and on the job training. It could be argued formal schooling pays for itself as previous studies have revealed a positive relationship between income and education (Becker 1992). It is important for an individual to consider the costs of education before the decision to stay at school/college is made. Lazear (1998) suggests two main costs involved with formal schooling that should be considered by individuals. The first expense is the direct costs of education (tuition fees, books, living expense etc.). The more important cost is the opportunity cost. For example, an individual may give up a placement at a firm that offers substantial compensation in order to gain an MBA qualification. While the individual studies for the MBA many years of salary will have been sacrificed and the individual needs to decide whether the acquisition of the MBA will allow for higher compensation in the future therefore offsetting the opportunity costs of having returned to education.

Although education can have an effect on productivity, too much education (i.e. atypical prolonged higher education) will most likely not increase productivity by too much, if at all. An individual should therefore leave education and pursue potential job opportunities at the point throughout their lives when further education is unlikely to increase their prospects and productivity. Otherwise the opportunity cost of staying in education will be higher than the compensation received from finding a job later on after schooling has ceased (Lazear 1998).

On the job training takes two forms, both of which have different implications for productivity and hence earnings. The first type of training is general training. General training raises productivity at the firm providing the training as well as other firms in the same industry (Lazear 1998). The second type of training is firm specific training. Firm specific training makes a worker more productive at the firm where they are employed, but will have no effect upon their productivity should the worker leave and seek employment elsewhere (Lazear 1998). Becker (1964) argues that firms have much stronger incentives to invest in firm specific training as it can solely be of benefit to an employee’s current firm. This would likely reduce the chance of a
worker quitting their position once they had been trained, as their skills would not be applicable elsewhere.

Generally, the investment into human capital is considered to have a positive impact upon individual productivity as the individual becomes more knowledgeable and consequently enhances their cognitive and or non-cognitive ability. Cognitive psychology is concerned with a variety of mental activities such as knowledge representation, memory, thinking and perception, all of which are related to human information processing and problem solving (Shuell 1986). Carroll (1993, p10), defines a cognitive ability as one “that concerns some class of cognitive tasks” where a cognitive task is “any task in which correct or appropriate processing of mental information is critical to successful performance” (Carroll 1993, p10). Much early research on the acquirement of skill and investment in human capital has focussed almost exclusively on measures of cognitive abilities such as, academic achievements (qualifications), exam scores and IQ tests (Heckman and Rubinstein 2001).

Non-cognitive skills refer to a broad range of human characteristics and personality traits such as, self discipline, tenacity, reliability, motivation, confidence and perseverance all of which are known to influence levels of success (Heckman 1999). Non-cognitive skills are increasingly considered to be as important, or even more important than cognitive skills in explaining academic and employment outcomes (Heckman et al 2006). Recent studies across a wide variety of disciplines have established a relationship between non-cognitive skills and academic outcome (Jencks 1979; Bowles and Gintis 2002; Farkas 2003; Heckman et al 2006). Furthermore research has revealed that investing in the development of non-cognitive skills could potentially yield high returns in future educational and employment outcomes and helps reduce the attainment gap between advantaged and disadvantaged youths (Heckman et al 2006).

It could be said all employees bring a certain level of skill and experience to their performance of a specific task, however it is the accumulation of educational qualifications and relevant experiences in the field that give rise to the differentiation in the level of reward (compensation) necessary to secure and retain certain workers. In addition, the growing number of specialist fields that require workforces with
certain skills, knowledge and experience has increased the need for employee reward systems that go beyond simple market clearing systems (Lazear 1998).

2.5.3 Moral Hazard and Incentive Contracts

Moral hazard is a concept which could be defined as a:

Form of post contractual opportunism that arises because actions that have efficiency consequences are not freely observable and so the person taking them may choose to pursue his or her private interests at others’ expense. (Milgrom and Roberts 1992, p167)

In other words moral hazard could describe situations where individuals may perform inefficient actions due to personal interests, which would then have knock on consequences on efficiency. The efficiency problem this creates is that these extra benefits enjoyed by the insured are not always worth their costs (Milgrom and Roberts 1992). The concept of moral hazard originated in the insurance industry (Pauly 1968; Arrow 1970b) and it is a problem that came to light because insured individuals will often alter their behaviour in order to enjoy the benefits of their insurance. Ultimately, acquiring insurance reduces an individual’s incentives to avoid the insured against event or accident, giving rise to moral hazard. In cases where the moral hazard problem is strong insurers often offer limited or even no insurance (Stiglitz and Driffill, 2000). The moral hazard problem is effectively an information problem and Milgrom and Roberts (1992) argue that it is a problem that has been created due to the difficulty and cost of monitoring and enforcing appropriate behaviour.

One potential method to combat the problem of moral hazard is to use incentive contracts. Incentive contracts are designed in a special way, intended to make an individual accountable or at least partly responsible for the consequences of their actions (Milgrom and Roberts 1992). In the insurance industry, incentive contracts were devised so that insured individuals would only be partially covered, therefore exposing the insured individual to financial risk in the event of a loss (Shavell 1979). By exposing the insured individual to some financial risk, the individual becomes more motivated to prevent loss, thereby negating potential moral hazard (Stiglitz and Driffill, 2000). Moral hazard can also cause issues when devising employee contracts. When firms hire or promote employees they may sometimes devise incentive
contracts in order to provide the worker with incentive and also insulate people from risk (Milgrom and Roberts 1992).

Incentive contracts are commonly offered in the labour market in order to allow workers to be solely responsible for their performance, consequently enabling workers to be in control of future compensation levels and promotions (Lazear 1998). However, according to Milgrom and Roberts (1992) workers may on occasion be reluctant to take the risk of being in complete control of their compensation, therefore these contracts are offered to workers at a cost. Efficient contracts must balance the costs of bearing the risk against the incentives that are gained as a result (Milgrom and Roberts 1992).

2.5.4 Contract Design and Incentive Contracts

Some workers working in mid-level positions may have a low chance of promotion but also a low chance of being fired. In these scenarios motivating workers can become problematic, therefore contracts can and may contain unique incentives. A worker’s wage profile can take many forms as illustrated in figure 2.p below, which represents a worker general wage model.
A worker who puts a lot of effort in will produce output $V$. As the worker’s experience increases output is also expected to rise up to a point and then decline as the worker nears retirement age $T$ (Lazear 1998). Alternatively, a worker can put little effort into their work and produce output $V^*$ (a lower output throughout their working career). The path labelled $\text{Alt}$ represents a worker’s alternative use of time. As workers near retirement age the best alternative use of time is likely to be leisure, the point $T$ therefore represents the retirement age. The steepest wage profile is the curve $W$. This curve is drawn such that the present value of $W$ from time zero to time $T$ is exactly equal to the present value of $V$ from time zero to time $T$. A worker with profile $V$ would always earn the value of their output. A worker with profile $W$ however would almost never receive the exact value of their output. Until time $T_0$, the worker would receive less than their worth. After time $T_0$ the worker would receive more than their worth, but over the entire lifetime, their wage payments would add up to the value of their output (Lazear 1998).
The reason for presenting these wage profiles in figure 2.p is to clearly illustrate that not every wage profile is the same. Although the present values of W and V are the same (at time T0), steeper wage profiles (curve W) will lead to greater effort levels and greater motivation for workers (Lazear 1998). Lowenstein and Sicherman (1991) claim that workers have a preference for steeper profiles as this provides workers with favourable anticipation of regular wage increases rather than the trepidation of steady or falling wages. The design of a contract with wage profile W, would therefore provide workers with more incentive to continue working with high effort levels up until retirement age T than the flatter wage profile V (even though total compensation would be the same with both curves).

2.5.5 Tournament Contracts

Sometimes contracts are designed in a “winner takes all” format. This is more commonly referred to as the tournament theory. Tournament theory first arose in labour economics literature (Lazear and Rosen 1981) and it is prevalent in management research, where it has been commonly employed to explain compensation structures (Leonard 1990; Main et al 1993; Eriksson 1999; Messersmith et al 2011). Tournament theory has since expanded to a wide range of other disciplines and its principles have been applied to the likes of lawyers (Price 2003) and drug dealers (Levitt and Dubner 2009).

The idea of a tournament is that firms induce effort from employees by effectively pooling some portion of wages from all the employees at one rank into wages at the next highest rank, giving each opportunity to win promotion to that rank (Connelly et al 2014). Tournaments are therefore conceptualised as contests that are designed to entice an optimal level of effort from several actors, each vying for a prize that will be awarded based on relative rank (i.e. performance) (Lazear 1999, Connelly et al 2014). The prize is optimal, when it maximises the productive output of the tournament including all participants (Lazear and Rosen 1981, Knoeber 1989, Knoeber and Thurman 1994, Connelly et al 2014).

According to Szymanski (2003), tournament theory has its roots in the seminal work of Tullock (1980), which explores “rent seeking” behaviour by “contestants” for public funds. Rent seeking behaviour refers to the behaviour and actions of economic
agents extracting a return as a factor of production that is not a directly connected to their productivity, but more a consequence of their control over resources as a result of regulations or other constraints (Downward et al 2009).

Tournament theory consists of two main components. Firstly, prizes (pay level of promotions) are fixed in advance and are independent of absolute performance (Lazear and Rosen 1981). So if a vacancy became available for the position of CEO\(^\text{11}\) and there were only two contenders for this vacancy, it would not matter how much of a better candidate the eventual winner was, all that matters is that the winner was a better candidate. Secondly the prizes will not be awarded because the eventual winner has been better over a specific period of time, but will be awarded instead because the winner was better at that particular moment in time. In other words, absolute performance is irrelevant but rather relative performance is what matters (Lazear and Rosen 1981; Lazear 1998).

Tournament theory maintains that relative rank order prizes are superior to pay for performance compensation mechanisms, because they motivate a broader base of employees striving for promotion, rather than focus on a single individual. In competitive industries where firms may require a small group of “specialists” to perform specific duties and to provide a solid foundation on which the firm can be built around, tournament theory suggests the preferred choice to guide recruitment. Instead of the steady upward progress in the relationship between salary and experience which is present in other labour market theory (represented in figure 2.n above), tournament theory is perceived to contain “big jumps” where the “jumps” have represented a promotion and the “big” represents a “prize” (boost in salary) resulting from the promotion.

Tournament theory carries the assumption that there are distinct differences between the very best, the ‘winners’ and the next best, ‘the losers’ (Lazear 1998; Connelly et al 2014). The differences in core ability between the winners and losers may be worth a lot to employers and the argument of tournament theory is that these differences are rewarded substantially. A newly promoted CEO may have just acquired a big increase in salary and it could be argued that this role requires the largest pay rise as there is no more progression an individual can make within the company. It could also be argued

\(^{11}\) A chief executive officer, the highest-ranking person in a company or other institution, ultimately responsible for taking managerial decisions (Oxford Dictionary 2010).
however that this salary is not designed to motivate the CEO and that the newly promoted individual’s performance is completely irrelevant, but rather this salary increase is intended to create competition amongst the several subordinates who were competing for this role when it was previously available.

Although tournament theory creates incentives (Lazear and Rosen 1981; Lazear 1998, 1999; Fredrickson et al 2010; Connelly et al 2014), there is a limit to the rewards a company can offer (Lazear 1998; 1999). The optimal level of incentive is attained as a result of trading off the value of additional effort against the compensation necessary to induce workers to supply that level of effort. In other words, it may be in some cases more useful or productive for firms to settle for a lower equilibrium effort level than to pay them lots more to raise effort only a proportionally small amount (Lazear 1998). At some point, it simply does not pay to induce greater effort levels from a firm’s point of view.

If prize spreads are too small, contestants will not have enough incentive to compete causing the total productive output of the tournament to drop. Similarly a prize spread too high can also be detrimental to tournament efficiency as it induces so much effort that contestants must be broadly compensated (Connelly et al 2014). Tournament design therefore involves strategically choosing optimal prize spreads that maximise productive output of the tournament. From an employer’s point of view the value of offering large spreads depends on the value of the extra output produced. Naturally, if the value of the additional output exceeds the value of the large spread in wage then it will be deemed worth it. However, even if large spreads fail to produce output of equal or greater value, the large spread may still need to be offered as employers are competing with each other in competitive labour markets. Therefore in labour markets where supply may be short, employers may be forced to offer large spreads (Lazear 1998).

In organisations that require some degree of team effort, managers may decide to compress the prize rewards. Conversely if the organisation believes that only a few members of the entire workforce are likely to produce the most value then managers may set these prize spreads far apart. Generally with tournament style pay systems, individuals are encouraged to compete for these prizes with the large spreads providing motivation, as opposed to emphasising group co-operation (Lazear 1998).
As mentioned earlier in section 2.4.1 of this chapter, luck can occasionally play an important role in an individual’s productivity and performance and hence could also influence who ultimately wins a tournament (Lazear and Rosen 1981; Lazear 1991, 1998; Connelly et al 2014). Luck, good or bad can result from one of two main factors, production uncertainty and measurement error (Lazear 1991, 1998). Production uncertainty refers to when a worker has put in a lot of high effort into a task but their output has been low. This low level of output can on occasion be down to bad luck on the employee’s part as it could have been caused by an uncontrollable or unanticipated event. Luck however can also be positive, as on occasion a worker may be able to profit from the same unforeseen or unanticipated event.

Measurement error refers to when a worker may have been exerting high levels of effort but may have been unjustly or inappropriately ignored or perceived to have exerted medium effort. A superior may even then attribute high effort to another worker who has in fact produced very little. Both of these types of luck will reduce the value of effort as they reduce the probability of a worker being promoted. Ultimately, tournament pay styles are designed to reward workers for their level of effort, however if luck plays a prominent role in the decision of promotions then the link between effort and rewards will be broken, inevitably causing worker effort to decline.

2.5.6 Effort and Efficiency Wages

The efficiency wage theory revolves around the notion that managers will strategically offer work contracts that will achieve more efficient employment in the medium term. The efficiency wage hypothesis states that a worker’s productivity is positively correlated to the wage they receive. In such cases, the firm may find it profitable to pay wages in excess of market clearing (Katz, 1986).

There are four main features of efficiency wage mechanisms that could cause increased productivity. Firstly, offering higher wages reduces ‘shirking’ (deliberate slacking), particularly in cases where individual effort is difficult to observe (Shapiro and Stiglitz 1984). Secondly, efficiency wages can raise worker morale, which in turn can trigger a raise in the worker’s effort level (Akerlof 1982). Furthermore, efficiency wages can reduce the number of voluntary resignations and hence reduce labour
turnover costs (Salop 1979). Finally, offering higher wages will inevitably attract a greater number of applicants (Weiss 1980) enlarging the pool of candidates from which the firm can select the most suited applicant.

The efficiency wage theory carries a disciplinary aspect, which helps to prevent ‘shirking’ (Perkins and White 2008). Shirking is a term used to describe situations where employees may be deliberately withholding their full effort levels (Jones 1984). The argument for the efficiency wage theory is that by firms offering higher than normal wages, not only will the employee feel inclined to work harder but also their performance will be conditioned by the fear of losing their employment which currently offers above market pay.

Firms may choose to offer wages above the market clearing level to simply motivate the workforce and create a ‘feel good’ factor amongst employees (Lazaer 1998). Consequently, this creates the impression that the employer is a ‘good’ employer that cares about the welfare of its employees. In turn, this could result in workers becoming more industrious and more willing to adapt to future changes that may include the introduction of technology and hence retraining programmes.

Occasionally, employers worry that their employees’ will use their human capital to secure work at enhanced pay rates with rival or competing firms. In cases where these employees were trained or retrained by the firm (on the job training) the loss of a worker comes at a significant cost to the business (Salop 1979). Turnover costs consist of direct costs such as formal orientation programs or indirect costs such as the lowered productivity during the initial ‘settling in’ period (Salop 1979) and when a worker leaves, these turnover costs must be borne again (on the replacement). Labour turnover and hence its associated costs are usually greater than paying above market rates, thus paying higher rewards from the off is a rational response by an employer in order to sustain an on-going relationship. In a sense, efficiency wages are an investment by the firm in an employee’s firm specific skill and knowledge, designed to maintain a degree of loyalty and minimise opportunistic employee behaviour (Lazear 1998).

The efficiency wage model also introduces the sorting effect, which refers to a method of recruitment. In those organisations where production is complex and requires certain levels of skill and knowledge, specific high quality levels of workers
will be needed. Firms may therefore find it efficient to use their above market pay rates to signal their intent to attract only the top, above average quality candidates for vacancies. As successful candidates will be highly skilled or experienced to begin with, managers will not be required to waste time and effort supervising them, unlike workers with low ability, that would require supervision in order to ensure the required level of organisational output. Friedman (1977) refers to this as “responsible autonomy” and states a big problem associated with this management strategy is that because managers devote time and effort supervising low skilled workers, their own productivity suffers. Rubery (1997) argues self discipline on an employee’s part, coupled with the offer of above market pay rates may ultimately be a much less costly option for the firm than funding additional supervision, providing firms with further incentive for the use of efficiency wages.

2.5.7 Team Production

Teams exist because some firms believe their productivity can be increased by working in teams as opposed to working individually, or because of the nature of the product. In order for individuals to work together in teams efficiently, firms must know how to allocate and set up teams and motivate team members. For instance, along an assembly line, workers will be assigned a specific role and all the work will take place under one roof. The output from this assembly line will be attributed to all workers on the line, similarly when something goes wrong with the product all members of the assembly team suffer.

Measuring a team’s production function is not that difficult as team output can be easily observed, however what can be difficult is observing individual output within the team (Meyer 1994). Sometimes it may not even be possible to observe the output of any given worker, which then opens the door to the free rider effect (Alchian and Demsetz 1972; Holmstrom 1982; Blair and Stout 1999). The free rider effect defines scenarios where workers may hide behind the ability and successes or failures of the group as a whole. These situations will inevitably lead to a dilution of incentives (Lazear 1998).

Firms should only implement teamwork in cases where the advantages to the firm outweigh the disadvantages (Lazear 1998). The reason for this is that teamwork can
lead to two main problems. If workers are paid on the basis of team output, free riders will not put forth enough effort. If workers are paid on the basis of individual output, they will ignore the rest of the team. This is because a worker who is paid dependent on individual performance will have incentive and motive to base their decisions in a selfish manner to the detriment of their team and ultimately the firm.

The free rider effect that teams are susceptible to, can be controlled to a certain extent by controlling the size of the team (Kandel and Lazear 1992). In partnerships as there are only generally a few team members, each worker has significant information on what work each other partner is doing. On this basis, if one partner shirks, the partnership would either breakdown altogether or the shirker would be expelled from the partnership. In larger teams peer monitoring becomes much more difficult as a shirker would have less of an effect on any other member of a team (Lazear 1998). As a result, incentives to punish the shirker are reduced.

In some firms teamwork can increase productivity because of the transfer of knowledge that can occur between workers (Shaiken et al 1997). Sometimes individual workers within a team can be required to perform separate tasks. Although these individuals have different roles there can on occasion be a small area within their roles that overlap and therefore coincide. In cases such as these workers are able to work together and in the process transfer knowledge, ultimately making each of them more competent (Lazear 1998).

Team production can sometimes be increased or at least encouraged by firms offering some incentives. Firms could offer teams bonus payments (on top of standard pay), which would increase workers compensation. These rewards are typically offered in one of two ways, either as explicit rewards or implicit rewards (Lazear 1998). Explicit rewards include things like team bonuses or profit sharing plans. Team bonuses will generally be given to a small team for performing a specific function efficiently. Small team incentives offer employees a more direct ‘line of sight’ between performance and rewards than in large teams and are therefore more powerful motivators (Vroom 1964; Lawler 1981; Wilson 1990; DeMatteo et al 1998). In other words, teams must be small for bonus payments to have effect. Teams that are too large will give rise to the free rider effect and the incentive for each individual to perform will therefore decrease.
Profit sharing refers to bonus payments that are made out of company profits. This gives some small incentive to all members of a team (even if it’s a large one) to perform as their own compensation could be increased if the firm has a good year. These bonus payments will however normally not be shared in equal amounts (as the team bonuses described above most probably would be) but instead are allocated to workers proportionally to their annual salary (Lazear 1998). This type of reward scheme is stated in terms of an incentive scheme but in truth it does not offer much motivation to workers (Lawler 1981; DeMatteo et al 1998), as their own individual performance is likely to have a very small effect on a company’s performance. This scheme is therefore more of a risk spreading tactic. The company profit sharing system is a way in which shareholders can share risk with workers in a firm as their own returns will not fall as much in downturns nor rise as much in boom periods as they would if worker wages were constant (Lazear 1998).

Implicit rewards also tend to be in the form of workers taking a share of the company profits however the proportion allocated to them is much larger than the payments made explicitly. This is because these implicit reward payments tend to be for workers who can single-handedly have a large effect on the company’s profits (Lazear 1998).

The rotation of team members can potentially lead to an increase in team productivity (Lazear 1998; Bennett 2003; Origo and Pagani 2008; Saravani and Abbasi 2013). Rotating team members can increase individual knowledge and experience (Noe et al 2009) and improve human capital (Origo and Pagani 2008). Job rotation also allows management to evaluate a worker’s capabilities and therefore enables them to allocate workers to those departments to which they are best suited (Lazear 1998). When teams are switched, the period of the switch should be fairly long before the next rotation. Workers require time in order to build up relationships and allow communication channels to open so knowledge transfer can actually take place. Splitting up teams too soon could also have a negative effect on team morale (Lazear 1998).
2.5.8 Summary

As detailed above, there are many factors that can influence individual and team productivity. The acquirement of knowledge and experience in the form of human capital has a positive impact upon individual productivity as in most cases it enhances an individual’s cognitive and non-cognitive ability. In addition, the design of employment contracts also impact levels of production as the rewards and incentives these contracts offer together with potential post contractual issues such as moral hazard, collectively manipulate effort levels exerted by individuals. Furthermore, although there are many difficulties in attempting to determine individual productivity within a team, teamwork can raise overall efficiency and hence productivity due to the transfer of knowledge. Section 2.6 below moves the discussion on towards the peculiar economics of sport (Neale 1964) and explains how the features of labour economics detailed above play a significant role in influencing the production levels within sporting contests.
2.6 Economics of Professional Sport

Economists have been expressing a greater level of interest in professional team sports. This greater interest has been partly generated by the ever-increasing levels of commercial activity within sports (Slack 2004) and partly due to the belief that the characteristics of sport and in particular professional team sports lend themselves to economic analysis (Downward and Dawson 2000). The purchase and sale of players, as well as payments to players in order for them to perform require financing decisions. Clubs must co-ordinate match schedules since they cannot produce in isolation and spectators must be made aware of where matches occur. Potential spectators need to be accommodated in purpose built stadiums enabling payment to be extracted from them whilst restricting access to the sport to non-payers (Downward et al 2009).

Characteristics such as these indicate that rather like the production, distribution and consumption of other goods and services, professional team sports can be viewed as an economic process. Professional team sports comprise of leagues, in which clubs will compete on the field in accordance with the rules of the sport and by arrangement through a fixture list, set by the leagues. Success in professional team sports is usually determined by the rank order of clubs in the league at the end of a season after points have been allocated for their performance in their set of fixtures (Downward and Dawson 2000). This applies to sports at both amateur and professional level, however what differs at professional level is the money that changes hands in the production, distribution and consumption of the sport (Downward and Dawson 2000).

In amateur sports participants can be viewed as consumer-producers as voluntary time, membership fees and other expenses are combined in order to fund the supply of a sports event in which they typically participate (Downward et al 2009). In the early days of baseball in the US, contests between clubs would only occur infrequently, often coinciding with large social events and typically between relatively close communities (Szymanski and Zimbalist 2005). However in order for baseball to grow and develop as a sport, formalization was required which would define the nature of the sport (outline and set rules) and organise contests. In economic terms this meant that baseball contests needed to be defined and made sufficiently homogenous in
order to create a potential large-scale demand, which could then be supplied (Downward et al 2009).

When shifting from voluntary to professional sports, changes in the form of consumption and production occur due to the formalization of the sport (Downward et al 2009). Formalization refers to the extent to which mechanisms such as rules, regulations, policies and procedures govern the operation of a sport organization during a shift from amateurism to professionalism (Slack and Parent 2006). The formalization of sport has three major consequences. Firstly, the formalization of a sport gives rise to the development of organized competition. For example, the formation of the Football Association (FA) in 1863 allowed the formal unification of clubs within one national structure, whereas previously clubs were governed by more informal county-based organisations. This unification allowed the formation of knockout cup competitions where local rivals could compete, raising public interest in the sport and hence prices.

Formalization also provides incentive for clubs to pay players in order to extract revenue from paying spectators. According to Downward et al (2009) increasing demand for homogenous sports’ contests raises the incentive for, and possibility of organizing and developing teams for commercial motives. In order to access payment by spectators, stadiums are built, which in turn gives rise to the movement towards professionalism, which is that it provides incentive to pay players, i.e. to attract the best talent to a team and enhance the prospects of success.

Finally, formalization leads to the formation of leagues within which teams can regularly compete in order to extract more regular revenue from paying spectators. In the history of English football this final movement towards professionalism occurred with the formation of leagues where professional players were paid as a means to extract more regular gate receipts and fully utilise facilities. In the case of the English Football League, the FA retained control of football in terms of development, rule interpretation and the parameters of commercial activity (Downward et al 2009). In addition, the FA controlled movement of players between clubs, their salaries, dividend payments to shareholders and ticket prices. This governance allowed the league to be developed, which then soon flourished and expanded (Downward et al 2009).
There are distinct similarities in old team sports such as baseball in the US and rugby and football in England in the shift to professionalism. Consumer producers that previously operated in an environment where clubs were effectively self-financed from members have developed in to hierarchical organisations, governing activities in an environment where professional producers and distinct consumers (gate paying spectators) exist (Downward et al 2009).

Despite the similarities however, the specific approaches to professionalism across these sports had different characteristics. In baseball, organisation developed from commercially minded owners looking to protect their own economic interests. In England, leagues in football and rugby developed under the governance of one structure that each aspired to cover professional and amateur levels of the game. As opposed to closed league formats adopted in the US, English leagues have consisted of a vertical structure of competition (adopted throughout Europe) that includes access to professional status (Downward et al 2009).

A further distinction between the US and Europe are the objectives of club owners, where the motives of owners in the US appear to be driven by the acquisition of profits rather than ensuring financial survival and club success as in Europe. Throughout the history of football and rugby league the absence of profits are accepted as a feature of business reality (Szymanski and Kuypers 1999; Downward and Jackson 2003). In contrast, US sports have tended to make profits (Fort 2000; Szymanski and Zimbalist 2005).

2.6.1 The Role of Sports Clubs

Sports clubs are examples of club goods, where individuals will choose to become part of the collective group to consume-produce both the sports activity and membership to enhance their utility (Downward et al 2009). The club then acts to internalize the externality of non-rival benefits by controlling access to the sport, facilities, and equipment and reputation benefits provided by the club.

Of specific importance to sports clubs is that the consumption-production of the sports activity requires the sharing of production costs in supply, not only with regard to running or administration of the club but, more fundamentally, sport is a prime example of joint production, where individuals are competing against others.
Competitors need to cooperate to facilitate sport, either within the club or across clubs as part of a sports system. Team sports extend this basic principle to the case in which opponents operate as groups rather than as individuals. It takes two teams to produce an individual sporting contest and many teams to produce a tournament (Kesenne 2007). As mentioned above competitors need to cooperate to facilitate sport and this is a basic principle that pertains to individual sports as well as extending to team sports. Equations 2.5a and 2.5b below present these relationships in the form of production functions.

\[ Ct_{wc} = Ct_{wc} (L, C, CO_A, CO_B) \]

(Equation 2.5a)

\[ Ct_{12} = Ct_{12} (L_{c1}, C_{c1}, CO_{c1}, CO_{c2}) \]

(Equation 2.5b)

Equation 2.5a depicts a sports production function within a club (subscript “wc”) where the contest (Ct) requires the club’s land (L), capital (C) and the efforts of two competitor (CO) members, A and B. Equation 2.5b depicts a sports production function between two different clubs, c1 the home team and c2 the away team. Here, the contest (between opposing club members) is a function of the land and capital of one of the clubs (depending on which team is the home team), in this case c1 and the efforts of both competitors from each club c1 and c2.

As demonstrated by equations 2.5a and 2.5b above the production and hence supply of a sporting contest in professional team sports requires joint contributions from two independent competitors (Jeanrenaud and Kesenne 2006). Alternatively, in team sports this production process could be described as an inverted joint production process where two production processes, one by each competing team, is required in order to supply and produce a single game (Kesenne 2014). This principle can therefore be extended to football where clubs combine labour, capital and land (players and stadiums) in order to supply teams, which enable the production of the saleable product, the game itself (Downward et al 2009). As emphasised by Kesenne (2014), team independence is a necessary requirement in this production process as it maintains contest legitimacy, ensuring sporting contests are perceived to involve teams that are competing to win.
2.6.2 Sports Competitions as Economic Contests

Sports contests can be understood as economic tournaments (Szymanski 2003). As outlined earlier in section 2.5.5 tournaments are concerned with economic activity whose output is assessed in relative rather than absolute terms and it is because tournaments produce a rank ordering of outcomes that makes them applicable to sport (Downward et al 2009). In a sporting contest of any type, (a knock out competition, a league championship, an athletic race) the winner will be the one that finished the contest above their opponents in accordance with a ranking rule that links the performance of the competitor to their assigned rank order. Competitions of this type measure competitor contribution in an absolute manner.

In many cases, however, the performances of competitors can be measured in relative terms i.e. how one competitor performed against another in a specific context. The precise manner in which this takes place can vary and usually follows either a knock out competition, a round robin league or some hybrid version. The reason why the round robin competition and its hybrids are more likely can be understood as a direct result of economic incentives (Downward et al 2009). These specific competitions require pooling together the better talent (restricting the scale of the competition), attracting better quality competitors and rewarding genuine effort rather than luck (which is more likely in knockout competition with randomly matched competitors).

As mentioned in section 2.5.5 earlier, rent seeking behaviour refers to economic agents extracting a return as a factor of production that is not connected to their productivity but rather a consequence of their control over resources as a result of regulations. Monopoly power could be viewed as rent seeking, particularly if the monopoly exists as a result of regulation. Traditional forms of labour market in sports, where reserve clauses and maximum wage caps have existed could also be viewed as rent seeking behaviour as these regulations restrict player mobility and wages, relative to what voluntary exchange would produce.

In the MLB, between 1880 and 1976, players’ contracts included a reserve clause, which ensured that upon expiry of the contract, their team retained the option to renew the player’s contract for another year. This clause served to limit a player’s freedom of movement by binding them to their present employers. This effectively
created a monopsony in the player’s labour market, as each contracted player could only negotiate with one potential buyer of his services. Salary caps impose a maximum amount payable to a player and can also lead to the monopsonistic exploitation of players’ particularly when players’ receive salaries below their value to the team. Both of these regulations ensured rents were directed away from players and towards teams (until they were challenged and subsequently relaxed). The principle-agent relationship also provides opportunities for rent seeking behaviour. Ultimately, rent seeking is inefficient as it is associated with the transfer of incomes at the expense of potential welfare gains (Downward et al 2009).

Tournament theory can indicate how hard a contestant is likely to work relative to their opponents, adding to the potential for their success and the overall quality of the competition as this effort is balanced against opponents. This process involves an economic trade-off between the competitor, the agent who needs to be enticed in to taking part in the contest and the organiser of the contest, who seeks an economic return from organising the activity, but also needs to employ the competing agents. In addition, tournament theory indicates that a balance needs to be achieved between relative contestant effort and randomness in affecting outcomes because of joint production (Downward et al 2009).

Ultimately the greater the value of a tournament, the greater the efforts of the contestants will be and consequently the higher the quality of the tournament and potential economic return (Downward et al 2009). This explains why tournament designs have shifted from knock out competitions to multi stage formats and leagues. A tournament that rewards effort rather than luck such as a round robin style tournament or league will likely be of greater quality and corresponding economic value than the traditional knock out competition where the element of chance has a significant bearing upon success. League fixtures are repeated trials that are more likely to reveal genuine efforts that underpin commercial viability (Downward et al 2009).

It follows that tournament organisers have incentives to promote effort in such a way as to please spectators. Downward et al (2009) reveal, therefore, several examples of how tournament regulators have adjusted tournament rules in order to elicit greater efforts, such as the amendment of the ruling awarding three points for a win as
opposed to two in football. Similarly, in rugby union competitions such as the Heineken cup, bonus points are awarded for scoring more than four tries or for maintaining losses to within seven points. These are both attempts made by tournament regulators to provide greater incentive for attacking play and therefore raise the overall quality of the spectacle in order to induce greater spectator interest.

2.6.3 Peculiar Economics of Sport

Another reason why tournament organisers might make rule changes is to promote the uncertainty of a contest. Seminal contributions to literature in the economics of professional team sports such as Rottenberg (1956) and Neale (1964) have emphasised the importance of a key externality known as the uncertainty of outcome hypothesis. The uncertainty of outcome hypothesis (UOH) was first introduced by Rottenberg (1956) who argued that, other things being equal, the closer the competition between two competing teams in any sporting event, the greater the interest in the event and therefore the greater the attendance and revenue, thereby illustrating the joint production element in sports.

Neale (1964) further emphasised the importance of UOH and provided a boxing example now commonly known as the Louis- Schmeling paradox. Neale (1964) states that Joe Louis:

Wants to earn more money, to maximise his profits. What does he need in order to do so? Obviously a contender, and the stronger the contender the larger the profits from fighting him… since doubt about the competition is what arouses interest… Pure monopoly is a disaster. Joe Louis would have no one to fight and therefore no income. (Neale 1964 p. 1-2)

In other words, Neale (1964) proposed that earnings for heavyweight champion Joe Louis would increase if he fought against an evenly matched opponent rather than a relatively weak contender. This quote from Neale (1964) also illustrates the joint nature of production in sports, where contests depend on the exertion of efforts and economic value is likely to increase in cases where greater joint effort occurs between two evenly matched opponents with comparable skills. In theory, public interest in sport and hence revenue and attendance will increase when teams are as closely competitive as possible, all other things equal. For example, close competition between two teams (team A and team B) in a league system would benefit the entire
league as a whole as not only would attendance at fixture AB increase but also attendances involving other teams’ in the league, team C and team D (hence the externality). On the other hand, domination of the league by team A would likely reduce interest and attendance at games involving the remaining teams. Although, attendance in games involving team A may increase, in the long run team A would also suffer if the overall standard of competition declined.

This highlights the potential importance of competitive balance to sports leagues management if maintaining competitive balance can increase uncertainty of outcome (Dobson and Goddard 2001). On this basis traditionally league organisers have regulated the product and labour markets as discussed below in section 2.6.4. Finally, Neale (1964) also suggested that one of the major externalities in professional team sports was the benefit that leagues bestow on the media. Neale (1964) termed this externality the ‘league standing effect’. The league standing effect refers to the manner in which the media earns revenue by reporting sports results without having contributed to the cost of producing the sport (Downward et al 2009). According to Neale (1964) interest in the league standings of teams is jointly produced with games and is not confined to active supporters. Whilst the games could occur without any media input, the media cannot benefit from interest in league standings without the existence of the league (Downward and Dawson 2000). Sports leagues therefore have an externality that is profitable for media sources and which now leagues capitalise on by internalising the externality. In recent years, sports leagues have acted collectively to sell the broadcasting rights for live matches and highlights to suppliers of cable and satellite (digital) television (BskyB for example). These forms of media require explicit subscription (unlike traditional free-to-air broadcasting) meaning access to the sport is restricted.

2.6.4 Market Structure in Sport

As outlined in section 2.2 earlier, in a perfectly competitive model all firms in an industry are assumed to be profit maximisers that sell an identical product to fully informed utility (satisfaction) maximising consumers. The corresponding sporting industry would be the league (market) of profit maximising clubs (firms) that all supply equally competitive fixtures (product) to fans (consumers) that seek to maximise their utility or enjoyment from viewing the fixture.
The perfectly competitive model, in this context, would assume that fans and clubs are small relative to the total number of clubs and fans, meaning their individual decisions to buy and sell tickets to watch fixtures cannot affect industry level activity as a whole. Furthermore, it would be assumed that there are no restrictions to entry or exit from the industry, implying unprofitable clubs are free to leave the league, while other clubs are free to enter the league and compete for profits. Clubs would also be able to decide their own quantity of output and could therefore produce more or less fixtures to the league as required. In addition, it would be assumed that all sporting fixtures are identical and there are no differences in quality between fixtures or the identity of fans to particular teams. A fan will therefore only choose one particular fixture over another because of differences in ticket price, however clubs are ‘price takers’ and have to accept prices determined by the market as a whole as they are unable to set their own. These are unrealistic assumptions, which is why it has been argued in the past that sports leagues are examples of monopolies (Neale 1964).

The fundamental difference between the monopoly model and the perfectly competitive model is that in a monopoly, only one firm supplies the market. As professional sports are often provided by one league (the sole supplier), a league can be viewed as a monopoly. In complete contrast to the perfectly competitive model, where clubs are unable to influence ticket price on the market, a monopoly league can alter ticket prices and therefore can adjust the volume of ticket sales for each of its fixtures. As the sole supplier of the product, the monopoly league faces the total market demand curve and is therefore a ‘price maker’.

If ticket prices are lowered in order to increase ticket sales then the MR earned by the sale of these additional tickets will be lower than the AR earned on previous ticket sales. Consequently, MR is less than AR for a monopoly or any form of market that is not perfectly competitive and where firms can control market price. Because monopoly firms face the market demand curve they are able to set prices according to what consumers are prepared to pay to see fixtures.

A key implication of monopoly supply is that because of the lack of competition and inability of clubs to enter the league or rival leagues to be formed to supply the sport, supernormal profits can always be earned even in the long run. This is because on each ticket sold there is a mark up above average cost (P>AC). These profit margins
monopolists are able to earn ultimately results in lost consumer surplus and lost producer surplus, meaning there is economic inefficiency in the monopoly case relative to the perfectly competitive case.

According to Neale (1964) the industry of team sports shows several characteristics of a natural monopoly. Furthermore, Neale (1964) claimed that sports would gravitate towards monopolies providing there is feasible economic or sporting basis for competition in the sport (Downward and Dawson 2000). Neale (1964) illustrates the history of sporting leagues is broadly consistent with the predictions of the natural monopoly thesis. Fort and Quirk (1995) reveal that by the end of 1994, monopoly leagues characterised all four major US team sports (American football, baseball, basketball and ice hockey). Neale (1964) argues monopoly profits will always attract competing leagues however, these will usually be short lived since existing leagues will resist entry by increasing labour costs and or reducing ticket revenue and income from broadcasting rights (Downward and Dawson 2000).

Historically, rival leagues have either merged as has happened in team sports such as, American football and English football, when the Football League absorbed the remains of the Southern League in 1920 after initially acquiring its most successful clubs. Alternatively, rival leagues can co-exist as separate entities whose members do not compete on the field but come together to produce world champions. In baseball, two rival leagues, the American League and the National League co-operate in order to produce the annual and highly lucrative World Series (Downward and Dawson 2000).

Although Neale’s (1964) predictions seem to be consistent with many outcomes of the development of sporting leagues, his argument fails to outline the rationale for, and description of mechanisms by which leagues have developed and operated (Downward and Dawson 2000). Neale’s (1964) argument essentially implied overall co-operation in matters of league management, however in sports, this is not always the case. The evolution of leagues can be down to the pursuit of particular interests rather than a perceived common good (Downward and Dawson 2000).

Sloane (1971) argued that a sporting league and its constituent clubs could be more accurately viewed as a cartel rather than a multi-plant firm. Sloane (1971) argued teams’ have control over their output relative to the league. In a cartel model, the club
behaves as a firm rather than a plant, although it is still unable to offer a saleable product without the league and is still subject to uncertainty of outcome. Consistent with Sloane’s argument is the fact most teams are independent entities that have control and make their own decisions on matters such as investment, the level of production, ticket prices and merchandising (Downward and Dawson 2000; Dobson and Goddard 2011). Sloane (1971) argued that Neale’s argument overemphasised the mutual interdependence of two competitors in the production of a sporting contest and labelled the club as the relevant economic decision maker.

Whilst cooperation of clubs within a cartel is necessary, clubs can still attempt to pursue a private advantage by breaking ranks with a membership as a whole (Downward et al 2009). A recent example of this is the formation of the Premier League in the UK, where the largest clubs broke away from the English Football League as a means to acquire a larger share of media and broadcasting revenue. According to Downward et al (2009) the cartel theory explains sports leagues better than the monopoly model as teams within the league have to cooperate in order to generate and distribute revenues, but frequently have differing interests.

The objectives of the professional team sports firm are widely debated. Differing objectives can lead to differing outcomes in terms of the distribution of talent amongst clubs in the league, total league revenue, player salaries and ticket prices (Kesenne 2007). The most common firm objective in economic theory is profit maximisation and analysts of professional sports in the United States often assume profit maximisation is also the main aim of professional sports clubs (Rottenberg 1956; Neale 1964; El Hodiri and Quirk 1971; Quirk and Fort 1992 and Vrooman 1995). Although professional sports clubs in North American sporting leagues are more business-like than their European counterparts some US economists have argued profit maximisation is an inappropriate assumption as team owners may privately pursue other objectives (Quirk and Fort 1992; Zimbalist 2003).

Sports economists in Europe have also raised doubts about profit maximisation being a realistic objective in professional sports (Kesenne 2007). Sloane (1971) rejected the profit maximisation objective in favour of utility maximisation, which he argued was more applicable to football. Sloane (1971) observed that many owners of European football clubs considered ownership of a club as an act of consumption rather than
investment. The implication of this assumption for professional team sports owners was that they consumed their resources in order to give them satisfaction rather than profit (Downward and Dawson 2000). As consumers, Sloane (1971) argued that many owners of European football clubs acted as if they were maximising a utility function where other variables such as playing success, stadium attendances, competitive balance and community building appeared as arguments (Kesenne 2007). As argued by Dobson and Goddard (2011), most chairmen and directors of football clubs have achieved success in business in other fields. Their motives for investing in clubs may therefore include a desire for power, prestige or simply because of their sporting enthusiasm. It is therefore sensible to view the objective of a football club as one of utility maximisations subject to a financial solvency constraint (Dobson and Goddard 2011).

Sloane’s (1971) proposed team utility function is given in equation 2.5c below.

\[ U = U \left( P, A, X, \pi_R - \pi_0 - T \right) \quad \text{subject to } \pi_R \geq \pi_0 + T \quad (2.5C) \]

Where \( U \) = utility attained; \( P \) = playing success; average attendance; \( X \) = league health (uncertainty and competitive balance); \( \pi_R \) = recorded profit; \( \pi_0 \) = minimum acceptable after tax profit and \( T \) = taxes. The utility maximising model has implications that differ to those that follow from the profit maximising assumptions of Rottenberg (1956) and Neale (1964). More specifically, the weighting of \( P \) is heavy relative to \( X \) and \( \pi_R \) (Dobson and Goddard 2011). As also outlined by Kesenne (2007), the crucial variable in Sloane’s (1971) utility function is \( P \) as sports clubs are above all, interested in winning. In order to make Sloane’s (1971) utility maximising model more operational Kesenne (1996) introduced win maximisation as the sole objective.

Kesenne (1996) argued that the behaviour of club owners and managers across Europe suggested that their main objective was to maximise their winning percentages, subject to a breakeven constraint ensuring no losses were made. According to Kesenne (1996) club owners will estimate their expected total season revenue and then pursue the best players they can afford within this budget, such as:

\[ \text{Max } W \]

\[ \text{Sub } R - C = 0 \quad (2.5d) \]
Where, \( W \) is the win percentage of a team, \( C \) is the total season cost and \( R \) is the total season revenue. The revenue function was assumed to be concave in the winning percentage. Rascher (1997) proposed yet another variation of Sloane’s (1971) utility maximisation model and assumed sports clubs maximised a linear combination of profits and wins.

Theoretically, an increase in any one of the factors identified by Sloane would make the management, owners and supporters happier. The more closely the explanatory variables are correlated with each other the more probable it is that the predictions of the utility maximising and profit maximising models resemble one another ultimately making it difficult to distinguish between the objectives of clubs (Downward et al 2009).

### 2.6.5 Cross Subsidisation in Professional Team Sports Leagues

The sub-sections above introduced and discussed the peculiar economics of professional sport and discussed whether leagues are monopolies or cartels. What is apparent in professional team sports is that that there is a need to co-ordinate the activities of clubs’ within a league. This section focuses on cross subsidisation policies employed in leagues in order to maintain competitive balance. The economic rationale for cross subsidisation in professional team sports lies in the market structure of sporting leagues.

As suggested above, sporting leagues can be viewed as cartels of competing producers of sports. Within these cartels clubs need to compete in sporting terms, but also co-operate to ensure that the sport is managed effectively, giving the impression that they are monopolies (Downward and Dawson 2000). To attain these ends governing bodies of sporting leagues have traditionally set sporting terms upon which clubs meet in competition and have directly influenced economic aspects of competition. League associations have for example influenced admission price structures, negotiated television broadcasting deals and sponsorship arrangements and have controlled the mobility of players (transfers between clubs).

As stated by Downward and Dawson (2000), these policies are an unusual phenomenon in economic terms. In other industries, an attempt to regulate the terms under which its members operate has been ruled as acting against public interest.
These ideas have become enshrined in legislation such as the *Sherman Act* (1890), the *Clayton Act* (1914) in the USA and similar legislations in the UK such as the *Monopolies and Restrictive Practices Act* (1948) and the *Restrictive Trade Practices Act* (1956). Despite these legal frameworks however there has been little government intervention in sporting markets.

There are two main ways in which league authorities have intervened in the management of club finances in order to promote cross subsidisation between clubs. Traditionally, league authorities have targeted sporting labour markets and the distribution of club revenue. By targeting the sports labour market, league authorities are influencing the price and availability of inputs rather than the output (sporting results), which may seem unusual however player salaries comprise of a large portion of costs incurred by professional sports teams in the US and Europe (Scully 1989; Szymanski and Kuypers 1999). Policies aimed at players will therefore have a large direct financial effect on clubs (Downward and Dawson 2000). Furthermore, unlike other forms of production it is the labour (the players), in the production function of sports teams that ultimately determine success or failure on the field of play. Consequently, both the resources of clubs as well as their results can be influenced through policies aimed at the sporting labour market (Downward and Dawson 2000).

League authorities have attempted to influence club finances, club personnel (talent) and club results by implementing three major types of labour market policy instruments: drafting systems; salary caps; and reserve option arrangements (Downward et al 2009). A well known example of the drafting system is the ‘rookie draft’ in American football, introduced by the NFL in 1936. The NBA introduced a similar system in basketball in the 1950s and Major League Baseball (MLB) introduced a draft in the 1960s. The NHL also introduced a draft system in the National Hockey League a few years later.

In the NFL the main source of recruitment is college football, which comprises of amateur players. The drafting system in operation in the NFL dictates the order in which professional teams can sign new talent (rookies). The reverse draft ensures that every year professional teams are able to sign new talent and those that finished the lowest in the league the previous season get the first option. In other words, the least successful teams of the previous season get the first option on new recruits and hence
the opportunity to enhance their stock of playing talent at a price that is probably discounted on the player’s genuine market value (Downward et al 2009). The intention of this drafting system is to reallocate talent between teams and ensure competitive balance.

Whilst drafting systems target the physical reallocation of sporting talent and therefore have an indirect financial implication for clubs, salary caps directly target the financial cost of talent (Downward and Dawson 2000). Salary caps impose a maximum amount that clubs can spend on talent. The NBA introduced a salary cap in 1980 and the NFL and baseball also introduced salary caps in the mid 1990s. Salary caps can either involve fixing the costs of an individual by establishing a maximum wage, represent a maximum absolute amount clubs can spend on players, or be expressed as a percentage of team or league turnovers. The intended implication and justification for these salary cap policies is that, in principle they make the best talent affordable to all teams and therefore prevents teams from hoarding talent and thus aid competitive balance (Downward et al 2009).

The final form of labour market policy employed by sporting leagues is the reserve option clause. A famous example of this is the option clause that was introduced in baseball in 1880 and still exists in a relaxed form for rookie players. Essentially, the reserve clause in baseball tied players to their club for their entire playing careers. When players signed for a club, the reserve clause gave the club the option to renew the player’s contract upon expiration. Players had little choice but to accept these contracts as unless their clubs released them they would not be able to seek further employment within the sport. A similar ‘retain and transfer’ system was implemented throughout European soccer. The justification for the reserve option clause was that it prevented the financial powerhouses from buying up all of the available talent in the league (Downward and Dawson 2000). Combined, therefore, policies such as these have targeted the physical availability of players as well as costs associated with them.

The other main form of cross subsidisation adopted by sports leagues is the redistribution of revenue. The aim of distributing this revenue is to ensure weaker teams have an income stream with which they are able to fund the acquisition of superior playing talent. Traditionally, the redistribution of revenue has focussed on
gate receipts (money earned by spectators viewing the game live in stadiums). Initially in baseball, a revenue split of 50:50 was in operation, where the home and away teams would share revenue equally. Over time however, the away teams share have steadily fallen and many arrangements now exist throughout sports in the USA that ensure home teams receive a larger share of revenues (Downward and Dawson 2000). The NFL operates a 60:40 split in favour of the home club and similarly generous arrangements have applied in European football. In England between the 1920s and 1980s an 80:20 split in favour of home clubs existed. Gate revenues have also been shared via a “pool” system, whereby a predetermined share of league revenues go into a central fund and are distributed proportionally as has been the case in English football and rugby league in the UK.

In the USA, TV revenue obtained from local broadcasters was not shared with the visiting teams (Fort and Quirk 1995) although revenue from national broadcasters was shared. This particular system evolved in a complicating manner when the American Football League (AFL) and National Football League (NFL) merged largely due to national TV contracts held by the AFL and not the NFL. According to Downward and Dawson (2000) the most radical changes to TV financing occurred outside of the US with the growth and emergence of BskyB satellite TV in Europe and Australia. However, this system of cross subsidisation has broken down and failed in some sports in Europe due to the formation of breakaway leagues such as the Premier League in football and Super League in rugby. The growth of BskyB’s funding of these breakaway leagues has substantially reduced the extent of cross subsidies (Downward and Dawson 2000). Consequently, a huge financial gulf now exists between those clubs competing within the Premier League and Super League and those clubs outside these competitions.

2.6.6 Key Concepts

As illustrated above, one important feature of the professional team sports industry is the joint production of contests. One team cannot play a football match on their own – they require an opponent. Furthermore sport is about competition. If playing strengths between two opponents are far apart, then the stronger team wins without much competition, meaning the product (the fixture) decreases in appeal (Kesenne 2007). Another vital characteristic of professional team sports is therefore that a certain
degree of competitive balance is required in order to increase the appeal of fixtures (product) and sell them. This is why the uncertainty of outcome hypothesis is a key component in the economics of professional team sports.

The sports economics theory of joint production provides the foundation upon which this thesis is built. Subsequent chapters explore and discuss the productivity of teams and players in professional football before a series of empirical investigations are conducted, analysing production in German Bundesliga football at various levels. Team, positional unit and individual player productivity is determined via the use of match specific performance statistics. A rigorous methodology is devised that investigates whether team, positional unit and individual player productivity is jointly determined with the corresponding productivity of their opponent(s).

Despite the theory outlined in this section, very little previous research has actually analysed joint production within football matches in detail. Subsequent chapters will examine the extent to which player and team performances are interdependent and hence jointly produced (if at all). Moreover, joint production in football is examined at a match level by decomposing collective team, positional unit and individual player efforts. Ultimately, at each of these levels any evidence found confirming interdependency between performances would support the theory of joint production in sports economics. On the other hand, if simultaneity between performances is not discovered the theory of joint production will be rejected.
2.7 Conclusion

This chapter has outlined the key economic theories upon which this study has been based. Consequently, it has been shown that the conventional economic scope of analysis is concerned with an optimal allocation of resources. Specific attention has been paid to production and how levels of production vary for firms depending upon the market structure within which they operate.

Shifting focus away from the behaviour of firms, section 2.4 outlined several factors that could impact the productivity of an individual within a firm. It was shown how levels of effort exerted and hence productivity can be influenced to varying extents by, the acquirement of human capital and the related enhancement of cognitive and non-cognitive ability, the potential problem of moral hazard and the offer of specially designed incentive and tournament contracts.

Section 2.5 of this chapter highlighted the differences that exist between the economic characterisation of professional team sports and usual economic policy recommendations. The process of formalisation in sport highlights the shift in the economic organisation and motivational dimensions of sports clubs. The uncertainty of outcome hypothesis, a key externality, was presented as a central theme in the economics of professional team sports and its importance to the success and evolution of sporting leagues was highlighted. Furthermore, this section emphasised the joint production element in sports that stems from this production externality.

Chapter three below presents a detailed literature review that identifies factors of individual and team production across various professional sports, before shifting focus towards factors of production specific to professional football. Furthermore, insights into the economic significance of the key externality of the uncertainty of outcome hypothesis are presented. Chapter three also presents a review of previous literature on the use of production functions and their associated efficiency frontiers in the analyses of team performance in football. Focus then shifts towards the joint element of production in professional football. Factors that impact this joint production are outlined and the magnitude of their impact is discussed. Specific attention is devoted to the presence and quality of the opponent and hence their role in
the joint production process. Subsequent chapters present empirical analyses investigating the impact of opponents within this joint production process.
Chapter Three

Production in Team Sports

3.1 Introduction

This chapter explores the process of production in team sports. Section 3.2 begins by outlining the role of contest organisers in the production of a contest and more specifically their influence on the exertion of effort levels and hence overall quality of the contest through contest design. Specifically, studies examining the impact of contest designs upon the overall quality of the contest across various sports are reviewed.

Section 3.3 links the concept of contest design to the joint production element in team sports, detailed in section 2.5 in the previous chapter. It outlines the uncertainty of outcome hypothesis (UOH). The UOH holds that spectators are attracted to evenly matched sporting events and it has become a central principle in the economic analysis of team sports. In particular, investigations into the relationship between uncertainty and spectator attendance (the main focus of research on UOH) are reviewed and the important role of the opponent in the production of a contest in team sports is further emphasised.

Section 3.4 shifts the focus towards the use of production functions and their associated production frontiers in the analysis of team performance in sports. The sporting production function first proposed by Rottenberg (1956) is outlined and various playing inputs, non-playing inputs and outputs across various sports are identified and adopted methodologies are detailed.

Section 3.5 highlights the general lack of attention awarded to the impact of the opponent by reviewing the relatively small minority of studies that have acknowledged the impact of opponents upon performance. This section also highlights the limitations of these studies.

Section 3.6 focuses specifically on the non-playing input of managerial contribution and discusses the varying impact a head coach or manager can have across team
sports. As illustrated in this section, it is widely accepted that a manager or head coach of a team will impact performance.

Section 3.7 outlines the potential advantages of utilising a disaggregated data set in the estimation of a sporting production function and presents a review of studies across various sports that have employed disaggregated match specific data. In particular the capture of in-game variation and in-between game variation enables certain aspects of play to be accounted for and hence the benefits of utilising a similar data set for the purposes of this study are discussed. Finally, section 3.8 concludes the chapter and provides a summary of the gaps in previous research that the literature review has identified and outlines what is taken forward for investigation in subsequent chapters.
3.2 Tournament Theory and Contest Design

Section 2.5.1 in the previous chapter outlined how sporting contests can be seen as economic tournaments. A key underpinning of tournament theory is that relative performance is more influential in deciding the winner of the tournament rather than absolute performance. Sports stars, particularly in individual sports are often paid based upon their relative performance (Frick 2003). According to Szymanski (2003), sporting contests provide good examples of incentive – mechanism design. Tournament organisers will offer incentives in order to induce greater effort from competitors and hence increase the quality of the competition and potential economic return (Downward et al 2009).

The composition of pay varies between individual and team sports. As illustrated by Frick (2003) sports performers in individual sports acquire a sizeable proportion of their income via performance related (variable) pay, whereas performers participating in team sports (i.e. football) acquire fixed salaries and are therefore guaranteed a certain amount of income irrelevant of performance. According to Szymanski (2003), the implications of these alternative forms of income distribution can be explained by simple economic theory. One of these implications is the influence upon effort levels of performers and this particular topic is one that has been researched in great detail in the past.

Seminal work by Ehrenberg and Bognanno (1990) on the 1984 US and 1987 European PGA tours revealed that player performance was positively correlated with the value of prizes after controlling for course difficulties, weather and the quality of other competitors. In other words as more money became available, competitors raised their effort level. Fernie and Metcalf (1999) also investigated the influence of income on performance in a study on horse racing and revealed that British jockeys that had signed contracts guaranteeing fixed income showed deteriorating performances as opposed to when they possessed contracts that offered no guaranteed fixed income.

In other sports, broadly consistent results are obtained. Lynch and Zax (2000) investigated the prize structure of 135 professional road running contests held between 1993 and 1995 and revealed that the races that rewarded runners with large
prizes recorded faster finishing times. Lynch and Zax (2000) did however acknowledge that this might have been because these races attracted the faster runners (elite performers), not necessarily because they encouraged all runners to exert more effort and run faster.

Similarly, Maloney and McCormick (2000) investigated runners participating in 115 races held in the US between 1987 and 1991 and revealed that the more concentrated the prize money in a race the higher the effort level exerted by the runner, resulting in faster times. Conversely, the less concentrated the payout of a race the slower the field became. Ultimately, Maloney and McCormick (2000) concluded that an increase in the financial prizes on offer, would inevitably lead to an increase in performance, although they also acknowledged this could be due to either the participants raising effort levels (incentive effect) or due to the elite performers participating in the race (selection effect).

Lallemand et al (2008) examined prize rewards and their incentive effects in women’s official WTA tennis tournaments between 2002 and 2004 and revealed that the larger the prize spread in tournaments the greater the amount of effort induced from players. Lallemand et al (2008) discovered that a doubling of the prize money would lead to an increase in performance by 0.8 and 1.2 games ceteris paribus and therefore identified a positive and significant relationship between prize spread and a female tennis player’s performance. However, Lallemand et al (2008) also discovered that the final outcome of a match was more linked to players’ abilities and intrinsic talent than players’ incentives to adjust their effort according to their chances of success.

Frick (2003) investigated selection and incentive effects across three golf tournaments in the US, each with varying levels of prize money and prestige and revealed that the better players tended to play more in senior tournaments as they have higher pay rewards. Frick (2003) also revealed that players that participated in PGA tournaments enjoyed significant premiums over other players playing in lower ranked tournaments.

In addition to offering incentives to participants, tournament organisers also have to maintain a degree of competitive balance for the benefit of spectators (consumers) demonstrating why designing an optimal contest can prove extremely difficult for
event organisers. Szymanski (2003) covers various contest designs and highlights several potential difficulties faced by event organisers in the design of optimal contests. Tournament organisers have to deliberate over design issues such as deciding upon the optimal number of entrants or competitors in a tournament, the optimal prize structures in a tournament and the competitive balance within a tournament.

The difficulties associated with designing an optimal contest were demonstrated by the introduction of the UCI professional tour (a type of league system consisting of the best teams) in cycling in 2005. Rebeggiani and Tondani (2008) observed the behaviour of cycling teams after the introduction of the UCI professional tour in 2005 and concluded that the oligopolistic design of the pro tour encouraged non-competitive behaviour. This study revealed that 50% of teams achieved the vast majority of their points in the races made in the country where the team is affiliated. In other words teams would put greater effort in those races that were organised in the team’s home country. This pattern of non-competitive behaviour displayed by race teams meant that the introduction of the UCI pro tour ultimately failed to satisfy its primary aim, which was to increase overall competition amongst cycling teams (Rebeggiani and Tondani 2008).

Conversely, in football, overall competition amongst teams increased after the award of three points for a win rather than the original two points. Garciano and Huerta (2006) investigated the impact of this rule change in Spanish top division football since its implementation in 1995 and their results revealed a reduction in the number of draws from 29.75% in the 1994-95 season to 25.5% in the 1998-99 season. In addition, it was revealed that the margins of victory narrowed after the implementation of the ruling as less matches were won by a margin of two or greater goals (Garciano and Huerta 2006).

Garciano and Huerta (2005) suggested the rule change from two points for a victory to three points also resulted in more closely competed games and perhaps a more fluid

---

12 Contest designs discussed by Szymanski (2003):
- Symmetric winner takes all contest
- Multiple prizes in symmetric contests
- Asymmetric two person contests
- Asymmetric contests with more than two players
- Dynamic contests

---
attacking approach to games as the number of games won by a single goal rose significantly from 115 in the 1994-95 season to 153 games in the 1998-99 season. Vital attacking statistics such as shots on goal and corner kicks increased, however, so did the number of fouls committed and number of yellow cards so the net effect was no more goals being scored (Garciano and Huerta 2005).

Garciano and Huerta (2005) revealed that the award of three points for a win rather than two led to a substantial increase in both the offensive efforts of a team but also the tendency to commit more fouls and break up play. Garciano and Huerta (2005) also claimed that teams that deliberately set up these sabotage strategies ultimately depressed the attendance at their rivals home stadium, presumably because fans predict a boring match. According to Garciano and Huerta (2005) the award of three points for a win affected Spanish football in more negative ways than positive, as teams conducted more sabotage tactics after the new ruling was enforced. Therefore the cost of increasing incentives in football was to encourage more effort of the wrong kind.

The research reviewed in this section demonstrates the fine balance between the offer of incentives and the appropriate level of competition event organisers have to factor in to the design of an optimal contest. Furthermore, this section highlights attempts made by tournament organisers to elicit greater effort levels and hence improve the overall quality of the competition. According to Downward et al (2009) the driver of these attempts is the joint production of sporting contests, which itself can be understood as consisting of two key dimensions, the quality and uncertainty of a sporting contest. Section 3.3 presented below expands upon the joint production element in sports contests and reviews previous literature examining the uncertainty of outcome hypothesis (UOH), which has become a central organising principle in the economic analysis of professional team sports.
3.3 Uncertainty of Outcome

Rottenberg (1956) first detailed the uncertainty of outcome hypothesis (UOH), before it was expanded by Neale (1964) and Sloane (1971). The UOH holds that fan demand is positively related to uncertainty of outcome and it is a central tenet to the economic analysis of sports leagues (Owen and King 2014). UOH maintains that spectators are more likely to be attracted to sporting encounters involving evenly matched opposition and sports fans are risk lovers Dawson and Downward (2005).

According to Dawson and Downward (2005) there are three main types of uncertainty, short term uncertainty, measuring the outcome of a particular match, medium term uncertainty, measuring the outcome of a seasonal championship and long term uncertainty, which refers to a series of championships (Szymanski 2003). A substantial empirical literature has examined all three of these types of uncertainty across a variety of sports, some of which are discussed in the subsections below. Subsection 3.3.1 presents some of the literature that has examined the UOH in sports other than football, while subsection 3.3.2 presents the literature that has focussed specifically on football.

3.3.1. Uncertainty of Outcome in North American Sports

It can be argued that UOH and the concept of competitive balance are positively related. The more uncertain outcomes become the greater the level of competitive balance (Lee and Fort 2008). UOH implies that declining competitive balance in a league causes fan interest in perennial losing teams to wane, threatening the economic viability of those teams. In addition, even surviving teams might suffer because of reduced revenues as general interest in the sport declines.

Since the seminal work of Rottenberg (1956) on baseball, many economists have investigated competitive balance in sports using various different measures of uncertainty. More recent studies have been able to test UOH on a game-by-game (short term) and seasonal basis (medium term) as more data has became available. American sports have been the focus of many of these studies and the results obtained from some of these studies are summarised below.
In a study on baseball, Knowles (et al 1992) utilised pre match betting odds (probability of a home win) as a measure of match (short-term) uncertainty during the 1988 MLB (Major League Baseball) season. It was revealed that short-term uncertainty of outcome was a significant determinant of attendance in MLB games (Knowles et al 1992). Similarly, Rascher (1999) also utilised pre match betting odds as a measure of uncertainty during the 1996 MLB season and concluded that short-term uncertainty had a positive influence on attendance at MLB games.

Schmidt and Berri (2001) examine seasonal (medium term) uncertainty and competitive balance (long-term uncertainty) in the MLB from the 1903 to the 1998 seasons. A Gini coefficient is employed as the measure of uncertainty and Schmidt and Berri (2001) revealed competitive balance to have a significant impact on league attendance. The direction of the impact however was revealed to vary with the time period considered in estimating the Gini coefficient.

MLB fans responded negatively to an improvement in competitive balance within a given season, however improvements in competitive balance in the previous 3 to 5 years resulted in a positive reaction from fans. Schmidt and Berri (2001) concluded MLB fans had distaste for increased uncertainty within a season however over time will react negatively to persistent competitive imbalance.

Meehan et al (2007) examine the relationship between attendances and short-term uncertainty in the MLB throughout the 2000-02 seasons utilising the differences between the winning percentages of home and visiting teams as their measure of match uncertainty. Meehan et al (2007) reveal baseball fans generally prefer to attend games in which the outcome is in doubt, therefore supporting UOH.

Lee and Fort (2008) investigated all three aspects of uncertainty (short term, medium term and long term) at the game level, playoff determination level and across several seasons in baseball and examined the impact of each on average attendances. The measure for game uncertainty utilised in the study was the dispersion of winning percentages at the end of each season. Playoff uncertainty was measured by the difference in winning percentages between first and second place finishers in each league and average differences in winning percentages between the division winners

---

13 A Gini coefficient is a conventional economic measure of inequality (Schmidt and Berri 2001). The coefficient provides a single unit measure for the degree to which a relationship deviates from equality (Schmidt and Berri 2001).
and runners up. Consecutive season uncertainty was measured by calculating the correlation between a team’s winning percentage in a season and their average win percentage over the previous three seasons. Short-term and long-term uncertainties were found to have no impact on match attendances, however playoff uncertainty (medium-term uncertainty) was revealed to have a significant impact on league average attendance per game.

Welki and Zlatoper (1994) utilise betting market information as a measure of uncertainty in American Football and find evidence to support the positive association between attendances and short-term uncertainty in the NFL during the 1986-87 NFL season. Conversely, Coates and Humphreys (2012) used point spreads (betting information) as a proxy for uncertainty for over 5,000 NFL games played throughout the seasons 1985 to 2008 and revealed no relationship between match uncertainty and game attendance. Instead, Coates and Humphreys (2012) argued NFL fans prefer to attend games where the home team is expected to win and are averse to attending games with uncertain outcomes or games where the home team is expected to lose.

Depken and Wilson (2005) investigate UOH at an amateur college football level and provide evidence to confirm that long-term uncertainty (competitive balance) does pertain to college football (attendances). According to Depken and Wilson (2005) the rationale for UOH in professional sports holds in organised amateur sports even though players are not paid to play as teams (schools), are still interested in generating profits from sporting events. Depken and Wilson (2005) claim that as competitive balance declines so does fan attendance, suggesting UOH is an important component to a league’s financial success.

In a study on the NHL (Ice-Hockey), Coates and Humphreys (2012) used betting odds as a measure of uncertainty for 6,054 regular season NHL games between the 2005-06 season to the 2009-10 season and revealed that attendance increased when the home team was expected to win and fell when two teams were closely matched, therefore opposing the traditional UOH. Similarly, Jones and Ferguson (1988) also tested for the impact of short-term UOH on attendances in the NHL 1977-78 season and found no evidence to support the UOH.
Some evidence has been found to support UOH in basketball. Rascher and Solmes (2007) used win percentages\(^{14}\) to measure match uncertainty throughout the 2001-02 season and revealed fans of the NBA prefer true contests where the outcomes of games are uncertain. According to Rascher and Solmes (2007) competitive balance is of vital importance to the continued success of the NBA.

As demonstrated above, North American studies have generally revealed a mixed response with regards to the uncertainty of outcome hypothesis at game level, seasonal level and championship level across various team sports. The relationship between uncertainty and demand has also been investigated in football and the sub section 3.3.2 below presents the main findings of some of these studies.

### 3.3.2 Uncertainty of Outcome in European Football

Hart et al (1975) measured match uncertainty for four English football clubs between the 1969/70 season and the 1970/71 season using the log difference in league positions as a measure of uncertainty. Hart et al (1975) failed to find conclusive evidence that match attendances were positively correlated to matches with uncertain outcomes. Conversely, Peel and Thomas (1988) used betting odds (probability of a home win) to measure match uncertainty of English football league matches in the 1981-82 season and found evidence in support of UOH\(^{15}\).

Forrest and Simmons (2002) also utilised adjusted bookmaker odds (that accounted for bias in favour of the heavily supported clubs) to measure match uncertainty for English Football League matches throughout the 1997-98 season. Forrest and Simmons (2002) found evidence to confirm match attendances were positively related to matches with uncertain outcomes and therefore supported UOH.

In an investigation on the English Premier League (EPL), Buraimo and Simmons (2008) used pre match betting odds as a measure of match uncertainty throughout six seasons between 2000-01 and 2005-06 and revealed that an increase in uncertainty of outcome is associated with reduced gate attendance in complete contradiction to the

---

\(^{14}\) Winning percentages were used to create a number of variables. The differences between home and away team winning percentages, the probability that the home team wins a game and this probability squared. These variables allowed for the estimation of the optimal probability that the home team wins (Rascher and Solmes 2007).

\(^{15}\) Dobson and Goddard (2011) argue that by incorporating their measure of uncertainty in linear rather than quadratic form, Peel and Thomas (1988) actually measure team quality rather than uncertainty of outcome therefore their results offer only weak support of UOH (Szymanski 2003).
traditional UOH. According to Buraimo and Simmons (2008) fans in the EPL are predominantly supporters of the home team and would rather see their team play a much inferior team and win, rather than attend a closely contested match.

Buraimo and Simmons (2008) discovered a U-shaped relationship between gate attendances and home win probability. Fans of the EPL are believed to have preference to pay to watch games that have either a very high or very low probability of a home victory. The attraction of the low probability win outcome for home fans is due to “David vs Goliath” effect, in which home fans wish to be present on the rare occasion David beats Goliath (Buraimo and Simmons 2008).

Benz et al (2009) used quantile regressions to investigate the impact of match uncertainty upon match attendances in the German Bundesliga throughout five consecutive seasons between 1999 and 2004. Uncertainty measures included the absolute difference in league standings, points per game records of all teams and German bookmaker odds. Ultimately match uncertainty was deemed of secondary importance in influencing attendance demand. According to Benz et al (2009) a team’s reputation and current league position are more important determinants of attendance.

Jennett (1984) analysed Scottish Premier League football between the seasons 1975-76 and 1980-81 using the championship or relegation significance of each game throughout a season as the measure of uncertainty and found evidence to support UOH. Dobson and Goddard (1992) adopt Jennett’s (1984) model in order to analyse match attendances for teams in the English Football League between 1989 and 1991 and discover a positive impact of uncertainty for fans of the home team but no significant effect for fans of the away team. Furthermore, García and Rodríguez (2002) assess the impact of mid-term uncertainty upon the attendances of Spanish La Liga matches between 1992 and 1996. Using a measure of the likelihood of winning the championship (product of points behind and remaining games) for two competing teams in a fixture as the measure of uncertainty, García and Rodríguez (2002) revealed a significant positive effect.

Pawlowski and Anders (2012) also assess the impact of medium term uncertainty upon attendances in German Bundesliga football in the 2005/2006 season and measure uncertainty by adopting an index assessing the championship or Champions
League qualification significance of each fixture. In cases where either the home or away team had a theoretical chance of winning the championship, attendance was revealed to increase. Conversely, the competition between teams competing for the reward of champion’s league qualification had no impact upon attendance figures.

Recently, some studies have shifted emphasis away from the impact of uncertainty on stadium attendance figures and instead towards the impact of uncertainty upon the size of television audiences. Forrest et al (2005) utilise data on subscribers of satellite TV in order to investigate the impact of match uncertainty upon the size of the television audience in televised top level English football matches played between 1993 and 2002. Uncertainty was measured using a composite index involving league form and home advantage and Forrest et al (2005) ultimately discovered final audience figures to be positively correlated with matches with uncertain outcomes, therefore supporting UOH.

Alavy et al (2010) also investigated on the relationship between uncertainty of outcome and the viewership of TV audiences of English football matches played and broadcasted between 2002 and 2005. Uncertainty was measured by calculating the probabilities of different match outcomes and Alavy et al (2010) revealed TV audiences value uncertainty of outcome and have a preference for an evenly contested game where either side could conceivably win, in line with UOH.

Empirical studies explicitly examining the impact of long-term uncertainty outcome on stadium attendances or TV viewership in football are lacking however anecdotal evidence contradicts the traditional UOH (Pawlowski and Budzinski 2013). Over the long term, season aggregate attendances have increased in some European leagues even though they have become dominated by a small minority of teams. For example, in the German Bundesliga, the slightly increasing competitive imbalance has been accompanied by increasing attendance figures for the league, overall (Pawlowski et al 2010). According to Pawlowski and Budzinski (2013) such findings suggest that competitive balance is of minimal importance for stadium and TV demand, therefore challenging the relevance of UOH in professional Football.

As demonstrated throughout this section the joint element of production in team sports carries an element of uncertainty, the importance of which is disputed across sports. However this section also highlights that irrelevant of whether it is for match
results (short-term success) seasonal rewards (medium-term success) or domination (long-term success), competition between at least two teams is always required. The opponent therefore is an important part of the production process. Section 3.4 below reviews literature that outlines the production process across various team sports.
3.4 Team Performance: Production and Efficiency in Team Sports

As discussed earlier in section 2.5 production in team sports requires the joint cooperation of at least two teams. This requirement that clubs cooperate in order to produce the output (the contest) is almost unique to team sports and it is why Rottenberg (1956) first proposed the notion of a sporting production function when he defined the product (the fixture) as a function of the players of one team (a factor of production) and the players of the other team (another factor) as well as additional factors such as team management, transportation and stadiums.

Two strands of literature exist with regards to team performance in sports. The first strand of research employs average production functions that model team outputs as a function of playing and non-playing inputs in order to inform teams of their weaknesses and potential areas for improvement. The second strand of literature consists of the associated production frontier analyses and efficiency measurements that enable teams to assess whether they are making full use of their financial and sporting resources, relative to their potential (Carmichael and Thomas 2014). Both strands of literature are reviewed below in order to illustrate the distinct agendas between the two strands of research in terms of assessing the relationship between team performance and production and team performance and efficient use of resources in team sports.

3.4.1 Production Functions in Team Sports

The core economic process within any professional sports team is the sporting production process in which ‘raw’ playing talent is transformed into player and team performance and, in turn, match outcomes. Formally, this process can be modelled as a sporting production function. (Gerrard 2001, European Sport Management Quarterly, 1: 3 p220)

One of the first empirical applications of production functions in sport was presented by Scully (1974) on monopoly exploitation in baseball. Scully (1974) aimed “to crudely measure the economic loss to the players due to the restrictions of the reserve clause” (Scully, 1974, p915) and set out to accomplish this by obtaining estimates of the marginal revenue product of pitchers and hitters performing in the MLB, before
comparing these estimates to player earnings in order to determine what percentage of a player’s MRP was retained by him.

Scully (1974) proposed a causal chain (recursive) model consisting of two equations, the first of which directly determined the marginal effect of player performances upon team success and the second the marginal effect of team success upon team revenues. Scully (1974) adopted an OLS regression estimation method and applied it to both equations in order to estimate these marginal effects, which in turn enabled Scully (1974) to estimate player MRP.

The first equation of Scully’s (1974) model consisted of the dependent variable PCTWIN (a calculation of the percentage of team wins), which Scully hypothesised to be a measure of team success, and several other independent variables. The independent variables included measures of player performance such as the team slugging average (TSA) and team strikeout-to-walk ratio (TSW). TSA was hypothesised to be a measure of batting (attacking) prowess and measured the number of bases gained per match whereas TSW measures a pitchers ability to control pitches and was therefore used a measure of the pitching (defensive) contribution to team wins.

In the second stage of Scully’s (1974) model, PCTWIN (now specified as an explanatory variable) and other independent variables that accounted for local population, team fan base, stadium location and racial prejudice were used to determine team revenues. Scully (1974) then used the coefficients obtained for TSA and TSW (from the first stage) in conjunction with the coefficient obtained for PCTWIN (from the second stage) to estimate average MRP. In order to convert these average MRP estimates to MRP per player Scully (1974) assumed that the team’s performance was the sum of the individual’s performances. Furthermore, Scully’s estimations did not incorporate a team’s full roster but only twelve out of fifteen hitters (that were expected to have played in a match) and eight out of ten pitchers (expected to have played in a match). Scully (1974) then presented ‘gross’ MRPs that did not account for costs associated with placing players on the field (i.e. cost of training players, cost of equipment and transport costs) before calculating ‘net’ MRPs after the deduction of these costs.
Finally, Scully (1974) estimated player salaries that would have been paid to players (in various categories of skill) in order to determine whether players were remunerated in accordance with their MRP or if they were exploited. These estimates were obtained from separate salary regressions for 148 players (hitters and pitchers) performing in the 1968 and 1969 seasons. Scully’s (1974) results ultimately suggested baseball players were being exploited to a fairly high degree.

Scully’s player performance model has been criticised by others. Medoff (1976) for instance argues Scully’s (1974) model was incorrectly specified as the recursive model did not account for simultaneity between win percentage and revenues in the two equations. Consequently, Medoff (1976) conducted a follow up study utilising a Two Stage Least Squares estimator and revealed similar results that confirmed the exploitation of baseball players. Furthermore, Gerrard (2001) highlights the limitations of Scully’s model by claiming the model is only applicable in those team sports where comprehensive player performance data can be captured easily, such as batting and fielding games like baseball and cricket.

In another study on baseball, Zech (1981) employed a Cobb-Douglas production function16 in order to identify those factors that contributed to team success in the MLB. Once identified, Zech (1981) empirically estimated the effect of these factors and used the results of the empirical testing to construct a measure of the league’s most valuable player (MVP).

Zech (1981) summarised player skills into four main categories, hitting, running, defence and pitching. Measures used to account for these player skills included, players batting averages (as measures of hitting frequency), number of homeruns (as measures of player power), a teams stolen base total (as a measure of team speed), a players fielding percentage (as a measure of catches taken), total chances taken (as a measure of difficult catches taken) and pitchers strikeout to walk ratio (as a measure of pitcher ability). In addition, Zech (1981) also incorporated two measures of head coach ability in the form of a manager’s lifetime won-lost percentage and the number of years managed in the major leagues.

16 A Cobb-Douglas function is a common function used by economists when turning to empirical work (Gartner 2006). The function is used to describe the relationship between two or more inputs (such as capital and labour) and the amount of output produced.
Zech (1981) established a baseball production function and then uses it to compute each player’s marginal product by calculating each player’s contribution to team victories. The top ten most valuable players’ (MVPs) in the league computed by Zech (1981) differed significantly from the official voting list of sportswriters (who selected MVP’s for each season using an informal weighting system for each of the player’s different skills). Zech (1981) suggests this discrepancy may have been due to player publicity and personality playing a role in the sportswriters’ weighting schemes.

Zech (1981) tested his model using data from 26 MLB teams competing in the 1977 baseball season and tested the model four times employing all possible combinations of defensive and managerial ability in order to ensure no multi-collinearity. Ultimately, Zech (1981) revealed that no measures of a team’s defensive skills (player fielding percentage and total chances taken) were significant and the contribution of the manager was also revealed to be insignificant. Furthermore, hitting was revealed to contribute almost 6 times as much to team success than pitching and the number of homeruns contributed about as twice as much as the number of stolen bases to a team’s success.

Similar methods of modelling the sporting production process implemented by Scully (1974) and Zech (1981) have since been applied to other sports. Schofield (1988) estimated production functions in cricket using data from the 1981 to 1983 cricket seasons in the County Championships and John Player League in England. Similarly to the studies detailed above Schofield (1988) specified output to be team success and a function of player performance and weather. Player performance (measured by batting, bowling, captaincy and fielding variables) was further composed of a sub-set of variables namely: coaching skills, team management skills, player availability, scouting, form, experience, the quality of practice and training facilities. Schofield (1988) therefore adopted a recursive system in which endogenous variables of player performances and team success were sequentially determined.

Schofield’s (1988) study involved collecting cross sectional data from seventeen competing teams and data was initially normalised to account for random variation in the mean caused by external factors that had an influential impact on input and output variables. Team success (dependent variable) was measured in two different ways, the
first measure was the number of wins achieved throughout a single cricket season and the second measure was the number of points acquired. Schofield’s (1988) analysis was broken into three parts; the first part assessed the general importance of batting and bowling and the second and third parts assessed the relative importance of different types of batting and bowling skills.

Batting performance was measured by calculating the number of runs scored per game and per over and the bowling variables included in the production function included the calculation of a bowling average (runs conceded/wickets taken) a composite measure of both attacking and defensive bowling skills. Furthermore separate measures of attacking bowling such as the number of wickets taken per game and number of overs bowled per wicket taken as well as measures of defensive bowling such as the number of runs conceded per over were also included as bowling inputs. Using these calculated values Schofield (1988) developed a linear production function using OLS and ultimately concluded that generally bowling had a greater impact upon match success than batting.

Bairam et al (1990) also constructed a cricket production function in their study on cricket played in Australia and New Zealand. Although, this study was largely based on Schofield’s (1988) study it differed in a few key areas. Output was defined as points percentage rather than number of wins or points achieved. Furthermore, two measures of bowling ability were utilised in this study, the first a measure of attacking bowling, balls per wicket and the second measure of defensive bowling, the number of runs scored by the opposition. Runs scored per wicket was another dimension that was included and reflected attacking and defensive bowling. Runs per innings and runs per over were two calculated values that represented the batting performance of teams.

Bairam et al (1990) chose to omit the element of fielding in order to avoid multicollinearity. Pooled data was collected and normalised by taking ratios of the variables used to their seasonal means and multiplying the ratios by 100. This was done to account for random events such as differences in weather conditions that could have influenced both input and output variables. Bairam et al (1990) revealed that both batting and bowling variables were important in explaining match success.
In order to maximise the probability of match success, Bairam et al (1990) concluded that New Zealand should adopt a combination of attacking batting and attacking bowling strategies. In a comparative analysis with England it was revealed that relative to England, batting was marginally more important than bowling for New Zealand. The results obtained for Australian cricket suggested the probability of match victory was maximised by adopting an attacking batting strategy and a defensive bowling strategy (Bairam et al 1990).

Carmichael and Thomas (1995) formulated a production function for rugby league football in the form of a recursive model in which both player performances and team success were determined sequentially. Team success (the output) was stated as a function of three measures of attacking performance, tries for, kicked goals for (conversions following a try or penalty) and drop goals for and three measures of defensive performance, tries against, kicked goals against and drop goals against. Carmichael and Thomas (1995) assumed these six performance inputs were influenced by variables such as player fitness, inherent ability, player experience, team organisation and coaching skill.

Output was measured by the percentage of games won during the season (expressed relative to the divisional average). Attacking performance was measured by the average number of tries, kicked goals and drop goals scored per game and defensive performance was similarly measured by the average number of tries, kicked goals and drop goals scored against. Carmichael and Thomas (1995) included height and weight variables as a proxy for player strength and measured experience by squaring the average player age. Inherent ability was proxied by the percentage of full international players and the percentage of players from overseas. The average number of appearances made per squad member measured team organisation and the ability of the coach was proxied by the number of years spent as a full time rugby coach prior to the season of investigation.

Carmichael and Thomas (1995) tested their model on cross sectional data collected for 35 rugby league clubs competing in the 1990-1991 season. Data was collected for 1,214 individual players and team input and output measures were also calculated as seasonal totals. In conclusion Carmichael and Thomas (1995) stated successful performance in rugby league was more dependent on defensive performance than
attacking performance and suggested a disaggregated data set with more player specific data could potentially produce improved analysis of production functions in rugby league.

Carmichael et al (2000, 2001, 2011) and Carmichael and Thomas (2005) have conducted production function studies on English Premier League football (EPL). Carmichael et al (2000) employed disaggregated player performance statistics for matches played by teams during a single season (1997-98) in order to estimate a production function for match performances. The output was specified as the match score and was expressed as a goal differential as this measurement allowed for the easy interpretation of ordinary least squares (OLS) estimates (Carmichael et al 2000). The independent (input) variables consisted of various attacking and defensive play actions and team characteristics. Measures of play included; shots on target, % of successful passes, the number of tackles made, clearances, blocks, interceptions, dribbles won/lost, the number of free kicks conceded, yellow and red cards. Measures of the goalkeepers’ performance such as % of successful distributions and number of catches/drops were also included as independent variables. Ultimately, the results obtained by Carmichael et al (2000) emphasised the influence of attacking skills such as accurate shooting and passing as well as the importance of defensive skills as reflected in tackles, clearances and blocks made upon match outcomes.

In a follow up study Carmichael et al (2001) utilised aggregated match plays for each team over a full league season (1997-98) in order to examine team performance and efficiency by estimating a season based production function. Carmichael et al (2001) adopted a multiple equation (recursive) system where output (overall team success) expressed as percentage points was determined sequentially. The results obtained from this analysis emphasised the importance of defensive aspects of play and identified shots on the opposition goal, goals scored and accurate passing to be of particular significance with regards to output determination.

Carmichael and Thomas (2005) use match play statistics from the 1997-98 EPL season in order to examine the home field effects (advantages) on within match performance of home and away teams. A variety of team play measures are included as inputs in a team production function in which the number of shots and goals scored by each team in each match are specified as outputs. The model utilised by
Carmichael and Thomas (2005) incorporates a recursive system, with a team’s success in terms of match plays determined sequentially. Team play measures included play actions such as dribbles, runs, passes, fouls, tackles, clearances, blocks and saves made by the goalkeeper. The regression results revealed the statistical significance of attack related team play measures was stronger for the home team and the significance of defence related measures were stronger for the away team, implying attacking play is more important at home and defensive play more important away.

Carmichael et al (2011) study team performance in the EPL over the seasons 1998-99 to 2004-05 and investigate the link between playing success and commercial success. Employing a data set that combines financial measures and performance data, Carmichael et al (2011) based their empirical analysis on three behavioural equations. The first equation is a sports production function where league success (output) was interpreted as a function of playing inputs (performance statistics) and non-playing inputs such as managerial contribution and home support. The second equation was a standard revenue equation in which revenue was specified as a function of output and managerial specific inputs. Equation three was a hedonic wage-price or earnings function that related player performances to wage expenditure. Carmichael et al (2011) revealed on-field success to be directly related to player skills and abilities and revenue was positively influenced by on field success.

In a study on the EURO 2004 international football tournament, Carmichael and Thomas (2008) test an aggregated production function model for competing teams using a match play data set. Carmichael and Thomas (2008) constructed four play variables designed to capture the quantity and quality of attacking and defensive plays and included these variables as independent variables in a regression where the dependent variable was a team’s average goal difference throughout the tournament. In addition to the estimated average production function, Carmichael and Thomas (2008) also estimated a production possibility frontier in order to examine team efficiency. The use of production frontiers in order to measure team efficiency is also evident in literature on team performance. The sub section 3.4.2 below describes the

---

17 Hedonic models are often used to value assets by breaking them down into their component parts and obtaining estimates of the contributory value of each component (Brooks 2008).
production frontier approach and presents the findings of some studies that have adopted this approach.

3.4.2 Production Frontiers and Efficiency Measurements in Team Sports

Two distinct approaches have been adopted to measure team efficiency, the econometric SFA\(^{18}\) (stochastic frontier) approach that builds on concepts of regression analysis and the deterministic non-parametric frontier approach (DEA)\(^{19}\) that consists of mathematical programming techniques. As for production function analyses, efficiency studies have covered match level performance and seasonal team performance in leagues, hybrid cup tournaments as well as assessing managerial and coaching efficiency.

Barros and Leach (2006, 2007) use SFA to measure efficiency scores for twelve English Premier League teams between the years 1999 and 2003. The cost function consisted of three factor input prices (labour and two capital inputs) and three outputs (points, attendance and turnover). An average (time-invariant) efficiency score of 88\% is reported, suggesting that the average club operating at maximum efficiency could reduce costs by around 12\% without affecting outputs.

Similarly, Kern and Süßmuth (2005) use SFA to estimate team production functions for teams competing in the German Bundesliga. The ex-ante inputs (pre-seasonal estimates) of player and coach wages are transformed during the production process of a season into ex-post (actual) pecuniary revenues and sporting success. The output measure takes into account performance across numerous domestic and European competitions and they report a wide variation in team efficiency scores.

In standard SFA, teams are assumed to have the same technological possibilities, however if technological possibility between teams differ technical efficiency may be overstated. In order to overcome this inaccuracy in estimation some studies employ a variant of SFA known as the random frontier model (Greene 2005). The random

\(^{18}\) SFA is a regression based approach used to evaluate performance efficiencies and it is regarded as an alternative to DEA. SFA decomposes the usual error term into two components, an inefficiency component and a random component, that measure things like measurement error and environmental influences and is therefore useful for identifying factors that influence performance (Cooper et al 2007).

\(^{19}\) DEA is used for evaluating performance of various kinds of entities engaged in different activities in different countries. DEA involves mathematical programming techniques that can handle large numbers of variables and (relationships between these variables). This relaxes the requirements that one encounters when having to choose limited inputs and outputs because the techniques employed otherwise encounter difficulties (Cooper et al 2007).
frontier model accounts for heterogeneity as it separates technical inefficiency from technological differences between teams. Barros and Garcia-Del-Barrio (2008) compare the efficiency scores obtained for several cost function specifications (including a stochastic frontier model and a random frontier model) in their analysis of twelve teams competing in the EPL between 1999 and 2004. According to Barros and Garcia-del-Barrio (2008) the random frontier model is a better representation of the data than the standard SFA.


Deterministic frontiers have also been employed to estimate production and cost frontiers based on team sports data, fitted using the DEA linear programming technique (Dobson and Goddard 2011). Studies on North American sports utilising this methodology include Mazur (1994), Anderson and Sharp (1997) and Volz (2009) for MLB and Einolf (2004) for NFL and MLB.

In football, Haas (2003b) uses DEA to measure the productive efficiency of EPL teams during the 2000-01 season using team input measures such as, team payroll and coach salary (as proxies for playing talent and manager contribution) and outputs such as league points and revenues. DEA frontiers were fitted based on both constant returns to scale (CRS) and variable returns to scale (VRS) assumptions concerning the production technology.

Using EPL data from the seasons 1998-99 to 2002-03, Barros and Leach (2006b) combined sporting and financial variables in order to estimate a DEA cost frontier. Outputs were league points, turnover and attendances and inputs consisted of the number of players, net assets, payroll and expenditure on the stadium. Efficiency scores for EPL teams were estimated using both CRS and VRS specifications and Barros and Leach (2006b) revealed mixed efficiency scores for teams.

Haas et al (2004) estimated a DEA frontier using data on the German Bundesliga for the 1999-00 season. Player and manager payroll were used as input measures and
outputs were league points, total revenues and a ratio of attendance to stadium capacity. CRS and VRS specifications were estimated and Haas et al (2004) revealed efficiency scores were not correlated to league position, indicating performance and efficiency are not synonymous.

In a study on American Major League Soccer (MLS) Haas (2003a) estimated a DEA frontier using data from the 1999-00 season. Input measures included player and coach payroll and the outputs consisted of league points, absolute number of spectators and revenues. Efficiency scores were revealed to be highly correlated with league performance. When inefficiency was decomposed into technical inefficiency and scale inefficiency, it was shown that the suboptimal scale of production largely explains inefficiency.

Espita-Escuer and Garcia-Cebrian (2004) used DEA to measure the efficiency of teams in the conversion of attacking moves during a match into sporting success by applying CRS and VRS specifications to Spanish La Liga data from the 1998-99 season to the 2000-01 season. Input variables included the number of players used, the number of attacking moves, the number of minutes during which the teams had possession of the ball and the number of shots and headers (all measured throughout the course of the season). The Output was the number of league points achieved throughout the season. The results revealed that efficient teams did not always correspond with those that finish highest in the league, therefore implying that highly placed inefficient teams could have achieved the same results with fewer resources or could have achieved more with the same resources (Dobson and Goddard 2011).

Bosca et al (2009) used DEA to compare attacking and defensive efficiency in Italian and Spanish football between the three seasons from 2000-01 to 2002-03. Attacking inputs were shots on target, attacking plays created, balls kicked into the opposing penalty area and minutes of possession. Defensive inputs were the inverse of the attacking inputs. The outputs for attacking and defensive production were the number of goals scored and the number of goals conceded. Bosca et al (2009) revealed teams in Spain were more closely matched in terms of ability than teams in Italy. In the Italian league, the best ranked teams scored more goals, conceded fewer goals and on average, obtained more points. In Spain, the correlations between these indicators were revealed to be smaller. Furthermore, Bosca et al (2009) revealed league rankings
were highly correlated with measures of defensive efficiency in Italy, whereas the opposite was true for Spain.

As demonstrated above, one of two approaches has generally been adopted for the measurement of team performance in football, the production function approach and the production frontier approach. In particular, the literature reviewed in this section has highlighted the general lack of attention awarded to the opponent and their impact upon observed performances. In previous literature specific to production in football, Carmichael et al (2000) and Carmichael and Thomas (2005) both include variables accounting for opponent interactions highlighting their importance in the production process, however only at a team level.

Even at a more disaggregated level studies that have analysed technical and physical aspects of play in football have generally overlooked the impact of the opponent. Out of 44 articles reviewing technical ability in football, Mackenzie and Cushion (2013) critically reveal that only 8 studies included the opposition in their analysis. A similar pattern was revealed for performance analysis research concerning the physical demands of football, where out of 15 articles reviewed only 3 acknowledged the opposition. Section 3.5 below presents a review of a seemingly minority of studies analysing technical and physical aspects of play in football that has actively measured (and hence acknowledged) opponent interactions.
3.5 Studies Directly Measuring Opponent Interactions

This section reviews research that has focussed on specific technical or physical aspects of play whilst including measures for the performance of the opponent. The aims and objectives of the studies reviewed in this section vary, however all are specific to football and all account for opponent interactions.

Gerisch and Reichelt (1993) conducted a computer and video based analysis of one-on-one situations (duels) between two opposing players participating in the first leg of a European cup semi final between FC Bayern Munich and Red Star Belgrade. Player’s and their opponent in the duel were identified using player numbers and it was revealed that in a total of 250 duels in the first game between the two sides, each team won 125 duels each. Although the number of duels won was even amongst both teams it was not divided equally amongst individual players. Some players playing in key positions for Munich (identified to be a central defender, two wide players and both strikers) were revealed to have low success in their duels, which were ultimately cited as the reason for why Red Star Belgrade created more chances and eventually won the tie.

Yamanaka et al (2002) performed a computerised notational analysis of three World cup 1998 games played between Japan, Argentina, Croatia and Jamaica with specific focus on the three fixtures in which Japan played. Playing patterns were analysed and technical actions such as dribbling, passing and shooting were compared amongst the four competing teams. In total 32 playing actions were categorised in to ‘time’, ‘place’, ‘player’ and ‘action’ and included in the analysis and performance data was entered after repeatedly viewing videotaped recordings of matches. The football pitch was divided into six areas horizontally and three areas vertically and the frequency of each play action per area was recorded. Yamanaka et al (2002) concluded that Japan did not utilise pitch spaces as well as Argentina, Croatia and Jamaica (contributing to their loss in each fixture against them) and that Japan should look to play more backward passes from the midfield area that enable them to retain possession and build attacks.

More recently, Lago and Martin (2007) conducted an investigation on the importance of possession (of the ball) using data on 170 matches played in the 2002-03 season in
the Spanish soccer league. In particular, four variables were examined, evolving match status, venue, the identity of the subject team and the identity of their opponent in each match. Nineteen dummy variables were created in order to identify the 20 teams in the league with Real Madrid the reference team (as they finished the season as champions). The results obtained from a linear regression analysis revealed that all four variables were statistically significant and together explained most of the variance in possession. Furthermore, the identity of the opponent was revealed to be of importance as the worse the opponent the greater the amount of possession.

Taylor et al (2008) examined the effects of match location, quality of the opponent and match status on the technical aspects of performance for a single professional British football team using data on forty matches played in the 2002-03 season. A computerised notational analysis system was used to notate the 40 match sample with 13 on the ball behaviours (play actions) and their corresponding outcomes (successful or unsuccessful). The quality of the opposition was dichotomised into two categories, “strong” or “weak” depending on whether the opponent finished in the top half (position 1-12) or bottom half (position 13-24) of the division upon conclusion of the season.

The independent and interactive effects of the three situation variables on the play action incidence were examined through log linear modelling. The results indicated that incidences of all technical play actions, with the exception of “set pieces” were influenced by at least one of the three situation variables with both independent and interactive effects discovered. The independent and interactive effects of the situation variables on play action outcomes were investigated through logit modelling, the results of which revealed that the situation variables had no influence on the outcome of play actions.

Lago (2009) examined the effect of match location, opponent quality and match status on possession strategies for a professional Spanish football team (RCD Espanyol). A computerised match analysis system was used to notate (post-event) 27 matches played throughout the 2005-06 season. Quality of the opponents was measured by the distance between end of season league rankings of competing teams. A linear regression analysis revealed that time spent in possession of the ball was greater when playing against weaker opponents. Each unit of distance between the end of season
ranking between competing teams increased or decreased the team’s possession by 0.2%.

Redwood-Brown et al (2012) analyse the impact of different standards of opponents on an observed team’s performance at a team level, positional unit level and individual player level. 29 Premier League matches were analysed during the 2010-11 season for 18 performance indicators. The standard of the opposing team was categorised as “top”, “middle” or “bottom” depending upon their final league position. The participating (observed) team was categorised in the “middle” category and 18 players from this team’s squad were selected for the purposes of the study. A one-way ANOVA analysis was conducted that assessed the observed team’s performance behaviour along with five positional units (centre-back, full-back, central midfield, wide midfield and centre forward) and individual player performance behaviour.

At team level, successful passes were revealed to be significantly higher against middle teams than top or bottom teams and interceptions were also revealed to be significantly higher against middle standard teams than top teams. In general the observed team performed better when playing against teams of similar calibre (other middle ranked teams) rather than top or bottom teams. Performance indicators at the positional unit and individual player level were also revealed to be sensitive to the standard of opponent. Furthermore Redwood-Brown et al (2012) state differences in individual player performances are not always evident at a positional unit or team level.

As can be seen above there are some studies that account for the quality of the opponent in football, however as stated by Tenga et al (2010) these studies tend to be limited by univariate data analyses or by small sample sizes, and are hence not representative of a population (Mackenzie and Cushion 2013). In a league consisting of twenty teams, a traditional season would consist of 380 fixtures. However, none of the studies reviewed above surpass 170 matches (study conducted by Lago and Martin 2007) and even this sample consisted of games played between teams across different Spanish leagues. Furthermore, although these studies have accounted for

20 Analysis of variance (ANOVA) tests are used to compare several samples (three or more) in terms of some numerical dependent variable. ANOVA tests determine whether or not any significant differences exist between the samples (O’Donoghue 2012).
opponent interactions, absolute measures of their ability (league position) have been utilised meaning the impact of the manager (in terms of specific match tactics and team selections) and relative team form have been ignored.

In various team sports it has been generally accepted that manager or head coach of a team has some influence upon overall performance (Audas et al 1999; Scully 1994; Hadley et al 2000; Dobson and Goddard 2011). At a team level, tactics and strategy could potentially decide between the winners and losers of a contest. At an individual player level coaching, training and motivating could all play a part in improving player ability. Section 3.6 presented below shifts the focus towards another factor of production in the form of a manager’s contribution.
3.6 Impact of a Manager upon Performance

The degree of impact a coach/manager can have upon performance is disputed amongst previous research and appears to vary across sports. For example, Porter and Scully (1982) claim the contribution of a coach to team performance in baseball is comparable to that of an individual star player. Conversely, in football, Barros and Leach (2006a) claim the impact of a manager can be limited without the presence of an exogenous shock (a billionaire takeover for example). Although the magnitude of a coach or manager’s importance appears to vary across sports, there is a general consensus amongst the previous literature that the ability and skill of the coach does have at least some bearing upon team and individual performances.

The extent of the impact of a coach or manager has been examined across various sports and is prevalent in previous literature. Hadley et al (2000) utilised a poisson regression model in order to assess the performance of coaches in the NFL using a data set ranging from the 1969-70 season to the 1992-93 season. The mean performance index for the entire sample of NFL teams was revealed to be 0.641. Hadley et al (2000) stated that a team would require 10 wins out of a traditional 16 game season in order to qualify for the post season playoffs. In a hypothetical team with sufficient talent to achieve these 10 wins, an efficient coach with average performance would be expected to win 6.41 games. The difference between an average coach and the best was therefore revealed to be approximately 3.5 game wins per season. Hadley et al (2000) also discovered that successful NFL teams that were managed well, often turned out to be to most consistent performers throughout the 22-year sample period and it was concluded that an efficient coach was a vital component of the production process (Hadley et al 2000).

Further examples of North American studies analysing coaching performance includes, Porter and Scully (1982) whose results indicated that managerial skill made a substantial contribution to the production process. Scully (1994) employs a stochastic frontier to estimate managerial efficiency across baseball, basketball and American football and discovers a strong positive correlation between managerial efficiency and average managerial survival time (job tenure).
In a follow up study, Scully (1995) developed a model of managerial departure in the MLB, NBA and NFL and revealed career efficiency to be positively related to career length and average efficiency to increase over time in all three sports. In addition, involuntary terminations were found to be highly sensitive to league standings in baseball and basketball. In other studies, Hofler and Payne (1996, 1997) examine coaching efficiency in the NFL and NBA and Ruggiero et al (1996) calculate efficiency scores for 24 coaches with four or more full seasons employment in the MLB.

Studies on football have also recognised the role of the manager in the production process. Dawson and Dobson (2002) state that the manager of a football club is required to perform important functions similar to that of his business counterpart such as, planning, controlling, leading and organising. Furthermore some studies investigating the impact of a manager in football have discovered that in particular, poor team performances tend to trigger managerial dismissals (Audas et al, 1999; Salomo and Teichmann, 2000; Frick et al, 2010) suggesting football managers are an important factor in achieving team success and are hence held accountable when their team fails to achieve.

Audas et al (1999) employed match level data on English professional football over a 25-year period in order to separately estimate voluntary (resignation) and involuntary (dismissal) job termination hazard functions. In particular the use of match data enabled the investigation between short-term fluctuations in team performance and managerial survival.

Audas et al (1999) revealed voluntary resignations were more likely from individuals who had already accumulated significant experience prior to their current appointment, as they were more likely to receive job offers from other clubs. Similarly managers that had previous playing experience and possessed international caps were more likely to receive job offers and hence quit their current position voluntarily.

Audas et al (1999) suggested that involuntary terminations could reflect managerial ability as short-term fluctuations in team performance were revealed to strongly influence the involuntary termination of managers. In addition, job insecurity was shown to increase over time suggesting increased commercial pressures have raised
the demand for success as well as the penalties for failure. Recent match results and other team performance measures were revealed to be strongly significant in the involuntary termination case. Voluntary termination cases were also sensitive to recent match results but to a much lower extent. Managerial human capital characteristics were largely insignificant in the involuntary termination case but significant in the voluntary termination case.


Comparing five different models, it was revealed that estimates of coaching efficiency were highly dependent on the model specification and the estimation procedure. According to Dawson et al (2000) the explanatory power of the five different models suggested that two thirds of the variation in win percentage was accounted for by differences in playing talent between teams. The remaining one third of variation was concluded to be either due to statistical noise or coaching inefficiency.

Estimates of coaching efficiency were found to be very sensitive to the use of wages as input measures (Dawson et al 2000). Results obtained from the analysis suggested that the estimated sporting production function for English association football was little affected by divisional coverage of their sample or by the alternative measures of team performance. Furthermore, Dawson et al (2000) revealed individual scores for coaching efficiency were not highly correlated with team performance.

In conclusion, Dawson et al (2000) stated that coaching efficiency was only partially correlated to team performance. It was suggested that coaching efficiency should be based instead on success achieved relative to the available playing talent rather than match outcome, thereby suggesting that the most successful coaches may not necessarily be the most efficient.

Salomo and Teichmann (2000) employed logit regressions to model managerial departure using data on the German Bundesliga between the seasons 1978-79 and 1997-98. The primary reasons for managerial dismissal are revealed to be poor
performance in the most recent matches, a breakdown in relations between the manager and directors and the intensity of media speculation. The appointment of a new board president was also revealed to be positively correlated with the probability of manager dismissal and the managers of teams that were exposed to intense media coverage (especially when on poor form) were also more vulnerable to termination.

Dawson and Dobson (2002) employed a production frontier analysis to measure managerial performance amongst managers of professional English clubs between the 1992-93 season and 1997-98 season. Specifically, Dawson and Dobson (2002) investigated the relationship between managerial performance and human capital. Human capital in general, and certain characteristics such as having prior affiliation with the club and achieving international recognition as a player were revealed to be especially important with regards to managerial efficiency (Dawson and Dobson 2002).

Barros et al (2009) employed duration models to analyse the length of time coaches survived (avoided dismissal) in the German Bundesliga over a 22-season period from 1981-82 season to the 2002-03 season. According to Barros et al (2009) head coaches of successful teams are more likely to survive, whereas the head coaches of teams with relatively high wage bills are more susceptible to dismissal. Barros et al (2009) claim that the Bosman ruling has also contributed to an increased probability of survival for head coaches in the German Bundesliga, possibly due to a number of reasons such as, top quality coaches becoming more scarce, terminating coach contracts becoming more expensive and finding quality managers midseason becoming much more difficult.

Frick et al (2010) employed a random parameter logit model in order to investigate the determinants of dismissal for head coaches in the German Bundesliga. 114 different coaches were assessed over a range of seasons starting from the 1981-82 season and ending in the 2002-03 season. Frick et al (2010) discovered that the probability of a head coach resigning or being fired was positively related to the most recent performances. The probability of a coach being fired was revealed to be negatively related to his individual efficiency, his career win percentage, his level of experience as a head coach in the Bundesliga and the number of points won in the current season, suggesting a coach with a proven track record and vast experience is
given more leeway (by fans or their superiors) at the club in the case of poor performance.

According to Frick et al (2010) managerial efficiency considerably reduces the probability of a manager being dismissed. Furthermore, it is suggested that the best way for a head coach to survive in the German Bundesliga is to sign for a less ambitious team where the expectations of the fans and management is lower rather than an expensively assembled team, where the high level of expectation (and hence lower tolerance of failure) means head coaches are more vulnerable.

The literature review presented in this section illustrates that managers are generally accepted to have some bearing upon performance levels. Studies have revealed a positive correlation between involuntary termination and poor team performance (Audas et al 1999; Frick 2010) and more specifically poor performance in the most recent matches (Salomo and Teichmann 2000) suggesting managers are held accountable for underperforming teams.

So far this chapter has discussed how contest designers can manipulate production in sports in section 3.2. Section 3.3 emphasised the importance of the opponent in the production of a sporting contest. Section 3.4 detailed the production function process in team sports and its general use in the analysis of team performance as well as having identified several team performance measures utilised as inputs and outputs. This section has discussed another factor of production in the form of managerial contributions.

The focus in section 3.6 below shifts towards estimating production functions through the utilisation of disaggregated match specific data. Some previous research has supported the use of game level data in order for greater accuracy in estimations. As emphasised in chapter two, with a club good, outputs are related (simultaneity between opponents) and therefore the use of disaggregated match data could potentially capture this relationship and offer a richer explanation of the determinants of performance. Section 3.7 below explores the reasons why match specific data is considered to provide a more accurate assessment of production in sports and discusses why a similar approach could be beneficial for the purposes of this thesis.
3.7 Disaggregated Analysis

Since the early work of Scully (1974) and Zech (1981) there has been a persistent tradition of production function analyses in the economics of team sports. A sporting production function can serve as a basis for understanding the determinants of players’ salaries, as well as helping to identify team strengths and weaknesses (Carmichael et al 2000). A production function in football can therefore influence decisions made by team managers (and club owners) in the transfer market, including which type of player should be acquired to strengthen and improve the team and similarly which players need to be shed (Carmichael et al 2000). Zak et al (1979) in a study on baseball also stress the importance of accurate production function estimations as the use of performance data can identify those areas in which a team’s best opportunities for improvement lie. Leard and Doyle (2010) reiterate that analysing the relationship between inputs and sporting success empirically can have a major impact on player salaries and player and coach terminations.

Much previous research into the production of sports teams has consisted of the aggregated use of match statistics in order to provide seasonal based estimations of production. The aggregation of data however ‘masks’ important information relating to actual productivity, leading to inaccurate estimations (Leard and Doyle 2010). This section presents a review of research that has utilised disaggregated game play data in order to avoid the problem described above and then discusses the potential advantages the use of a similar approach could have upon this study.

Some previous research has indicated the use of disaggregated game play data is more suitable in estimating the relationship between inputs and sporting success in the production process. Subtle variations between games, which are not accounted for in studies that employ the use of aggregated data, can be preserved and hence accounted for by the use of disaggregated game specific data. Leard and Doyle (2010) in a study on the NHL claimed that aggregating game level data into seasonal data would remove important game level variation in play statistics and therefore obscure the identification of their effect on performance.

Carmichael et al (2000) in their study of football also express the value of disaggregated data by stating data of this type allows for greater focus on the on-field
actions which directly contribute to individual match outcomes, which can otherwise be distorted, or concealed, by season aggregated statistics. Furthermore, McGoldrick and Voeks (2005) state the use of aggregated data can lead to the loss of important information and implies the existence of multi-collinearity. In their study on NBA and WNBA basketball, McGoldrick and Voeks (2005) compared differences between the use of absolute measures and relative measures of output (probability of winning versus the ratio of final scores) and deemed the use of relative measures more appropriate.

Zak et al (1979) employed disaggregated match data collected from the 1976-77 NBA season to estimate a production frontier in order to determine levels of team efficiency and identify determinants of team performance. Depkin and Wilson (2004) also employed disaggregated game-specific data from 2,222 NHL games played in seasons 1998-99 and 1999-00 in order to investigate the impact of second referees on arena attendances and national TV audiences. In another study on the NHL, Shmanske and Lowenthal (2007) utilised a disaggregated data set in order to analyse team performance in overtime.

A distinct advantage of disaggregated data sets is they allow the capture of in-game variation and in-between game variation (Leard and Doyle 2010). Studies that have utilised disaggregated game play data stress that this type of data is more capable of reflecting the impact of an opponent on observed performance. As stated by Leard and Doyle (2010), the marginal effects of additional effort cannot be assumed constant from game to game since each game involves a different opponent. This variation between games therefore could potentially be accounted for by utilising a disaggregated data set. Furthermore a team’s ability to score goals depends on its ability relative to its opponents thereby illustrating the importance of relative performance measures.

The results obtained from the literature reviewed in this section offers strong support for the use of disaggregated game data. Results obtained in these studies suggest that match level data is more capable of capturing the effects of opponent interactions. The utilisation of disaggregated match data for the purposes of this study could therefore result in more accurate production function estimations and hence, the
relationship between inputs and sporting success could be reflected in a more appropriate and precise manner.
3.8 Conclusion

Section 3.2 highlighted how the production of a sporting contest and the quality of the contest is partly influenced by tournament organisers through contest design. The offer of incentives, the number of entrants and the structure of a tournament (i.e. rules and format) all have to be finely balanced in order to ensure an optimal contest is produced.

Furthermore what is of great importance in the production of a sporting event in team sports is the presence and contribution of the opponent. As stated in section 3.3 the joint production element in team sports carries with it the element of uncertainty. Consequently the UOH has since become a central tenet in the economic analysis of sports teams. Two teams are required to produce a sporting contest and it was this that led Rottenberg (1956) to first propose the concept of a sporting production function.

The literature reviewed in section 3.4 outlined the use of production functions (and frontiers) in the analysis of team performance. Furthermore this section highlighted that regardless of the measurement approach adopted, relatively little attention has been given to the impact of the opponent(s) performance in football. Opponent interactions have only rarely been specified as inputs to the production process. Furthermore when the impact of the opponent has been acknowledged absolute measures of quality have been used (such as final league position) which do not account for the game-to-game variation caused by the relative performance of varying opponents.

As revealed in section 3.5 studies specifically analysing technical and physical aspects of performance in football have also generally overlooked the performance of the opponent (MacKenzie and Cushion 2013). Similarly to production function studies, the minority of studies that have acknowledged the opponent have usually included absolute measures of opponent quality such as final league position, even though relative measures are more appropriate as in game variation and in-between game variation can be captured (McGoldrick and Voeks 2005; Leard and Doyle 2010).

In addition, many studies reviewed in section 3.5 suffer from small sample sizes (Tenga et al 2010) questioning the overall representative nature of these studies
(MacKenzie and Cushion 2013). Furthermore, only Redwood-Brown et al (2012) assessed player performances whilst in direct competition with an opponent at a more isolated individual player level, however only for a sample of 18 players, over only 29 games for only one observed team (also questioning how representative of a population this study is).

Section 3.6 outlined the role of a manager in the production process and emphasised the impact a manager can have upon performance levels and hence team success. Managerial impact, a non-playing input in the production process appears to have been understated in production function studies in football. Carmichael et al (2000; 2001), Carmichael and Thomas (2005) and Carmichael et al (2011) all acknowledged that managerial input can have some bearing upon match results yet only Carmichael et al (2011) included an active measure of managerial impact in their model.

In summary, the literature reviewed throughout this chapter has identified gaps and a few limitations of previous research. Generally the impact of the opponent has been ignored and on the rare occasion measures for the quality of the opponent are included, absolute measures have been employed that ignore aspects such as relative team performance and hence the game-to-game variation caused by the regularly changing opponents. The relative performance of an opposing team and the impact of their performance upon an observed team’s performance (and hence the assumed simultaneity) at a match level in football has not been previously addressed. This study will advance previous literature by estimating match based production functions, using match specific play variables whilst accounting for the relative performance of the opponent and hence the simultaneity between opposing teams’ performances.

Furthermore, of all studies reviewed in this chapter only Redwood-Brown et al (2012) analyse player performance at a more isolated individual player level\(^\text{21}\) (rather than team level) however only for a sample of 18 players over only 29 games. A production function estimating the degree of simultaneity between the relative performances of individual players in direct competition has not been previously attempted. This study will partially fill this gap by estimating match based production functions, using match specific play variables whilst accounting for the relative

\(^{21}\) It should be noted that this is not a production function study. A one-way ANOVA methodology is adopted to analyse team/positional unit/individual player performance behaviour.
performance of the opposing player and hence the simultaneity between the performances of individual players in direct competition.

Finally, sporting production function studies investigating the degree of simultaneity between the performances of different positional units and their opposing team(s) have not been conducted previously. This study will partially fill this gap by estimating match based production functions, using match specific play variables whilst accounting for the relative performance of the opponent and hence the simultaneity between the performances of positional units and their opposing team(s) (collectively).

Chapter four below details the methodology adopted in order to fulfil the objectives of this thesis (outlined in chapter 1) and address the gaps in previous literature detailed above. The chosen methodology for the empirical analyses conducted throughout this study, along with model specifications and justifications for the adopted methodology are outlined in subsequent chapters.
Chapter 4

Methodology

4.1 Introduction

According to Blaikie (2003), methodology refers to the analysis of methods of how research problems should be undertaken. A methodology should incorporate all principles, theories and values that underpin a particular approach to research. The methodology should therefore discuss the theoretical perspectives that underpin a research problem, how these theories are generated and what criteria should be met in order for research to be relevant to a particular field (Blaikie 2003). This chapter will outline the methodological and philosophical foundations of this study and review the key theoretical approaches that affect the research design. In addition, the data used to conduct this study will be described and the selected econometric process will be analysed and then justified in accordance to this research’s aims and objectives. Key performance indicators (KPI) of playing positions will then be determined via a comprehensive review of previous literature in the field of performance analysis in football and the justification for the inclusion of these KPI’s in subsequent empirical chapters will be offered.

Section 4.2 below defines ontology, epistemology and research paradigms and examines the ontological and epistemological stance adopted, which ultimately provided the philosophical foundations of this thesis. Section 4.3 details the research design and strategy adopted for this study and offers justification for these choices as well as outlining the two stage least squares regression (2SLS) model used to conduct all the empirical analyses throughout this study. Section 4.4 offers a detailed description of the data used to conduct this study and in particular highlights the difficulty academics face in obtaining such unique data, hence, this section also outlines how this unique data set has been fully utilised. Section 4.5 details how key performance measures used in the empirical chapters of this current study were identified and why they merited inclusion in the empirical analyses conducted in
subsequent chapters. Finally section 4.6 concludes this chapter and provides a
discussion of the methodological underpinnings of this study.
4.2 Philosophical foundations

When undertaking research it is extremely important to consider several different research paradigms with regards to ontological and epistemological stances. Hatch and Cunliffe (2006) emphasise how different paradigms can encourage researchers to study the same phenomena in various different ways and they stress how a particular phenomena can be described from several totally different perspectives. These varying perspectives highlight how different kinds of knowledge can be derived from observing the same phenomena.

It could be said that ontology and epistemology are to research what ‘footings’ are to a house, as they form the foundations of the whole structure (Grix, 2010). If a researcher is to present clear, precise and logical work and engage and debate with another’s work then one has to know the core assumptions that underlie their work and inform their choice of research questions, methodology, methods and even sources.

There are several reasons for wanting to have a clear and transparent knowledge of the ontological and epistemological assumptions that underpin research. Firstly, knowledge of this type enables one to understand the interrelationship of key components of research, including methodology and methods. Secondly, transparent and clear understanding of ontological and epistemological assumptions avoids confusion when discussing theoretical debates and approaches to social phenomena. Finally, understanding of ontological and epistemological assumptions enables us to be able to recognise others, and when required to defend our own positions (Grix 2010).

Ultimately, these philosophical parameters can help describe a researcher’s beliefs, assumptions and perceptions with regards to reality and truth (Grix 2010). In other words these parameters can influence the way in which research is undertaken which then goes on to influence results gained and conclusions made. It is therefore important to discuss and understand these parameters in order to ensure that the eventual research method and approach chosen is correctly aligned to the aims and nature of the research hypothesis and also to ensure that a researchers potential bias are understood and minimised.
According to Blaikie (2000), these parameters must be carefully considered by a researcher as it is extremely important that the choices a researcher makes are connected to and in alignment with the original research question. Without this consideration, incompatible methods and philosophical stances may be adopted which would ultimately question the credibility of the research due to a lack of rationality and coherence.

4.2.1 Ontology

According to Blaikie (1993, p6) ontology is “the science or study of being” and he develops this description for the social sciences to encompass “claims about what exists, what it looks like, what units it is made of and how these units interact with each other” (Blaikie 1993, p6). In other words, ontology simply refers to the study of the nature of reality, and more specifically, whether this reality actually exists (objective reality) or whether this reality is a fabrication of the human mind (subjective reality).

Broadly speaking, ontological positions are often contained under the terms ‘objectivism’ or ‘constructivism’ (Grix 2010). Objectivism refers to an ontological stance, which asserts that social phenomenon and their meanings have an existence that is independent of social actors and they are indeed facts that have an independent existence (Walliman 2006). Objectivism is also referred to as realism and a realist ontological position can be divided into shallow realist, conceptual realist, cautious realist, depth realist, idealist and subtle realist positions22.

Constructivism on the other hand is an ontological stance revolving around the belief that social phenomena and their meanings are being continually accomplished by

---

22 Shallow Realism: A shallow realist will observe what exists and assume that is all that exists. In other words “what you see is what is there” (Blaikie 2007, p14).
Conceptual Realism: Reality is seen to have an existence independent of the human mind however this reality can only be known through human thought and reason. This reality is not directly observable as it is a structure of ideas, separate from the thoughts of any one individual but shared by human beings collectively (Blaikie 2007).
Cautious Realism: As with a shallow realist, an independent external reality is assumed. This reality is however impossible for humans to perceive due to their imperfect senses. Researchers therefore have to be ‘cautious’ and critical of their work because of these human frailties (Blaikie 2007).
Depth Realist: Reality consists of three levels. The empirical level that we experience through the use of senses, the actual level referring to the existence of events regardless of whether anyone observed them or not and the real level referring to the structures and processes that generate events. As reality consists of these three levels, reality has an ontological depth independent of our knowledge of it and hence observable phenomena can be explained with reference to underlying structures and mechanisms (Blaikie 2007).
Idealism: “Reality is what human beings make or construct” (Blaikie 2007, p16). Social reality consists of a meaning giving process and it is these meanings and interpretations created by social actors that constitute their daily lives.
Subtle Realism: Attempts to overcome deficiencies of other realist positions. Assumes external reality but rejects the notion that we have direct access to this reality even though knowledge is based on assumptions and is constructed by individuals (Blaikie 2007).
social actors. Constructivism implies that social phenomena and categories are not only produced through social interaction but that they are in a constant state of revision (Bryman 2001), therefore social knowledge can only be inter-determinate (Walliman 2006). Dependent on which ontological position is adopted then, the social world can either be regarded as having an external reality to social actors or a reality that is interpreted differently by individuals (Bryman 2008).

To further illustrate the importance of the ontological position a researcher adopts and to emphasise the difference between an objectivist and a constructivist approach, a good example would be to think of a report formulated on the daily activities taking place in a work environment. Upon reading the report it should be considered whether this report depicts what is actually going on, or whether the report simply depicts what the author believes is going on. As one can imagine there is no easy answer to the question above and difficulty is only increased when considering intricate phenomena such as power, control and culture (Hatch and Cunliffe 2006). With phenomena such as these, observers are tasked with attempting to determine whether these phenomena really exist or are simply a creation of the mind. Consequently all researchers will follow a set of ontological assumptions which will impact their view of what is real and whether existence of an object is attributed to experience and interaction with it (constructivism), or if this object exists regardless of any experience of, or with it (objectivism).

The ontological position a researcher adopts is therefore considered to be a great importance because:

It is only with a clear understanding and recognition that a wide variety of ontological positions actually exist and can lead to different research results, that we can begin to engage with other scholars’ work. Grix (2010, p60)

Taking into consideration all of the above, this study adopted an objectivist (cautious realist) approach. As this study is based on economic theory, which demands that any relationship between two variables (as the hypothesis suggests) must be observable and quantifiable (Asteriou and Hall 2011) the objectivist ontological stance adopted is logical.
Researchers will all follow a number of ontological assumptions, which affects their view and understanding of what is real and whether this existence is attributed to one set of rules over another. If these (underlying) assumptions are not considered or identified then a researcher risks dismissing certain aspects of the social phenomena as they are implicitly assumed and taken for granted and therefore not open to debate or question. As researchers’ hold differing views on what constitutes reality, another question is then raised which is how can this reality be measured? The answer lies in epistemological assumptions.

4.2.2. Epistemology

Epistemology can be described as a branch of philosophy that addresses the question of the sources, nature and limits of knowledge (Klein 2005). Another description of epistemology is offered by Hatch and Cunliffe (2006) whom define epistemology as “knowing how you can know” and develop this by asking the questions, how is knowledge generated? And what differentiates good knowledge from bad knowledge?

It can be seen from the definitions offered above that epistemology is quite a broad topic within the field of philosophy, however, generally speaking epistemology is about how we know things. According to (Grix 2010) epistemology focuses on knowledge gathering processes and is concerned with developing new models or theories that are better than competing models and theories. Blaikie (1993) states, epistemology is a set of claims or assumptions about the ways in which knowledge of reality can be acquired, what can be known, how it is known and what conditions must be satisfied in order for it to be described as knowledge. Or in other words, epistemology is a theory of knowledge, which represents the criteria that should be met in order for reality to be distinguished from beliefs (Blaikie 2003).

According to (Grix 2010), epistemological positions are often divided between those based on foundationalism and those based on anti-foundationalism. Foundationalist views consist of the belief that reality exists independently of our knowledge of it, and that there are central values that exist that can be rationally and universally grounded. Anti-foundationalists on the other hand do not believe that the world exists independently of our knowledge of it, but rather human actors socially and
discursively construct reality. They also believe that there are no central values that can be rationally and universally grounded.

Sometimes, contrasting epistemological stances can be adopted and depending upon which epistemological stance is actually taken, this will most probably influence the type of methodology the researcher eventually employs. This highlights, firstly, the way in which a researcher’s ontological and epistemological positions can lead to different views of the same social phenomena, and secondly, the inter-dependent relationship between ontology, epistemology and methodology. Once this relationship is understood, it becomes much clearer why a researcher’s philosophical stance is of such importance. A researcher that holds a particular ontological stance may be influenced in their epistemological choices, which will then go onto influence their chosen methodology.

There are a number of specific epistemological positions, which include, empiricism, rationalism, falsificationism, neo-realism, constructionism, and conventionalism (Blaikie 2007) each of which will now be briefly described. Empiricism is described as knowledge that has been gained by sensory experience (Walliman 2006), and revolves around the notion that any idea that is not confirmed through experience is not valid and cannot be considered as scientific. Rationalism in contrast, is described as knowledge gained by reasoning (Walliman 2006), and revolves around the notion that social reality consists of the observed as well as the unobserved structures that are innate to human beings. Falsificationism, otherwise known as the hypothetico-deductive strategy was championed by Karl Popper (1959) and was proposed in order to account for the weaknesses of empiricism. This stance implies that for any theory to be regarded as scientific, it must be possible, at least in principle to falsify it, by using evidence to challenge it (Blaikie 1993).

Keat and Urry, (1975, p5) state that in neo-realism ‘a scientific theory is a description of structures and mechanisms which causally generate the observable phenomena, a description which enables us to explain them’. This view of causation highlights the independence of an event and its associated structures and mechanisms (Blaikie 2007). Although both rationalism and neo-realism entail the idea of an underlying reality and both contribute causal powers to this reality the former is a reality of shared innate ideas while the latter is an external independent reality. Hence, not only
are the conceptions of reality different, the influences on the inhabitants of this reality are found to be in different places (Blaikie 2007).

Constructionism consists of the belief that knowledge is the outcome of people having to make sense of their encounters with other people and the physical world. Therefore, contrary to both empiricism and rationalism, constructionism rejects the notion that knowledge is discovered from an external reality or that it is produced by reasons independently of such a reality (Blaikie 2007). Constructionism is regarded to have two different forms, constructivism and social constructionism. Constructivism refers to the ‘meaning-giving activity’ of the individual mind (Blaikie 2007) therefore implying that knowledge is derived from an individual’s meaning giving activity to everyday life. Social constructionism, refers to the collective, inter subjective shared knowledge derived from social (rather than individual) meaning giving activity in explaining social processes.

In conventionalism, similarly to constructionism, reality is regarded to be a human creation. However, whilst in the latter, reality is assumed to be produced by social actors as they conduct their everyday lives, in the former reality is assumed to be an invention of the scientist. In other words a conventionalist is of the belief that scientific theories are created by scientists as a convenient tool to deal with the world (Blaikie 2007), that can be justified by the production of desired results whilst placing very little emphasis on whether they are true representations of the empirical world. According to a conventionalist, scientific theories go beyond the available data in determining what is true and what is not. As a result, whatever is regarded as reality is a consequence of the theory that is used, theories do not therefore describe reality but just determine what the scientist deems to be real.

In this study, the epistemological approach of falsificationism is adopted. Falsificationism implies that theories exist before observations take place. The occurrence of observations therefore can be used to test pre-existing theories, in an attempt to accept true ones or reject false ones and not as a means to develop them. Falsificationism is hence a logical fit as the research question implies a hypothesis, motivated by economic theory, to be examined empirically.

The inter-dependency of ontological and their related epistemological positions can form what Blaikie (2000) refers to as research paradigms. These philosophies are
effectively therefore, different research approaches that incorporate both an ontological and (the aligned) epistemological set of beliefs and assumptions. The next section outlines some of these key paradigms.

4.2.3 Research Paradigms

“A paradigm is an integrated set of assumptions, beliefs, models of doing good research, and techniques for gathering and analysing data” Neuman (2007, p41). There are three broad paradigms that are available to researchers with regards to the philosophy of the social and human sciences and these are known as the positivist, post-positivist and interpretivist positions (Grix 2010), all of which indicate how ontological and epistemological assumptions are often combined in research (Marsh and Furlong 2002). Each paradigm determines the overall strategy adopted for a particular study in relation to the hypothesis being tested and holds different assumptions in interpreting social inquiry.

Positivism has proved to be the most dominant research paradigm in the last century and in fact more recent paradigms use it as a marker against which they seek to differentiate themselves (Grix 2010). Positivism entails ontology of “an ordered universe made up of atomistic, discrete and observable events” (Blaikie 1993, p94). In other words, this paradigm revolves around the belief that only that which can be observed, (experienced by the sense) can be regarded as real and therefore worthy of the attention of science. Positivism is based on a realist, foundationalist epistemology that lays great emphasis on explanation in social research as opposed to understanding. Many positivists assume there is no dichotomy between what we see and how things really are and that the world is real and neither mediated by our senses nor socially constructed (contrary to realism and interpretivism).

The attractiveness of an approach seeking the precision, exactitude and power of prediction promised by the natural sciences is understandable. The human sciences can be messy, people unpredictable and factors leading to events hard to unravel. Positivism attempts to overcome this messiness by seeking rules and laws with which to render the social world understandable (Grix 2010). Many large-scale bureaucratic agencies, companies and people in the general public favour the positivist approach
because it emphasises getting a measure of ‘hard facts’ in a numerical form (Neuman 2007).

Positivism places an emphasis on empirical theory in the production of knowledge, hence it rejects normative questions (such as questions of values and trust) and believes that social science can be value free (Grix 2010). Consequently, the positivist approach revolves around the principle that, knowledge can be gained by the gathering of facts that constitute the basis for law or law like principles.

The vast majority of positivist studies are therefore quantitative and experiment based (Neuman 2007). The positivist approach consists of the advocating of value-free science, the seeking of precise quantitative measures, the testing of causal theories with statistics and belief in the replicating of studies, as other studies that find similar results will allow confidence to grow that the workings of social reality have been captured accurately and therefore scientific knowledge will increase (Neuman 2007).

According to Blaikie (1993), interpretivism entails an ontology in which social reality is regarded as the product of processes by which social actors together negotiate the meaning for actions and situations. Put another way this paradigm holds the view that human social life is not based on objective, hard and factual reality, but instead on the ideas, beliefs and perceptions humans hold about reality (Neuman 2007). Interpretivism is therefore regarded as the polar opposite of the positivist approach. Whereas positivists seek objectivity and tend to model their research on the natural sciences, interpretivists believe in subjectivity and believe there is a clear distinction between the natural and social world and therefore call for the use of methodologies that are more appropriate for the subject being studied (Grix 2010).

The interpretivist paradigm maintains that the view of the world around us is the creation of the mind. In other words the world can only be experienced through our perceptions, which are influenced by our beliefs and preconceptions (Walliman 2006), therefore we are not neutral observers. From an interpretivists view, social reality is not some ‘thing’ that may be interpreted in different ways, it is those interpretations, hence in contrast to physical reality, social reality is pre-interpreted (Blaikie 1993). Unlike the natural sciences, the researcher is not observing phenomena from outside the system, but will in fact study the social world from within, as they are inextricably attached to the social reality being researched.
(Walliman, 2006), and therefore use methodologies different to those used to study natural sciences (Grix 2010).

In the epistemology of the interpretivist stance, knowledge is derived from everyday concepts and meanings (Blaikie 1993). Interpretivist positions are therefore based on anti-foundationalist epistemologies and subscribe to the view the world does not exist independently of our knowledge of it (Grix 2010). Interpretivists in general, do not attempt to provide causal explanations in the social world, as they are more about understanding the social world. Researchers of this paradigm are therefore quite sceptical of positivist attempts to produce precise quantitative measures of objective facts as they view social reality as fluid and subject to regular change (Neuman 2007). Researchers of this paradigm tend to place emphasis on the meaning in the study of social life and emphasise the role that language plays in constructing ‘reality’ (Grix 2010) and therefore tend to favour qualitative rather than quantitative data.

Critical realism is sometimes referred to as post positivism and unlike the previous two approaches detailed above (which are polar opposites), this paradigm is an umbrella term under which more specific approaches to social enquiry fall. Critical realism can be seen as the reconciliatory approach which combines the ‘how’ understanding (linked to interpretivism) with the ‘why’ explanation (linked to positivism) approaches in an attempt to bridge the gap between the two extremes (Grix 2010). According to Neuman (2007) the critical approach emphasises the multi-layered nature of social reality. In other words, on the surface there appears to be illusion, myth, manipulation and distorted thinking but a realist will look beyond the surface into a deeper, hidden level where a “real” objective reality lies (Neuman 2007).

A critical realist will believe in the existence of a natural order in social events and discourse, akin to the beliefs of a positivist. However unlike a positivist, a critical realist will claim that this underlying order cannot be discovered through merely observing a pattern of events but instead through the process of interpretation, while doing practical and theoretical work in the social sciences (Walliman 2006). Critical realism is therefore believed to be different from positivism in two important respects. Firstly, realists have a much more open conception of ontology. A realist will assume that ontological mechanisms are highly complex, often structural and crucially not
always directly observable (Grix 2010). This therefore means that sometimes explanation must be derived from interpreting causal links from observable outcomes. Another major difference between positivism and critical realism is that as realism incorporates agents, explanations must sometimes require an understanding of the differential meanings, which agents infer upon their actions (Grix 2010). In other words, as agents themselves are active in the explanation process and are likely to interpret situations differently, the search for an explanation will require the incorporation of the notion of hermeneutics.

Similarly to how a critical realist shares some beliefs with a positivist, they will share much in common with an interpretivist (Blaikie 1993), such as the shared perception that interpretive understanding is an important feature of social science (Grix 2010). A critical realist however will be distinguished from pure interpretivists in their belief that there are underlying objects and structures at work that generate social events and formulate concepts and theories (Walliman 2006). For a critical realist, the objects and structures in society are understood to have causal powers, allowing for causal statements and the identification of causal mechanisms, contrary to the belief of an interpretivist (Grix 2010).

The three research paradigms detailed above are regarded as the three broadest ones and most research, whether explicitly stated or not will take place within one of them. It is however important to understand that these paradigms are not clear cut and according to Grix (2010) much of the best research takes place on the borders between them. Bearing this in mind this study is more suited to and will hence adopt another approach (which sits on the rationalist side of positivism) otherwise known as critical rationalism.

Critical rationalism is a rationalist side of positivism, which applies a deductive research strategy (Grix 2010) and adopts the position that ‘the natural and social sciences differ in their content but not in the logical form of their methods’ (Lewis-Beck et al 2004, p377). Working within this paradigm, theories are only regarded useful if they can generate testable (preferably falsifiable) hypotheses, hence this approach rejects the positivist position in favour of a different logic of explanation based on methods of trial and error where theories are tested against ‘reality’ (Blaikie 1993). According to Grix (2010), this type of approach is commonly adopted in
sociology, political science and economics and one of the main assumptions it carries is that all political actors maximise their own personal utility, or self-interest, when choosing between alternatives.

The founder of this approach was Karl Popper (1959) and although Popper shared the view that scientific knowledge is the most certain and reliable form of knowledge available to humans he rejected the idea that observations are the foundations of scientific theories (Blaikie 1993) and was therefore critical of the traditional positivist approach. According to Popper, critical rationalism is a search for truths, however he argued that we can never establish theories to be true, all we can do is eliminate those theories that are false (Blaikie 1993). In other words what distinguishes critical rationalism from other paradigms is that here knowledge can never be justified in an ultimate sense and the investigation into what is true and real is an on-going process. As a result, the critical rationalist approach tends to be associated with quantitative research and according to Gratton and Jones (2004), focuses on the derivation of predictions of behaviour based on existing theory, which can be tested through the collection and examination of data. As stated above this study adopted a cautious realist ontology and the epistemology of falsificationism, meaning the current research fitted within the critical rationalist paradigm.
### 4.3 Research Design and Strategy

This section will focus on the next phase of the social inquiry process which involves the examination of several research strategies in an attempt to discover and then justify the most suitable research strategy that will allow the research question presented in this study to be answered. The research strategies to be explored are the Inductive, Deductive, Retroductive and Abductive strategies.

According to Walliman (2006), inductive reasoning starts from a specific observation and derives a general conclusion from it. Inductivists will make several observations, collect much data and then produce generalisations, which they hope can describe patterns of life (Blaikie 2000). As these observations are made, concepts will become more refined, empirical generalisations will be made and preliminary relationships will be identified, so in a sense induction involves building theory from the ground up (Neuman 2007) and is often an empiricist’s adopted strategy.

Walliman (2006) states that three main conditions must be satisfied for a generalisation to be accepted by an inductive. First, there must be a large number of observation statements. Secondly, the observations must be repeated under a large range of circumstances and conditions. Finally no observation statement must contradict the derived generalisation. There remains criticism of inductive reasoning such as, the ambiguity surrounding how large the number of observations have to be and also the uncertainty of how large a range of circumstances and conditions they must be repeated under in order that true conclusions can be reached. Blaikie (2007) reveals other limits of this strategy by claiming the induction strategy is particularly effective at answering ‘what’ questions but limited when it comes to ‘why’ questions.

Deductive reasoning is more of a rationalist’s approach and begins with general statements, which through logical arguments deduce specific conclusions (Walliman 2006). Whereas, inductivists look for evidence to confirm their generalisations, deductivists attempt to refute their hypothesis by falsifying it (Blaikie 1993). This research strategy therefore adopts a different starting point and works in the reverse order to the inductivist strategy detailed above (Blaikie 2007).

In the deductivist approach research is guided by its preceding theory. These theories therefore, are considered to be speculative answers to perceived problems and are
tested by observation and experiment (Walliman 2006). If the observations match the
theory, it becomes possible to confirm the possible truth of the theory (particularly if
further tests produce similar results). However theory can also be falsified and totally
rejected by making observations which are inconsistent with its statements (Walliman
2006). In this case, a theory will either need to be modified or rejected outright
allowing the testing of other candidate theories (Blaikie 2007). In this sense, science
is seen as trial and error, when one theory is rejected, another is proposed and tested
and only the fittest theory survives (Walliman 2006). Rigorous tests must therefore be
applied in order to distinguish the strong theories from their weaker counterparts.

In order to make the method of selection by elimination work, and to
ensure that only the fittest theories survive, their struggles for life must be made severe for them (Popper 1957, p134).

In order for a theory to be tested, it must first be expressed as a hypothesis. The nature
of a hypothesis means it must be falsifiable. This therefore means that it is logically
possible to make true observational statements that conflict with the hypothesis and
thus can falsify it. According to Blaikie (1993) a good theory is falsifiable because it
makes a definite claim about the world and the more falsifiable the theory is the better
it is. This act of falsifying a theory however, will inevitably results in the researcher
having to start from scratch (Walliman 2006). This strategy also has its limits and is
considered to be more appropriate for ‘why’ questions as opposed to ‘what’ questions
(Blaikie 2007).

Alleged deficiencies in the two research strategies above have led to the formation of
two alternate strategies known as retroduction and abduction. Retroduction has been
advocated for use in both the natural and social sciences, whereas abduction, a less
common strategy is considered the appropriate method of theory construction in
interpretive social science (Blaikie 1993).

“Retroduction is a process of working back from data, to an explanation, by the use of
creative imagination and analogy” (Blaikie 2007, p9). Retroduction begins with an
observed regularity, which is explained by locating the underlying structure or
mechanism that is responsible for producing the observed regularity. In order for a
researcher to discover the previously unknown underlying structure or mechanism
they must construct a hypothetical model of it, and then attempt to establish its
existence through observation and experiment (Blaikie 2007). This strategy provides an alternate method to answering ‘why’ questions (Blaikie 2007).

Abduction is fairly distinguished from the other three strategies and is used in interpretivism and approaches that use interpretive ontological and epistemological elements (Blaikie 1993). The aim here is to describe and understand social life from a social actors perspective, in terms of their motives and understanding and repeatedly test a developed theory (Blaikie 2007). This strategy is also different from the others in the sense it can be used to answer both ‘what’ and ‘why’ questions (Blaikie 2007).

The research strategy adopted for this study will be the deductive approach. This current study uses pre-existing economic theory that sports teams are jointly responsible for production as a framework to develop the hypotheses, which can then be tested via quantitative methods. The deductive approach is appropriately suited to studies of a quantitative nature and allows for an investigation into team, positional unit and individual player performances whilst in direct competition with opponents. The quantitative method utilised in the subsequent empirical chapters of this thesis are detailed in the following sections.

4.3.1 Regression Analysis

Regression analysis is a method of analysis that is fundamentally based on the concept of correlation. There are considered to be two main uses for this type of analysis. The first is to test hypotheses and examine the relationship among two or more variables (Schroeder et al 1986). The other main use of regression analysis is to predict a change in a dependent variable in response to changes in one or several independent variables whose values are known (Hair et al 1995). There are two main types of regression analysis depending upon the number of predictor variables (explanatory variables) there are present. When only one explanatory variable is involved in the analysis the regression method is known as ‘simple regression’. When the problem involves the use of two or more independent variables the method is known as ‘multiple regression’ (Hair et al, 1995).

A classic linear regression model is one that allows a researcher to examine the nature and form of relationships among two or more variables (Asteriou and Hall 2011). In a simple regression model a researcher first has to determine which variable is affecting
the other. In other words one variable will be dependent on the other and hence, the
two variables are known more commonly as the dependent variable (usually denoted
by Y) and the independent variable (usually denoted by X).

The equation 4.1 below illustrates the simple regression relationship described above:

\[ Y_i = \alpha + \beta X_i + u_i \] (4.1)

Where:

\( i \) = An index of observation

\( Y \) = the dependent variable

\( \alpha \) = Constant term/intercept, (Value of Y when X = 0)

\( \beta \) = Slope/gradient/regression coefficient (unit change in Y value for every 1 unit
change in X value)

\( X \) = Independent Variable/predictor = variable used to predict Y (must be correlated
with Y)

\( u \) = Error term (represents other underlying factors that influence Y values)

As stated, regression analysis aims to summarise the relationship between the
independent (X) and dependent (Y) variables. This summary is achieved by
estimating the best straight line (linear) approximation of the relationship between X
and Y (Schroeder et al 1986). In order to obtain the line of best fit, a means of
summing or averaging the distances between the actual and estimated values (from
the regression line) is needed and this is generally referred to as the sum of squared
values. The regression line, which is then chosen, is commonly obtained via a process
known as ordinary least squares regression (OLS), which simply chooses the
regression line that minimizes the sum of squared deviations of observed values from
those estimated by the regression line. These residual errors are estimates of the
random disturbance term (Schroeder et al 1986). The descriptive statistic that
measures the degree of linear association between the regression variables is called \( R^2 \)
and this statistic measures closeness as a percentage of total variation in the dependent
variable explained by the regression line (Schroeder et al 1986).

In economic research the use of a multiple regression analysis is deemed more
appropriate than the simple regression procedure described above because other
factors can affect the dependent variable. The principles of multiple regression remain the same, however the multiple regression model differs in the sense that it measures the effects of several independent variables concurrently. In cases of multiple regression where there are two or more explanatory variables, the estimated coefficient on any independent variable, estimates the effect of that variable whilst holding the other independent variable(s) constant (Schroeder, et al 1986).

In multiple regression analysis, $R^2$ (the coefficient of multiple determination) measures the percentage of variation in the dependent variable, which is explained by variations in the independent variables taken together (Schroeder et al 1986). The calculation of $R^2$ is presented in equation 4.2 below:

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

(4.2)

In equation 4.2 above, SST represents the total sum of squares, SSE represents the explained sum of squares and SSR represents the residual sum of squares (Wooldridge 2009). Although the $R^2$ value is very useful it is not always the sole determinant of the ‘goodness’ of a multiple regression model. This is because the addition of independent variables in a multiple regression process will never decrease the value of $R^2$, even though the additional variable itself may have no bearing on the variations in the dependent variable (Schroeder et al 1986). It is for this reason that some analysts prefer the use of what is known as the adjusted $R^2$ value to express the explanatory power of the regression. The adjusted $R^2$ value is “a goodness of fit measure that penalizes additional explanatory variables by using a degrees of freedom adjustment in estimating the error variance” (Wooldridge 2009, p835). Put another way this $R^2$ value is simply an adjusted value of $R^2$ that has accounted for the presence of additional independent variables used in the regression.

**4.3.2 Hypothesis Testing**

Hypothesis testing refers to a procedure that allows a researcher to draw inferences of a regression analysis of a sample to the entire population. The theory of hypothesis testing provides a method, which recognises that if an inferential statement is based on a sample, we as researchers can never be 100% certain of the validity of the inference about the population. This procedure therefore, allows for the probability
that an incorrect conclusion has been drawn and allows us to define the likelihood of making such an incorrect inference (Schroeder et al 1986).

The hypothesis testing procedure involves the formation of a null and alternate hypothesis of the $\beta$ parameter of interest in the regression equation. When a researcher can reject the null hypothesis (usually $\beta = 0$), in favour of the alternate, the regression coefficient is said to be significant (or significantly different from zero) at a stated probability (normally, 1, 5 or 10%). However before this conclusion can be reached the null and alternate hypothesis have to be examined through test values. Test values are constructed values that depend on two factors, firstly the estimated variability of the estimates of $\beta$ from sample to sample (standard error) and secondly a probability distribution.

The standard error of a regression coefficient is a measure of the amount of variability that would be present among different $\beta$ estimates obtained from different samples drawn from the same population (Schroeder et al 1986). As statistical theory allows us to make inferences from sample to population, it will allow us to estimate how much variability there would be amongst all these different estimates by taking just information from one sample. In other words statistical theory allows us to estimate the standard error for a population from information on just one sample.

The measure of standard error allows us to make inferences about how sensitive the estimates of $\beta$ are to a change in sample composition without taking another sample. Large standard errors are undesirable not only because they cast doubt on the estimate, but also because the magnitude of the test value depend positively on the size of the standard error (Schroeder et al 1986).

As stated above, throughout the hypothesis testing procedure, a researcher will also require the use of a probability distribution. One of these probability distributions is known as the t-distribution. This t-distribution allows a researcher to make statements regarding the probability of obtaining an estimate with a given degree of closeness to the assumed, null hypothesised, value of $\beta$ (Schroeder et al 1986). Ultimately, the t-distribution reveals the appropriate test statistic, which can then be used to test hypotheses (Wooldridge 2009).
The statistical inference made about the population parameter from its estimate depends on the size of the test value, which in turn depends on the standard error of the estimated coefficient and on the size of the appropriate t-statistic. Put another way, the larger the standard error is, the more difficult it becomes to reject the null hypothesis, which could then lead to completely different inferences being drawn about the population.

In testing, t-ratios are used and are calculated by dividing the estimated coefficient by the estimated standard error (Schroeder et al 1986). In this case the t-ratio simply needs to be compared with the t-statistic in order to determine whether or not to reject the null or alternate hypothesis. If the absolute value of the t-ratio exceeds the equivalent t-statistic, the null hypothesis is rejected at the specified level of significance.

In some cases of multiple regression analysis, a researcher may opt to test a hypothesis about all or a set of the regression coefficients simultaneously. This is common particularly in cases where the null hypothesis has not been rejected for individual coefficients, however a researcher still feels that the independent variables in unison will affect the dependent variable. In such cases, the test for the simultaneous equality of all regression coefficients equalling zero is done through the use of the F-statistic (Schroeder et al 1986). The null hypothesis in such cases will therefore be that the regression coefficients taken in unison are equal to zero ($\beta_1 = \beta_2 = 0$).

The F-statistic, similarly to the t-statistic relies on the sample and a probability distribution known as the F-distribution. Just as a t-ratio can be computed and compared to a t-statistic, an F-ratio can also be computed and compared to an F-statistic obtained from a distribution table. The F-ratio reflects the degree of explanatory power of all parameters in a multiple regression model (Wooldridge 2009).

As stated above, regression analysis particularly in economics can be used for the purposes of prediction. This process generally involves predicting the value of a dependent variable using known values of independent variables. However, according to Schroeder et al (1986), results obtained for the purposes of prediction from a regression analysis may not be particularly useful even if a number of independent
variables are revealed to be significantly related to a dependent variable. In cases where a regression analysis is conducted for the purpose of prediction, specific attention must be given to the $R^2$ and t-ratio statistics. A small $R^2$ value would indicate that only a small proportion of the total variability in the dependent variable is accounted for by the independent variables, suggesting there are numerous other unobserved or random factors affecting the size of the dependent variable, hence questioning the validity of the prediction made. Similarly a low t-ratio for regression coefficients implies considerable uncertainty about the true population regression coefficients, again therefore, questioning the credibility of the prediction made.

So far the last two subsections have outlined the regression procedure and the different uses of this type of analysis by researchers. The whole regression analysis however does consist of several assumptions, which if violated cause several difficulties. These assumptions are now detailed further in the subsection below.

### 4.3.3 Assumptions

There are several assumptions upon which a regression analysis is based and this section will highlight each of these assumptions as well as discuss what effect a violation of these assumptions can have upon results obtained. Several procedures for the detection of these violations and subsequent corrections are also detailed.

The most obvious assumption upon which a regression analysis is based is the one of linearity, which refers to the dependent variable being a linear function of a specific set of independent variables and an error term (Asteriou and Hall 2011). In other words, it is assumed that an equation of the correct functional form is being used and that all the variables have been carefully selected. There are also several assumptions made concerning the error term namely homoskedasticity, the independence of the error term and the normality of the error term (Hair et al 1995). Other assumptions made throughout the regression procedure are that the variables are measured accurately and the explanatory variables are independent of each other. The data must also constitute a random sample. These assumptions are now explained in turn in greater detail below.

With regards to model specification, certain assumptions are made regarding the omission of relevant variables, the inclusion of irrelevant variables and the functional
form of the model. Omitting relevant variables from the model would lead to the formation of ‘biased’ (therefore, incorrect) estimates of the regression coefficients. Similarly, by including variables which are not relevant, although would lead to unbiased coefficient estimates, would also lead to inflated standard errors of the relevant variables, particularly if there is correlation between a relevant and irrelevant variables, leading to incorrect conclusions being drawn from the analysis (Schroeder et al 1986).

Another assumption made with regards to the data is that all independent variables present in the model are not related to each other. This problem of correlation between independent variables is referred to as multicollinearity. While regression coefficients estimated using correlated independent variables are unbiased, the presence of multicollinearity ensures that the standard errors of these coefficients are larger than they would be in the absence of correlation between independent variables (Schroeder et al 1986). This increase in standard error will ultimately mean t-ratios will be smaller and therefore increase the likelihood that incorrect conclusions are made throughout the hypothesis testing procedure.

There are several ways multicollinearity can be detected and hence accounted for however they all tend to be difficult to interpret and are often prone to misuse (Wooldridge 2009). The simplest of these methods to account for the presence of multicollinearity is to increase the sample size. By increasing the sample size, standard errors can be reduced and the $R^2$ between two independent variables can be reduced therefore attacking the multicollinearity problem directly (Thomas 1997). In the absence of more data however, there are two other methods available to a researcher for detecting and dealing with multicollinearity. These methods involve the use of the variance inflation factor (VIF), or tolerance statistics.

The VIF is a measure of the inflation of the variance of regression coefficients caused by the presence of multicollinearity. In other words the VIF statistic will indicate if an independent variable is correlated to another. The interpretation of this VIF statistic is considered to be somewhat problematic as it is achieved via a scale of 0 through to 10+. The closer this VIF statistic is to zero the weaker the correlation between independent variables is assumed to be. If the VIF statistic should exceed the ‘cutoff’ of 10 or above however then the presence of multicollinearity is deemed to be
extreme. Wooldridge (2009, p99), expresses concern regarding the generally accepted ‘cutoff’ point of 10 and describes it as an ‘arbitrary’ value, whilst Schroeder, et al (1986) also state that the method of looking for high correlations between variables is far from foolproof.

The tolerance statistic for detecting multicollinearity is similar to the VIF statistic in the sense that both statistics reveal the extent to which an independent variable is explained by the other independent variables. The tolerance statistic therefore, reveals the amount of variability of the selected independent variable that is not explained by the other independent variables (Hair et al 1995). A very small tolerance statistic will represent high collinearity and the cutoff threshold for tolerance is 0.10 (which corresponds to VIF values above 10). Hair et al (1995) state that depending upon the degree of collinearity present a researcher has several options including simply dropping or substituting the highly correlated predictor variables from the analysis or using more complex regression procedures such as ridge regressions23.

Throughout the regression procedure several important assumptions are also made regarding the error term in the equation. The simplest of these assumptions is that the error terms are normally distributed (have a mean of zero). A second assumption regarding the error term is that there is no autocorrelation present, otherwise meaning that the error term for one particular observation is in no way correlated with error terms from other observations. The presence of autocorrelation will ultimately lead to underestimation of the standard error and inflated t-ratio statistics causing incorrect conclusions to be drawn from the hypothesis testing procedure (Schroeder et al 1986).

A common method for the detection of autocorrelation is the Durbin-Watson coefficient, which is valid when a constant term is present in the regression model and there are no lagged variables and serial correlation is assumed to be first order only (Asteriou and Hall 2011). When autocorrelation has been detected a technique called generalised least squares (GLS) is one commonly used to overcome the problem, which is based on OLS regression but uses transformed variables (Schroeder et al

---

23 According to Hair et al (1995), ridge regressions can be employed as a means to overcome the problems multicollinearity presents. In the presence of multicollinearity least squares estimates will be unbiased, but their variances will be large meaning they will not represent true value. By adding a degree of bias to the regression estimates, ridge regressions reduce the standard errors, ensuring estimates are more reliable.
Alternatively, robust estimation methods are also appropriate for fairly arbitrary forms of serial correlation and heteroskedasticity (Wooldridge 2009).

Yet another assumption regarding the error term is the assumption of homoskedastic errors. Several different methods are available to a researcher for the detection of heteroskedasticity such as, a series of Lagrange multiplier tests (LM tests) like the Breusch-Pagan, Harvey-Godfrey, Park tests or the more complex Goldfeld-Quandt test. Generalised least squares can also be used to circumvent the effects of heteroskedasticity (Asteriou and Hall 2011).

This assumption of homoskedasticity assumes that the variance in an error term is not related to any factor (observation) present in the analysis. When the variance of the error term changes for every observation this assumption is violated and heteroskedasticity is present (Asteriou and Hall 2011). Similarly to autocorrelation, violation of this assumption leads to inaccurate standard errors of regression coefficients, therefore biasing hypothesis test results (Schroeder et al 1986).

The presence of heteroskedasticity can be encountered when using dummy variables. On occasion, dummy variables may sometimes be employed in regression analyses, which unlike other continuous variables can only assume a limited number of values. These dummy variables can usually only assume the values 0 or 1 and are therefore referred to as dichotomous or binary variables and can be specified within an econometric model as an independent variable or a dependent variable. The use of dummy variables is deemed appropriate in situations where the theory implies that behaviour differs between different time periods, or between two groups within a cross section (Schroeder et al 1986).

In order to overcome the difficulties heteroskedasticity presents, robust estimation methods can be used which adjust standard errors so that they remain valid even in the presence of heteroskedasticity of unknown form (Wooldridge 2009). This means no matter what the level of heteroskedasticity present in the population, robust estimation will still report valid statistics. Although robust estimation is not equal to OLS under classical assumptions, it is less sensitive to the violation of these assumptions, in particular the assumption that requires the disturbances to be normally distributed (Thomas 1997).
Chapter two outlined the joint production element present in sports where sport was defined as a club good and it was demonstrated that the production of a fixture is dependent on the participation of two teams. Any output is therefore jointly determined by the combined efforts of the subject and their opponent. It follows that any analysis of production in team sports would really account for the possibility (that occurs elsewhere too) that unlike the assumptions of the discussion above, that the error terms are purely random variables, interdependence (correlation) is likely between error terms and independent variables (Asteriou and Hall 2011). This is a common problem when simultaneous phenomena are being investigated (Schroeder et al 1986).

In cases where two variables are jointly determined, the variables are referred to as endogenous. Those variables that are not determined by the specified model are referred to as exogenous variables. In cases of simultaneity a reduced form equation is formulated by specifying each endogenous variable in terms only of the exogenous variables, the parameters of the model and the stochastic error term (Asteriou and Hall 2011). In the presence of simultaneity normal OLS will lead to the formation of biased and inconsistent estimators (Thomas 1997), therefore a different approach is required. According to Thomas (1997), one of the best methods of overcoming problems of contemporaneous correlation, whether it results from simultaneity or measurement error is that of instrumental variable estimation.

Instrumental variable (IV) estimation involves the inclusion of a new variable that must satisfy certain properties. The two properties this IV must consist of are, that it must be uncorrelated with the error term, whilst at the same time correlated with the endogenous explanatory variables in the equation (Wooldridge 2009). As long as these two conditions are met an IV estimator will remain consistent, however it should be noted that IV estimators are not unbiased estimators and in cases where sample sizes are small the level of bias may be substantial (Wooldridge 2009).

As the correlation between an endogenous explanatory variable and the IV increases, the sampling variance of the corresponding coefficient decreases. It is therefore clear that the better the instrument used, the more precise the estimates will be. Conversely,
a weak correlation between the endogenous explanatory variable and the IV will lead to a sizeable bias in the estimator (Wooldridge 2009). Similarly if there is any correlation between the IV and the error term (which cannot be observed) and weak correlation between the endogenous explanatory variable and the IV, estimates obtained will be inconsistent. This provides further justification of empirically evaluating the correlation between the endogenous explanatory variable and the IV. The IV estimator will always have a larger asymptotic variance than an OLS estimator, but that is to be expected due to the addition of another variable (the IV) and hence another source of uncertainty.

Unlike ordinary OLS where the $R^2$ measure reflects the percentage of variation explained, this is no longer true in the context of IV estimation. In cases where an endogenous explanatory variable is correlated with the error term, the variation in $Y$ can no longer be decomposed into SSE (sum of squared errors) and SSR (sum of squared residuals) therefore $R^2$ has no natural interpretation (Wooldridge 2009).

In order to account for the presumed contemporaneous correlation between the performance of an observed team and the performance of their opposition the econometric model chosen to conduct this study was the Two Stage Least Squares (2SLS) estimation. A 2SLS estimation technique is used commonly to obtain estimates of parameters in a system of simultaneous relationships (Thomas 1997). The basic idea behind the 2SLS estimation method is to replace a stochastic endogenous regressor (one that suffers from simultaneity bias, measurement error or omitted variable bias), with a non-stochastic variable which is a variable that is independent of the error term (Asteriou and Hall 2011). This procedure is conducted in two stages (hence the name **two stage least squares**) and is a technique highly regarded for its simplicity, ease of use and desirable statistical properties (Schroeder et al 1986).

A basic 2SLS model, similar to those that will be used in the subsequent empirical chapters of this thesis, is presented in equations 4.3 and 4.4 below:
\[ Y_{Bjt} = \alpha_1 + \alpha_2 PV_1 + \alpha_3 PV_2 + \ldots + \alpha_n PV_z + \alpha_{n+1} CV_1 + \alpha_{n+2} CV_2 + \alpha_{n+3} CV_3 + \alpha_{n+4} Z_{Cjt} + \epsilon_{jt} \]  

(4.3)

\[ Z_{Cjt} = \alpha_1 + \alpha_2 PVO_1 + \alpha_3 PVO_2 + \ldots + \alpha_d PVO_z + \alpha_{k+1} IV_1 + \alpha_{k+2} IV_2 + \alpha_{k+3} IV_3 + \alpha_{k+4} IV_m + \epsilon_{jt} \]  

(4.4)

Where:

\( Y_{Bjt} \) = Dependent Variable B, from match j, in season t.

\( \alpha_{i (i \neq 1)} \) = Slope Coefficients

\( PV_1 \) = Performance Variable 1

\( PV_2 \) = Performance Variable 2

\( PV_z \) = Representative of all additional player performance variables

\( CV_1 \) = Control Variable 1

\( CV_2 \) = Control Variable 2

\( CV_3 \) = Control Variable 3

\( Z_{Cjt} \) = Performance of the opponent (C), in match j, at time t

\( PVO_1 \) = Opponent Performance Variable 1

\( PVO_2 \) = Opponent Performance Variable 2

\( PVO_z \) = Representative of all other opposing player performance variables

\( IV_1 \) = Instrumental Variable 1 (unique to opponent)

\( IV_2 \) = Instrumental Variable 2 (unique to opponent)

\( IV_3 \) = Instrumental Variable 3 (unique to opponent)

\( IV_m \) = Representative of all additional instrumental variables (unique to opponent)
\[ \varepsilon_{jt} = \text{Error term} \]

Equation 4.3 presented above represents the structural equation (stage 2 regression) derived from the economic theory in sports production detailed earlier in chapter 2. Equation 4.4 above presents the reduced form equation (stage 1 regression), which measures the performance of the opponent. The reduced form equation (equation 4.4) is a linear equation expressing an endogenous variable as a function of exogenous variables and unobserved errors (Wooldridge 2009). The second stage of the 2SLS procedure involves using the fitted values obtained from the first stage as instruments for the endogenous regressor(s) in the original structural equation, the equation derived from economic theory and reasoning (Asteriou and Hall 2011).

The 2SLS procedure can be used in complex regression models where there are many endogenous variables and any number of instruments (Wooldridge 2009). There are however two certain conditions that must be satisfied in order for the 2SLS procedure to be accurate. The first condition is known as the order condition. The order condition requires that there are at least an equal number of excluded exogenous variables, as there are included endogenous explanatory variables in the structural equation (Wooldridge 2009). If this condition is not satisfied then the equation will not be identified, otherwise meaning that it will not be possible to solve the equation for a unique solution in terms of the instrumental variables estimator (Wooldridge 2009). This is also required to produce tests of the validity of the instruments. If there are fewer instruments than required, the equation is deemed under-identified and there is no econometric technique that can solve this problem. If the equation is satisfied by equality (equal number of instruments and endogenous variables) then a standard IV estimator can yield a solution. In cases where there are more instruments than endogenous variables the order condition is satisfied with inequality (deemed over-identification) and the 2SLS technique must be utilised in such cases in order to derive unique estimates.

The order condition is necessary, but not sufficient. In other words, once this condition has been satisfied another condition known as the rank condition must also be satisfied in order to conclude the identification procedure (Asteriou and Hall 2011). The rank condition for the identification of a structural equation is satisfied if, and only if, the reduced form equation contains at least as many exogenous variables
(as there are endogenous explanatory variables in the structural equation) with non-
zero coefficients that are excluded from the structural equation (Wooldridge 2009).
Indirect evidence on the suitability of the instruments used can be obtained in over-
identified equations via the use of a $\chi^2$ distribution test discussed in greater detail
below.

4.3.5. Testing for Endogeneity

When the explanatory variables within a structural equation are exogenous (rather
than endogenous), the 2SLS estimator can have very large standard errors meaning
2SLS is less efficient than the standard OLS estimator (Wooldridge 2009). In such
cases, tests can be conducted that determine whether 2SLS estimation is even
necessary and these tests are generally referred to as tests of endogeneity. In order to
determine which estimation procedure is more suitable Hausman (1978) suggested
directly comparing OLS estimates with 2SLS estimates and determining whether the
differences are statistically significant. If 2SLS and OLS differ significantly, then the
dependent variable would be considered endogenous (Wooldridge 2009) meaning the
2SLS procedure should be used for a more accurate estimation.

The STATA 12.0 software used to conduct all statistical analyses throughout this
study offers the Durbin $\chi^2$ and Wu-Hausman tests of endogeneity and these work in a
similar manner by estimating the model via both OLS and 2SLS and comparing the
resulting coefficient vectors (Baum et al 2003). The difference between the Durbin
and Wu-Hausman tests of endogeneity is that the Durbin test uses an estimate of the
error term’s variance based on the model assuming the variables being tested are
exogenous, while the Wu-Hausman test uses an estimate of the error variance based
on the model assuming the variables being tested are endogenous.

In the case of a robust estimation, Wooldridge’s (1995) robust score test and a robust
regression based test are reported. Both of these tests can tolerate autocorrelated and
heteroskedastic errors. The null hypothesis for all four of these tests is that the
variable is exogenous, so a significant test statistic would lead to the rejection of the
notion of exogeneity (meaning the variable being tested should be treated as
endogenous) therefore indicating 2SLS IV estimation is more appropriate than OLS.
4.3.6 Testing Overidentification Restrictions

As mentioned earlier in this chapter an instrumental variable must satisfy two specific conditions: it must be uncorrelated with the error term and it must be correlated with the endogenous explanatory variable (Wooldridge 2009). Even in models with additional explanatory variables the second requirement can be tested using a t test, in cases where there is only one instrumental variable or an F test, in cases where there are multiple instruments as in the reduced form equation 4.4 presented above. In the case of simple IV estimation, the prior requirement of the IV being uncorrelated to the error term cannot be tested. However in cases where there are more instruments than required, this requirement can be tested for some of the Instrumental variables (Wooldridge 2009).

 Procedures capable of testing this initial requirement involve comparing different IV estimates of the same parameter and these tests are more generally referred to as tests of overidentifying restrictions (Wooldridge 2009). The general idea of these tests is to use the additional instruments (ones that overidentify the model) to consistently estimate the parameters. Summing the number of instruments and then deducting the number of endogenous variables from the total number of instruments calculates the number of overidentifying restrictions. For example, if a structural equation contained only one endogenous variable and the reduced form equation consisted of three instrumental variables, the number of overidentifying restrictions would be 2 (3-1). In cases, where the number of overidentifying restrictions is two or more comparing several IV estimates can become difficult so test statistics can be more easily computed based on 2SLS residuals (Wooldridge 2009). If the model is exactly identified, tests of overidentifying restrictions cannot be performed.

STATA 12.0 software used to conduct all analyses throughout this study offers the default Sargan $\chi^2$ (1958) and Basmann $\chi^2$ (1960) tests of overidentifying restrictions, which both assume that the error term is independent and identically distributed. If the errors are not independent and identically distributed then these tests are no longer valid and the model must be refitted to estimate heteroskedasticity-robust standard errors. In the case of robust estimation, Wooldridge’s (1995) robust score test is reported.
Tests of overidentifying restrictions such as these detailed above actually test two different things simultaneously. Firstly, these tests of overidentifying restrictions determine whether the instruments are uncorrelated with the error term and secondly, these tests also determine whether the equation has been mis-specified and whether or not one or more of the excluded exogenous variables (instruments) should be included in the structural equation. Obtaining significant test statistics therefore would indicate either an invalid instrument or an incorrectly specified equation.

4.3.7. Testing Explanatory Power of Instruments

As stated above, for an instrument to be valid it must satisfy two requirements. Firstly, it must be sufficiently correlated with the endogenous regressors whilst also being uncorrelated with the error term. It is important that selected instrumental variables are not weakly correlated to the endogenous regressors as if this is the case, 2SLS estimators will be biased towards OLS estimators and any inferences made from the obtained results can be severely misleading. As Hahn and Hausman (2003) state, this problem of weak instruments cannot be corrected by simply adding more excluded exogenous variables (instruments) as the biases of instrumental variables estimators increase with the number of instruments.

STATA 12.0 offers several test statistics able to determine the strength or weakness of instrumental variables such as, $R^2$, adjusted $R^2$, partial $R^2$, F statistic and the minimum eigenvalue statistic. $R^2$ statistics are available which fit the first stage regression using OLS and determine the proportion of the total sample variation in the dependent variable that is explained by the independent variables (Wooldridge 2009). An adjusted $R^2$ statistic is also available which is a goodness of fit measure that penalises additional explanatory variables by using a degrees of freedom adjustment in estimating the error variance (Wooldridge 2009). In 2SLS estimation however, Wooldridge (2009) states these statistics can occasionally be misleading. In cases when a dependent variable is strongly correlated to an included exogenous variable but only weakly correlated to additional instruments then these statistics would still be large even though a weak instrument problem is present. Consequently, Bound et al (1995) promote the use of the partial $R^2$ statistics as a better gauge of the explanatory power of instrumental variables as this filters out the effects of the strongly correlated exogenous variable.
As mentioned earlier, F statistics can test for the joint significance of the instrumental variables in a reduced form equation. If the associated p-value for the F statistic is revealed be of non-significance then the instrumental variables are deemed to have no significant explanatory power for the dependent variable after controlling for the effects of included exogenous variables. Hall et al (1996) state that simply having an F statistic that is significant at the typical 5% or 10% levels is not sufficient. In cases where there is only one endogenous regressor an F statistic should be significant and also exceed the value of 10 in order to make a reliable inference (Staiger and Stock 1997; Stock et al 2002).

The Cragg and Donald (1993) minimum eigenvalue is also useful for testing the weakness of instruments, particularly when used in conjunction with the Stock and Yogo (2005) tests for weak instruments. In cases where there is only one endogenous regressor, the minimum eigenvalue will be identical to the F statistic. Stock and Yogo (2005) discuss two characterisations of weak instruments. Firstly, weak instruments cause instrumental variables estimators to be biased and secondly, the hypothesis tests of parameters estimated by instrumental variables estimators may suffer from severe size distortions. The null hypothesis of each of these Stock and Yogo (2005) tests is that the instruments are weak, however prior to performing these tests, a choice has to be made between the largest relative bias of a 2SLS estimator willing to be tolerated or the largest rejection rate of a 5% Wald test willing to be tolerated. If the test statistic exceeds the critical value, the null is rejected (Stock and Yogo 2005).

The “2SLS relative bias” critical values reported by STATA represent the critical values for the test that the instruments are weak based on the bias of the 2SLS estimator relative to the bias of the OLS estimator. The “2SLS size of nominal 5% Wald test” critical values reported by STATA represent the critical values relating to Stock and Yogo’s (2005) second characterisation of weak instruments. Here, instruments are designated weak if a Wald test at the 5% level can have an actual rejection rate of no more than 10%, 15%, 20% or 25% (Stock and Yogo 2005).

In cases where there is more than one endogenous regressor, a Shea (1997) partial $R^2$ measure can be calculated that takes into account the inter-correlations among instruments. As the bias of instrumental variable estimators increases with the number of instruments used, Shea’s (1997) adjusted partial $R^2$ statistic is often utilised as this
adjusts the degrees of freedom for the number of instruments included. A “low” value for the Shea (1997) partial $R^2$ measure depends on the specifics of the model being fit and the data being used.


4.4 Data

The data used in this study was a secondary data set supplied by Opta statistics, one of the world’s leading sports statistics organisations. Opta provide detailed match statistics that quantify and qualify every touch of the ball by each player during active play, based on live and recorded analysis of every match (Carmichael et al 2000). This recorded information is fed into a custom designed database from which a comprehensive and diverse assortment of information can be extracted for each match. Opta provide such data on a wide variety of sports commercially to various clients in the gaming industry, broadcasters, news media, clubs, club sponsors and governing bodies.

As mentioned in the introduction in chapter 1, much research that explicitly relates sports team output to inputs has been conducted primarily on American sports, particularly in the sport of baseball. What makes baseball an ideal context for the estimation of sporting production functions is the readily available wealth of detailed and specific statistics that enable individual contributions to be more easily identified and categorised (Carmichael et al 2000). This data richness however does not characterise many other team sports. According to Gerrard (2007) the two main difficulties with obtaining detailed performance data in football are down to confidentiality and expense. Gerrard (2007) states detailed performance data is usually only provided to teams (and the other users stated above) on a commercial basis, hence rendering it either unavailable or prohibitively expensive for academic researchers.

In 2002, Opta made four seasons worth of Premier League data ranging from the 1998-1999 season to the 2001-2002 season available in published football yearbooks. Each yearbook contains season total statistics for 25 player statistics for every outfield player and 14 player statistics for goalkeepers (Gerrard 2007). Previous studies on production in football (Carmichael et al 2000; 2001) have been based upon this publicly available data set.

The investigations conducted in this study are also based upon four seasons worth of data provided by Opta, however this data set in contrast to the four yearbooks publicly released has been privately obtained. The privately obtained data set used in this study
also differs from the data described above as this data set is unique to those professional players’ that have performed in the highest level of football in Germany, the Bundesliga (1). In addition, this data set is much more contemporary, ranging from the 2007-2008 season to the 2010-2011 season, is match specific (game level) and more importantly, much more detailed. Whereas the publicly released Opta yearbooks only consisted of data on 25 player statistics for outfield players and 14 player statistics on goalkeepers, the data set used in this study contains data on 200 player statistics for outfield players and over 70 player statistics for goalkeepers.

Crucially, the data set utilised in this study discloses important information on team formations and specific player positions within these formations. The data therefore reveals exactly which team formations were used as well as which position each player performed in within this formation for each team, in each fixture played throughout all four seasons. This positional data provides two main advantages over previous studies that this study exploits. Firstly, this team formation and player positional data for each individual player reveals exactly where each player performed on the field of play making it possible to isolate the contribution of any given set of players to a match. Secondly, this data on team formations and player positions within the formation allows the identification and hence measurement of opposing players’ performances. Ultimately, these two advantages combine in order to enable the estimation of players’ contributions whilst having accounted for the performance of identified opponents.
4.5 Identification of Key Performance Indicators

As discussed previously in chapter 1, the fluid nature of football can make it difficult to determine how much a player has contributed to the match outcome. Factors such as player skill, player ability, team morale, ability of the coach, refereeing performance and luck can all play a part (Carmichael and Thomas 2008). The previous literature has revealed some common performance attributes to be of particular importance to specific playing positions. This sub section provides a comprehensive review of this literature in order to identify the most important performance attributes for each positional unit of players. The identification of these key performance indicators (KPI’s) filter out those player statistics of most importance and therefore those that merit inclusion as performance measures in the subsequent empirical analyses.

There are a number of performance measures that at a collective team level (and therefore not position specific) appear to significantly impact match results. For example, Carmichael et al (2001) reveal the importance of passing accuracy in achieving positive results. Positive match outcomes are also positively related to attributes such as accurate shooting and successful tackling and blocking. Torgler (2004) also reveals a strong positive correlation between the number of shots on target and the probability of winning a match. Match outcomes have been revealed to be negatively correlated to actions of indiscipline and pressure (Carmichael et al 2000). Torgler (2004) in an analysis on team performances in the FIFA World Cup stated red cards received in games led to a decrease in the chances of winning a match by 29.6%.

In their analysis of performance at EURO 2004 Carmichael and Thomas (2008) revealed that the eventual winners Greece performed unexceptionally in an attacking sense by recording low averages for shots on target and total number of shots, however they did record a healthy shot to goal conversion highlighting their efficiency in front of goal. The author’s model also showed that Greece had performed very well defensively recording high averages for tackles made. These results ultimately suggested that Greece’s success was attributed to organised and
disciplined defending and potency in front of goal when rare opportunities presented themselves.

The literature on production functions in European football has identified a recurring pattern of certain key attributes that appear to have a large bearing on the match result, namely these attributes relate to accuracy such as passing and shooting accuracy and statistics concerned with maintaining and regaining possession, such as tackling and interceptions.

Many further studies have highlighted statistics relating to possession of the ball and general build up play, prior to a goal being scored as being the most important performance variables (Hook and Hughes 2001; Lago- Penas et al 2010). Hughes and Franks (2005) have emphasised the importance of passing sequences prior to goals being scored. Other key performance indicators that have been identified include statistics such as pass completion, corners, on and off target crosses, shots on and off target, dribbles, assists, fouls (for and against), yellow and red cards, offsides (committed and received) and length of passes (Lago and Martin 2007; Lago- Penas et al 2010).

Physical attributes that could affect productivity have also previously been identified by Bloomfield et al (2005) who identified significant differences in age, body mass and the body mass index between players in different positions. These results suggest that players of particular shapes and sizes would perhaps be more suited to specific positions depending upon the level of physical demand that position requires.

The performance measures described so far are general measures used to analyse team performances rather than individual player performances. As this study analyses performance of players playing in specific positions, performance measures more specific to these positions had to be identified. KPI’s for defenders consist mainly of statistics relating to the actions of preventing shots and regaining the ball such as intercepting, blocking, marking, tackling, positioning, passing, clearances, yellow and red cards, fouls committed and received and ground and air duels, tackles won/lost and headers won/lost (Lago and Martin 2007; Dellal et al 2011; McHale et al 2012; Redwood Brown et al 2012). The most common performance attribute the previous research revealed in relation to defensive performance was the action of tackling. Blocking, interceptions and clearances were also recognised as important to
successful defensive performance and therefore these four KPI’s were included in the empirical analyses conducted throughout this study.

The performance of midfielders appears to be related to actions such as regaining the ball and redirecting it to the forwards in order to create opportunities for striker and themselves. Much of the previous literature identifies KPI’s such as assists, goals scored, shot accuracy, passes to opposition, number of successful/unsuccessful passes, successful/unsuccessful dribbles, final third entries (forward runs) and penalty area entries (Lago- Penas et al 2010; Dellal et al 2011; Redwood Brown et al 2012; McHale et al 2012).

The KPI’s relating to the performance of attacking players or strikers were fairly easy to determine. The vast majority of literature concur that the main KPI for attacking players is the number of goals scored (Hughes and Franks 2005; Lago and Rafael 2006; Lago- Penas et al 2010; McHale et al 2012). Other statistics of importance in relation to attacking performance are those associated with power and accuracy. The number of goals scored in and out of the penalty area with feet, the number of goals scored in and out of the penalty area with the head, goals scored from free kicks or penalties, successful and unsuccessful free kicks, shots off target in and out of the area, assists, offsides, total shots and percentage of shots on target have all been used as previous measures of attacking performance.

Ultimately, the four KPI’s selected for the measurement of the attacking players’ performance consisted of the most recurrent KPI’s for attacking players as identified by the previous literature. The number of goals scored and number of shots on target has been revealed to be of great importance with regards to an attackers performance (Hughes and Franks 2005; Lago and Martin 2007; Lago- Penas et al 2010; McHale et al 2012). In addition, passing and chance creation has also been revealed to be an important part of the attacking side of football (Hughes and Franks 2005; Lago and Martin 2007; Lago- Penas et al 2010; McHale et al 2012) therefore the four KPI’s for attackers were, the number of goals scored, the number of shots on target, the number of assists and the number of successful passes.
4.6 Conclusion

This chapter began by presenting and discussing in detail the ontological and epistemological positions available to researchers before providing a justification for the critical rationalist approach ultimately adopted. A critical rationalist philosophical approach aligned with the ontology of cautious realist and the epistemology of falsificationism was concluded to satisfy a deductivist research strategy of testing existing theories against reality through a process of conjecture and refutation. The chapter then presented a thorough background on the analytical methods of regression analysis, multiple regression analysis and the selected two stage least squares (2SLS) regression analysis, which was used to test the formed hypotheses.

Section 4.4 described the unique data set used in this study and the advantages this data set provided, more specifically enabling the identification of subject players and their opposing players. The data set therefore made it possible to isolate the contribution of a set of players’ to a match. Consequently, this study set out to exploit this unique data set and explore the impact of the performance of the opponent(s) upon the performance of an observed team or specific set of players. Key performance indicators were identified to aid this investigation and the analyses of data is addressed and presented in the following chapters.

Chapter 5 below presents the first of three empirical chapters and investigates the impact of an opponent at a collective team level, aggregating the performances of all eleven starting players and up to three substitute players and assessing the impact of this collective team performance upon the performance of their opposing team and overall match result. Chapter 6 investigates the impact of the performance of an opponent at a more isolated individual player level by assessing the simultaneity between the performances of individual players in direct competition. Finally, chapter 7 investigates the impact of the performance of a positional unit of players (defenders and attackers) upon the performance of their opposing team as a collective.
Chapter Five - Investigating Interdependency Between the Performances of Two Teams in Direct Competition in the German Bundesliga

5.1 Introduction

This chapter investigates the simultaneity between the relative performances of opposing teams. Firstly, an investigation is conducted in order to determine the level of interdependency between the relative performance(s) of the away teams and the relative performance(s) of their opposing home teams. Another investigation is then conducted in order to determine the level of interdependency between the relative performance(s) of the home teams and the relative performance(s) of their opposing away teams. As stated by Leard and Doyle (2010), the method of splitting the data into home and away groups and estimating separate models for the home and away teams has been commonly utilised in previous studies that have analysed performance in sport.

Section 5.2 below provides a detailed description of the sample, including the data used, a detailed list of all observed teams and a detailed description of the chosen performance and instrumental variables along with a justification for their selection. Section 5.3 of this chapter offers a detailed outline of the analysis strategy and model specification. Section 5.4 reports the first set of results obtained when analysing the impact of the relative performances of away teams upon the relative performances of the home team. Section 5.5 reports the second set of results obtained when analysing the impact of the relative performances of home teams upon the relative performances of the away teams. Section 5.6 offers a discussion with regards to the empirical results obtained in all analyses conducted in this chapter.

Section 5.7 presents a conclusion, which provides a summary of those results of most importance and contains some generalised implications. It is concluded that team performances between an observed team and their opponent are not interdependent.
However, a team’s recent past performance and their general quality are revealed to have a significant impact upon their opponents’ performances. The significance of these team effects emphasise that the outcomes of a football match are essentially linked to team and collective influences that are not easily separated into particular individual efforts.
5.2 Organisation of Data and Variables

The dependent variables throughout the analyses conducted in this chapter were the result obtained by the home team, HTRESULT and the result obtained by the away team, ATRESULT, expressed as the number of points gained in the match. In the games of football covered in this analysis the reward for a win is 3 points, with 1 point for a draw and 0 points for a loss (Carmichael et al 2001). In previous studies on team performance in European football, team performance has often been measured using the match result as the output. For example, Carmichael et al (2000) measured team performance using match results expressed as goal differentials, with positive differentials representing wins, negative differentials representing losses and 0 differentials representing drawn games. Similarly, in a follow up study on team performances, Carmichael et al (2001) used aggregated points attained as a percentage of total points available (over the course of the season) as their output.

Table 5.2.1 below reveals a list of all Bundesliga teams that were included in the sample (team-game observations), which covered the seasons from 2007-08 to 2010-11. The established teams that regularly compete in European competition and battle for domestic honours such as FC Bayern Munich, Werder Bremen, Hamburger SV, VfL Wolfsburg, VfB Stuttgart, Bayer Leverkusen, Hannover 96, Eintracht Frankfurt and Borussia Dortmund all featured the maximum 136 times each. Each of these teams played in 68 home fixtures and 68 away fixtures over the four season sample period.

Other teams present for only three of the four seasons played in 51 home fixtures and 51 away fixtures and were therefore observed a total of 102 times and these teams included Hertha BSC, VfL Bochum, FC Nuremburg, Borussia Monchengladbach and TSG 1899 Hoffenheim. Some teams including Karlsruher SC, FC Energie Cottbus, Arminia Bielefeld, SC Freiburg and FSV Mainz 05 spent only two seasons in Bundesliga 1 and hence only amassed 34 home games and 34 away matches, totalling 68 team-game observations. Some teams only played in the Bundesliga for a solitary

24 Theoretically, one might argue that the dependent variable used throughout the analyses is of an ordinal nature, however throughout the analyses in this chapter (and again in chapter 7) linear estimators were adopted for two main reasons. Firstly, the use of linear estimators ensure that the interpretation of coefficients is much more straightforward (Downward et al 2014). Secondly and more importantly, estimators to control for endogeneity are much better developed and defined (Downward et al 2014). It is this latter issue that is of great significance for the analyses as a central feature of this study is the simultaneity between performances of direct opponents.
season throughout the four season sample period and played only 17 home games and 17 away games such as Hansa Rostock, MSV Duisburg and FC Kaiserslautern, totalling 34 observations each.

One of the fixtures, involving the teams FC St Pauli and FC Schalke 04 played on the 1st of April 2011 was excluded from the sample as data for this fixture was only partially complete as the match was abandoned part way through the game and the result was awarded to the away side who were leading by a score of 0-2 at the time of abandonment. As a result the eventual sample size consisted of a total of 1,223 match fixtures and 2,446 team-game observations.

<table>
<thead>
<tr>
<th>Team Represented in Final Sample</th>
<th>Number of Recorded Team-Game Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayern Munchen</td>
<td>136</td>
</tr>
<tr>
<td>Werder Bremen</td>
<td>136</td>
</tr>
<tr>
<td>FC Schalke 04</td>
<td>135*</td>
</tr>
<tr>
<td>Hamburger SV</td>
<td>136</td>
</tr>
<tr>
<td>VfL Wolfsburg</td>
<td>136</td>
</tr>
<tr>
<td>VfB Stuttgart</td>
<td>136</td>
</tr>
<tr>
<td>Bayer Leverkusen</td>
<td>136</td>
</tr>
<tr>
<td>Hannover 96</td>
<td>136</td>
</tr>
<tr>
<td>Eintracht Frankfurt</td>
<td>136</td>
</tr>
<tr>
<td>Hertha BSC</td>
<td>102</td>
</tr>
<tr>
<td>Karlsruher SC</td>
<td>68</td>
</tr>
<tr>
<td>VfL Bochum</td>
<td>102</td>
</tr>
<tr>
<td>Borussia Dortmund</td>
<td>136</td>
</tr>
<tr>
<td>FC Energie Cottbus</td>
<td>68</td>
</tr>
<tr>
<td>Arminia Bielefeld</td>
<td>68</td>
</tr>
<tr>
<td>FC Nuremberg</td>
<td>102</td>
</tr>
</tbody>
</table>
Table 5.2.2 below provides a list of all variables and their definitions. The dependent variables represent the output, the number of points gained in each match (match result). The endogenous variables represent the opponents’ player performance measures (aggregated team-match totals). The instrumental variables consist of variables unique to each team and help to identify the causal relationship between the dependent and endogenous variables. Finally, the control variables consist of variables used to control for team effects such as the relative quality of teams, the relative form of teams and other effects that impact the quality of football played within the league in-between seasons such as promotions, relegations, player transfers and managerial changes.
Table 5.2.2: List of all variables used throughout the analysis and their definitions

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
</tr>
<tr>
<td>HTRESULT</td>
<td>Home Team Match Points Acquired</td>
</tr>
<tr>
<td>ATRESULT</td>
<td>Away Team Match Points Acquired</td>
</tr>
<tr>
<td><strong>Endogenous/Performance Variables</strong></td>
<td></td>
</tr>
<tr>
<td>GLSAWAY</td>
<td>Number of Goals Scored by Away Team</td>
</tr>
<tr>
<td>SOTIGLSAWAY</td>
<td>Number of Shots on target (inc goals) by Away Team</td>
</tr>
<tr>
<td>TSPAWAY</td>
<td>Number of Total Successful Passes by Away Team</td>
</tr>
<tr>
<td>TWAWAY</td>
<td>Number of Tackles Won by the Away Team</td>
</tr>
<tr>
<td>SMAWAY</td>
<td>Number of Saves Made by the Away Team GK</td>
</tr>
<tr>
<td>GLSHOME</td>
<td>Number of Goals Scored by Home Team</td>
</tr>
<tr>
<td>SOTIGLSHOME</td>
<td>Number of Shots on target (inc goals) by Home Team</td>
</tr>
<tr>
<td>TSPHOME</td>
<td>Number of Total Successful Passes by Home Team</td>
</tr>
<tr>
<td>TWHOME</td>
<td>Number of Tackles Won by the Home Team</td>
</tr>
<tr>
<td>SMHHOME</td>
<td>Number of Saves Made by the Home Team GK</td>
</tr>
<tr>
<td><strong>Instrumental Variables</strong></td>
<td></td>
</tr>
<tr>
<td>AVHHTAWAY</td>
<td>Average Height of the Away Team in Metres</td>
</tr>
<tr>
<td>RFOOTAWAY</td>
<td>Number of Right Footed Players in the Away Team</td>
</tr>
<tr>
<td>LFOOTAWAY</td>
<td>Number of Left Footed Players in the</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>BFEETAWAY</td>
<td>Number of Two Footed Players in the Away Team</td>
</tr>
<tr>
<td>TPLAPPAWAY</td>
<td>Total Number of Previous Bundesliga Appearances by the Away Team</td>
</tr>
<tr>
<td>AVAGEAWAY</td>
<td>Average Age of the Away Team (Nearest Year)</td>
</tr>
<tr>
<td>AVHGHTHOME</td>
<td>Average Height of the Home Team in Metres</td>
</tr>
<tr>
<td>RFOOTHOME</td>
<td>Number of Right Footed Players in the Home Team</td>
</tr>
<tr>
<td>LFOOTHOME</td>
<td>Number of Left Footed Players in the Home Team</td>
</tr>
<tr>
<td>BFEETHOME</td>
<td>Number of Two Footed Players in the Home Team</td>
</tr>
<tr>
<td>TPLAPPHOME</td>
<td>Total Number of Previous Bundesliga Appearances by the Home Team</td>
</tr>
<tr>
<td>AVAGEHOME</td>
<td>Average Age of the Home Team (Nearest Year)</td>
</tr>
</tbody>
</table>

**Control Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEASON1</td>
<td>Season 1 dummy variable (2007-08)</td>
</tr>
<tr>
<td>SEASON2</td>
<td>Season 2 dummy variable (2008-09)</td>
</tr>
<tr>
<td>SEASON3</td>
<td>Season 3 dummy variable(^{25}) (2009-10)</td>
</tr>
<tr>
<td>GDHOME</td>
<td>Accumulated Goal Difference of Home Team (at time (t-1))</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>Accumulated Goal Difference of Away Team (at time (t-1))</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>% Points of the Home Team (at time (t-1))</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>% Points of the Away Team (at time (t-1))</td>
</tr>
</tbody>
</table>

---

\(^{25}\) Season 4 (2010-11) was used as the reference
5.3 Analysis Strategy

In order to investigate simultaneity between performances at a team level a pooled cross sectional data set was constructed using the exclusive data provided by Opta and supplementary data obtained from several editions of transfermarkt magazines. As detailed in chapter 4 earlier, OLS will result in biased and inconsistent estimators in cases of simultaneity. One of the best methods of overcoming this contemporaneous correlation is instrumental variable (IV) estimation (Thomas 1997). Consequently, a 2SLS instrumental variable estimator was employed for each of the analyses conducted in this chapter and this section details the model specifications.

As this study was based upon the sports economics theory of joint production it was assumed that match results obtained by both the home and away teams’ would be influenced by the performances of their own players as well as the performances of players on the opposing team. The purpose of this analysis was to measure how much impact the relative performances of opposing players had upon the match result achieved by the observed team. Initially, the impact of the away teams’ relative performances upon the home teams’ relative performances were investigated and the results obtained from this analysis are presented in sections 5.4 and 5.5. A further investigation was conducted in order to determine the impact of the home teams’ relative performances upon the relative performances of away teams and the results obtained from this analysis are also presented in sections 5.4 and 5.5.

5.3.1 Impact of away team performance upon home team performance

Equation 5.1 below presents the model used to conduct this analysis. The dependent variable $Y_1$ is the number of points acquired by the home team (HTRESULT) in match j at time t (game-week). The Z variables consist of explanatory variables that measure the performance of the home team and other exogenous independent variables that control for inter-season effects. $Y_2, Y_3, Y_4, Y_5$ and $Y_6$ represent the endogenous player performance variables that encapsulate the performance of the opposing team, which in this case is the away team.

As illustrated in equation 5.1 below the output in this production function is the home teams’ match results. The result achieved by the home team is presented as a function
of the home team’s performance and the away team’s performance. The away team’s performance will however not only affect the match result obtained by the home team (dependent variable) but will also itself be affected by the match result obtained by the home team (as the match result obtained by the home team encapsulates the performance of the home team in relation to the away team). As there is an assumed dual dependency between match result obtained by the home team and the performance of the away team an IV estimation strategy was necessary in order to overcome this simultaneity. The 2SLS model used to conduct this analysis is presented below.

\[ Y_{1jt} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \ldots + \beta_{n-1} Z_{n-1} + \beta_{n} Y_2 + \beta_{n+1} Y_3 + \beta_{n+2} Y_4 + \beta_{n+3} Y_5 + \beta_{n+4} Y_6 + u \]  

(5.1)

\[ Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \ldots \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \ldots + \pi_{n+6} IV_6 + \nu \]  

(5.2)

\[ Y_3 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \ldots \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \ldots + \pi_{n+6} IV_6 + \nu \]  

(5.3)

\[ Y_4 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \ldots \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \ldots + \pi_{n+6} IV_6 + \nu \]  

(5.4)

\[ Y_5 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \ldots \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \ldots + \pi_{n+6} IV_6 + \nu \]  

(5.5)

\[ Y_6 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \ldots \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \ldots + \pi_{n+6} IV_6 + \nu \]  

(5.6)

The explanatory (Z) variables in equation 5.1 include five team performance measures for the home team, goals scored by the home team (GLSHOME), shots on target including goals scored by the home team (SOTIGLSHOME), total successful passes by the home team (TSPHOME), tackles won by the home team (TWHOME) and saves made by the home team goalkeeper (SMHOME). In equations 5.2 to 5.6 the
endogenous variables $y_2, y_3, y_4, y_5$ and $y_6$ represent goals scored by the away team (GLSAWAY), shots on target including goals scored by the away team (SOTIGLSAWAY), total successful passes by the away team (TSPAWAY), tackles won by the away team (TWAWAY) and saves made by the away team goalkeeper (SMAWAY). All player performance variables were aggregated in order to provide team-match totals.

The goals scored variable has been shown to be of great match winning potential (Lucifora and Simmons 2003) and therefore positively affects match results (Carmichael and Thomas 2008). Goals scored are also deemed to be a key performance indicator for attacking players in much previous research on player performance (Hughes and Franks 2005; Lago and Martín 2007; Lago-Penas et al 2010; McHale et al 2012).

Previous research has also revealed match results are positively correlated to shots on target, meaning the more accurate attacking players are the greater the chances of winning the game (Carmichael et al 2001; Torgler 2004; Lago and Martín 2007; Lago-Penas et al 2010). Total successful passes made by a player measures their ability to retain possession of the ball for the team by accurately directing the ball towards a team member. Possession of the ball has been revealed in many studies as being a key determinant of team success and various aspects of the passing attribute have been extensively reviewed such as, passing accuracy, passing range, longevity of passing sequences and recovery of possession (Carmichael et al 2000; Hughes and Churchill 2005; Jones et al 2004; Lago-Penas et al 2010; Vogelbein et al 2014).

The variable for tackles won is a measure of a player’s ability to regain or win possession of the ball and has been included as a measure of performance in many previous studies (Carmichael et al 2000, 2001; Carmichael and Thomas 2005; Tiedemann et al 2011). Tackles won is also a measure commonly used in soccer performance rating systems such as the EA sports player performance index (McHale et al 2012).

Finally, saves made refer to the actions of goalkeepers, in particular the number of saves they make as a consequence of the opposition taking shots on target. The number of saves made is generally regarded as one of the most important key performance indicators for a goalkeeper (Oberstone 2010). Welsh (2004) claims the
goalkeeper is the most important and most pressurised player on the pitch as it is their sole duty to protect the goal and hence, prevent the opponents from scoring. Welsh (2004) also states that the performance of the goalkeeper is judged most commonly on only two criteria, the number of mistakes they make and the number of saves they make. The goalkeeper is the most specialised position (Frick 2007) and is the only position to be stipulated in the laws of the game (Welsh 2004).

All five performance variables detailed above were selected as the independent variables for this analysis, as opposed to other potential measures, due to the widespread importance given to these measures in previous research on performance analysis in football. The same five performance variables for the opposition (in this case the away team), represented the endogenous variables in this analysis.

Several variables were also included in the model presented in equation 5.1 above. Four of these control variables were included in order to capture team effects and more specifically control for relative team quality and relative team form. Previous research has revealed that momentum can impact performance within individual games as well as build up through a series of games and therefore lead to winning or losing streaks (Shmanske and Lowenthal 2007; Leard and Doyle 2010; Arkes and Martinez 2011). The variables PERPTSHOMETEAM and PERPTSAWAYTEAM were therefore included in the model in order to account for each teams relative form (and hence any momentum). These variables reflected relative league performance of the home and away teams, as the total number of points achieved expressed as a percentage of the maximum number of points possible (if all previous games were won). Both variables were lagged by one match-day and therefore represented the percentage points acquired by the home and away teams upon conclusion of their last match (and hence prior to the observed match). Percentage points measures have been used as measures of team success in previous research (Bairam et al 1990; Carmichael 2001).

Two additional control variables were also included that accounted for relative team quality. These variables were GDHOME and GDAWAY, which represented the accumulated goal differences of the home and away teams, upon conclusion of the last match (and hence prior to the observed match). These goal difference variables were therefore also lagged by one match-day and were calculated by subtracting the
number of goals conceded by the number of goals scored. Measures of this type have been previously used as proxies for measures of relative team quality (Carmichael et al 2000; Shmanske and Lowenthal 2007).

In addition, three dummy variables were included in order to control for the changes in the general quality of the German Bundesliga from season to season. As outlined earlier in chapter one, the German Bundesliga comprises of eighteen teams and at the end of every season a minimum of two teams are replaced (and possibly an additional third) due to the promotion and relegation legislation. The general quality of football played within the German Bundesliga would therefore differ from season to season due to different teams competing against each other.

Aside from promotions and relegations, other occurrences such as managerial changes and player transfers would also have an impact on the general quality of football played in the German Bundesliga. As outlined in chapter 3 earlier, the ability of the manager has been widely accepted to have a significant effect on the success and performance of the team as a whole and individual players within the team (Audas et al 1999; Dawson et al 2000; Salomo and Teichmann 2000; Dawson and Dobson 2002; Frick et al 2010). Throughout the four season sample period between season 2007-2008 and season 2010-2011 managerial turnover in the Bundesliga was regular, occurring on 48 occasions including off/pre-season and within season changes.

Another factor that can have an impact on the quality of football played in the German Bundesliga are the player transfers that occur at the start of, during or at the end of a season. Player turnover is likely to strengthen or weaken a squad and particularly the starting eleven. Carmichael et al (2000) argue the entire purpose of the production function in sport is to identify areas of weakness and therefore provide important information with regards to which players need to be shed and which players need to be acquired in order for a team to progress. Bearing this in mind it would then be reasonable to assume that teams losing key players or acquiring key players would have a significant impact upon both their own performance and the general quality of the league as a whole.

The six instrumental variables present in equations 5.2 to 5.6 were included in order to ensure the model was correctly specified (overidentified) and hence satisfied the rules of the 2SLS IV estimation procedure. These instrumental variables included, the
accumulated number of previous Bundesliga appearances made by the away team (TPLAPPAWAY), the number of right footed players in the away team (RFOOTAWAY), the number of left footed players in the away team (LFOOTAWAY), the number of players able to use both feet in the away team (BFEETAWAY)\textsuperscript{26}, the average player height of the away team (AVHGHHTAWAY) and the average player age of the away team (AVAGEAWAY).

For the two instruments referring to player height and player age, average measurements were used in this analysis. The reason for this was because throughout a football match, although only 11 players can start the game, up to three substitutions are permitted for either team. Although, 3 substitutions are allowed for both teams, their use is at the discretion of the manager of the team. Consequently, as a varying number of players may have represented each team, a minimum of 11 and a maximum of 14 throughout a match, averages were utilised rather than an aggregated sum.

Previous league experience is positively correlated to player salary (and hence their value) and has a strictly linear influence on annual income (Frick 2011). In addition, the most recent appearances (made in the last season) have been shown to matter more than experience gained in prior seasons (Frick 2011). The three dummy variables for a player’s preferred foot provided a measurement of whether a player was a right footed or a left footed player, or a player able to use both feet. Bryson et al (2009) state a player’s ability to play with both feet can have a direct impact on their performance, for example, the number of goals they score or passes they complete. Furthermore, Bryson et al (2009) revealed that after controlling for performance, two footed players playing in top leagues across Europe commanded a 15.4% premium, whereas two footed players playing in the Bundesliga received a 13.2% premium. Similarly, Frick (2007) also revealed that players with a preferred left foot command a 15% premium and players that are able to play with both feet receive more than a 50% premium, across all five big European leagues (Frick 2007).

\textsuperscript{26} The data set included information on the use of substitutions. This feature of the data was exploited in order to aid identification. Although a maximum of three substitutions are permitted in football, they are not obligatory (discretion of the team manager). The number of players that contributed to the match result therefore will have ranged between 11 players (the starting eleven) and up to 14 players (if all 3 substitutes were made during the game). Because the number of players involved in a fixture varied from game to game, multicollinearity was avoided in all of the empirical analyses conducted in this chapter.
A player’s age is a variable that also measures player experience (similar to the previous league experience variable), however the experience captured by the age variable was not restricted to experience gained in the German Bundesliga but experience gained in all levels of football throughout a player’s career. A player’s age therefore also takes into account, professional club experience gained abroad, under 21 international experience and senior international experience, all of which have been shown to be positively correlated to player salaries and transfer fees (Carmichael et al 1999; Lucifora and Simmons 2003; Frick 2011).

Finally, a measure of player height was also included in the model as an instrumental variable. Bloomfield et al (2005) states that physical player attributes can be reflective of the roles they are required to perform on the field and Frick (2007; 2011) reveals significant differences in player salaries depending upon which position the player plays in. Although the salary premiums between different positions could be explained by performance characteristics such as goals, assists etc, a player’s physical attributes could at least partially explain their efficiency at conducting these play actions.

5.3.2 Impact of home team performance upon away team performance

Equation 5.7 below presents the model used to investigate the impact of the home teams performance upon the performance of the away team they are competing against. The dependent variable $Y_7$, represents the number of points acquired by the away team in match $j$ at time $t$ (ATRESULT). The $Z$ variables represent the explanatory (aggregated player performance) variables that measure the performance of the away team and $Y_8, Y_9, Y_{10}, Y_{11}$ and $Y_{12}$ represent the endogenous variables that encapsulate the performance of the opposing team, which in this analysis is the home team.

$$Y_{7jt} = \beta_0 + \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \beta_4Z_4 + \ldots + \beta_nZ_n + \beta_{n+1}Y_8 + \beta_{n+2}Y_9 + \beta_{n+3}Y_{10} + \beta_{n+4}Y_{11} + \beta_{n+5}Y_{12} + u$$

(5.7)

$$Y_8 = \pi_0 + \pi_1Z_1 + \pi_2Z_2 + \pi_3Z_3 + \pi_nZ_n + \ldots + \pi_{n+1}IV_1 + \pi_{n+2}IV_2 + \ldots + \pi_{n+6}IV_6 + \upsilon$$
\[ y_9 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \cdots + \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \cdots + \pi_{n+6} IV_6 + \nu \]  

(5.8)

\[ y_{10} = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \cdots + \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \cdots + \pi_{n+6} IV_6 + \nu \]

(5.9)

\[ y_{11} = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \cdots + \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \cdots + \pi_{n+6} IV_6 + \nu \]

(5.10)

\[ y_{12} = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_n Z_n + \cdots + \pi_{n+1} IV_1 + \pi_{n+2} IV_2 + \cdots + \pi_{n+6} IV_6 + \nu \]

(5.11)

The explanatory variables in equation 5.7 include five team performance measures for the away team, goals scored by the away team (GLSAWAY), shots on target including goals scored by the away team (SOTIGLSAWAY), total successful passes by the away team (TSPAWAY), tackles won by the away team (TWAWAY) and saves made by the away team (SMAWAY). In equations 5.7 to 5.12 the endogenous variables \( Y_8, Y_9, Y_{10}, Y_{11} \) and \( Y_{12} \) represent goals scored by the home team (GLSHOME), shots on target including goals scored by the home team (SOTIGLSHOME), total successful passes by the home team (TSPHOME), tackles won by the home team (TWHOME) and saves made by the home team goalkeeper (SMHOME). All player performance variables were aggregated in order to provide team-match totals.

Six instrumental variables present in equations 5.7 to 5.12 were included in order to ensure the model was correctly specified. These instrumental variables included, the accumulated number of previous Bundesliga appearances made by the home team (TPLAPPHOME), the number of right footed players in the home team (RFOOTHOME), the number of left footed players in the home team (LFOOT HOME), the number of players able to use both feet in the home team (BFEETHOME), the average player height of the home team (AVHGHHTHOME) and the average player age of the home team (AVAGEHOME).
5.4 Results: First Stage Regression Results

Table 5.4.1 below presents the results obtained from the first stage reduced form regressions where the endogenous player performance variables for the away team represent the dependent variables. Table 5.4.2 reveals the first stage reduced form regression results obtained where the endogenous player performance variables for the home team represent the dependent variables.

As outlined in chapter 4, OLS inferences can be faulty in the presence of heteroskedasticity. Heteroskedasticity can often be encountered when using dummy variables (Schroeder et al. 1986) or large sample sizes (Wooldridge 2009). In order to overcome the difficulties heteroskedasticity presents, robust estimation methods can be used that will adjust standard errors so that they remain valid even in the presence of heteroskedasticity (and serial correlation) of unknown form (Wooldridge 2009). Therefore, no matter what the level of heteroskedasticity present in the population, robust estimation will still report valid statistics. As the models presented in the previous section consisted of multiple dummy variables robust standard error estimation occurred throughout the analyses in this chapter.\(^{27}\)

\(^{27}\) Robust standard error estimation was also conducted in subsequent empirical chapters for the same reasons.
Table 5.4.1 First Stage Regression Estimates Obtained When Assessing the Impact of the Performance of the Away team on the Performance of the Home Team

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>GLS Away Coef</th>
<th>GLS Away t</th>
<th>SOTGLS Away Coef</th>
<th>SOTGLS Away t</th>
<th>TSP Away Coef</th>
<th>TSP Away t</th>
<th>TW Away Coef</th>
<th>TW Away t</th>
<th>SMA Away Coef</th>
<th>SMA Away t</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEASON 1</td>
<td>-0.369942</td>
<td>-0.42+</td>
<td>-0.0785786</td>
<td>-0.87+</td>
<td>-12.03226</td>
<td>-2.19**</td>
<td>-0.702212</td>
<td>-1.60+</td>
<td>-0.014716</td>
<td>-0.52+</td>
</tr>
<tr>
<td>SEASON 2</td>
<td>0.045609</td>
<td>0.50+</td>
<td>0.691506</td>
<td>0.73+</td>
<td>-5.920878</td>
<td>-1.06+</td>
<td>-0.7202247</td>
<td>-1.80*</td>
<td>-0.0112918</td>
<td>-0.38+</td>
</tr>
<tr>
<td>SEASON 3</td>
<td>0.0403876</td>
<td>0.43+</td>
<td>0.456958</td>
<td>0.47+</td>
<td>27.068</td>
<td>4.77***</td>
<td>-0.9810897</td>
<td>-2.39**</td>
<td>-0.0096916</td>
<td>-0.35+</td>
</tr>
<tr>
<td>GD HOME</td>
<td>-0.0124741</td>
<td>-3.84***</td>
<td>-0.0126488</td>
<td>-3.77***</td>
<td>-0.5467938</td>
<td>-2.52**</td>
<td>0.0150873</td>
<td>0.95+</td>
<td>0.0013298</td>
<td>1.14+</td>
</tr>
<tr>
<td>GD AWAY</td>
<td>0.062474</td>
<td>1.53</td>
<td>0.062464</td>
<td>1.48+</td>
<td>0.8184924</td>
<td>3.22***</td>
<td>0.0226613</td>
<td>1.24+</td>
<td>0.0004197</td>
<td>0.34+</td>
</tr>
<tr>
<td>PERP SHOME TEAM</td>
<td>-0.019121</td>
<td>-4.62***</td>
<td>-0.0104677</td>
<td>-4.55***</td>
<td>0.1989182</td>
<td>1.27+</td>
<td>0.0149314</td>
<td>1.28+</td>
<td>-0.0002629</td>
<td>-0.36+</td>
</tr>
<tr>
<td>PERP TS AWAY TEAM</td>
<td>0.0123957</td>
<td>5.41***</td>
<td>0.0127129</td>
<td>5.11***</td>
<td>-0.2094751</td>
<td>-1.39+</td>
<td>-0.0036089</td>
<td>-0.29+</td>
<td>-0.0004756</td>
<td>-0.59+</td>
</tr>
<tr>
<td>GISH HOME</td>
<td>0.0547997</td>
<td>2.01**</td>
<td>0.0588396</td>
<td>2.10**</td>
<td>8.973572</td>
<td>5.36***</td>
<td>-0.3124986</td>
<td>-2.36**</td>
<td>-0.9526756</td>
<td>-63.00***</td>
</tr>
<tr>
<td>SOTGISHOME</td>
<td>-0.0061995</td>
<td>-0.49+</td>
<td>-0.0054913</td>
<td>-0.42+</td>
<td>-0.2371022</td>
<td>-2.61***</td>
<td>-0.181679</td>
<td>-2.81***</td>
<td>0.9613028</td>
<td>165.83***</td>
</tr>
<tr>
<td>TS PHOME</td>
<td>0.0015267</td>
<td>4.05***</td>
<td>0.0013929</td>
<td>3.35***</td>
<td>-0.3565049</td>
<td>-13.99***</td>
<td>0.0087518</td>
<td>4.54***</td>
<td>-0.000128</td>
<td>-0.97+</td>
</tr>
<tr>
<td>TW HOME</td>
<td>-0.0159827</td>
<td>-2.76***</td>
<td>-0.0149167</td>
<td>-2.51**</td>
<td>1.222347</td>
<td>3.36***</td>
<td>0.1879867</td>
<td>6.99***</td>
<td>0.0003524</td>
<td>0.19+</td>
</tr>
<tr>
<td>SM HOME</td>
<td>0.0299747</td>
<td>1.78*</td>
<td>1.034205</td>
<td>58.40***</td>
<td>1.972459</td>
<td>1.85**</td>
<td>-0.0360688</td>
<td>-0.48+</td>
<td>0.0033546</td>
<td>1.66+</td>
</tr>
<tr>
<td>AV GHTAWAY</td>
<td>0.522524</td>
<td>2.52**</td>
<td>0.9053947</td>
<td>2.64***</td>
<td>24.12378</td>
<td>0.90+</td>
<td>1.706212</td>
<td>1.11+</td>
<td>-0.030487</td>
<td>-0.38+</td>
</tr>
<tr>
<td>RFOOTAWAY</td>
<td>0.217119</td>
<td>3.78***</td>
<td>0.2203817</td>
<td>3.71***</td>
<td>-8.39104</td>
<td>-2.05***</td>
<td>0.172065</td>
<td>0.69+</td>
<td>0.0122688</td>
<td>0.59+</td>
</tr>
<tr>
<td>LFOOTAWAY</td>
<td>0.3997151</td>
<td>3.55***</td>
<td>0.2029737</td>
<td>3.40***</td>
<td>-9.045998</td>
<td>-2.17**</td>
<td>0.5345784</td>
<td>2.08**</td>
<td>-0.001406</td>
<td>-0.07+</td>
</tr>
<tr>
<td>BFEE TAWAY</td>
<td>0.394174</td>
<td>3.48***</td>
<td>0.2148448</td>
<td>3.55***</td>
<td>-7.517337</td>
<td>-1.81**</td>
<td>0.3037114</td>
<td>1.14+</td>
<td>0.0256304</td>
<td>1.17+</td>
</tr>
<tr>
<td>TPLAPP AWAY</td>
<td>0.0033117</td>
<td>2.75***</td>
<td>0.0034126</td>
<td>2.72***</td>
<td>0.5274881</td>
<td>6.70***</td>
<td>-0.0008068</td>
<td>-1.47+</td>
<td>-0.0023242</td>
<td>-0.62+</td>
</tr>
<tr>
<td>AVAGE AWAY</td>
<td>-0.0691089</td>
<td>-2.69***</td>
<td>-0.0747781</td>
<td>-2.80***</td>
<td>-3.82232</td>
<td>-2.12**</td>
<td>-0.96661</td>
<td>-0.81+</td>
<td>0.0041034</td>
<td>0.53+</td>
</tr>
</tbody>
</table>

Notes: *p<.10, **p<.05, ***p<.01
The R-Squared value obtained is a relatively low 0.1873 (18.73%) and the adjusted R-Squared 0.1751 (17.51%), however as explained in chapter 4 earlier both of these values have no natural interpretation in an IV estimation such as the 2SLS procedure (Wooldridge 2009). In addition, the recorded F statistic for each first stage regression revealed that all included variables were jointly significant.

As can be seen from table 5.4.1, the three control variables SEASON1, SEASON2 and SEASON 3 all appear to have no significant impact upon any of the two attacking performance variables that measure the number of goals scored and the number of shots on target. These three control variables however, do have an impact on the total number of successful passes and the number of tackles won. The variable SEASON1 is significant at the 5% significance level and negatively correlated with the number of successful passes made, indicating regular carelessness and inaccuracy amongst away teams in the German Bundesliga season 2007-08 (in comparison to the reference season 2010-11). SEASON3 is highly significant and positively correlated with the number of successful passes indicating passing accuracy was better in season 2009-2010 than the reference season 2010-11.

The SEASON2 dummy variable is negatively correlated to the number of tackles won by the away team at the 10% significance level, suggesting away teams in the season 2008-09 performed this specific defensive duty less capably than away teams in the 2010-2011 season. The SEASON3 dummy variable is also negatively signed and significant at the 5% level indicating tackling success deteriorated in the subsequent 2009-10 season (but was much improved in the reference season 2010-2011).

The control variables GDHOME and GDAWAY, which are proxies for relative team quality, also appear to have more significance for attacking aspects of football rather than the defensive aspects. The relative team quality of the home team is, as expected, negatively correlated to the number of goals scored by the away team, the number of shots on target by the away team and the number of successful passes made by the away team at the 1% and 5% significance levels.

GDAWAY, a measure of the relative team quality of the away team is positively correlated to the number of successful passes made by the away team at the 1% significance level. Relative team quality of the away side did not significantly impact goals scored or shots on target by the away side. Neither, GDHOME or GDAWAY
had any significant impact on the number of tackles won by the away team or the
number of saves made by the away team.

The remaining two control variables PERPTSHOMETEAM and
PERPTSAWAYTEAM, which are proxies for relative team form, are also revealed to
significantly impact attacking play rather than defensive play. The relative form of the
home team is negatively correlated to the number of goals scored by the away team
and the number of shots on target by the away team, both at the 1% significance level.
Conversely, the relative form of the away team is positively correlated to the number
of goals scored by the away team and the number of shots on target by the away side
also at the 1% significance level.

The player performance variables reported in table 5.4.1 reveal contrasting results.
The number of goals scored by the home team significantly impact all five player
performance variables of the away team, having a positive impact on attacking
aspects of play such as the number of goals scored, shots on target and total
successful passes, but having a negative impact upon defensive aspects of play such
as the number of tackles won and saves made. The level of significance for this
variable also varies with it being highly significant at the 1% level for the number of
successful passes and saves made by the away team, but only significant at the 5%
significance level for goals scored, shots on target and tackles won by the away team.

The variable SOTIGLSHOME which measures shots on target (including goals) was
revealed to have no significant impact on the attacking aspects of play by the away
side, such as their number of goals scored or their number of shots on target. The
number of shots on target by the home team did however impact the number of
successful passes made, the number of tackles won and the number of saves made all
at the 1% significance level. SOTIGLSHOME was as expected, positively correlated
to the number of saves made by the away goalkeeper and negatively correlated to the
number of tackles made by the away team and the number of successful passes made
by the away team.

Table 5.4.1 reveals the variable TSPHOME is correlated to all outfield performance
variables for the away team at the 1% significance level, but is of no significance to
the number of saves made by the away goalkeeper. This indicates the maintenance of
possession is more a duty for outfield players rather than the goalkeeper, which is to
be expected. The number of successful passes made by the home team has a positive impact upon the number of goals scored by the away team, the number of shots on target by the away team and the number of tackles won by the away team. TSPHOME is however as expected negatively correlated to the variable TSPA\text{WAY}.

The number of tackles won by the home team (T\text{WHOME}) is of significance for all outfield performance variables for the away team but has no impact on the number of saves made by the away side. As expected, the number of tackles won by the home side is negatively correlated to the number of goals scored by the away team and the number of shots on target by the away team. Conversely, the number of tackles won by the home team is positively correlated to the number of successful passes made by the away team and the number of tackles won by the away team. The relationship between the variable T\text{WHOME} and the three variables GL\text{SAWAY}, TSPA\text{WAY} and TW\text{AWAY} were all statistically significant at the 1\% significance level, however the positive correlation between T\text{WHOME} and SOTIGLSAWAY was revealed to be significant at a 5\% significance level only.

The final performance variable, SM\text{HOME}, which measures the number of saves made by the home team goalkeeper is as expected only statistically significant for attacking play variables from the away team. The number of saves made by the home team goalkeeper is positively correlated to the number of goals scored by the away team, the number of shots on target by the away team and the number of successful passes made by the away team, at the 10\%, 1\% and 5\% significance levels respectively.

Table 5.4.1 reveals that the instrumental variables unique to players from the away teams are of much greater significance with attacking play variables than the defensive play variables for the away team. The average height of the away team has a significant positive impact upon the attacking performance variables that measure the number of goals scored by the away team (GL\text{SAWAY}) and the number of shots on target by the away team (SOTIGLSAWAY).

A player’s preferred foot on the away team is shown to have a highly significant positive impact (1\% significance level) on the number of goals scored by the away team and the number of shots on target by the away team, however has a negative impact on the total number of successful passes made by the away team, indicating
players playing away often misplace passes. Out of the three instrumental variables measuring a player’s preferred foot only LFOOTAWAY was revealed to be of significance with regards to the number of tackles won by the away team. This result indicates that players on away teams had greater success in tackling with their left foot. A player’s preferred foot (on the away team) has no bearing on the number of saves made by the away team goalkeeper.

The accumulated number of previous Bundesliga appearances and the average age of the away team, which are measures of league and career experience are also revealed to be highly significant and positively impact the number of goals scored by the away team, the number of shots on target by the away team and the total number of successful passes made by the away team. Neither instrumental variable TPLAPPAWAY or AVAGEAWAY had any impact on the number of tackles made by the away team or the number of saves made by the away team goalkeeper.
<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>GLSHOME</th>
<th>SOTIGLHOME</th>
<th>Dependent Variable</th>
<th>TSPHOME</th>
<th>TWHOME</th>
<th>SMHOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coef</td>
<td>t</td>
<td>Coef</td>
<td>t</td>
<td>Coef</td>
<td>t</td>
<td>Coef</td>
</tr>
<tr>
<td>SEASON1</td>
<td>0.068582</td>
<td>0.63+</td>
<td>0.0868107</td>
<td>0.74+</td>
<td>-.6308717</td>
<td>1.08+</td>
</tr>
<tr>
<td>SEASON2</td>
<td>0.089318</td>
<td>0.85+</td>
<td>0.1068337</td>
<td>0.99+</td>
<td>-2.27982</td>
<td>-0.41+</td>
</tr>
<tr>
<td>SEASON3</td>
<td>-1.39</td>
<td>-1.39+</td>
<td>-1.11+</td>
<td>30.98824</td>
<td>5.47***</td>
<td>-0.503806</td>
</tr>
<tr>
<td>GDHOME</td>
<td>0.18021</td>
<td>3.75***</td>
<td>0.0176303</td>
<td>3.55***</td>
<td>.8300556</td>
<td>3.40***</td>
</tr>
<tr>
<td>GDAY</td>
<td>-0.0170661</td>
<td>-4.01***</td>
<td>-0.0177496</td>
<td>-4.01***</td>
<td>-.8343942</td>
<td>3.63***</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>0.010424</td>
<td>3.33***</td>
<td>0.0106137</td>
<td>3.43***</td>
<td>-0.572235</td>
<td>-0.36+</td>
</tr>
<tr>
<td>PERPTSAYTEAM</td>
<td>-0.0063157</td>
<td>-2.12**</td>
<td>-0.0060838</td>
<td>-2.01**</td>
<td>-1.915466</td>
<td>-1.17+</td>
</tr>
<tr>
<td>GLSAWAY</td>
<td>0.0713684</td>
<td>1.89+</td>
<td>0.0926079</td>
<td>2.41**</td>
<td>8.626421</td>
<td>4.30***</td>
</tr>
<tr>
<td>SOTIGLSAWAY</td>
<td>0.004174</td>
<td>0.08+</td>
<td>0.0106332</td>
<td>-0.58+</td>
<td>-4.603582</td>
<td>-4.97+</td>
</tr>
<tr>
<td>TSPAWAY</td>
<td>0.0022421</td>
<td>4.94***</td>
<td>0.0021315</td>
<td>4.55***</td>
<td>-3.607753</td>
<td>-14.38***</td>
</tr>
<tr>
<td>TWAWAY</td>
<td>-0.0254565</td>
<td>-3.82***</td>
<td>-0.0268523</td>
<td>-3.9***</td>
<td>1.683706</td>
<td>2.78***</td>
</tr>
<tr>
<td>SMAWAY</td>
<td>0.0040771</td>
<td>0.24+</td>
<td>1.04804</td>
<td>55.51***</td>
<td>3.159396</td>
<td>3.38***</td>
</tr>
<tr>
<td>AVHGHOMETEAM</td>
<td>2.744447</td>
<td>1.06+</td>
<td>2.943001</td>
<td>1.06+</td>
<td>122.936</td>
<td>0.88+</td>
</tr>
<tr>
<td>RFOOTHOMETEAM</td>
<td>0.246902</td>
<td>4.16***</td>
<td>0.2352748</td>
<td>3.86***</td>
<td>-3.556137</td>
<td>-1.03+</td>
</tr>
<tr>
<td>LFHOOTHOME</td>
<td>0.157454</td>
<td>2.34**</td>
<td>.11839</td>
<td>1.79+</td>
<td>-4.115717</td>
<td>-1.09+</td>
</tr>
<tr>
<td>BFETHOMETEAM</td>
<td>0.2686519</td>
<td>4.30***</td>
<td>0.2349222</td>
<td>3.36***</td>
<td>-1.922381</td>
<td>-0.53+</td>
</tr>
<tr>
<td>TPLAPPHOMETEAM</td>
<td>0.0039116</td>
<td>2.87***</td>
<td>0.0038925</td>
<td>2.78***</td>
<td>0.7573311</td>
<td>9.60***</td>
</tr>
<tr>
<td>AVAGEHOME</td>
<td>-0.0293805</td>
<td>-0.67+</td>
<td>-0.0144008</td>
<td>-0.45+</td>
<td>-8.345209</td>
<td>-5.06***</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
As can be seen from table 5.4.2 above the three dummy variables, SEASON1, SEASON2 and SEASON 3 had no significant impact upon the attacking performance variables. SEASON 3 was positively correlated to the total number of successful passes made by the home team and highly significant suggesting home teams maintained possession well throughout the 2009-2010 season (in comparison to the reference 2010-11 season). As these results are consistent with those presented for the variable TSPAways in table 5.4.1 it appears that turnover of possession in German top-level football between two performing teams is a common occurrence.

The dummy variable SEASON1 is revealed to have a positive impact on the number of saves made by the home goalkeeper and this result is significant at the 5% significance level. This suggests either, that home team goalkeepers performed better in the season 2007-08 than subsequent seasons in the sample period and or that home defenders performed worse in this particular season (as more shots were allowed).

The control variables GDHOME and GDAWAY both had a significant impact upon the attacking performance of the home team but were insignificant with regards to the defensive performance of the home team. As expected, the relative team quality of the home team is positively correlated to the number of goals scored by the home team, the number of shots on target by the home team and the number of successful passes completed by the home team at a 1% significance level. In contrast, but also as expected, the relative team quality of the away team is negatively correlated to the number of goals scored by the home team, the number of shots on target by the home team and the number of successful passes made by the home team, also at a 1% significance level.

PERPTSHOMETEAM and PERPTSAWAYTEAM are revealed to significantly impact attacking performance variables such as the number of goals scored by the home team and the number of shots on target by the home team but have no bearing on any other performance variables. As expected, the relative form of the home team has a highly significant (1% significance level) and positive impact upon both, the number of goals scored by the home team and the number of shots on target by the home team. Conversely, the relative form of the away team has a highly significant (1% significance level) and negative impact upon the number of goals scored by the home team and the number of shots on target by the home team.
The performance variable GLSAWAY has a significant impact upon all performance variables for the home team. The number of goals scored by the away team has a positive impact upon all attacking performance variables for the home team, the number of goals scored by the home team (at the 10% significance level), the number of shots on target by the home team (at the 5% significance level) and the number of successful passes made by home team (at the 1% significance level). GLSAWAY, however was negatively correlated to the defensive duties of the home side and was revealed to be significant for both the number of tackles won by the home team and the number of saves made by the home team at the 5% and 1% significance levels respectively.

The number of shots on target by the away team is revealed to significantly impact two performance variables for the home team. SOTIGLSAWAY is positively correlated to the number of saves made by the home team goalkeeper at the 1% significance level as one would expect. The number of shots on target by the away team is also revealed to negatively impact the total number of successful passes made by the home team also at the 1% significance level.

The total number of successful passes made by the away team is highly significant (1% significance level) and positively correlated to all four-performance variables for the outfield players of the home team. The number of successful passes made by the away team has no bearing on the number of saves made by the home team goalkeeper. The total number of passes made by the away team however was positively correlated to the number of goals scored by the home team, the number of shots on target by the home team, the number of successful passes made by the home team and the number of tackles made by the home team. The results obtained for the variables measuring the number of total successful passes of both the home and away teams are consistent suggesting that maintaining possession is a very important part of outfield play.

The number of tackles won by the away team is also significantly correlated with all four performance variables for outfield players of the home team. As expected, the number of tackles won by the away team is negatively correlated to the number of goals scored by the home team and the number of shots on target by the home team both at the 1% significance level. The number of tackles won by the away team is positively correlated to the number of successful passes made by the home team and
the number of tackles won by the home team, both at the 1% significance level. The number of tackles won by the away team is insignificant with regards to the number of saves made by the home team goalkeeper. The results for this variable are also consistent suggesting the action of tackling is a very important aspect of team performance.

The number of saves made by the away team goalkeeper (SMAWAY) is revealed to have a positive impact on the number of shots on target made by the home team at the 1% significance level. The number of saves made by the away team goalkeeper also has a positive impact on the number of successful passes made by the home team and this is also significant at the 1% significance level. The variable SMAWAY also has a positive impact on the number of tackles won by the home team at a 10% significance level. SMAWAY has no bearing on the number of saves made by the home team goalkeeper or the number of goals scored by the home team. The results obtained for this variable in conjunction suggest home teams put the away team goalkeeper under more pressure in attacking areas than the away team do the home team goalkeeper.

In contrast to the results obtained for the variable AVHGHTAWAY in table 5.4.1, the instrumental variable AVHGHTHOME is revealed in table 5.4.2 to have no significant bearing upon any performance variable of the away team. This implies that away teams utilised their physicality to their advantage more often than home teams. The remaining five player instruments, RFOOTHOME, LFOOTHOME, BFEETHOME, TPLAPPHOME and AVAGEHOME all unique to the home team had greater impact upon the attacking side of performance rather than the defensive side of performance, analogous to the results presented for the same five instrumental variables measuring the same characteristics of the away team in table 5.4.1.

The three instrumental variables RFOOTHOME, LFOOTHOME and BFEETHOME, measuring a player’s preferred foot had a highly significant impact upon the number of goals scored by the home team and the number of shots on target by the home team. None of these three variables however had any impact on the number of successful passes made by the home team, in contrast to the results presented in table 5.4.1. These results in conjunction therefore suggest the home team were less careless in possession. Table 5.4.2 reveals that a home player’s preferred foot had no impact
upon the number of tackles made by the home team, or the number of saves made by
the home team goalkeeper.

The two instruments TPLAPPHOME and AVAGEHOME reveal contrasting results. The total number
of accumulated previous Bundesliga appearances between the home team had a significantly positive impact on attacking aspect of play such as, the
number of goals scored by the home team, the number of shots on target by the home
team and the number of successful passes made by the home team. The average age
of the home team however, had a significantly negative impact upon the number of
successful passes made by the home team whilst also having a negative impact upon
the number of tackles made by the home team and the number of saves made by the
home team goalkeeper. The results obtained for this variable in table 5.4.1 and 5.4.2
in conjunction suggest a player’s career experience impacts home teams more in a
defensive manner and away teams more in an attacking manner.
5.5 Second Stage Regression Results

Table 5.5.1 below presents the two sets of results obtained for the second stage regression estimates from the 2SLS IV estimation procedure. Table 5.5.1 presents the results obtained when analysing the impact of the performance of the away team upon the performance of the home team as well as the results obtained when analysing the impact of the performance of the home team upon the performance of the away team. Tables 5.5.2 and 5.5.3 present the results obtained from the tests of overidentifying restrictions and the tests of endogeneity.

---

28 Tests for the explanatory power of instrumental variables (detailed in chapter 4) were not conducted for the analyses in this chapter as the inclusion of 5 endogenous variables meant these statistics could not be estimated (The STATA 12 software utilised throughout the empirical analyses conducted in this thesis failed to calculate any values in this case). Stock and Yogo (2002) calculated critical values for up to 3 endogenous regressors and between 3 and 30 instruments. If the number of endogenous variables or instruments exceeded these limits, critical values would not be shown.

29 Standard error estimation results are reported as well as robust standard error test scores.
Table 5.5.1 Second Stage Regression Results

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coef</th>
<th>z</th>
<th>Coef</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLSAWAY</td>
<td>-1.449458</td>
<td>-0.47+</td>
<td>1.2851556</td>
<td>0.91+</td>
</tr>
<tr>
<td>SHTSONTGTINCGLSAWAY</td>
<td>.882409</td>
<td>0.29+</td>
<td>-.6581556</td>
<td>-0.46+</td>
</tr>
<tr>
<td>TOTSUCCPASSAWAY</td>
<td>-.0022293</td>
<td>-0.93+</td>
<td>-.0008424</td>
<td>-0.78+</td>
</tr>
<tr>
<td>TACKWONAWAY</td>
<td>.0548805</td>
<td>0.77+</td>
<td>-.0050204</td>
<td>-0.32+</td>
</tr>
<tr>
<td>SAVESMADEAWAY</td>
<td>-.3409199</td>
<td>-0.21+</td>
<td>1.058998</td>
<td>1.28+</td>
</tr>
<tr>
<td>SEASON1</td>
<td>.018498</td>
<td>0.15+</td>
<td>-.0249468</td>
<td>-0.20+</td>
</tr>
<tr>
<td>SEASON2</td>
<td>.0122249</td>
<td>0.09+</td>
<td>.0110225</td>
<td>0.14+</td>
</tr>
<tr>
<td>SEASON3</td>
<td>.1599191</td>
<td>1.83*</td>
<td>-.0202269</td>
<td>-0.20+</td>
</tr>
<tr>
<td>GDHOME</td>
<td>.0067644</td>
<td>1.69*</td>
<td>.0059627</td>
<td>1.46+</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>.0039463</td>
<td>0.78+</td>
<td>.0092821</td>
<td>2.57**</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>.0127389</td>
<td>4.25***</td>
<td>-.0093369</td>
<td>-3.81***</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>-.0065248</td>
<td>-2.15**</td>
<td>.00083267</td>
<td>4.13***</td>
</tr>
<tr>
<td>GLSHOME</td>
<td>.2745967</td>
<td>0.17+</td>
<td>.4973879</td>
<td>0.62+</td>
</tr>
<tr>
<td>SHTSONTGTINCGLSHOME</td>
<td>.3248499</td>
<td>0.21+</td>
<td>-1.041366</td>
<td>-1.28+</td>
</tr>
<tr>
<td>TOTSUCCPASSHOMESTEAM</td>
<td>-.0028446</td>
<td>-3.16***</td>
<td>.0001117</td>
<td>0.08+</td>
</tr>
<tr>
<td>TACKWONHOME</td>
<td>-.0145539</td>
<td>-0.93+</td>
<td>.0380055</td>
<td>0.51</td>
</tr>
<tr>
<td>SAVESMADEHOMESTE</td>
<td>-.896833</td>
<td>-0.29+</td>
<td>.677204</td>
<td>0.46+</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
Table 5.5.1 above reveals the results obtained by the second stage regression from the structural equation where the dependent variable is HTRESULT, which represents the number of points acquired by the home team. Four out of seven control variables are statistically significant and these variables include the dummy variable for the 2009-2010 season (SEASON3), the goal difference of the home team (GDHOME), the percentage points of the home team (PERPTSHOMETEAM) and the percentage points of the away team (PERPTSAWAYTEAM). The only performance variable revealed to be significant was the total number of successful passes made by the home team. This therefore means that out of a total of ten performance variables included in the model, five (explanatory) for the home team and five (endogenous) for the away team, only the number of successful passes made by the home team had any impact upon the home teams match result.

The dummy variable SEASON3 was statistically significant at the 10% significance level and positively correlated to the match result of the home team, indicating home teams performed better and hence acquired more points in the season 2009-10 than the reference season 2010-11. The variable GDHOME, which controls for the relative quality of the home team is positively correlated with the match result for the home team and is also only significant at the 10% significance level.

The variables PERPTSHOMETEAM and PERPTSAWAYTEAM, which control for relative team form, are statistically significant at the 1% and 5% significance levels respectively. The percentage points of the home team (PERPTSHOMETEAM) is as expected, positively correlated to the home team result, meaning the relative form of the home team is positively related to the match points obtained by the home team. Conversely, the variable measuring the relative form of the away team, PERPTSAWAYTEAM is negatively correlated to match result of the home team. Table 5.5.1 above also reveals that the total number of successful passes made by the home team (TOTSUCPASSHOME) is negatively related to the result achieved by the home team. This result may imply that home teams enjoy lots of possession and engage in lots of passes, but do not necessarily put them together in important parts of the pitch. Unproductive or superfluous passing has been known to contribute towards negative match outcomes (Collet 2013).
Table 5.5.1 also presents the results of the second stage regression, where the variable ATRESULT is the dependent variable. As can be seen none of the five player performance variables for the home team (endogenous variables) are revealed to be of any significance. Similarly none of the five player performance variables for the away team (explanatory variables) are revealed to be of any significance.

None of the three dummy variables, SEASON1, SEASON2 or SEASON3 are statistically significant. Three of the four remaining control variables however, are statistically significant and these variables include GDAWAY, PERPTSHOMETEAM and PERPTSAWAYTEAM. The goal difference of the away team (GDAWAY), which accounts for the relative quality of the away team is significant at the 5% significance level and is positively correlated to the team result (points attained) by the away side.

The two control variables that account for relative form, PERPTSHOMETEAM and PERPTSAWAYTEAM are both significant at the 1% significance level and are negatively and positively signed respectively. The percentage points of the home team are therefore negatively correlated to the number of points acquired by the away team and the percentage points of the away team are positively correlated to the number of points acquired by the away team.

The suitability of the instrumental variables was verified by conducting Sargan $\chi^2$, Basmann $\chi^2$ and the Score $\chi^2$ (robust) tests of over-identifying restrictions. In a model where there are more instruments than the number of endogenous regressors, the model is deemed to be over-identified and in such cases tests for over-identifying restrictions can be run revealing the validity of the instruments. In cases where a model consists of an equal number of instruments and endogenous regressors, the model is deemed to be just identified and the Sargan, Basmann and Score $\chi^2$ tests for over-identifying restrictions cannot be performed.

Table 5.5.2 below reveals the results of the tests of over-identifying restrictions using the Sargan (1958), Basmann (1960) and Wooldridge (1995) score $\chi^2$ tests. Each of these tests, not only test to see if the instruments are correlated to the error term but also test to see if the equation is correctly specified. In other words each test will also check if any of the exogenous variables should be included in the structural equation,
therefore a significant test statistic would represent either an invalid instrument or an incorrectly specified structural equation. As can be seen from table 5.6.2 below the Sargan, Basmann and Score tests for over-identifying restrictions reveal test statistics much higher than 0.05 therefore the null hypothesis that the instruments are valid, cannot be rejected at the 5% significance level. All twelve instrumental variables therefore are valid and neither equation has been incorrectly specified.

Throughout the analysis it was assumed that the performance of the home team was affected by the performance of the away team. Similarly, it was assumed the performance of the away team was affected by the performance of the home team. In order to confirm endogeneity, the Durbin $\chi^2$, Wu-Hausmann $\chi^2$ and Robust Score $\chi^2$ tests were conducted and a robust regression estimated and their results are presented in table 5.6.3 below.

If the endogenous regressors did turn out to be exogenous then it would have meant OLS estimations would be a more appropriate method of testing instead of the 2SLS adopted. Table 5.6.3 below reveals the results of the test of endogeneity and it can be seen that the null hypothesis of all tests that the performance variables are exogenous can be rejected, as all test statistics are highly significant. The performance variables of the away team were therefore correctly treated as endogenous in the analysis investigating the impact of the away team’s performance on the home team. Similarly the performance variables of the home team were also correctly treated as endogenous in the analysis investigating the impact of the performance of the home team on the away team’s performance.
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>ATRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sargan $\chi^2$ Test Score</td>
<td>0.006557</td>
<td>1.14308</td>
</tr>
<tr>
<td>Sargan $\chi^2$ (p-value)</td>
<td>0.9355</td>
<td>0.2850</td>
</tr>
<tr>
<td>Basmann $\chi^2$</td>
<td>0.006455</td>
<td>1.12637</td>
</tr>
<tr>
<td>Basmann $\chi^2$ (p-value)</td>
<td>0.9360</td>
<td>0.2886</td>
</tr>
<tr>
<td>Score $\chi^2$</td>
<td>0.00956</td>
<td>1.03404</td>
</tr>
<tr>
<td>Score $\chi^2$ (p-value)</td>
<td>0.9221</td>
<td>0.3092</td>
</tr>
</tbody>
</table>
Table 5.6.3 Tests of Endogeneity

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>ATRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Durbin $\chi^2$ Test Score</td>
<td>11.8929</td>
<td>10.8676</td>
</tr>
<tr>
<td>Durbin $\chi^2$ (p-value)</td>
<td>0.0363</td>
<td>0.0441</td>
</tr>
<tr>
<td>Wu-Hausman F Score</td>
<td>2.35678</td>
<td>2.15177</td>
</tr>
<tr>
<td>Wu-Hausman F (p-value)</td>
<td>0.0385</td>
<td>0.0467</td>
</tr>
<tr>
<td>Robust Score $\chi^2$ (5)</td>
<td>11.5962</td>
<td>11.5705</td>
</tr>
<tr>
<td>Robust Score (p-value)</td>
<td>0.048</td>
<td>0.0412</td>
</tr>
<tr>
<td>Robust Regression (F)</td>
<td>2.74523</td>
<td>2.33987</td>
</tr>
<tr>
<td>Robust Regression (p-value)</td>
<td>0.0179</td>
<td>0.0398</td>
</tr>
</tbody>
</table>
5.6. Discussion

With regards to the impact of the away teams performance on the home teams performance the results obtained suggest that team effects have much more impact than specific, though aggregated player performance variables. Table 5.5.1 revealed the variables GDHOME, PERPTSHOMETEAM and PERPTSAWAYTEAM were all statistically significant influences on the number of points achieved by the home team. GDHOME, measuring the relative quality of the home team was revealed to be positively correlated to the match result achieved by the home team as expected. The relative form of the home team captured by the variable PERPTSHOMETEAM was revealed to be positively correlated to the number of points acquired by the home team. The relative form of the away team (PERPTSAWAYTEAM) was negatively correlated to the match result obtained by the home team. These results in conjunction therefore confirm that the relative form of the home team going into a match is positively associated with the number of points obtained by the home team (in the observed match). Similarly, the relative form of the away team (carried into a match) had a negative impact upon the number of points acquired by the home team.

The SEASON3 dummy variable is positively correlated to the result achieved by the home teams suggesting home teams were quite successful during the 2009-2010 season. The only other variable of any significance in table 5.5.1 was the player performance variable TOTSUCPASSHOME. The total number of successful passes made by the home team is revealed to be negatively correlated to the match result obtained by the home team. At first this result appears surprising, but much previous research on team possession reveals that possession in the right areas of the field (attacking areas) is of most importance (Bate 1998; Lago-Penas et al 2010; Collet 2013). This result may simply imply that home teams, although engaging in lots of passes, did not necessarily put them together in important parts of the pitch. Collet (2013) in a study across European leagues revealed that unproductive or superfluous passing contributed to negative match outcomes. Carmichael et al (2001) also stress that quantity of possession is not enough to determine team output, but instead quality of possession in all its manifestations are of greater importance. As the variable TOTSUCPASSHOME was highly significant this result also suggests that home
teams enjoyed the larger share of possession throughout the four season sample period.

Table 5.5.1 also presents similar results obtained when assessing the impact of the home teams performance upon the away teams performance. None of the ten player performance variables (five for the home team and five for the away team) were of any significance. The only three variables of significance were three of four variables that controlled for relative team quality and relative team form, GDAWAY, PERPTSHOMETEAM and PERPTSAWAYTEAM. Similarly to the relative quality of the home team having a positive effect upon the home teams result, the relative quality of the away team also has a positive impact upon the away teams result. Percentage points acquired by the home and away teams were negatively and positively signed respectively and both had the expected impact upon the match result achieved.

Generally the results obtained from the second stage regressions estimated in this chapter lead to the suggestion that with regards to team results (output), individual player actions are not of such great importance as opposed to actual team effects. Relative team form going into a match is highly significant (1% level) and has the expected impact upon team results. This implies the existence of momentum effects in football similar to those discovered in other team sports such as ice hockey (Shmanske and Lowenthal 2007; Leard and Doyle 2010) and basketball (Arkes and Martinez 2011).

Goal difference, which captured the relative quality of teams, is also revealed to have a significant impact upon match results (and is also signed as expected), but only at the 5% and 10% significance levels. Relative team quality therefore, although impacts match results, does not have as profound an effect on the match result as relative form.
5.7 Conclusion

The purpose of the analyses conducted in this chapter was to determine the extent of the interdependency between the performances of two opposing teams. As such the most influential performance variables in determining match outcome (as identified by previous literature) were treated as endogenous. Post estimation tests confirmed that performance measures of the opponent were correctly treated as endogenous and the tests of over-identification along with the first stage regression results confirmed the validity of the instrumental variables used and the specification of the models.

Ultimately, the second stage regression results revealed that the performance variables of the opponent were insignificant and therefore had no impact upon an observed team’s match result. The second stage regression results however, also revealed that team effects such as relative team quality and relative team form (prior to the observed match) have more impact on the opposing team’s result than individual player actions, which perhaps helps to emphasise that the outcomes of a football match are essentially linked to team and collective influences that are not easily separated into particular individual efforts. In addition the impact of these team effects appear to be much more prevalent for attacking match contributions such as the number of goals scored and the number of shots on target rather than defensive match contributions, suggesting that the most important and decisive match actions are actually those that occur in the attacking third of the pitch.

The results obtained for the instrumental variables throughout the analyses imply that the average height of the away team, as well as the preferred feet of the players on the away team and the previous Bundesliga experience accumulated by the away team all impact the away teams performance in a positive manner. Specifically, these instrumental variables revealed strong positive correlation with attacking performance variables such as the number of goals scored and the number of shots on target made by the away team. In particular, the total number of previous Bundesliga experience appeared to consistently impact attacking play actions and successful maintenance of possession in a positive manner, suggesting experience at the very top level of German football is of vital importance, particularly when playing away. The fact that these instrumental variables appeared to have more significance with attacking play
actions for the away teams suggests that players performing in attacking units were selected partly on the basis of their physical characteristics when playing away.

Generally, the results obtained for the first stage regressions throughout the analyses suggest that the team effects, individual player contributions (accumulated to team-match totals) and physical characteristics of players (accumulated to team characteristics) are more important towards attacking match contributions rather than defensive match contributions. The results obtained for the second stage regressions, imply that team effects such as relative quality and relative form contribute more to positive team results than individual match actions, whilst the quality and form of the opposition has the opposite impact on the observed team’s ability to acquire a positive result. Team performance is therefore very much dependent upon both their own recent past performance as well as the recent past performance of their opposing team.
Chapter 6 - Investigating Interdependency Between the Performances of Two Players in Direct Competition in the German Bundesliga

6.1 Introduction

The purpose of the investigation conducted in this chapter is to determine whether the performances of individual players in direct competition are interdependent. Chapter 5 presented the results obtained from the analysis of collective team performance and revealed that performances of two opposing teams were not jointly determined. Chapter 5 also revealed that the relative quality and relative form of the opponent carried into a game had a linear impact upon the observed teams performance. This chapter examines player performance at a more isolated player level by examining specific individual players performing in either attacking or defensive positions.

Section 6.2 provides a detailed description of the sample, performance variables and instrumental variables used throughout the analysis. Section 6.3 details the analysis strategy and model specification. Section 6.4 reports the results obtained from the analysis investigating the impact of the performance(s) of defenders upon the performances of their directly opposing attackers. Section 6.5 presents the results obtained from the analysis investigating the impact of the performance(s) of attackers upon the performances of their directly opposing defenders. Section 6.6 discusses the results obtained and offers suggestions as to their implications and provides reasoning.

This chapter concludes with section 6.7, which reviews those results of most importance and summarizes the findings from the analyses conducted. The results confirm that the performances of defenders are statistically significant and negatively correlated to the performance of their opposing attackers, however the performances of attackers has no bearing on the performance of their opposing defenders. It can be concluded therefore that assuming productivity adjusted minutes played encapsulates
performance, the performance of a defender and their opposing attacker are interdependent. The results confirm that the performances of defenders and attackers at an isolated individual player level are jointly determined, supporting the sports economic theory of joint production.
6.2. Organisation of Data and Variables

In chapter 5, the dependent variable used throughout the analyses conducted was the match result (expressed as the number of points acquired) of the observed team. In order to assess the relationship between the performances of players at a more individual level, the designated output was the productivity adjusted minutes played by each player. Productivity adjusted minutes played simply refers to the minutes played by an observed player in a match, having accounted for the minutes played by his direct opponent on the field of play. Previous research shows that playing time has a positive effect on salary (Hakes and Sauer 2006) and that the variance in player salary is largely explained by the variance in individual performance (Frick 2011). On this basis the minutes played by players was assumed to encapsulate player performance and was therefore utilised as the measure of a player’s output. The results presented in sections 6.4 and 6.5 reveal how much of an impact the performance of players in either defence or attack directly affects the performance of their opponent(s).

In order to assess the impact of players’ performances against their opponents, players first had to be matched to opponents to attempt to replicate the typical environment present in sports such as baseball, where individual player contribution to match outcome has been more commonly measured (Scully 1974: Zech 1981: Macdonald and Reynolds 1994: Krautmann and Oppenheimer 2002: Bradbury 2007). To allow this match up, only players with a set of plausible opponents i.e. in fixtures where team formations allowed the match up of individual players with opposing players were selected. For example if two opposing teams, team X and team Z both played a standard 4-4-2 formation, then each of 20 outfield players would in theory be directly up against only one opponent as figure 6.2.1 below illustrates.
Figure 6.2.1 - Team X and Team Z Lining up Against Each Other in a Typical 4-4-2 Team Formation

As can be seen from figure 6.2.1 above, team X, represented by the X’s are matched up against their direct opponents (team Z’s players) represented by the Z’s. Theoretically, each player represented by an X in figure 6.2.1 above could be competing against the opposing Z directly in front of them, therefore allowing the match up of one player with their sole opponent on the field of play. One might argue that in sports like football, one on one situations regularly occur as argued by Gerisch and Reichelt (1993) who utilised video feeds to analyse one on one situations in a two match sample and found there were 250 match ups in total. Unlike the study conducted by Gerisch and Reichelt (1993) however, this study utilised a much larger data set consisting of 1,183 fixtures that recorded player-starting positions. This study
therefore focuses on the area of the pitch that one-on-one match ups might occur and hence identifies which player in the one-on-one belongs to which position on the field of play. The actual areas of interest from figure 6.2.1 for this particular study are presented in figure 6.2.2 below.

Figure 6.2.2 - Typical 4-4-2 VS 4-4-2 Team Formation Match Up

Figure 6.2.2 above presents four players from either side up against four players from the opposing team. In figure 6.2.2 the two horizontal rows of letters (X’s and Z’s) closest to the baseline represents the four players comprising of the defending unit of each team. Consequently, the four opposing players these defenders are matched up against represent the attacking unit of the opposing team. Unlike other players the four central midfield players are matched with players belonging to the same
positional unit (i.e. opposing central midfielders) this meant that any analysis on their performances would lack any natural and clear interpretation and they were therefore excluded from the analyses. However, as central midfielders are often used to attack as well as defend (Frick 2007) it was assumed that their impact upon defensive and attacking performance would be equal.

In some previous studies that have attempted to analyse an individual player’s performance a solitary index value has been used as the unit of measurement, such as the index value in ‘Puntos Marca’ (Del-Barrio and Pujol 2007) in Spanish football or the IVG statistic available in ‘Tutto Campionato’ (Montanari 2008) for the Italian Serie A. There are however, a few areas of concern when using index values such as these mentioned above, the first being that the users of such statistics most likely only have a vague idea of the weightings given to the performance criteria included and of the complex calculations involved in deriving the value. Secondly, these index values are usually in the form of weighted averages (McHale et al 2012) and therefore mask important individual contributions to a game. Thirdly, these index values given to players appear to be calculated independently of the performance of their opponents.

The unique match up presented in figure 6.2.2 above, addresses each of these concerns as it allows the measurement of a players performance whilst in direct competition with an identified opponent. In turn, the identification of the opponent means recorded individual player performance statistics for both players (observed and opponent) can be used, therefore eliminating both the need to use averages or to attach weightings to play actions.

More detailed diagrams of all of the team formation match ups present in the four-season sample period are presented in appendix three. Each player match up within these formation variations is highlighted (within elliptical shapes) therefore revealing precisely how all players were matched to an opponent and how the eventual sample was selected.

As stated earlier in chapter 4, four seasons of data was utilised and exhausted in order to fulfil the purposes of this study. The data provided a total of 1,224 fixtures, however only 1,189 fixtures could be used in the analyses. In total, 34 fixtures were excluded as they either, allowed no match up of any players due to the deployment of unorthodox formations or they only allowed matching of central midfield players to
their midfield opponents. These 34 excluded fixtures included 5 formation variations where no match up of players was allowed. Specifically, these five formation variations included those matches where two teams lined up against each other in the following formations, 3-4-3 VS 4-2-2-2, 4-3-3 VS 4-1-4-1, 4-3-3 VS 4-5-1, 4-3-3 VS 4-4-1-1 and 4-3-3 VS 4-5-1. The remainder of the 34 excluded fixtures consisted of 6 further formation variations that only allowed central midfield players to be paired to other central midfield opponents. These six formation variations included those matches where teams lined up in the 4-1-2-1-2 VS 3-4-3, 4-4-2 VS 3-4-3, 3-5-2 VS 3-5-2, 4-3-3 VS 4-3-2-1, 4-3-3 VS 4-3-3 and 4-4-1-1 VS 4-5-1 formations. All formation diagrams presented above have been based on detailed formation diagrams provided by Opta and the corresponding starting eleven formation data recorded. In addition, one fixture was excluded, which was FC St Pauli Vs. FC Schalke 04 played on April 1\textsuperscript{st} 2011. This was due to the data being only partially available as the game was abandoned, and the result awarded to the away side.

To further screen the data goalkeepers and central midfield players were also removed from the sample. Goalkeepers are the most specialised players on the field (Frick 2007) and their ability to handle the ball is restricted to their own penalty area clearly marked on every field of play. Goalkeepers are the last line of defence (McHale et al 2012) and are most commonly called into action on occasions where the defensive unit protecting them has been breached. Consequently, goalkeepers are only called upon sporadically and they are therefore never constantly up against a member of the opposition, meaning a match up similar to that presented in figure 6.2.2 is never possible, regardless of team formations.

Central midfield players were removed from the sample as they are in direct competition with other central midfielders. A central midfielder Vs. central midfielder analysis would therefore lack natural interpretation. Due to central midfielders having both attacking and defensive responsibilities (Frick 2007), their impact was assumed to be equal upon the performances of their fellow defenders and attackers. Due to the omission of goalkeepers and central midfield players, the actual sample used in this analysis consisted only of players from the remaining positional units, namely the attacking unit (consisting of forwards and wingers) and the defensive unit (consisting of central defenders and full backs).
To control for potential bias on the rare occasion players were played out of position (and hence in unfamiliar roles), a few additional match ups were also excluded from the sample. For example, Hamit Altintop of FC Bayern Munich was matched up to an opponent in 28 fixtures, however played 27 of these matches in his usual right sided midfield position and 1 match in the right back position (most likely due to injuries and/or suspensions for the usual right back). This match up, where Hamit Altintop was played out of position was excluded from the sample as were similar match ups involving other players, playing out of their usual position. A player’s usual position was determined by the position in which the player appeared most often throughout the four season sample period. In total, only 9 match ups were affected by this bias and therefore dropped from the final sample.

In addition, any substitutions made by teams throughout the four seasons were deliberately excluded from the sample. The reason for this was that if substitutions were included the total time played by defensive and attacking players would be almost, if not exactly the same (as most observations for time played would be 90 minutes). This would inevitably conceal important information regarding a specific player’s performance. Throughout the analyses conducted in this chapter the time played (in minutes) was viewed to be a function of their productivity and the objective was to determine if this productivity was affected by their opponents productivity. Any substitutions made throughout a game would therefore reveal important information regarding the substituted players productivity in relation to the productivity of their direct opponent.

Injuries picked up by players during a game also had to be accounted for. As injuries could conceivably be picked up at any point during a standard 90-minute match it could have been extremely difficult to distinguish between substitutions that were forced (injuries) and substitutions that were voluntary (tactical). However, Del Corral et al (2008) states only 4% of substitutions occur in the first half of games as they are usually forced and the majority of tactical substitutions occur around the 60th minute mark as managers want to persist with the starting 11 for as long as possible. To control for injuries picked up by players leading to forced substitutions, all players that played less than 45 minutes of the game (one half), except those players that had been red carded were also excluded from the sample. Ultimately, this meant a total of 148 player observations (74 match ups) were excluded as these players were assumed
to have picked up injuries early on in the game, leading to a forced substitution. Once all selection was completed the sample consisted of 5,411 player observations.

TABLE 6.2.1 below provides a detailed list of all variables used throughout the analyses.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent /Endogenous Variables</strong></td>
<td></td>
</tr>
<tr>
<td>DEFTP</td>
<td>Defender Time Played in minutes</td>
</tr>
<tr>
<td>ATTTP</td>
<td>Attacker Time Played in minutes</td>
</tr>
<tr>
<td><strong>Performance Variables</strong></td>
<td></td>
</tr>
<tr>
<td>DEFTACKWON</td>
<td>Number of Tackles Won by a Defender</td>
</tr>
<tr>
<td>DEFTOTCLE</td>
<td>Number of Total Clearances made by a Defender</td>
</tr>
<tr>
<td>DEFBLO</td>
<td>Number of Blocks made by a Defender</td>
</tr>
<tr>
<td>DEFINT</td>
<td>Number of Interceptions made by a Defender</td>
</tr>
<tr>
<td>ATTGLS</td>
<td>Number of Goals Scored by a Attacker</td>
</tr>
<tr>
<td>ATTSHTSONINCGLS</td>
<td>Number of Shots on Target including Goals Scored by an Attacker</td>
</tr>
<tr>
<td>ATTASSTS</td>
<td>Number of Assists made by an Attacker</td>
</tr>
<tr>
<td>ATTSUCPASS</td>
<td>Number of Successful Passes made by an Attacker</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
</tr>
<tr>
<td>DEFYC</td>
<td>Number of Yellow Cards Obtained by a Defender</td>
</tr>
<tr>
<td>DEFRC</td>
<td>Number of Red Cards Obtained by a Defender</td>
</tr>
</tbody>
</table>
DEFVDUM  Venue Dummy for Defenders
ATTYC  Number of Yellow Cards Obtained by an Attacker
ATTRC  Number of Red Cards Obtained by an Attacker
ATTVDUM  Venue Dummy Variable for Attackers
SEASON 1  2007-2008 Dummy Variable
SEASON 2  2008-2009 Dummy Variable
SEASON 3  2009-2010 Dummy Variable

*Instrumental Variables*
DEFAGE  Defender Age in years
DEFHEIGHT  Defender Height in Metres
DEFRFOOT  Defender with a Preferred Right Foot
DEFLFOOT  Defender with a Preferred Left Foot
DEFPREVLAPP  Previous Bundesliga Appearances made by a Defender
ATTAGE  Attacker Age in Years
ATTHEIGHT  Attacker Height in Metres
ATTRFOOT  Attacker with a Preferred Right Foot
ATTLFOOT  Attacker with a Preferred Left Foot
ATTPREVLAPP  Previous Bundesliga Appearances made by Attacker

---

30 Season 4 (2010-2011) was used as the reference season
31 Variables representing defenders and attackers able to use both feet (DEFBFEET and ATTBFEET) were deliberately omitted from this analysis in order to prevent multicollinearity from affecting estimations. A simple correlation matrix revealed high correlation between the variables DEFBFEET and DEFLFOOT and DEFRFOOT. Similarly ATTBFEET was found to be highly correlated with ATTRFOOT and ATTLFOOT.
6.3 Analysis Strategy

In order to investigate simultaneity between performances at an individual player level a pooled cross sectional data set was constructed using the exclusive data provided by Opta and supplementary data obtained from several editions of transfermarkt magazines. A 2SLS instrumental variable estimator was employed for each of the analyses conducted in this chapter. The model specifications are outlined in the subsections 6.3.1 and 6.3.2 below.

6.3.1 Impact of defenders performance on attackers performance

Equation 6.1 below presents the model utilised when investigating the interdependency between defenders’ performances and the performances of their opposing attackers. In equation 6.1, the dependent variable $Y_1$ represents the productivity adjusted minutes played by the attackers (ATTTP), the $Z$ variables are explanatory variables and $Y_2$ represents the endogenous variable, the time played by the defenders (DEFTP).

\[
Y_1 = \beta_0 + \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \ldots + \beta_4Y_2 + u 
\]  
(6.1)

\[
Y_2 = \pi_0 + \pi_1Z_1 + \pi_2Z_2 + \pi_3Z_3 + \pi_4IV_1 + \pi_5IV_2 + \ldots + \pi_9IV_6 + \upsilon 
\]  
(6.2)

In the model presented above the independent variables in the structural equation (equation 6.1) consisted of four attacking play variables, (ATTGLS) goals scored, (ATTSHTSONINCGLS) shots on target including goals scored, (ATTASSTSS) assists made and (ATTSUCPASS) total successful passes. The goals scored variable has been shown to be of great match winning potential (Lucifora and Simmons 2003) and positively affects match results (Carmichael and Thomas 2008). Goals scored were also deemed to be a key performance indicator for attacking players in much previous research on player performance (Hughes and Franks 2005; Lago and Martín 2007; Lago-Penas et al 2010; McHale et al 2012).

Another action of high match winning potential is the action of taking shots on target. Studies have revealed match results to be positively correlated to shots on target, meaning the more accurate attacking players are the greater the chances of winning the game (Carmichael et al 2001; Torgler 2004; Lago and Martín 2007; Lago- Penas
An assist refers to the action of passing the ball to the eventual scorer of a goal so it is a measure of both passing accuracy and chance creation. The variable assist was therefore another key performance indicator for attacking players in much previous literature (Hughes and Franks 2005; Lago and Martín 2007; LagoPenas et al 2010; McHale et al 2012).

Finally, total successful passes made by a player measures their ability to retain possession of the ball for the team by accurately directing the ball towards a team member. Possession of the ball has been revealed in many studies as being a key determinant of team success and various aspects of the passing attribute have been extensively reviewed such as, passing accuracy, passing range, longevity of passing sequences and recovery of possession (Carmichael et al 2000; Hughes and Churchill 2004; Jones et al 2004; Lago-Penas et al 2010; Vogelbein et al 2014). All four of these performance variables were selected as the independent variables for this analysis, as opposed to other potential measures, due to the widespread importance given to these variables across both the quantitative and qualitative research on performance analysis in football.

In addition six other control variables were also included in equation 6.1, two of which were ATTRC (attacker red cards) and ATTYC (attacker yellow cards). Red card offences are serious offences such as, tackles from behind that endanger opponents, denying obvious goal scoring opportunities and racist/offensive gestures (Dawson et al 2007). Red cards lead to immediate dismissal from the field of play (as punishment) and therefore suspend the minutes played by the player. Picking up red cards is generally believed to weaken a team (Carmichael and Thomas 2008) and have been shown to have a significantly negative effect upon match results (Carmichael et al 2000; Torgler 2004).

Yellow cards are more often awarded to players for less serious offences such as, time wasting, simulation (diving), removing their shirt (often in celebration) and the intentional holding or pulling of opponents (Dawson et al 2007). Yellow cards, although will not directly affect minutes played may still influence tactical decisions made by managers. The awarding of a yellow card is most often viewed as an indication of aggression, ill-discipline and desperation (Carmichael and Thomas 2008). Picking up a solitary yellow card is regarded a warning but receiving two of
these yellow cards results in the player receiving a red card and being removed from the field of play. Consequently, a manager could realistically be more likely to substitute a yellow carded player in order to protect the interests of the team. Furthermore, it would be reasonable to assume that a player receiving a yellow card would impact a player’s behaviour throughout the rest of the match.

Another control variable included in the model was a dummy variable controlling for whether the fixture was being played at home or away. Home sides have been shown to have an advantage over their opposition (Nevill et al. 1996; Tucker et al. 2005; Pollard 2008), due to a number of factors such as referee bias (Sutter and Kocher 2004; Dawson et al. 2007; Pollard 2008) and higher effort levels from home players (Leard and Doyle 2010). Other factors such as the support received from the crowd and familiarity effects have also been recognised as being the source of home advantages (Nevill et al. 1996; Carmichael and Thomas 2005; Pollard 2008).

Three additional control variables SEASON 1, SEASON 2 and SEASON 3 were included in order to control for inter-season effects that may have had an impact upon the general quality of football played in the league, such as managerial changes or player transfers. As outlined in chapter 3, the ability of the manager has been widely accepted to have a significant effect on the success and performance of the team as a whole and individual players within teams (Audas et al. 1999; Dawson et al. 2000; Salomo and Teichmann 2000; Dawson and Dobson 2002; Frick 2010).

In the model presented above, the reduced form equation (equation 6.2) consists of the endogenous variable $Y_2$ (DEFTP), which is expressed as a function of four defensive play variables, three control variables and five instrumental variables. The defensive play variables include, DEFTACKWON (tackles won), DEFTOTCLE (total clearances), DEFBLO (blocks made) and DEFINT (interceptions). The three control variables consist of, DEFYC (defender yellow cards), DEFRC (defender red cards), DEFVDUM (venue). The five instrumental variables consist of, DEFAGE (defender age), DEFHEIGHT (defender height), DEFRFOOT (right footed defenders), DEFLFOOT (left footed defenders) and DEFPREVLAPP (Number of previous Bundesliga appearances made by defender).

The variable DEFTACKWON is a measure of a player’s ability to regain or win possession of the ball and has been included as a measure of performance in many
previous studies (Carmichael et al 2000, 2001; Carmichael and Thomas 2005; Tiedemann et al 2011). Tackles won is also a measure commonly used in soccer performance rating systems such as the EA sports player performance index (McHale et al 2012).

The remaining three performance variables, clearances, blocks and interceptions are also commonly used in the performance analysis of defenders (Lago and Martín 2007; Dellal et al 2011; McHale et al 2012; Redwood-Brown et al 2012). Clearances refer specifically to the action of removing the ball from a danger area, most likely to relieve some pressure for the defending side. Making a block requires a player placing a part of their body, besides their hands in the direction of the ball in order to prevent the shot taken by their opponent to travel in the vicinity of their goal. Finally, interceptions refer to a player’s reading of the game and their anticipation of events as they deliberately cut out the intended target of a pass or regain the ball due to a misplaced pass. In addition, control variables for yellow cards obtained by defenders (DEFYC), red cards obtained by defenders (DEFRC) and a dummy variable for venue (DEFVFDUM) were included in the reduced form equation.

In order to satisfy the conditions of a 2SLS IV estimation procedure, the reduced form equation presented in equation 6.2 consisted of five instrumental variables all unique to individual defenders. These instrumental variables included, the number of previous Bundesliga appearances made by each defender (DEFPREVLAPP), right footed defenders (DEFRFOOT), left footed defenders (DEFLFOOT), the height of defenders (DEFHEIGHT) and the age of each defender (DEFAGE).

Previous league experience has been shown to be positively correlated to player salary and have a strictly linear influence on annual income (Frick 2011). In addition, the most recent appearances (made in the last season) have been shown to matter more than experience gained in prior seasons (Frick 2011). The two dummy variables for a player’s preferred foot provided a measurement of whether the player was a right footed or left footed player. It has been revealed that players with a preferred left foot command a 15% premium (Frick 2007).

A player’s age is a variable that also measures player experience (similar to the previous league experience variable), however the experience captured by the age variable is not restricted to experience gained in the German Bundesliga but
experience gained in all levels of football throughout a player’s career. A player’s age therefore also takes into account, all previous professional club experience, under 21 international experience and senior international experience, all of which have been revealed to be positively correlated to player salaries and transfer fees (Carmichael et al 1999; Lucifora and Simmons 2003).

Finally, the instrumental variable for player height was also included in the model as an instrumental variable. Bloomfield et al (2005) states that physical player attributes can be reflective of the roles they are required to perform on the field and Frick (2007; 2011) reveals significant differences in player salaries depending upon which position the player plays in. Although the salary premiums between different positions could be explained by performance characteristics such as goals, assists etc, a player’s physical attributes could at least partially explain their efficiency at conducting these play actions.

6.3.2 Impact of attackers performance upon defenders performance

When investigating the interdependency between attackers performances and the performances of their opposing defenders a similar model was utilised. The dependent variable in equation 6.3 below is DEFTP (productivity adjusted minutes played by the defenders) and Y2 the endogenous variable is represented by ATTTP (time played by the attackers).

\[ Y_1 = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \ldots + \beta_4 Y_2 + u \]  
\[ (6.3) \]

\[ Y_2 = \pi_0 + \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 + \pi_4 IV_1 + \pi_5 IV_2 + \ldots + \pi_9 IV_6 + \upsilon \]  
\[ (6.4) \]

In the structural equation 6.3, the independent variables are four defensive performance measures, DEFTACKWON, DEFTOTCLE, DEFBLO and DEFINT and six control variables DEFYC, DEFRC, DEFDUM, SEASON1, SEASON 2 and SEASON 3. Equation 6.4 represents the reduced form equation in which the independent variables consist of the four attacking performance variables, ATTGLS, ATTSHSONINCGLS, ATTASSTS, ATTSUCPASS, three control variables ATTYC, ATTRC, ATTVDUM and five instrumental variables each unique to attacking players, ATTAGE, ATTHEIGHT, ATTRFOOT, ATTLFOOT and ATTPREVLAPP.
6.4 Results: First Stage Regression Results

Table 6.4.1 and 6.4.2 below presents the results obtained from the first stage reduced form regressions. In table 6.4.1 the productivity adjusted minutes played by members of the defending unit is the dependent variable. In table 6.4.2 the productivity adjusted minutes played by members of the attacking unit is the dependent variable. Robust Standard errors were estimated for each analysis conducted throughout this chapter in order to account for the inclusion of the limited (dummy) variables.
Table 6.4.1 First stage regression estimation results (DEFTP)

Dependent Variable DEFTP (Time played by Defenders)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coef</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTGLS</td>
<td>-0.0157601</td>
<td>-0.05+</td>
</tr>
<tr>
<td>ATTSHTSONINCGLS</td>
<td>-0.0127427</td>
<td>-0.07+</td>
</tr>
<tr>
<td>ATTASSTS</td>
<td>0.1763269</td>
<td>0.43+</td>
</tr>
<tr>
<td>ATTSUCPASS</td>
<td>-0.0021603</td>
<td>-0.15+</td>
</tr>
<tr>
<td>ATTYC</td>
<td>0.2537194</td>
<td>0.63+</td>
</tr>
<tr>
<td>ATTRC</td>
<td>2.161489</td>
<td>2.86***</td>
</tr>
<tr>
<td>ATTVDUM</td>
<td>0.3690966</td>
<td>1.29+</td>
</tr>
<tr>
<td>SEASON1</td>
<td>-1.628529</td>
<td>-3.44***</td>
</tr>
<tr>
<td>SEASON2</td>
<td>-1.464379</td>
<td>-3.28***</td>
</tr>
<tr>
<td>SEASON3</td>
<td>0.2668005</td>
<td>0.7+</td>
</tr>
<tr>
<td>DEFTACKWON</td>
<td>0.909926</td>
<td>10.88***</td>
</tr>
<tr>
<td>DEFTOTCLE</td>
<td>0.595697</td>
<td>11.46***</td>
</tr>
<tr>
<td>DEFBLO</td>
<td>1.094236</td>
<td>7.82***</td>
</tr>
<tr>
<td>DEFINTE</td>
<td>0.7733102</td>
<td>11.16***</td>
</tr>
<tr>
<td>DEFYC</td>
<td>0.3237793</td>
<td>0.88+</td>
</tr>
<tr>
<td>DEFRC</td>
<td>-22.63462</td>
<td>-6.96***</td>
</tr>
<tr>
<td>DEFVDUM</td>
<td>-0.566924</td>
<td>-1.95*</td>
</tr>
<tr>
<td>DEFAGE</td>
<td>-0.0302371</td>
<td>-0.61+</td>
</tr>
<tr>
<td>DEFHEIGHT</td>
<td>-3.31822</td>
<td>-1.26+</td>
</tr>
<tr>
<td>DEFRFOOT</td>
<td>1.33423</td>
<td>2.44**</td>
</tr>
<tr>
<td>DEFLFOOT</td>
<td>1.662377</td>
<td>2.80***</td>
</tr>
<tr>
<td>DEFPREVLAPP</td>
<td>0.0008929</td>
<td>0.30+</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
The R-Squared value obtained is a relatively low 0.2448 (24.48%) as is the adjusted R-Squared 0.2416 (24.16%), however as explained in chapter 4 earlier both of these values have no natural interpretation in a 2SLS procedure (Wooldridge 2009). In addition, the recorded F statistic revealed that all independent variables used in this first stage regression were jointly significant (p-value 0.0000).

As can be seen from table 6.4.1, all four defensive performance variables are positively correlated with the dependent variable DEFTP and all four defensive performance variables were also highly significant. As expected, and in support of, Carmichael et al (2000), Lago and Martin (2007), Dellal et al (2011), McHale et al (2012) and Redwood Brown et al (2012), the defensive KPI’s of tackles won, interceptions, blocks and clearances are all very important in relation to a defenders performance.

Two instrumental variables, DEFRFOOT and DEFLFOOT, which represented a player’s preferred foot, are also highly significant (1% significance level) and both are positively signed. Table 6.4.1 above reveals the coefficient value for left footed defenders is greater than that for right footed defenders suggesting left footed defenders performed slightly better throughout the seasons 2007-08 to 2010-11.

The variable DEFRC, which measures red cards and hence proxies ill-discipline by defenders was highly significant and negatively correlated to the dependent variable time played. This was to be expected as red carded players are removed from the field of play, therefore preventing them from accumulating any more minutes. Conversely, the number of red cards obtained by attackers (ATTRC) had a positive impact upon the performances of defenders. This is not surprising as if an attacker receives a red card the defending unit has one less opponent to compete against. Vecer et al (2009) reveals that when a team receives a red card, their scoring intensity decreases by a factor of around 2/3, whereas the scoring intensity of the non-affected team increases by a factor of around 5/4, implying less work for the defenders on the non-affected team and greater work for the defenders on the affected team.

The variable DEFVDUM was negatively signed and significant at the 10% significance level. This variable was a control for the venue at which the defenders were playing (0= Home, 1= Away) and the result therefore implies defenders performed worse away. To an extent, this result could be explained by the supposed
advantages of playing at home such as the effect of the home crowd (Nevill et al 1996; Carmichael and Thomas 2005; Pollard 2008), referee bias towards the home team (Sutter and Kocher 2004; Dawson et al 2007; Pollard 2008), fatigue due to travelling away (Neave and Wolfson 2003) and increased effort levels from home teams (Leard and Doyle 2010) all contributing to improved home performance (and conversely worse away performance).

Finally, the SEASON1 and SEASON2 dummy variables were highly significant (1% significance level) and negatively correlated to the performances of defenders. These results imply that defenders performed worse in the 2007-08 and 2008-09 seasons in comparison to the 2010-2011 reference season.
Table 6.4.2 First Stage Regression Results (ATTTP)

**Dependent Variable ATTTP (Time Played by Attackers)**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coef</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFTACKWON</td>
<td>-.0021756</td>
<td>-0.02+</td>
</tr>
<tr>
<td>DEFTOTCLE</td>
<td>-.0110554</td>
<td>-0.21+</td>
</tr>
<tr>
<td>DEFBLO</td>
<td>-.5488428</td>
<td>-2.23**</td>
</tr>
<tr>
<td>DEFINT</td>
<td>-.1640796</td>
<td>-1.63+</td>
</tr>
<tr>
<td>DEFYC</td>
<td>-1.175087</td>
<td>-2.51**</td>
</tr>
<tr>
<td>DEFRC</td>
<td>5.411761</td>
<td>3.55***</td>
</tr>
<tr>
<td>DEFVDUM</td>
<td>.1598586</td>
<td>0.48+</td>
</tr>
<tr>
<td>SEASON 1</td>
<td>-.6099842</td>
<td>-1.16+</td>
</tr>
<tr>
<td>SEASON 2</td>
<td>-.5661156</td>
<td>-1.15+</td>
</tr>
<tr>
<td>SEASON 3</td>
<td>-.6499534</td>
<td>-1.43+</td>
</tr>
<tr>
<td>ATTGLS</td>
<td>1.212897</td>
<td>3.75***</td>
</tr>
<tr>
<td>ATTSHITSONINCGLS</td>
<td>2.442161</td>
<td>13.76***</td>
</tr>
<tr>
<td>ATTASSTS</td>
<td>1.317004</td>
<td>3.76***</td>
</tr>
<tr>
<td>ATTSCUCLASS</td>
<td>.6333247</td>
<td>30.17***</td>
</tr>
<tr>
<td>ATTYC</td>
<td>1.984845</td>
<td>4.33***</td>
</tr>
<tr>
<td>ATTRC</td>
<td>-11.82249</td>
<td>-3.09***</td>
</tr>
<tr>
<td>ATTVDUM</td>
<td>.6334825</td>
<td>1.89*</td>
</tr>
<tr>
<td>ATTAGE</td>
<td>.2610728</td>
<td>4.69***</td>
</tr>
<tr>
<td>ATTHEIGHT</td>
<td>30.68665</td>
<td>10.61***</td>
</tr>
<tr>
<td>ATTRFOOT</td>
<td>.4896025</td>
<td>1.23+</td>
</tr>
<tr>
<td>ATTLFOOT</td>
<td>-.609086</td>
<td>-1.20+</td>
</tr>
<tr>
<td>ATTPREVLAPP</td>
<td>-.0176847</td>
<td>-5.26***</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
Table 6.4.2 above reveals the results obtained from the first stage (reduced form) regression where the time played by the attacking unit is the dependent variable. All four attacking performance variables, goals, shots on target including goals, assists and total successful passes are statistically significant and positively correlated with the performance of attacking players, supporting the claims of Carmichael et al (2000; 2001), Hughes and Franks (2005), Lago and Martin (2007), Lago-Penas et al (2010), McHale et al (2012).

Table 6.4.2 also reveals both variables ATTRFOOT and ATTLFOOT, representing an attackers preferred foot, are insignificant. This is in complete contrast to the results presented in table 6.4.1 for the instrumental variables DEFLFOOT and DEFRFOOT, which were both revealed to have a significant positive impact on a defenders performance. These results in conjunction therefore suggest that a player’s preferred foot is more important for defensive play rather than attacking play.

Furthermore, table 6.4.2 also reveals that the variable ATTAGE, representing an attackers age (measuring their overall playing experience) is highly significant and has a positive impact upon their performance. Previous Bundesliga experience, captured by the variable ATTPREVLAPP is also revealed to be highly significant but has a negative impact upon the time played by attackers. The negative sign for ATTPREVLAPP may be partly due to attacking players seeking transfers to more lucrative leagues across Europe, after an initial period of success in Germany (Frick 2007). Overall these results suggest that playing experience (whether abroad or in the Bundesliga) has more of an influence on attacking play rather than defensive play, as neither DEFAGE nor DEFPREVLAPP were revealed to be of any significance in table 6.4.1.

Another notable difference between the results revealed in tables 6.4.1 and 6.4.2 is that yellow cards acquired by attackers (ATTYC) has a significantly positive impact upon the performance of attackers. This result contrasts to the results presented in table 6.4.1 where the number of yellow cards picked up by defenders was insignificant for the performance of defenders. Attackers being awarded yellow cards early in matches could explain this result.

Red cards obtained by attackers (ATTRC) were revealed to be negatively correlated with attacking performance and highly significant. This result was expected as the
attacking player is removed from the field of play, thereby preventing them from amassing any additional minutes played. Furthermore, the variable DEFRC also had a highly significant but positive impact upon the time played by their opposing attackers. This result is also to be expected as following a red card the probability of the non-affected team scoring a goal increases (Vecer et al 2009). The variable, DEFYC was revealed to have a significant and negative impact upon their opposing attackers performance. This result could be partially explained by defenders deliberately attempting to unsettle and frustrate their opposing attackers. Carmichael et al (2000) state that the “professional foul” (which often results in a yellow card) is now part of a defenders repertoire, however they also stress that a very thin line exists between aggressive, but legal defending and overly aggressive, illegal defending.

The number of blocks made by a defender is also revealed to have a negative impact upon their opposing attackers performance. This result is not surprising, as a defender getting the better of their opposing attacker would likely increase the chances of the attacker being substituted.

In contrast to the results presented in table 6.4.1, the instrumental variable ATTHEIGHT, measuring the height of attackers is statistically significant at a 1% significance level and is positively signed. The natural interpretation of this result therefore is that in German Bundesliga football, a player’s height has more bearing on the performance of an attacker rather than the performance of a defender (as DEFHEIGHT in table 6.4.1 was insignificant). Attacking players and in particular strikers are required to be physically stronger than players in other positions as they have been revealed to perform the most physical contact at high intensity (Bloomfield et al 2005). Furthermore, Bloomfield et al (2005) revealed that FA Premier League strikers and defenders were generally heavier and slightly taller than midfielders, suggesting a player’s physical characteristics such as their height and weight are connected to the demands of specific playing positions.

Finally, ATTVDUM controlled for the venue at which attackers were playing (0= Home, 1= Away) was revealed to be significant at the 10% level and positively signed. Attackers therefore performed better in away fixtures. This result corresponds with the results presented in table 6.4.1 where defenders were revealed to perform better when playing at home.
Table 6.4.3 below presents several statistics that clarify the explanatory power of the instrumental variables used throughout the analyses. Results obtained for default standard errors (where independent identical errors are assumed) as well as robust standard errors are reported. As can be seen the $R^2$, adjusted $R^2$ and the partial $R^2$ statistics are relatively low however this is no cause for concern as Wooldridge (2009) claims $R^2$ statistics can on occasion be misleading and therefore have no natural interpretation in a 2SLS procedure.

The reported F statistics in table 6.4.3 below reveal the joint significance of the coefficients of the instrumental variables used in the first stage regressions and Stock, Wright and Yogo (2002) state that this F statistic should be greater than 10 for any inferences based on a 2SLS procedure to be reliable when using one endogenous variable. As can be seen, all F statistics are greater than 10 and all are highly significant (p-values), meaning the instrumental variables have a highly significant explanatory power for the endogenous variables after controlling for the effects of other included exogenous variables.

The Cragg and Donald (1993) minimum eigenvalue statistic is another measure of the strength of instrumental variables, particularly when this statistic is used in conjunction with Stock and Yogo (2005) tests for weak instruments. As can be seen from table 6.4.3 below when choosing the largest relative bias of the 2SLS estimator willing to be tolerated the test statistics (eigenvalues) well exceed the critical values. Similarly, when selecting the largest rejection rate of a nominal 5% Wald test willing to be tolerated the test statistics also exceed the critical values. The fact that the eigenvalues for all endogenous variables were greater than the critical values of the two Stock and Yogo (2005) tests for weak instruments allows us to reject the null hypothesis of each of the tests, that the instruments are weak. Based on all the results presented in table 6.4.3 below, it can be concluded that the instruments used throughout the first stage regressions for both analyses are strong and therefore valid instruments.
Table 6.4.3 Explanatory power of Instrumental Variables

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>ATTTP</th>
<th>DEFTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Squared (Robust)</td>
<td>0.2444</td>
<td>0.1086</td>
</tr>
<tr>
<td>Adjusted R Squared (Robust)</td>
<td>0.2413</td>
<td>0.1050</td>
</tr>
<tr>
<td>Partial R Squared (Robust)</td>
<td>0.2415</td>
<td>0.1069</td>
</tr>
<tr>
<td>F Statistic (Robust)</td>
<td>96.5505</td>
<td>27.0361</td>
</tr>
<tr>
<td>F Statistic p-value</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Minimum Eigenvalue Statistic</td>
<td>142.992</td>
<td>53.7471</td>
</tr>
</tbody>
</table>

2SLS Relative Bias

<table>
<thead>
<tr>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.01</td>
<td>11.52</td>
<td>6.53</td>
<td>4.75</td>
<td>21.01</td>
<td>11.52</td>
<td>6.53</td>
<td>4.75</td>
</tr>
</tbody>
</table>

2SLS Size of Nominal 5% Wald Test

<table>
<thead>
<tr>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>43.27</td>
<td>23.24</td>
<td>16.35</td>
<td>12.82</td>
<td>43.27</td>
<td>23.24</td>
<td>16.35</td>
<td>12.82</td>
</tr>
</tbody>
</table>
6.5. Results: Second Stage Results

Table 6.5.1 below reveals the results obtained from the second stage regression equations. On the left hand side of table 6.5.1, the performance of the attacking unit is the dependent variable. On the right hand side of table 6.5.1, the performance of the defensive unit is the dependent variable.
<table>
<thead>
<tr>
<th>Endogenous Variables</th>
<th>Dependent Variable</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ATTTP</td>
<td>DEFTP</td>
</tr>
<tr>
<td>DEFTP</td>
<td>-.1341369</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-2.93***</td>
<td>-</td>
</tr>
<tr>
<td>ATTTTP</td>
<td>-</td>
<td>.0007714</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.04+</td>
</tr>
</tbody>
</table>

| Independent Variables | ATTGLS            | 1.494087          |
|                       | 4.58***           | -                 |
|                       | ATTSHTSONINCGLS   | 2.644781          |
|                       | 14.75***          | -                 |
|                       | ATTASSTS          | 1.220275          |
|                       | 3.42***           | -                 |
|                       | ATTSUCPASS        | .5681252          |
|                       | 28.14***          | -                 |
|                       | ATTYC             | 2.158191          |
|                       | 4.72***           | -                 |
|                       | ATTRC             | -11.87103         |
|                       | -3.12***          | -                 |
|                       | ATTVDUM           | .6671214          |
|                       | 1.95*             | -                 |
| DEFTACKWON            | -                 | .9119902          |
|                       | 11.03***          | -                 |
| DEFTOTCLE             | -                 | .5690065          |
|                       | 12.33***          | -                 |
| DEFBLO                | -                 | 1.040114          |
|                       | 7.64***           | -                 |
| DEFINF                | -                 | .7670349          |
|                       | 11.13***          | -                 |
| DEFYC                 | -                 | .3083748          |
|                       | 0.84+             | -                 |
| DEFRC                 | -                 | -22.75028         |
|                       | -6.97***          | -                 |
| DEFVDUM               | -                 | -.5566479         |
|                       | -1.92*            | -                 |
| SEASON 1              | -.5724373         |
|                       | -1.16+            |
|                       | -1.615436         |
|                       | -3.66***          |
| SEASON 2              | -.5906475         |
|                       | -1.24+            |
|                       | -1.508036         |
|                       | -3.38***          |
| SEASON 3              | -.6494344         |
|                       | -1.42+            |
|                       | .1853246          |
|                       | 0.49+             |

Notes: + not significant, *p<.10, **p<.05, ***p<.01
As can be seen from table 6.5.1 above the defenders performances had a highly significantly and negative impact upon the performance of their opposing attackers. This result can be interpreted therefore as the better the performance of defenders the weaker the performance their opposing attackers. Table 6.5.1 also reveals all four attacking performance variables are significantly correlated to the performance of attackers at the 1% significance level with the expected positively signed coefficients. The variable ATTGLS, which is a measure of goals scored, is statistically significant and has a positive impact upon the productivity adjusted minutes played by the attacking players. The variable ATTSHTSINCGLS, is a dual measure of shots on target and goals scored and is therefore a measure of both attacking accuracy and finishing ability and is also revealed to be highly significant and positively signed. In addition, the variables ATTASSTS and ATTSUCPASS, which are measures of assists made and total successful passes played are also revealed to be highly significant and positively signed.

As expected, the results for the performance variables are highly significant and all positively signed as previous research reveals these attacking actions are of great importance to individual and team performances as well as player remuneration. Goals scored increase likelihood of a transfer (Carmichael et al 1999) suggesting it is a highly sought after attribute. In addition, Frick and Deutscher (2009) state that goals scored in the last season have a positive linear influence on annual income suggesting it is the most recent performances that matter most. Goals scored also contribute most to the ‘superstar effect’ along with the action of assists, both of which are potentially match winning actions (Lucifora and Simmons 2003).

Effective and accurate shooting have also been shown to be positively correlated to team performance (Carmichael et al 2000) and a high shot to goal conversion rate is also positively related to success (Hughes and Churchill 2005; Carmichael and Thomas 2008). Lago- Penas et al (2010) concluded total shots and shots on goal were two of six play actions that discriminate between winning, drawing and losing a game. Shooting accuracy, referring to shots on and off target was also deemed a relevant variable in explaining the points gained by teams in the group stage of the 2006 World Cup (Lago 2007). Accurate and effective shooting therefore appears to be a pre-requisite for success and the results presented in table 6.5.1 support this notion.
The performance variables ATTASSTS and ATTSUCPASS were both positively signed and statistically significant at the 1% significance level. These two performance variables are measures of possession and the effectiveness of this possession in the form of chance creation. Maintaining possession of the ball and effectively making use of this possession has been repeatedly shown to have a positive effect on performance (Carmichael et al 2000, 2001; Jones et al 2004; Hughes and Churchill 2005; Lago-Penas et al 2010; Vogelbein et al 2014).

The variable ATTYC, the measure for yellow cards had a significantly positive impact on the minutes played by attackers. This result is surprising and unexpected however may be simply explained by yellow cards being picked up earlier on in games by attacking players. Red cards obtained by attackers are negatively correlated and highly significant to their productivity adjusted minutes played as expected. Table 6.5.1 also reveals that the variable ATTVDUM (Home =0, Away =1) that accounts for the venue of the match being played is positively signed and significant at the 10% significance level. This result therefore suggests attackers performed better in away fixtures throughout the seasons 2007-08 and 2010-2011.

As can be seen from table 6.5.1 the productivity adjusted minutes played by the attacking players is insignificant. This result implies that the performances of attackers have no significant bearing on the performances of their opposing defenders. This result contrasts to the results obtained in the analysis investigating the impact of defenders performances upon attacking performance where it was discovered that the performance of defenders have a significantly negative impact on the performance of attackers.

All four defensive performance variables presented in table 6.5.1 are positively signed and all significant at the 1% significance level. As expected, the performance of a defender, measured by their productivity adjusted minutes played is positively correlated to the defensive actions of, winning tackles, making clearances, blocking and intercepting.

The dummy variable for the match location DEFVDUM, is significant at the 10% significance level and is negatively signed. This result implies that defending units perform worse when playing away from home (and subsequently better at home). To an extent, this result could be explained by the supposed advantages of playing at
home (and subsequent disadvantages of playing away) such as the effect of the home crowd (Nevill et al 1996; Carmichael and Thomas 2005; Pollard 2008), referee bias towards the home team (Sutter and Kocher 2004; Dawson et al 2007; Pollard 2008), fatigue due to travelling away (Neave and Wolfson 2003) and increased effort levels from home teams (Leard and Doyle 2010).

Finally, the SEASON1 and SEASON2 dummy variables were highly significant (1% significance level) and negatively correlated to the performances of defenders. These results imply that defenders performed worse in the 2007-08 and 2008-09 seasons in comparison to the 2010-2011 reference season.

Table 6.5.2 presented below reveals the results obtained for the Sargan \( \chi^2 \), Basmann \( \chi^2 \) and the Score \( \chi^2 \) tests of over-identifying restrictions. In a model where there are more instruments than the number of endogenous regressors, the model is deemed to be over-identified and in such cases tests for over-identifying restrictions can be run to test whether the instruments are correlated to the error term but also test to see if the equation is correctly specified. In other words both tests will also check if any of the exogenous variables should be included in the structural equation, therefore a significant test statistic would represent either an invalid instrument or an incorrectly specified structural equation. As can be seen from table 6.5.2 below each of the tests for over-identifying restrictions reveal test statistics much higher than 0.05 therefore the null hypothesis that the instruments are valid, cannot be rejected at the 5% significance level. All instrumental variables used throughout the analyses conducted in this chapter are therefore valid and the equation has not been incorrectly specified.
Throughout the analyses it was assumed that player performance is jointly determined with the performance of their direct opponent. The productivity adjusted minutes played by attackers and the productivity adjusted minutes played by defenders were therefore both treated as endogenous variables. In order to confirm endogeneity the Durbin $\chi^2$, Wu-Hausmann $\chi^2$ and Wooldridge’s (1995) Score tests were conducted which would confirm whether these variables were in fact endogenous and not exogenous. If the endogenous regressor did turn out to be exogenous then it would have meant OLS estimations would be a more appropriate method of testing instead.

Table 6.5.3 below confirms simultaneity between the two variables, ATTTP and DEFTP when ATTTP is the dependent variable and DEFTP is the endogenous variable. The test statistics revealed for this analysis are highly significant meaning the null hypothesis of the Durbin, Wu-Hausman and Score tests of the variable being exogenous can be rejected. In this case the variables are correctly treated as endogenous.

Table 6.5.3 also reveals the lack of simultaneity between the variables DEFTP and ATTTP when DEFTP is the dependent variable and ATTTP is the endogenous

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>ATTTP</th>
<th>DEFTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endogenous Variable</td>
<td>DEFTP</td>
<td>ATTTP</td>
</tr>
<tr>
<td>Sargan $\chi^2$ Test Score</td>
<td>10.0739</td>
<td>16.4636</td>
</tr>
<tr>
<td>Sargan $\chi^2$ (p-value)</td>
<td>0.5237</td>
<td>0.1248</td>
</tr>
<tr>
<td>Basmann $\chi^2$</td>
<td>10.0498</td>
<td>16.4437</td>
</tr>
<tr>
<td>Basmann $\chi^2$ (p-value)</td>
<td>0.5259</td>
<td>0.1254</td>
</tr>
<tr>
<td>Score $\chi^2$</td>
<td>14.5602</td>
<td>16.8425</td>
</tr>
<tr>
<td>Score $\chi^2$ (p-value)</td>
<td>0.2035</td>
<td>0.1126</td>
</tr>
</tbody>
</table>
variable. The test statistics revealed for this analysis are insignificant meaning the null hypothesis of the Durbin, Wu-Hausman and Score tests of the variable being exogenous cannot be rejected. In this case the variables are exogenous.

<table>
<thead>
<tr>
<th>Table 6.5.3. Tests of endogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Endogenous Variable</td>
</tr>
<tr>
<td>Durbin $\chi^2$ Test Score</td>
</tr>
<tr>
<td>Durbin $\chi^2$ (p-value)</td>
</tr>
<tr>
<td>Wu-Hausman $\chi^2$ Test Score</td>
</tr>
<tr>
<td>Wu-Hausman F (p-value)</td>
</tr>
<tr>
<td>Robust Score $\chi^2$</td>
</tr>
<tr>
<td>Robust Score (p-value)</td>
</tr>
<tr>
<td>Robust Regression F</td>
</tr>
<tr>
<td>Robust Regression F (p-value)</td>
</tr>
</tbody>
</table>
6.6 Discussion

The results obtained from the second stage regression estimations confirm the interdependence between the performances of attackers and defenders in direct opposition. Table 6.5.1 reported that the performances of defenders in the Bundesliga had a negative and significant impact upon the performances of their directly opposing attackers, therefore confirming contemporaneous correlation between the performances of players at an isolated individual player level. Table 6.5.1 also reported that the performances of attackers had no bearing on the performances of their opposing defenders. The results obtained throughout the analyses in this chapter support the findings of Tiedemann et al (2010) who discovered defenders to be the most efficient performers in their study on the German Bundesliga.

Table 6.5.1 revealed that the variable DEFVDUM was negatively correlated to the performances of defenders, meaning defenders performed worse when playing away. One possible explanation for this result could be that teams often sacrifice defenders for attackers or midfielders late on in a match in order to try and win the game and acquire all three points. Similarly, defenders may also be sacrificed in cases when a team is chasing the game (behind in terms of scoreline) in an attempt to try and equalise and acquire a solitary point in reward for a drawn game. Del Corral et al (2008) revealed in a study of the Spanish Primera Division that attacking substitutions are made before defensive substitutions particularly in cases when a win has not been secured. It is logical then that an away team generally under less pressure than the home side (Del Corral et al 2008) would be more open to taking risks and therefore more willing to sacrifice defenders in order to gain a win and the maximum three points on offer.

The result for the variable DEFVDUM coincides with the results obtained for the variable ATTVDUM, which is positively correlated to the performances of attackers, meaning attackers perform better away from home. Both of these results in conjunction therefore emphasise the notion that team managers play, or at least attempt to play more expansive, attacking football in away fixtures. This result is reinforced somewhat by the findings of Jacklin (2005) and Koyama and Reade (2009) who claim away sides have greater incentive to attack due to rule changes in the game.
and greater TV coverage. Bosca et al (2009) also claim the most successful teams are those who are efficient in attack in both home and away matches. On the basis of both results obtained for the dummy variables DEFVDUM and ATTVDUM, it appears that German Bundesliga teams take more risks in away games as attackers accumulate significantly more productivity adjusted minutes when playing away.

The results presented in table 6.5.1 also reveal yellow cards (ATTYC) are positively correlated to attacking performance. This may seem surprising, however could be simply explained by attackers picking up yellow cards very early on in games. Although yellow cards can be awarded for a variety of offences, their awarding is generally associated with defensive actions like committing hard fouls (Garciano and Palacios-Huerta 2005) and pulling or holding opponents (Dawson et al 2007), hence Carmichael and Thomas (2005) used the number of yellow cards awarded as a measure of defensive weakness. In addition, Garciano and Palacios-Huerta (2005) claim defenders commit two out of every three fouls that lead to a yellow card. On this basis it would be safe to assume attackers are more likely to pick up yellow cards earlier on in games than defenders maybe due to them being less skilled in defensive actions for which yellow cards are more commonly awarded (Garciano and Palacios-Huerta 2005).

Alternatively, this result may be partially explained by the impact of match status upon managerial decisions. In cases when teams fall behind, recovery times for regaining possession decrease since trailing teams are in greater need to promptly regain possession of the ball, as this is a basic requirement for scoring goals (Vogelbein et al 2014). As scoring goals is a performance characteristic attributed to attacking players, a manager may on occasion be forced to leave an attacking player on the pitch irrelevant of whether or not the player has picked up a yellow card. On occasions when a team is trailing the positive match winning action of scoring a goal or providing an assist may outweigh the potential negative effect of gaining a second yellow card, accumulating to a red and hence being sent off.

The control variables for defender red cards (DEFRC) and attacker red cards (ATTRC) were both revealed to be negatively correlated with the performance of defenders and attackers. This would be expected as players are removed from the field of play and hence cannot accumulate any more minutes. The coefficients for
DEFRC and ATTRC however imply that red cards have a greater negative impact upon the performances of defenders. One possible explanation for this could be that defenders impose themselves on attackers in an aggressive manner in order to disrupt their opponent’s rhythm. Carmichael et al (2000) state that a “professional foul” is part of a successful defenders repertoire and conclude that there is a thin line between overly aggressive defending and the employment of dubious and illegal, but very effective ‘tactics’ that disturb the opponent’s pattern of play. Defenders could therefore have accumulated more yellow and red cards throughout the sample period explaining the difference in coefficient value.

The control variables SEASON1 and SEASON2 were also revealed to be negatively correlated to the performance of defenders. This result therefore implies that defenders performed worse in seasons 2007-08 and 2008-09 in comparison to the reference season 2010-11. Finally all eight performance measures (four for defenders and four for attackers) were all highly significant and positively signed. Successfully executed defensive performance actions such as tackling, making clearances, blocking and intercepting significantly impact defenders performances supporting the findings of (Lago and Martín 2007; Dellal et al 2011; McHale et al 2012; Redwood Brown et al 2012). Similarly, successfully executing attacking performance actions such as scoring goals, accurately shooting, creating chances and maintaining possession significantly impact attacking performance in support of (Carmichael et al 2000, 2001; Hughes and Franks 2005; Lago and Martín 2007; Lago- Penas et al 2010; McHale et al 2012).

It is also important to note the contribution of the style of play adopted by teams competing in the German Bundesliga to the results obtained in table 6.5.1. Vogelbein et al (2014) analysed teams competing in the German Bundesliga in season 2010-2011 and discovered that the most successful teams were those that recovered possession quickest. According to Vogelbein et al (2014) recovering possession of the ball as quickly as possible after losing possession is an important determinant of successful defensive performance. Bosca et al (2009) also revealed teams with the best attack had the best defence in the Italian league, giving credence to the saying “attack is the best form of defence”.
Carmichael et al (2001) also stress the importance of maintaining possession of the ball due to its dual purpose of being a source of attacks, whilst also preventing the opposition from creating any chances or threats on goal. This high pressing, ball-retaining style of play adopted in the Bundesliga therefore, minimises the oppositions chances of scoring (Vogelbein et al 2014). Logically, teams that are adopting this style of play, should expect their defenders to outperform their opposing attackers as these teams intend to dominate possession, thereby giving their defenders less defensive work to do.

Possible reasons for the difference in the results obtained could also be explained by unobserved factors such as team tactics and match status. In order to understand how these unobserved factors could have effected the results obtained, it is important to first outline the relationship between the performance criteria upon which defenders and attackers were assessed and the concept of possession.

As detailed in section 6.3 four performance variables were selected for each position throughout this analysis on the basis of their recurrence and evident importance in previous research. It is important however, to consider the affiliation of these performance variables with the concept of possession. In much previous research on performance analysis in football, possession of the ball has repeatedly been shown to be a major contributor towards overall success. Generally, it is accepted that those teams able to dominate possession are the ones that tend to be the most successful (James et al 2004; Hughes and Churchill 2005; Lago 2009; Lago-Penas et al 2010; Vogelbein et al 2014).

Furthermore, the importance of possession and in particular, regaining possession of the ball once it has been lost has also been examined and it has been shown that the most successful teams are those that record the lowest average recovery times (Vogelbein et al 2014). Similarly, many of the same studies conclude that match status impacts the level of possession a team has. In particular when a team is losing, they will have more possession (James et al 2004; Jones et al 2004; Lago 2009; Lago and Martín 2007;) and show greater urgency in recovering possession once it has been lost (Vogelbein et al 2014), as they attempt to salvage some reward from the match.

Following on from this, it becomes clear how much of an effect levels of possession can impact player performance. For defenders to perform important defensive actions...
such as making tackles, blocks, clearances and interceptions they require an opponent to be in possession of the ball. In direct contrast, when attackers perform important offensive actions such as score goals, record shots on target, assist and successfully complete passes the attacking players are themselves in possession of the ball.

There is therefore a distinct difference in the recording of successful play actions between the two different positions, as attackers are required to be in possession of the ball in order to successfully execute their play actions whereas defenders require an opponent to be in possession of the ball in order to successfully execute their play actions. One interpretation of this could be that when an observed team is in possession, their defenders become irrelevant and the emphasis is on the attackers to perform. It is only when an observed team loses possession that their defensive unit is called into action. Logically then, a team that attempts to or sets out to dominate possession makes their attacking players ‘work’ more than their defensive players.

The fact that the interdependency between defenders performances and attackers performances was only statistically significant in one of the two analyses (when attacking performance was the dependent variable and defenders performances endogenous) suggests that teams competing in the Bundesliga, welcome attacking success slightly more than defending success. This may be partly explained by the huge financial significance of winning games and hence competitions (Dobson and Goddard 2001; Ashton et al 2010). This inference is safe to make bearing in mind substitute player data (players coming on) was deliberately excluded from the analysis and the sample size consisted of an equal number of 5,411 attackers and 5,411 defenders. Consequently, as sample sizes were the same, a portion of the difference in the productivity adjusted minutes played by players from the two positional units must be explained by substitutions that occurred (players coming off). The fact that time played by defenders had a negative impact upon the time played by their opposing attackers suggests attacking substitutions are made earlier on in games than defensive ones supporting the findings of Del Corral et al (2008).
6.7 Conclusion

The purpose of the analyses conducted throughout this chapter was to investigate the interdependency between the performances of defenders/attackers and their opposing attackers/defenders. This chapter expanded on the previous chapter by investigating simultaneity at a more isolated individual player level by specifically analysing players performing in defensive and attacking positions.

Players playing in the positions of attack and defence in the German Bundesliga throughout the four season sample period were carefully matched up to opposing players (explained earlier in section 6.2). This match up was done in order to replicate the one on one environment sports stars find themselves in the North American sport of baseball, where player performance is more accurately measured (Scully 1974; Zech 1981; Macdonald 1994; Krautmann and Oppenheimer 2002; Bradbury 2007).

As the results have confirmed, the performances of defenders are statistically significant and negatively correlated to the performances of their opposing attackers, however the performances of attackers have no bearing on the performances of their opposing defenders. It can be concluded therefore that assuming productivity adjusted minutes played encapsulates performance, the performance of a defender and their opposing attacker are interdependent. The results confirm that the performances of defenders and attackers at an isolated player level are jointly determined, supporting the sports economic theory of joint production.

It is important to recognise that part of this result may be explained by unobserved factors such as style of play. For example, managers may be more risk averse when it comes to unsettling their defence, but may be more open to altering their attack as making attacking substitutions earlier in games (Del Corral et al 2008) suggests. This readiness to change the attack rather than the defence may be driven by the greater rewards associated with winning matches (Dobson and Goddard 2001; Ashton et al 2010).

Reflection should also be made on the role of possession and the effect this had upon results obtained. Possession of the ball is a vital component of team success (James et al 2004; Hughes and Churchill 2005; Lago 2009; Lago-Penas et al 2010; Vogelbein et al 2014) and levels of this possession are determined by the style of play in which a
team plays (Tenga et al 2010) and match status (James et al 2004; Jones et al 2004; Lago and Martín 2007; Lago 2009). In turn, possession of the ball and duration levels of this possession will undoubtedly affect player performance as depending upon possession, player’s will either become more able to perform play actions, (successfully or unsuccessfully) or less able to perform play actions.

The style of play of German Bundesliga teams also appears to be more adventurous when playing away as implied by the results obtained for the venue dummy variables. The results presented in table 6.5.1 reporting the estimations from the second stage results of the 2SLS procedure confirmed that the productivity adjusted minutes played by attackers increased with away fixtures, whereas the productivity adjusted minutes played by defenders decreased with away fixtures. Inferences made from this result leads to the suggestion that defenders are more readily substituted in away games, rather than home games most probably to be replaced by more attacking players in cases when a win has not been secured. Teams competing in the German Bundesliga appear to be more risk averse when playing at home rather than away which could be partly explained by the increased pressure home sides are under to perform (Del Corral et al 2008).
Chapter 7. Investigating Interdependency Between the Performances of Positional Units and their Opposing Team(s) in the German Bundesliga

7.1 Introduction

This chapter separately investigates simultaneity between the performances of defending units and then attacking units and the collective performances of their opposing team. Section 7.2 below provides a detailed description of the sample and the data utilised throughout the analyses conducted in this chapter. Section 7.3 presents the analysis strategy and model specifications for all analyses conducted throughout this chapter. Section 7.4 presents the set of first stage regression results obtained revealing which variables affected the performances of the attacking positional units and defensive positional units. Section 7.5 presents the second stage regression results obtained for the analyses revealing the impact of the performances of attacking and defensive positional units upon their opponents’ collective team performances. Section 7.6 presents the results obtained from additional OLS estimations revealing the impact the performances of defensive and attacking positional units had upon the collective team performance of their opponent. Section 7.7 offers a discussion providing possible explanations for the results obtained and discusses their ramifications.

Section 7.8 offers a conclusion summarising those results of most importance and contains some generalised implications. It is concluded that a successful German Bundesliga team must perform well in both defence and attack as this has a positive impact upon the observed team’s performance and an adverse impact upon the performance of their opponent. No evidence is found to suggest that the performances of positional units (defensive and attacking) and their opposing teams are interdependent. The performances of positional units are however revealed to have a
positive (linear) impact upon their own (observed) team’s performance and a negative impact upon the performance of their opposing team.
7.2. Data and Variables

The dependent variables throughout the analyses conducted in this chapter were the result obtained by the home team, HTRESULT, and the result obtained by the away team, ATRESULT, expressed as the number of points gained in the match. In the games of football covered in this analysis the reward for a win is 3 points, with 1 point for a draw and 0 points for a loss (Carmichael et al 2001). The reason for the choice of these as the dependent variables has been due to their common use as the measure of output in previous studies in European football (Carmichael et al 2000). Match results are presented as a function of the performances of specific positional units and several other independent variables that account for time and team effects.

The performances of positional units were measured in a similar manner to chapter 6, in which it was argued that the performance of a player could be captured by their minutes played. The performance of an entire unit of players’ (defenders and attackers) was therefore measured by summing the minutes played by each individual player within a specific positional unit in order to provide positional unit-match total minutes played.

Chapter 6 indicated that the minutes played by individual players performing in defence and attack were strongly correlated to individual player performance variables specific to each position and hence provided an indication of how well a player performed in relation to his opponent. Following on from this, if all representatives of a positional unit were included in the sample irrespective of whether they faced a direct opponent, the measure of the unit’s performance would more likely depend upon the number of players’ within the unit and therefore be a misleading indication of the unit’s performance. It was for this reason that only those players that could be theoretically said to be in direct competition with only one opponent were included as part of the positional unit observations.

A total of 1,224 fixtures were available across four seasons of data, however 34 of these fixtures were excluded due to the inability to match any defenders to any opposing attackers and vice versa. A further fixture was also excluded due to only partial data being available due to the abandonment of the match (FC St Pauli vs. FC Schalke 04, 1st April 2011). Matching up a defender to attackers was therefore
permitted in each of the remaining 1,189 fixtures and this matching up of players provided a total of 10,822 player observations, 5,411 for defenders and 5,411 for their opposing attackers.

As the dependent variables were the match results achieved by the home and away teams the 5,411 player observations for defenders and the 5,411 observations for attackers needed to be split between home and away fixtures. This led to players being organised into four groups, away attackers, away defenders, home attackers and home defenders. Consequently, the final sample consisted of 2,703 player observations for representatives of the home team defensive unit and therefore 2,703 player observations for the representatives of the away team attacking units (with whom they were directly matched too). Similarly, there were 2,708 player observations for representatives of the home team attacking unit and therefore 2,708 player observations for representatives of the away team defensive units (totalling 10,822 observations).

In order to balance the data set and allow the performance of each unit to be matched to the corresponding match result (dependent variable), match identification numbers were used to separate each of the 1,189 fixtures. Further segregation of these individual match fixtures then included incorporating all participating players from each of the four groups formed (away attackers, away defenders, home attackers and home defenders) and then aggregating this player data in order to acquire one-unit observation, per variable, per match (positional unit-match totals).

Minutes played by all players in each of the four groups were summed in order to acquire the total number of minutes played by the matched unit of players that started the fixture. Other variables such as the instrumental variables measuring the players preferred foot and the number of previous appearances made by the players representing each of the four units were also summed. The instrumental variables measuring the age of players and the height of players in each of the four units were included as average measurements. Table 7.2.1 below presents a list of all variables used throughout the analysis and their definitions.

The matching up of players (to opponents) was only possible when two competing teams played formations that theoretically allowed a direct opponent to be identified. Some games therefore, only allowed select members of the positional units of either
side to be matched and these match ups were not always equal between the home and away teams. This is better illustrated in figures 7.2.1, 7.2.2 and 7.2.3 below. As can be seen in figure 7.2.1 representing fixture n, a home defender (HD2) is matched to an away attacker (AA1) and this represents the only possible match up allowed by the two team formations in this fixture. Figure 7.2.2 represents fixture n+1 and here it is possible to match up four members of the away defensive unit, AD1, AD2, AD3 and AD4 with four members of the home attacking unit HA1, HA2, HM1 and HM4. In figure 7.2.2 it is also possible to match two defenders of the home defending unit HD1 and HD4 to two members of the away attacking unit AM1 and AM5. Finally, figure 7.2.3 below represents fixture n+2 where two home defenders HD1 and HD4 are matched up to two away attackers AM1 and AM4 and in the same fixture two away defenders AD1 and AD4 are matched up to two home attackers HM1 and HM4.

Figure 7.2.1 – Positional unit match-ups in fixture n
Figure 7.2.2 – Positional unit match-ups in fixture n+1
Figures 7.2.1, 7.2.2 and 7.2.3 presented above represent only three fixtures and consist of 5 observations for home defenders and 5 observations for away attackers with who they are in direct competition. These same three fixtures also consist of a total of 6 observations for away defenders and consequently 6 observations for home attackers with who they are in direct competition. The segregation of data into positional units and home and away venues therefore presented the problem of an unequal number of player match ups across the same number of fixtures. Ultimately, this meant that over the 1,189 fixtures included in the sample, the number of observations for the two positional units home defenders and away attackers would be equal to each other however, would be unequal to the number of observations for the remaining two positional units, away defenders and home attackers. Consequently, two pooled cross sectional data sets were constructed. Data set 1 allowed the investigation of the impact of home defenders on the away team result and then the impact of away attackers on the home team result (as these two positional units are in
direct competition). Data set 2 allowed the investigation of the impact of the away defenders on the home team result and then the impact of home attackers on the away team result (as these two positional units are also in direct competition). Data set 1 consisted of 1,091 positional unit-match observations. Data set 2 consisted of 1,113 positional unit-match observations.

Once all the individual player data was aggregated and converted into positional unit-match observations, corresponding match data was also included in both data set 1 and data set 2. Data set 1 therefore consisted of 1,091 match observations and data set 2 consisted of 1,113 match observations. As there are a total of 2,204 match observations (1113+ 1091) but only 1,189 fixtures, 215 fixtures (2204-1189) appeared in both data sets. This was simply because the team formations used in each of these 215 fixtures consisted of players from all four positional units being matched to a direct opponent akin to figure 7.2.2 presented above.

Match data incorporated into data sets 1 and 2 included the match result, goal differences at time t-1 (last gameweek) of both participating teams in the fixture and the percentage points at time t-1 (last gameweek) of both participating teams in each of the 1,189 fixtures. Three dummy variables controlling for the general quality of the league in each season were also incorporated into data set 1 and data set 2.
Table 7.2.1. List of all variables used throughout the analysis and their definitions

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variables</strong></td>
<td></td>
</tr>
<tr>
<td>HTRESULT</td>
<td>Home Team Result</td>
</tr>
<tr>
<td>ATRESULT</td>
<td>Away Team Result</td>
</tr>
<tr>
<td><strong>Endogenous Variables</strong></td>
<td></td>
</tr>
<tr>
<td>TPHDEF</td>
<td>Aggregated Time Played By The Home Defensive Unit</td>
</tr>
<tr>
<td>TPADEF</td>
<td>Aggregated Time Played By The Away Defensive Unit</td>
</tr>
<tr>
<td>TPHATT</td>
<td>Aggregated Time Played By The Home Attacking Unit</td>
</tr>
<tr>
<td>TPAATT</td>
<td>Aggregated Time Played By The Away Attacking Unit</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
</tr>
<tr>
<td>SEASON1</td>
<td>Season 1 (2007-08) Dummy Variable</td>
</tr>
<tr>
<td>SEASON2</td>
<td>Season 2 (2008-09) Dummy Variable</td>
</tr>
<tr>
<td>SEASON3</td>
<td>Season 3 (2009-10) Dummy Variable</td>
</tr>
<tr>
<td>GDHOME</td>
<td>Goal Difference Home Team (t-1)</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>Goal Difference Away Team (t-1)</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>Percentage Points Home Team (t-1)</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>Percentage Points Away Team (t-1)</td>
</tr>
<tr>
<td><strong>Instrumental Variables</strong></td>
<td></td>
</tr>
</tbody>
</table>

238
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDAGE</td>
<td>Average Age of the Home Defensive Unit</td>
</tr>
<tr>
<td>HDHEIGHT</td>
<td>Average Height of the Home Defensive Unit</td>
</tr>
<tr>
<td>HDRIGHTFOOT</td>
<td>The Total Number of Right Footed Players in the Home Defensive Unit</td>
</tr>
<tr>
<td>HDLEFTFOOT</td>
<td>The Total Number of Left Footed Players in the Home Defensive Unit</td>
</tr>
<tr>
<td>HDBOTHFEET</td>
<td>The Total Number of Dual Footed Players in the Home Defensive Unit</td>
</tr>
<tr>
<td>HDPREVLAPP</td>
<td>The Total Number of Previous League Appearances Made By the Home Defensive Unit</td>
</tr>
<tr>
<td>ADAGE</td>
<td>Average Age of the Away Defensive Unit</td>
</tr>
<tr>
<td>ADHEIGHT</td>
<td>Average Height of the Away Defensive Unit</td>
</tr>
<tr>
<td>ADRIGHTFOOT</td>
<td>The Total Number of Right Footed Players in the Away Defensive Unit</td>
</tr>
<tr>
<td>ADLEFTFOOT</td>
<td>The Total Number of Left Footed Players in the Away Defensive Unit</td>
</tr>
<tr>
<td>ADBOTHFEET</td>
<td>The Total Number of Dual Footed Players in the Away Defensive Unit</td>
</tr>
<tr>
<td>ADPREVLAPP</td>
<td>The Total Number of Previous League Appearances Made By the Away Defensive Unit</td>
</tr>
<tr>
<td>HAAGE</td>
<td>Average Age of the Home Attacking Unit</td>
</tr>
<tr>
<td>HAHEIGHT</td>
<td>Average Height of the Home Attacking Unit</td>
</tr>
<tr>
<td>HARIGHTFOOT</td>
<td>The Total Number of Right Footed Players in the Home Attacking Unit</td>
</tr>
<tr>
<td>HALEFTFOOT</td>
<td>The Total Number of Left Footed Players in the Home Attacking Unit</td>
</tr>
<tr>
<td>HABOTHFEET</td>
<td>The Total Number of Dual Footed Players in the Home Attacking Unit</td>
</tr>
<tr>
<td>HAPREVLAPP</td>
<td>The Total Number of Previous League Appearances Made By the Home Attacking Unit</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>AAAGE</td>
<td>Average Age of the Away Attacking Unit</td>
</tr>
<tr>
<td>AAHEIGHT</td>
<td>Average Height of the Away Attacking Unit</td>
</tr>
<tr>
<td>AARIGHTFOOT</td>
<td>The Total Number of Right Footed Players in the Away Attacking Unit</td>
</tr>
<tr>
<td>AALEFTFOOT</td>
<td>The Total Number of Left Footed Players in the Away Attacking Unit</td>
</tr>
<tr>
<td>AABOTHFEET</td>
<td>The Total Number of Dual Footed Players in the Away Attacking Unit</td>
</tr>
<tr>
<td>AAPREVLAPP</td>
<td>The Total Number of Previous League Appearances Made By the Away Attacking Unit</td>
</tr>
</tbody>
</table>
7.3 Analysis Strategy

Throughout the analyses conducted in this chapter it was assumed that the performances of an opposing team’s attacking and defensive positional units and an observed team’s collective performance would be interdependent. In order to investigate simultaneity between performances at this positional unit vs. team level two pooled cross sectional data sets were constructed using the exclusive data provided by Opta and supplementary data obtained from several editions of transfermarkt magazines. A 2SLS instrumental variable estimator was employed and this section details the model specifications.

Initially, the simultaneity between the performances of the away teams’ attacking and defensive units and the home teams’ performances was investigated and the details of these analyses are presented in the sub sections 7.3.1 and 7.3.2 below. Sub sections 7.3.3 and 7.3.4 below provide a detailed breakdown of the analyses investigating the simultaneity between the performances of the home teams’ attacking and defensive units and the away teams’ performances.

7.3.1. Impact of the away team’s attacking unit upon the collective performance of the home team

The investigation of the interdependency between the away team attacking unit and the home teams collective performance required the use of data set 1. Equation 7.3.1a below presents the model used to conduct this analysis where the dependent variable $Y_1$ is the number of points acquired by the home team in match $j$ at time $t$ (HTRESULT), the Z variables represent the performance of the home team’s defence as well as other independent variables that control for team effects and inter-season effects. $Y_2$ represents the endogenous variable TPAATT, which encapsulates the performance of the attacking unit of the away team.

As interdependency between the home teams collective performance and the performances of the away teams attack is assumed, an IV (instrumental variable) estimation strategy was necessary in order to conduct this analysis. The 2SLS technique was the chosen method of estimation and the model used to conduct this analysis is presented below.
Y_{1it} = \beta_0 + \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \ldots + \beta_nZ_n + \beta_{n+1}Y_2 + u \quad (7.3.1a)

Y_2 = \pi_0 + \pi_1IV_1 + \pi_2IV_2 + \ldots + \pi_6IV_6 + \upsilon \quad (7.3.1b)

The variable $Z_1$ in equation 7.3.1a represents the time played by the home teams defence (TPHDEF). The variable TPHDEF therefore, captures the performance of the home teams defensive unit, with whom the away teams attacking players were in direct competition against. The remaining $Z$ variables presented in equation 7.3.1a represent several control variables included in the model to account for team effects and inter-season effects. These control variables included, GDHOME, GDAWAY, PERPTSHOMETEAM and PERPTSAYTEAM, which all measured team effects such as relative quality and relative form. Each of these four control variables was lagged by one gameweek in order to capture relative form and relative quality going into match $j$. Finally, the $Z$ variables presented in equation 7.3.1a also consisted of three control variables accounting for inter-season effects.

The variable $Y_2$ represents the endogenous variable that measures the performance of the away teams attack TPAATT. Equation 7.3.1b presented above represents the first stage (reduced form) regression equation where the performance of the away team’s attacking unit ($Y_2$) is the dependent variable. The endogenous variable TPAATT is assumed to be representative of the away teams attacking performance in match $j$ at time $t$ and is specified to be a function of six instrumental variables all unique to the players representing the away team’s attacking unit.

All six of these instrumental variables measure six particular player characteristics, the aggregated number of previous Bundesliga appearances made by each member of the away team attacking unit, the total number of players in the away team attacking unit with a preferred right foot, the total number of players in the away team attacking unit with a preferred left foot, the total number of players in the away team attacking unit able to use both feet, the average player height of the away team attacking unit and the average player age of the away team attacking unit. The justification for the inclusion of these instrumental variables was provided earlier in chapter 6, however where as in chapter 6 these instrumental variables were disaggregated player match observations, in this and subsequent analyses these instrumental variables are aggregated to represent positional unit-match observations.
Conducting Sargan $\chi^2$ and Basmann $\chi^2$ and Score $\chi^2$ tests of over-identifying restrictions revealed the suitability of the instrumental variables used to carry out this analysis. Other measures such as the Cragg and Donald (1993) minimum eigenvalue statistic and a joint F statistic were also utilised in order to further validate the choice of instruments.

This analysis was based upon the assumption that the performance of the away teams attacking unit and the collective team performance of the home team are interdependent. Therefore, the Durbin $\chi^2$, Wu-Hausman $\chi^2$ and Robust Score $\chi^2$ tests of endogeneity (as well as a robust regression) were conducted in order to confirm the variable TPAATT was endogenous. If the assumed endogenous regressor were revealed to be exogenous then OLS (Ordinary Least Squares) estimation would become more appropriate (Wooldridge 2009). In this case, the model would then be adapted as presented in equation 7.3.1c below.

$$Y_{1jt} = \beta_0 + \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \ldots + \beta_{n+1}Y_{2} + u$$ (7.3.1c)

In equation 7.3.1c above all variables retain the exact same representation as those presented in equation 7.3.1a. The only difference in equation 7.3.1c is the variable $Y_2$ (TPAATT) is no longer assumed to be endogenous but is instead treated as an independent variable and is assumed to have a linear causal effect upon the dependent variable.

7.3.2. Impact of the away team’s defending unit upon the collective performance of the home team.

When investigating simultaneity between the away team’s defensive performance and the collective performance of the home team data set 2 was utilised. Equation 7.3.2a below presents the model used to conduct this analysis where the dependent variable $Y_3$ is the variable (HTRESULT), the number of points acquired by the home team in match $j$ at time $t$. The variable $X_1$ represents the performance of the home team’s attacking unit TPHATT (with whom the away teams defensive players were in direct competition against) and was included in the model as the sole performance variable. The remaining $X_2, X_3, X_4$, and $X_n$ variables represent independent variables that control for team effects and inter-season effects. $Y_4$ represents the endogenous
variable $\text{TPADEF}$, which captures the performance of the defensive unit of the away team.

The output in this specification is the home team result. The result achieved by the home team is a function of the home teams attacking performance and the away teams defensive performance. Similarly to the analysis conducted in sub-section 7.3.1 above, the performance of the away defence is presumed to affect the collective performance of the home team, as well as be affected by the performance of the home team. To account for this interdependency between the collective performance of the home team and the performance of the away team’s defence, the same IV estimation strategy as used in sub-section 7.3.1 was utilised in order to conduct this analysis.

\begin{align*}
Y_{3jt} &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_n X_n + \beta_{n+1} Y_4 + u \tag{7.3.2a} \\
Y_4 &= \pi_0 + \pi_1 IV_1 + \pi_2 IV_2 + \ldots + \pi_6 IV_6 + \nu \tag{7.3.2b}
\end{align*}

Team effects such as relative team quality and relative team form were accounted for by the variables GDHOME, GDAWAY, PERPTSHOMETEAM and PERPTSAWAYTEAM, measuring goal differences and percentage points of both the home and away teams (going into match $j$). Changes in the general quality of the league as a whole were captured by the inclusion of the SEASON1, SEASON2 and SEASON3 dummy variables (Season 2010-11 was used as the reference season).

The variable $Y_4$ represents the endogenous variable that measures the performance of the away teams defensive units $\text{TPADEF}$. Equation 7.3.2b presented above represents the first stage (reduced form) regression equation where the performance of the away team defensive unit ($Y_4$) is treated as the dependent variable. The endogenous variable $\text{TPADEF}$ is assumed to be representative of the away teams defensive performance in match $j$ at time $t$ and is specified to be a function of six instrumental variables all unique to the players representing the away teams defensive unit. The same six instrumental variables for the analysis conducted in sub section 7.3.1 were also utilised here however all measurements were unique to the players belonging to the away team defensive units.
The suitability of the instrumental variables used to carry out this analysis was assessed by conducting Sargan $\chi^2$, Basmann $\chi^2$ and Score $\chi^2$ tests of over-identifying restrictions and the Cragg and Donald (1993) minimum eigenvalue statistic and the joint F test. Similarly to the analysis presented in sub section 7.3.1, the Durbin $\chi^2$, Wu-Hausman $\chi^2$ and Robust Score $\chi^2$ tests of endogeneity (as well as a robust regression) were conducted in order to confirm the endogenous relationship between the performances of the away teams defending unit and the error term. If these tests of endogeneity failed to recognise any endogeneity then OLS (Ordinary Least Squares) estimation would become a more suitable estimation technique. In the event the assumption of interdependency between the performance of the away teams defending unit and the home teams collective performance was falsified, the model would be adapted to the multiple regression (OLS) model presented in equation 7.3.2c below.

$$Y_{3jt} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_n X_n + \beta_{n+1} Y_4 + u$$ (7.3.2c)

In equation 7.3.2c above all variables retain the exact same representation as those presented in equation 7.3.2a. The only difference in equation 7.3.2c is the variable $Y_4$ (TPADEF) is no longer assumed to be endogenous but is instead treated as an independent variable and is assumed to have a linear causal effect upon the dependent variable $Y_3$.

### 7.3.3. Impact of the home team’s attacking unit upon the collective performance of the away team.

Data set 1 was used to investigate interdependency between the performances of the home team’s attackers and the collective performance of the away team. The model was similar to the one utilised in sub section 7.3.1. The model used to conduct this analysis is presented below in equations 7.3.3a and 7.3.3b below.

$$Y_{1jt} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \ldots + \beta_n Z_n + \beta_{n+1} Y_2 + u$$ (7.3.3a)

$$Y_2 = \pi_0 + \pi_1 IV_1 + \pi_2 IV_2 + \ldots + \pi_6 IV_6 + \nu$$ (7.3.3b)

The only difference between the model presented in equations 7.3.3a and the model presented earlier in 7.3.1a is the exchange of three variables. The first variable to be exchanged is the dependent variable $Y_1$, which is in this case the variable
ATRESULT, the result achieved by the away team. The second variable to be exchanged in the model presented in equations 7.3.3a and 7.3.3b is the endogenous variable Y2, which is exchanged for the variable TPHATT measuring the performance of the home teams attacking unit. The final variable to be exchanged was the variable X1, the sole performance variable, which is in this case the variable TPADEF. This variable is a measure of the performance of the away teams defensive unit, to whom the home attacking players were in direct competition. The remaining variables in equation 7.3.3a were exactly identical to those in equation 7.3.1a. The Instrumental variables in equation 7.3.3b measured the same six player characteristics as those in equation 7.3.1b however all measurements were unique to players belonging to the home team’s attacking unit.

Tests of over identifying restrictions, the F test, the eigenvalue statistic and the tests of endogeneity were also conducted for the same reasons stated earlier. Similarly to the previous analyses conducted, if endogeneity was revealed to be absent, the model presented in equation 7.3.3a would be adapted to the OLS multiple regression model presented in equation 7.3.3c below. All variables in equation 7.3.3c below are the same as in equation 7.3.3a above however the variable Y2 is no longer an endogenous variable but another independent variable assumed to have a linear causal effect upon the dependent variable Y1.

\[ Y_{1jt} = \beta_0 + \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \ldots \beta_nZ_n + \beta_{n+1}Y_2 + u \]  

\[ Y_{3jt} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \ldots \beta_nX_n + \beta_{n+1}Y_4 + u \]  

7.3.4. Impact of the home team’s defensive unit upon the collective performance of the away team

This analysis required the use of data set 2 and the purpose of this analysis was to investigate interdependency between the performance of the home team’s defensive unit and the collective performance of the away team. The model used for this analysis is similar to the model presented in the sub section 7.3.2. The only difference between the model presented below and the model presented earlier in the sub section 7.3.2 earlier was the exchange of three variables. The model used to conduct this analysis is presented in equations 7.3.4a and 7.3.4b below.

\[ Y_{3jt} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \ldots \beta_nX_n + \beta_{n+1}Y_4 + u \]
\[ Y_4 = \pi_0 + \pi_1 IV_1 + \pi_2 IV_2 + \ldots + \pi_6 IV_6 + \nu \]  

(7.3.4b)

The first variable to be exchanged is the dependent variable \( Y_3 \), which is in this case the variable ATRESULT, the result achieved by the away team. The second variable to be exchanged in the model presented above is the endogenous variable \( Y_4 \), which is exchanged for the variable TPHDEF measuring the performance of the home teams defensive units. The final variable to be exchanged was the variable \( X_1 \), the sole performance variable, which is in this case the variable TPAATT. This variable is a measure of the performance of the away teams attacking units, to whom the home defensive players were in direct competition with. The remaining variables in equation 7.3.4a were identical to those in equation 7.3.2a. The Instrumental variables in equation 7.3.4b above measured the same six player characteristics as those in equation 7.3.2b however all measurements were unique to players belonging to the home defending units.

Tests such as the tests for over identifying restrictions, the F test, the minimum eigenvalue statistic and the tests of endogeneity were all conducted for this analysis for the same reasons above. The OLS multiple regression model below represents the model that would be the more appropriate estimation technique if the assumption of simultaneity between performances was revealed to be false. All variables in equation 7.3.4c below have the same representation as the variables in equation 7.3.4a, with the only difference being that the variable \( Y_4 \) is no longer assumed endogenous, but instead is treated as another independent variable assumed to be linearly correlated to the dependent variable ATRESULT.

\[ Y_{3jt} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_n X_n + \beta_{n+1} Y_4 + u \]  

(7.3.4c)

7.4. First Stage Regression Results: Impact of Positional Units Upon the Performance of the Opposing Team

In this section of the chapter all first stage regression results are reported in table 7.4.1. Tests were conducted to test the suitability of the instrumental variables used to conduct the analyses and the results of these tests are detailed below in table 7.4.2, as
well as other general statistics referring to the goodness of fit of the first stage regression estimates.
Table 7.4.1 First Stage Regression Estimates Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of the Opposing Team.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>TPAATT</th>
<th></th>
<th>TPADEF</th>
<th></th>
<th>TPHATT</th>
<th></th>
<th>TPHDEF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>t</td>
<td>Coef</td>
<td>t</td>
<td>Coef</td>
<td>t</td>
<td>Coef</td>
<td>t</td>
</tr>
<tr>
<td>TPHDEF</td>
<td>.0496883</td>
<td>1.11+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>-</td>
<td>.0001017</td>
<td>0.00+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.0705535</td>
<td>1.40+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPAATT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.0084811</td>
<td>0.30+</td>
</tr>
<tr>
<td>SEASON1</td>
<td>-1.381477</td>
<td>-0.72+</td>
<td>.5387041</td>
<td>0.35+</td>
<td>-.7837725</td>
<td>-0.41+</td>
<td>.3043103</td>
<td>0.18+</td>
</tr>
<tr>
<td>SEASON2</td>
<td>-.03489111</td>
<td>-0.02+</td>
<td>.5387041</td>
<td>0.04+</td>
<td>.2594318</td>
<td>0.15+</td>
<td>-2.069382</td>
<td>-1.29+</td>
</tr>
<tr>
<td>SEASON3</td>
<td>1.092111</td>
<td>0.59+</td>
<td>1.37269</td>
<td>0.96+</td>
<td>-.7251894</td>
<td>-0.39+</td>
<td>.7440764</td>
<td>0.52+</td>
</tr>
<tr>
<td>GDHOME</td>
<td>-.0368977</td>
<td>-0.52+</td>
<td>-.0419132</td>
<td>-0.65+</td>
<td>.0676738</td>
<td>1.02+</td>
<td>.0142935</td>
<td>0.25+</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>.0582053</td>
<td>0.78+</td>
<td>-.131441</td>
<td>-2.22**</td>
<td>.0226881</td>
<td>0.33+</td>
<td>.0340763</td>
<td>0.58+</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>.0019444</td>
<td>0.04+</td>
<td>-.0423906</td>
<td>-1.08+</td>
<td>.0067759</td>
<td>0.12+</td>
<td>.0527681</td>
<td>1.23+</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>.062027</td>
<td>1.08+</td>
<td>.0794141</td>
<td>1.95*</td>
<td>-.0278599</td>
<td>-0.52+</td>
<td>-.0171093</td>
<td>-0.41+</td>
</tr>
<tr>
<td>AATTAGE</td>
<td>.8783579</td>
<td>2.97***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AATTHEIGHT</td>
<td>14.99428</td>
<td>1.00+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 7.4.1 continued

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>AATTRIGHTFOOT</td>
<td>75.27</td>
<td>18.13***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AATTLEFTFOOT</td>
<td>75.66</td>
<td>17.87***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AATTBOTHFEET</td>
<td>75.14</td>
<td>18.12***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AATTPREVLAPP</td>
<td>-0.0038</td>
<td>-0.42+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADEFAGE</td>
<td>-</td>
<td>-</td>
<td>0.0595</td>
<td>0.27+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADEFHEIGHT</td>
<td>-</td>
<td>-</td>
<td>28.89</td>
<td>2.44**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADEFRIGHTFOOT</td>
<td>-</td>
<td>-</td>
<td>88.16</td>
<td>39.28***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADEFLEFTFOOT</td>
<td>-</td>
<td>-</td>
<td>88.10</td>
<td>34.88***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AEFBOTHFEET</td>
<td>-</td>
<td>-</td>
<td>87.96</td>
<td>35.72***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AEFPREVLAPP</td>
<td>-</td>
<td>-</td>
<td>-0.0033</td>
<td>-0.91+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HATTAGE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.1356</td>
<td>0.48+</td>
</tr>
<tr>
<td>HATTHEIGHT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.610</td>
<td>0.67+</td>
</tr>
<tr>
<td>HATTRIGHTFOOT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>75.45</td>
<td>16.11***</td>
</tr>
<tr>
<td>HATTLEFTFOOT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>73.24</td>
<td>15.79***</td>
</tr>
<tr>
<td>HATTBOTHFEET</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>75.08</td>
<td>16.48***</td>
</tr>
<tr>
<td>HATTPREVLAPP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.0122</td>
<td>-1.50+</td>
</tr>
<tr>
<td>HDEFAGE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.2319</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>HDEFHEIGHT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Table 7.4.1 continued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDEFRIGHTFOOT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HDEFLEFTFOOT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HDEFBOTHFEET</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HDEFPREVLAPP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
As can be seen from table 7.4.1 above, the instrumental variables are consistently significant across the performances of all four positional units, the away attack, away defence, home attack and home defence. In particular, the three instrumental variables measuring a player's preferred foot are revealed to be significant at the 1% significance level for each positional unit. In addition, the instrumental variable measuring the average age of the away attacking unit is also revealed to be significant at the 1% significance level. Conversely, the instrumental variable measuring the average age of other positional units were of no significance. The instrumental variable measuring the average height of playing units is revealed to significantly impact the performance of away defenders at a 5% significance level. All significant results for the instrumental variables for each of the four playing units are positively signed meaning all significant instrumental variables have a positive impact upon the performance of the four positional units.

Team effects such as relative team quality and relative team form appear to have much less impact upon the performances of positional units as they do upon the performance of the entire team (as revealed by the results obtained in earlier in chapter 5). Team effects however were revealed to significantly impact the performance of away defenders. The variable GDAWAY that measures the goal difference (relative quality) of the away team was revealed to have a negative impact on the performance of away defenders and was revealed to be significant at the 5% significance level. The variable PERPTSAWAYTEAM that measures percentage points of the away team (relative form) was also revealed to significantly impact the performance of away defenders, however this impact was positive and significant at the 10% significance level.

Table 7.4.2 below reveals several first stage goodness of fit statistics such as the R², adjusted R², partial R² and F statistics testing the joint significance of the instrumental variables used in order to obtain the first stage regression estimates. As can be seen in table 7.4.2 below the R² and adjusted R² statistics are relatively high for all four endogenous variables, however Wooldridge (2009) claims these statistics can sometimes be misleading and have no natural interpretation in a 2SLS procedure. Bound et al (1995), therefore promote the use of the partial R² statistic, which measures the correlation between the endogenous variable and the instrumental variables after ‘partialling’ out the effect of the dependent variable.
The reported F statistics in table 7.4.2 below reveal the joint significance of the coefficients of the instrumental variables used in the first stage regression and Stock, et al (2002) state that this F statistic should be greater than 10 for any inferences based on a 2SLS procedure to be reliable when using one endogenous variable. As can be seen, all F statistics are much greater than 10 and all are highly significant (p-values), meaning the instrumental variables have a highly significant explanatory power for the endogenous variables after controlling for the effects of other included exogenous variables.

The Cragg and Donald (1993) minimum eigenvalue statistic is another measure of the strength of instrumental variables, particularly when this statistic is used in conjunction with Stock and Yogo (2005) tests for weak instruments. As can be seen from table 7.4.2 below when choosing either the largest relative bias of the 2SLS estimator willing to be tolerated or the largest rejection rate of a nominal 5% Wald test willing to be tolerated, the test statistics (eigenvalues) well exceed the critical values. The fact that the eigenvalues for all endogenous variables were greater than the critical values of the two Stock and Yogo (2005) tests for weak instruments allows us to reject the null hypothesis of each of the tests, that the instruments are weak. Based on all the results presented in table 7.4.2 below, it can be concluded that the instruments used throughout the first stage regressions for all analyses are strongly correlated to the endogenous regressors and are therefore valid instruments.
Table 7.4.2. Statistics Examining the Explanatory Power Of the Instrumental Variables

<table>
<thead>
<tr>
<th>Endogenous Variable</th>
<th>TPAATT</th>
<th>TPADEF</th>
<th>TPHATT</th>
<th>TPHDEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>R Squared</td>
<td>0.9198</td>
<td>0.9496</td>
<td>0.9178</td>
<td>0.9520</td>
</tr>
<tr>
<td>Adjusted R Squared</td>
<td>0.9188</td>
<td>0.9489</td>
<td>0.9168</td>
<td>0.9514</td>
</tr>
<tr>
<td>Partial R Squared</td>
<td>0.3456</td>
<td>0.5935</td>
<td>0.3384</td>
<td>0.6089</td>
</tr>
<tr>
<td>F Statistic (Robust)</td>
<td>57.3028</td>
<td>264.275</td>
<td>45.851</td>
<td>216.233</td>
</tr>
<tr>
<td>F Statistic p-value</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Minimum Eigenvalue</td>
<td>94.6913</td>
<td>267.205</td>
<td>93.584</td>
<td>279.171</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2SLS Relative Bias</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>5%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19.28</td>
<td>11.12</td>
<td>6.76</td>
<td>5.15</td>
<td>19.28</td>
<td>11.12</td>
<td>6.76</td>
<td>5.15</td>
<td>19.28</td>
<td>11.12</td>
<td>6.76</td>
<td>5.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2SLS Size of Nominal 5% Wald Test</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
</tr>
</thead>
</table>
7.5. Second Stage Regression Results: Individual Impact of Team Units Upon the Performance of the Opposing Team

In this section all results obtained for the second stage regressions for each of the four positional units are presented in table 7.5.1. Further tests of the suitability of the instruments and the specification of the model (tests of over-identifying restrictions) are conducted as these tests are not concerned with the first stage regressions, but the residual correlation with exogenous variables and are therefore concerned with the second stage regressions. The results obtained from these tests of over-identifying restrictions are presented in table 7.5.2. Tests of endogeniety were also conducted throughout the analysis and these results are also detailed below in table 7.5.3.
Table 7.5.1 Second Stage Regression Estimates Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of the Opposing Team.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>HTRESULT</th>
<th>ATRESULT</th>
<th>ATRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endogenous Variable</td>
<td>Coef</td>
<td>z</td>
<td>Coef</td>
<td>z</td>
</tr>
<tr>
<td>TPAATT</td>
<td>-.0063238</td>
<td>-2.72**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>-</td>
<td>-.004656</td>
<td>-2.89***</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHDEF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Independent Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPHDEF</td>
<td>.0057605</td>
<td>2.78**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>-</td>
<td>.004846</td>
<td>2.95***</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPAATT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SEASON1</td>
<td>.1056853</td>
<td>1.10+</td>
<td>.0574331</td>
<td>0.59+</td>
</tr>
<tr>
<td>SEASON2</td>
<td>.0818176</td>
<td>0.82+</td>
<td>.087955</td>
<td>0.89+</td>
</tr>
<tr>
<td>SEASON3</td>
<td>-.0427454</td>
<td>-0.44+</td>
<td>-.0539222</td>
<td>-0.56+</td>
</tr>
<tr>
<td>GDHOME</td>
<td>.0063525</td>
<td>1.74*</td>
<td>.0075552</td>
<td>2.13**</td>
</tr>
<tr>
<td>Variable</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>-0.0081019</td>
<td>-2.13**</td>
<td>-0.00521</td>
<td>-1.40+</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>0.0232355</td>
<td>9.39***</td>
<td>0.0231513</td>
<td>9.33***</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>-0.0153607</td>
<td>-6.16***</td>
<td>-0.0171966</td>
<td>-6.92***</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
Table 7.5.1 above reveals that all the assumed endogenous variables have a significant negative impact upon the match results (dependent variable) obtained by the opposing team. Team effects, similarly to the results obtained from the analysis presented earlier in chapter 5, are revealed to have a significant impact upon match results achieved by the home and away teams. The variables measuring relative team form, PERPTSHOMETEAM and PERPTSAWAYTEAM are significant at a 1% significance level for all four analyses and are signed as expected. The variables measuring team quality, GDHOME and GDAWAY are significant at the 5% or 10% significance level and are also signed as expected. GDHOME is consistently significant in all four analyses and therefore impacts both the home and away team match results achieved. The variable GDAWAY only impacts the match result obtained by the home team.

None of the three dummy variables accounting for changes in the quality of the league are shown to be associated with the match result achieved by either the home or away team. The performance variables measuring the performance of the corresponding members of the opposing unit with whom observed players were matched, all had a significantly positive impact upon the match result obtained by their respective teams. Table 7.5.1 reveals the performances of the home teams’ defence and the home teams’ attack are positively correlated to the match result obtained by the home team as expected, at the 1% significance level. Similarly, the performances of the away team defensive and attacking units have the expected positive impact upon the match result achieved by the away team.

Table 7.5.2 and table 7.5.3 below reveal the results obtained from the tests of over-identifying restrictions and the tests of endogeneity. These tests were conducted to ensure the validity of the instruments used in the analyses, ensure all four models were correctly specified and to verify the presence of the assumed simultaneity between the performances of positional units and their opposing teams.

Table 7.5.2 reveals the results obtained from the Sargan $\chi^2$, Basmann $\chi^2$ and Score $\chi^2$ tests of over-identifying restrictions. These tests are conducted in order to confirm the instruments are uncorrelated to the error term and therefore ensure that the instrumental variables are valid. Each of the tests of over-identifying restrictions also ensure the models used to conduct the analyses have been correctly specified by
determining whether or not an excluded exogenous variable should in fact be included in the structural equation. As can be seen from table 7.5.2 below all p-values are greater than 0.05 and therefore we cannot reject the null hypothesis (instruments are valid/model is correctly specified) therefore the instrumental variables are valid and the model is correctly specified.
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>HTRESULT</th>
<th>ATRESULT</th>
<th>ATRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endogenous Variable</td>
<td>TPAATT</td>
<td>TPADEF</td>
<td>TPHATT</td>
<td>TPHDEF</td>
</tr>
<tr>
<td>Sargan $\chi^2$ Test Score</td>
<td>3.34166</td>
<td>9.34905</td>
<td>1.28553</td>
<td>3.40635</td>
</tr>
<tr>
<td>Sargan $\chi^2$ (p-value)</td>
<td>0.6475</td>
<td>0.0959</td>
<td>0.9364</td>
<td>0.6376</td>
</tr>
<tr>
<td>Basmann $\chi^2$</td>
<td>3.30584</td>
<td>9.30119</td>
<td>1.26967</td>
<td>3.37004</td>
</tr>
<tr>
<td>Basmann $\chi^2$ (p-value)</td>
<td>0.6529</td>
<td>0.0976</td>
<td>0.9380</td>
<td>0.6431</td>
</tr>
<tr>
<td>Score $\chi^2$ (Robust)</td>
<td>3.26803</td>
<td>10.1187</td>
<td>1.27562</td>
<td>3.61037</td>
</tr>
<tr>
<td>Score $\chi^2$ (p-value)</td>
<td>0.6587</td>
<td>0.0719</td>
<td>0.9374</td>
<td>0.6068</td>
</tr>
<tr>
<td>Dependent Variable</td>
<td>HTRESULT</td>
<td>HTRESULT</td>
<td>ATRESULT</td>
<td>ATRESULT</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Endogenous Variable</td>
<td>TPAATT</td>
<td>TPADEF</td>
<td>TPHATT</td>
<td>TPHDEF</td>
</tr>
<tr>
<td>Durbin $\chi^2$ Test Score</td>
<td>.333795</td>
<td>1.68395</td>
<td>.106935</td>
<td>.449086</td>
</tr>
<tr>
<td>Durbin $\chi^2$ (p-value)</td>
<td>0.5634</td>
<td>0.1944</td>
<td>0.7437</td>
<td>0.5028</td>
</tr>
<tr>
<td>Wu-Hausman F Score</td>
<td>.330531</td>
<td>1.66984</td>
<td>.105889</td>
<td>.444741</td>
</tr>
<tr>
<td>Wu-Hausman F (p-value)</td>
<td>0.5655</td>
<td>0.1966</td>
<td>0.7449</td>
<td>0.5050</td>
</tr>
<tr>
<td>Robust Score $\chi^2$</td>
<td>.294379</td>
<td>1.7291</td>
<td>.098009</td>
<td>.39562</td>
</tr>
<tr>
<td>Robust Score $\chi^2$ (p-value)</td>
<td>0.5874</td>
<td>0.1885</td>
<td>0.7542</td>
<td>0.5294</td>
</tr>
<tr>
<td>Robust Regression</td>
<td>.292433</td>
<td>1.70894</td>
<td>.096381</td>
<td>.388905</td>
</tr>
<tr>
<td>Robust Regression (p-value)</td>
<td>0.5888</td>
<td>0.1914</td>
<td>0.7563</td>
<td>0.5330</td>
</tr>
</tbody>
</table>
In the analysis where the performance of the away defenders and their impact upon the performance of the home team was investigated the p-values for the Sargan $\chi^2$, Basmann $\chi^2$ and Score $\chi^2$ tests of overidentifying restrictions are revealed to be (0.0959), (0.0976) and (0.0719) respectively. Although the null hypothesis of the instruments being valid and the model being correctly specified cannot be rejected at the 5% significance level it would have to be rejected at the 10% level.

Throughout the analyses conducted in this chapter it was assumed that there is interdependency between the performances of the away attacking and defensive units and the collective team performance of their opponents, the home team. Similarly, it was assumed that there is interdependency between the performances of the home attacking and defensive units and the collective team performance of their opponents, the away team. Table 7.5.3 above reveals the results obtained from the tests of endogeneity, which test whether or not the assumed endogenous variables could have actually been treated as exogenous variables.

As the results in table 7.5.3 reveal, the hypothesised endogenous variables, TPAATT, TPADEF, TPHATT, TPHDEF are in fact revealed to be exogenous. As can be seen by the p-values obtained, the null hypothesis of these tests that the variables are exogenous cannot be rejected. These results imply that OLS estimation is more efficient and hence more appropriate than IV estimation in this particular instance (Wooldridge 2009). Section 7.6 below presents the results obtained from an OLS estimation assessing the impact of attacking and defensive positional units upon the collective performances of their opposing teams’.
7.6. OLS Estimation Results: Individual Impact of Positional Units Upon the Collective Performance of the Opposing Team

The results obtained in the previous section suggested that the performances of attacking and defensive positional units were incorrectly assumed endogenous and in fact were exogenous implying therefore that OLS estimation is more appropriate. Table 7.6.1 below presents the result obtained from an OLS estimation conducted in light of the results presented in section 7.5 above. The results show that the performances of positional units (and their matched opposing unit) have a significant impact upon the match result and are signed as expected, in this case however, they only have a linear correlation to the match result obtained by the opposition (as simultaneity is no longer assumed). As expected, the performance of the away attacking and away defensive units have a negative impact upon the performance of the home team, where as their matched opponents, the home defensive units and home attacking units have the expected positive impact upon home team performance. Conversely, the performance of the home team attacking and defensive units have a negative impact upon the performance of the away team, whilst their matched opponents the away defensive and the away attacking units have the expected positive impact upon the away teams performance.

Team effects also impact the performance of both the home and away teams and team form in particular impacts the performance of the observed team in a positive manner, but has an adverse affect upon the performance of the opposing team. Team quality also impacts the performance of the home and away teams however to a lesser extent than relative team form. None of the control variables, SEASON1, SEASON2 or SEASON3 are revealed to have any significant impact upon the performances of the home or the away teams’.

Table 7.6.2 presented below reveals the joint significance of all the variables included in the standard OLS regression model. The $R^2$ and adjusted $R^2$ are relatively low, however the purpose of this analysis was to investigate the impact of a positional unit of an observed team upon the overall performance of the entire opposing team. The low $R^2$ values are therefore to be expected, as it is reasonable to expect these independent variables to only account for a partial proportion of the total sample variation in the dependent variables. Robust estimation was conducted (as in the
analyses conducted in previous chapters) in order to account for the inclusion of multiple dummy variables and the associated violation of assumptions (i.e. heteroskedasticity).  

Readers should note that in cases of heteroskedasticity a generalized method of moments estimator (GMM) may also be employed instead of 2SLS estimation with robust standard errors. GMM estimations were obtained for all empirical analyses conducted throughout this thesis and the results are presented in appendix two. As can be seen, there is minimal difference, with regard to the significance of variables, between the GMM estimations and 2SLS estimations. 2SLS was ultimately chosen ahead of GMM for the purposes of this study because in cross sectional applications, it is more common to adopt the 2SLS, though with inference based on heteroskedastic robust standard error (Cameron and Trivedi, 2005).
Table 7.6.1 OLS Estimations Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of their Opposing Team.

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coef</th>
<th>t</th>
<th>Coef</th>
<th>t</th>
<th>Coef</th>
<th>t</th>
<th>Coef</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HTRESULT</td>
<td></td>
<td>ATRESULT</td>
<td></td>
<td>HTRESULT</td>
<td></td>
<td>ATRESULT</td>
<td></td>
</tr>
<tr>
<td>TPAATT</td>
<td>-.0052387</td>
<td>-3.60***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>-</td>
<td>-.0060236</td>
<td>-4.71***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.0047822</td>
<td>-3.56***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHDEF</td>
<td>-.0048173</td>
<td>3.56***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.0039861</td>
<td>-3.07***</td>
</tr>
<tr>
<td>SEASON1</td>
<td>.1069944</td>
<td>1.11+</td>
<td>.0592217</td>
<td>0.61+</td>
<td>-.1041606</td>
<td>-1.11+</td>
<td>-.1440941</td>
<td>-1.55+</td>
</tr>
<tr>
<td>SEASON2</td>
<td>.0789326</td>
<td>0.79+</td>
<td>.0863741</td>
<td>0.87+</td>
<td>-.0898642</td>
<td>-0.93+</td>
<td>-.0891091</td>
<td>-0.91+</td>
</tr>
<tr>
<td>SEASON3</td>
<td>-.0451225</td>
<td>-0.46+</td>
<td>-.0478914</td>
<td>-0.49+</td>
<td>-.021569</td>
<td>-0.23+</td>
<td>.0025455</td>
<td>0.03+</td>
</tr>
<tr>
<td>GDHOME</td>
<td>.0064</td>
<td>1.67*</td>
<td>.0074865</td>
<td>2.10**</td>
<td>-.00902</td>
<td>-2.68***</td>
<td>-.0085224</td>
<td>-2.46**</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>-.0081402</td>
<td>-2.13**</td>
<td>-.0054293</td>
<td>-1.45+</td>
<td>.0016804</td>
<td>0.47+</td>
<td>.0035756</td>
<td>0.97+</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>.0232821</td>
<td>9.37***</td>
<td>.0230904</td>
<td>9.22***</td>
<td>-.0198225</td>
<td>-8.24***</td>
<td>-.0193702</td>
<td>-8.15***</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>-.0154292</td>
<td>-6.15***</td>
<td>-.0170503</td>
<td>-6.82***</td>
<td>.0199829</td>
<td>8.66***</td>
<td>.019112</td>
<td>8.27***</td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01

SEASON4 (2010-11) was used as the reference season
Table 7.6.2. Summary Statistics for OLS Estimation

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>HTRESULT</th>
<th>ATRESULT</th>
<th>ATRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance Variables</td>
<td>TPAATT</td>
<td>TPADEF</td>
<td>TPHATT</td>
<td>TPHDEF</td>
</tr>
<tr>
<td></td>
<td>TPHDEF</td>
<td>TPHATT</td>
<td>TPADEF</td>
<td>TPAATT</td>
</tr>
<tr>
<td>R Squared</td>
<td>0.2761</td>
<td>0.2805</td>
<td>0.2728</td>
<td>0.2679</td>
</tr>
<tr>
<td>Adjusted R Squared</td>
<td>0.2701</td>
<td>0.2747</td>
<td>0.2669</td>
<td>0.2618</td>
</tr>
<tr>
<td>F Statistic (Robust)</td>
<td>88.83</td>
<td>92.35</td>
<td>83.41</td>
<td>78.07</td>
</tr>
<tr>
<td>F Statistic p-value</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
7.7 Discussion

The first stage regression results presented in section 7.4 revealed that each of the instrumental variables had a significant impact upon the performances of attackers and defenders except for the number of previous Bundesliga appearances. The three variables measuring a player’s preferred foot all consistently affected the performance of each positional unit at the 1% significance level suggesting that either most positional units included in the sample consisted of a fairly balanced number or right footed, left footed or two footed players or that all three variables are of relative equal importance with regards to the performance of a positional unit.

The average age of players, measuring their overall playing experience was only revealed to impact the away teams attacking unit, implying either that attacking players with more playing experience are more successful in away games rather than inexperienced attackers or that managers are more willing to play younger, more inexperienced attackers in home games rather than away games. Both of these explanations are feasible, Richardson et al (2012), state in their study on youth player migration that managers can give young players their debuts at 17, or 18 years of age and often in front of the home crowd in order to make them feel more comfortable. Similarly, home sides have been shown to have an advantage over their opposition in many previous studies due to a number of factors such as referee bias, higher effort levels from home players, home support and familiarity effects (Nevill et al 1996; Sutter and Kocher 2004; Tucker et al 2005; Carmichael and Thomas 2005; Dawson et al 2007; Pollard 2008; Leard and Doyle 2010). Consequently, it would be reasonable to assume away attackers are likely to be open to the same disadvantages, which are perhaps countered to an extent by fielding more experienced players.

Bloomfield et al (2005) claim that strikers and defenders are heavier and slightly taller than midfielders, suggesting a player’s physical characteristics such as their height and weight are connected to the demands of specific playing positions. This claim is further reinforced by the results revealed in section 7.4 that show that the instrumental variable measuring the average height of players’ significantly affected the performance of away defenders. The fact the average height of home defenders was found to be insignificant could simply be explained by defenders being called into
action more often in away games due to home teams being under more pressure to perform and under more pressure to entertain their own supporters by adopting more attacking, eye pleasing tactics (Wolfson et al 2005).

Team effects were revealed to have very little impact upon the performance of positional units, however similarly to the results presented in chapter 5, team effects had a much more profound impact upon the performance of a team as a collective. Table 7.5.1 revealed that relative team form in particular has a sizeable effect upon the match results obtained. The relative team form of the opponents was also revealed to have the adverse effect upon match results obtained by an observed team.

The performance of each positional unit was revealed to have an impact upon the collective performance of their opposing team, however only a linear impact was confirmed. The tests of endogeneity (table 7.5.3) failed to verify the endogenous nature of the positional unit performances and consequently an OLS estimation was conducted. The results from the OLS regressions presented in section 7.6 confirmed that the performances of positional units’ have a significant but linear impact upon match results obtained by their opposing teams.
7.8 Conclusion

The purpose of the analyses conducted throughout this chapter was to separately investigate the extent of the interdependency between the performances of attacking and defensive positional units and the collective performance of their opposing teams. Although, the initial assumption of simultaneity between these performances was ultimately proven to be incorrect, the performances of each positional unit was still revealed to have a significant positive linear impact upon the performance of their own team and a significant negative linear impact upon the collective performance of their opposing team.

Bosca et al (2009) claimed that defensive efficiency rather than offensive efficiency is more important in the Italian league and that those teams who have offensive efficiency both at home and away achieve success in the Spanish league. Contrary to these findings on the Italian and Spanish top level leagues the results obtained throughout the analyses in this chapter suggest success in the Bundesliga is dependent upon both offensive efficiency and defensive efficiency meaning Bundesliga teams should aim towards achieving a good balance of both.

In addition to positive performances of positional units having a positive impact upon their own team’s performance, the results obtained in this chapter also emphasise the dual effect the positive performance of positional units can have in the sense of negating the performance of the opposition. The performance of home defenders and home attackers was revealed to have a positive impact upon the performance of the home team, whilst having an adverse effect upon the performance of the away team. Similarly, the performance of the away team defenders and attackers was revealed to have a positive impact upon away team performance, whilst also having a negative impact upon the performance of the home team.

The results obtained in this chapter stress the importance of how a good performance from both the defensive and attacking positional units in unison can contribute to team success and simultaneously negate the opponent’s performance. Football is a team game and it appears that the combined efforts of all team members have a collective significant impact upon team success rather than the efforts of just one positional unit. In contrast to findings from studies on foreign leagues, strong
defensive performance and strong attacking performance seem to have relatively equal importance with regard to achieving successful match outcomes in the German Bundesliga.
Chapter Eight

Conclusions

8.1 Introduction

This chapter provides the concluding observations of this thesis. In section 8.2 below, the aims and objectives of this thesis are restated and a review of how they were fulfilled is presented. More specifically, a review of what each chapter covered and how it contributed to fulfilling the aims of this thesis is discussed. In section 8.3, the contributions of this thesis are noted. Section 8.3 also addresses the limitations of this study and provides recommendations for future research. Finally, section 8.4 provides some concluding observations that arose throughout the thesis.
8.2 Review of the Thesis

As outlined in the literature review in chapter 3, several factors have been proven to impact the performances of teams as well as the performances of individual players in football. Despite previous studies having identified several playing inputs and non-playing inputs the review of previous literature in chapter 3 highlighted the general lack of attention awarded to the performances of the opponent and the assumed simultaneity between the performances of direct competitors. In particular, previous production function studies in football have not included sufficient relative measures of the opponents’ performances and hence investigated the simultaneity between performances of the opponent and an observed team or player.

In addition, the literature review presented in chapter 3 also highlighted the team-orientated nature of much previous research in football. The performance of individual players and their impact upon the performances of their opponents therefore has also been frequently overlooked. This thesis aimed to fill these gaps in the literature by empirically investigating the level of simultaneity between the performances of two direct opponents at a collective team level, a more isolated individual player level and a positional unit vs. team level. The aim of this thesis therefore was to examine if, and to what extent, player productivity in football is affected, by the performance of their opponent(s).

To achieve this, a series of research objectives were set: First, to review the relevant economic, theoretical and empirical literature in order to establish what factors have been identified as relevant to the determination of team and player productivity. Second, to extend the existing empirical literature by organising and developing several data sub-sets that can be used to test for the impact of opponents and to apply an appropriate econometric strategy to perform the tests. Thirdly, to provide an accurate assessment of collective team, positional unit (attacking and defensive) and individual player performance in the German Bundesliga, over the course of four seasons, whilst specifically accounting for the relative performance of the opponent(s) and the simultaneity therein.

The analysis began in chapter 5 that sought to investigate simultaneity between two opponents’ performances at a collective team level. This analysis focussed on 1,223
match fixtures played in the German Bundesliga throughout the seasons 2007-08 and 2010-11. Play actions were aggregated in order to provide team-match totals for each of the 2,446 team-game observations.

The results obtained from a 2SLS IV estimation suggested that the performances of an opponent and an observed team are not interdependent, as aggregated play actions of the opponent were revealed to be insignificant. The results however also revealed that the recent past performance of an opponent did have a linear impact upon the performance of an observed team (in match _j_). In particular, relative team form going into a match was revealed to be highly significant, implying momentum effects are present in football. Proxies for relative team quality were also revealed to be significant and also had a linear impact upon the performances of the opponent.

The analysis proceeded with chapter 6 that sought to investigate simultaneity between two opponents’ performances at an individual player level. This analysis focussed on 5,411 player match-ups, where players performing in defence and attack were matched up to a sole opponent on the field of play. Players’ were matched to an opponent in fixtures that were played between the seasons 2007-08 and 2010-11. Disaggregated player-match specific performance variables were utilised for this analysis.

The results obtained from this 2SLS IV analysis confirmed that the performances of opposing players on the field of play are interdependent. Under the assumption that productivity-adjusted time played encapsulated player performance, the performances of defenders were revealed to have a highly significant and negative impact upon the performances of their opposing attackers. In contrast, the performances of attacking players did not have any significant impact upon the performances of their opposing defenders. Furthermore, the results obtained from this analysis confirmed that the venue in which the match is played impacts player performances at an individual player level.

Further analysis was conducted in chapter 7 where simultaneity between collective player performance in a positional unit and collective team performance was investigated. Two separate data sets were constructed for this analysis. The first data set consisted of 1,091 positional unit-match observations that enabled an investigation into the impact of the performances of home defenders and away attackers on the
performances of their opposing teams (collectively). The second data set consisted of 1,113 positional unit-match observations that enabled an investigation into the impact of the performances of away defenders and home attackers upon the performances of their opposing teams.

The results obtained from the post-estimation tests of endogeneity (conducted after a 2SLS IV estimation) revealed that the performances of positional units were incorrectly treated as endogenous variables. These results therefore implied that the performances of positional units and their opposing teams were not interdependent (i.e. simultaneity did not exist). An OLS regression analysis was subsequently conducted in order to investigate if the performances of positional units had any linear impact upon the performances of their opposing teams. This analysis revealed that the collective performances of players within a positional unit do have a linear impact upon the performances of their opposing teams. Similarly to the results obtained in chapter 5, team effects such as team form going into a match was revealed to be highly significant and had the expected impact upon the performance of their opposing team although only a linear correlation.
8.3 Research Study Contributions, Limitations and Recommendations for Future Research

This thesis is the first attempt to analyse the production of sport that specifically seeks to investigate the interdependency between the performances of two direct opponents at numerous levels. It therefore fills a gap and adds to the body of knowledge in sports economics. In particular, four main contributions are identified.

Firstly, the study adopted an innovative approach in using match specific performance measures of both opposing players and teams to investigate the productivity of teams and players. This enabled the relative performance of specific team and players to be investigated empirically in chapters 5, 6 and 7.

This meant that three unique outcomes could be identified for the sports economic literature. Each of these is a contribution of the thesis. The second contribution of the thesis thus, is that at a team level it was shown that relative form going into a match had a linear impact upon the match result obtained, whilst the relative form of the opponent had the opposite linear impact. The consistencies in these results therefore imply momentum effects are present in football as they are in ice hockey (Leard and Doyle, 2010) and basketball (Arkes and Martinez, 2012).

The third contribution of the thesis is that the analyses conducted in chapter 6 confirmed that the performances of individual players are jointly determined with the performances of their opponents. Specifically, defenders were revealed to have a significant and negative impact upon the performances of their directly opposing attackers. This study therefore confirmed the existence of simultaneity between the performances of individual players in football. As player performances between two direct opponents were revealed to be jointly determined this analysis offers support to the theory of joint production in sports economics. Furthermore, the results revealed in chapter 6 confirmed that the venue a match is played in impacts the performances of individual players. Specifically, attacking players were revealed to perform better when playing away and defenders revealed to perform worse when playing away.

The final contribution of the thesis is that it is the first to investigate simultaneity between the performances of specific positional units and their opposing teams collectively. The results ultimately failed to confirm interdependency. However, the
performances of opposing positional units did have a negative linear impact upon the performances of the observed teams. The performance of the opponents therefore was proven to be significant in each investigation conducted throughout this thesis (albeit recent past performance of the opponent in chapter 5).

To summarise, this study offers three empirical contributions to knowledge in the field of sports economics, through the application of an econometric methodology that accounts for the relative performance of teams and players. At a team level (team vs. team) it was discovered that production was not jointly determined, thus the empirical results obtained in chapter 5 did not support the theory of joint production in sports economics. Similarly, it was discovered that production was not jointly determined at a positional unit level (positional unit vs. team), thus the empirical results obtained in chapter 7 also failed to support the theory of joint production in sports economics. At an individual player level (player vs. player) however, this study discovered that production is jointly determined, therefore the empirical results obtained in chapter 6 supported the theory of joint production in sports economics.

Despite having made significant contributions to knowledge by providing insights into the impact of opponents upon performance, there are some limitations to the research that should be acknowledged as well as recommendations for future research. As suggested by Carmichael et al (2000), off the ball player movements can potentially impact team and player performance. Off the ball movements were not available in the data set, therefore the recording of this information could lead to the development of more sophisticated key performance indicators and improve the production function analyses. Furthermore, this research was restricted to the German Bundesliga. Future research should investigate the impact of opponents in other football leagues and in major football tournaments.

The results obtained in chapters 5 and 7 suggest team momentum effects are present in football as they have been proven to be in other team sports such as ice hockey and basketball (Leard and Doyle 2010; Arkes and Martinez 2012). This does however require further investigation as football, unlike ice hockey and basketball consists of drawn games where rewards are shared in equal proportion (a point for each competing team). In the case of a draw, it becomes difficult to determine which of the two competing teams would view a draw as a “good result” and an accumulation of a
point or a “bad result” and the loss of two points. This could therefore make it difficult to assess whether a momentum effect is negative or positive going into a subsequent fixture.

As stated above this research confirmed that the performance of the opponent is significant at a team and an individual player level in league competition. The performance of the opponent however, may not be as significant in knockout competitions where luck plays a bigger part in determining match outcome (Downward et al 2009). Future research therefore should look to investigate whether the performance of the opponent has any significance in determining match outcomes in knockout competition such as domestic/international cup competitions.
8.4 Concluding Remark

A series of robust models were constructed which enabled this thesis to address gaps in previous literature. In particular, this study addressed a gap in the literature by exploring whether the performance of an opponent had any impact upon the performance of a subject and whether these performances were interdependent. Ultimately, the research question was answered as it was revealed that the impact of an opponent’s performance is critical and substantial at a team level, positional unit level and an individual player level. At an individual player level, performances between an observed player and their direct opponent were revealed to be interdependent, thus confirming the existence of simultaneity between the performances of players. Moving forward, performance analysts should view the performance of the opponent as a key determinant of performance (for the subject) and look to incorporate the opponent’s performance in their analyses.
Bibliography


287


## APPENDIX ONE

### Table 7.6.3. Vif Statistics

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>HTRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VIF</td>
<td>VIF</td>
</tr>
<tr>
<td>TPAATT</td>
<td>8.16</td>
<td>-</td>
</tr>
<tr>
<td>TPHDEF</td>
<td>8.15</td>
<td>-</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>8.06</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>8.05</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>2.38</td>
<td>2.29</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>2.31</td>
<td>2.29</td>
</tr>
<tr>
<td>GDAWAY</td>
<td>2.07</td>
<td>2.03</td>
</tr>
<tr>
<td>GDHOME</td>
<td>2.00</td>
<td>2.05</td>
</tr>
<tr>
<td>SEASON1</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>SEASON2</td>
<td>1.47</td>
<td>1.48</td>
</tr>
<tr>
<td>SEASON3</td>
<td>1.51</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>HTRESULT (1)</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>TPAATT (2)</td>
<td>-0.0608</td>
<td>1.0000</td>
</tr>
<tr>
<td>TPHDEF (3)</td>
<td>-0.0116</td>
<td>0.9360</td>
</tr>
<tr>
<td>SEASON1 (4)</td>
<td>0.0327</td>
<td>-0.1104</td>
</tr>
<tr>
<td>SEASON2 (5)</td>
<td>0.0207</td>
<td>-0.0499</td>
</tr>
<tr>
<td>SEASON3 (6)</td>
<td>-0.0420</td>
<td>0.1789</td>
</tr>
<tr>
<td>GDHOME (7)</td>
<td>0.3113</td>
<td>-0.0048</td>
</tr>
<tr>
<td>GDAYAWAY (8)</td>
<td>-0.2624</td>
<td>0.1163</td>
</tr>
<tr>
<td>PERPTSHOMETEAM (9)</td>
<td>0.4438</td>
<td>-0.0071</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM (10)</td>
<td>-0.3799</td>
<td>0.0944</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>HTRESULT (1)</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>TPADEF (2)</td>
<td>-0.0166</td>
<td>1.0000</td>
</tr>
<tr>
<td>TPHATT (3)</td>
<td>0.0397</td>
<td>0.9351</td>
</tr>
<tr>
<td>SEASON1 (4)</td>
<td>0.0216</td>
<td>-0.0810</td>
</tr>
<tr>
<td>SEASON2 (5)</td>
<td>0.0236</td>
<td>-0.1149</td>
</tr>
<tr>
<td>SEASON3 (6)</td>
<td>-0.0323</td>
<td>0.1925</td>
</tr>
<tr>
<td>GDHOME (7)</td>
<td>0.3191</td>
<td>0.0890</td>
</tr>
<tr>
<td>GDAWAY (8)</td>
<td>-0.2413</td>
<td>-0.0425</td>
</tr>
<tr>
<td>PERPTSHOMETEAM (9)</td>
<td>0.4462</td>
<td>0.0360</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM (10)</td>
<td>-0.3712</td>
<td>-0.0237</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>ATRESULT (1)</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>TPHATT (2)</td>
<td>-0.0206</td>
<td>1.0000</td>
</tr>
<tr>
<td>TPADEF (3)</td>
<td>0.0280</td>
<td>0.9351</td>
</tr>
<tr>
<td>SEASON1 (4)</td>
<td>-0.0306</td>
<td>-0.0877</td>
</tr>
<tr>
<td>SEASON2 (5)</td>
<td>-0.0158</td>
<td>-0.1051</td>
</tr>
<tr>
<td>SEASON3 (6)</td>
<td>0.0159</td>
<td>0.1806</td>
</tr>
<tr>
<td>GDHOME (7)</td>
<td>-0.3104</td>
<td>0.1116</td>
</tr>
<tr>
<td>GDAWAY (8)</td>
<td>0.2367</td>
<td>-0.0371</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>-0.4311</td>
<td>0.0621</td>
</tr>
<tr>
<td>(9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>0.3852</td>
<td>-0.0365</td>
</tr>
<tr>
<td>(10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.6.4 Correlation Matrix 7.6.7

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATRESULT (1)</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPHDEF (2)</td>
<td>-0.0007</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPAATT (3)</td>
<td>0.0422</td>
<td>0.9360</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON1 (4)</td>
<td>-0.0370</td>
<td>-0.1105</td>
<td>-0.1104</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON2 (5)</td>
<td>-0.0180</td>
<td>-0.0677</td>
<td>-0.0499</td>
<td>-0.3275</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON3 (6)</td>
<td>0.0289</td>
<td>0.1801</td>
<td>0.1789</td>
<td>-0.3427</td>
<td>-0.3159</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDHOME (7)</td>
<td>-0.3042</td>
<td>0.0139</td>
<td>-0.0048</td>
<td>0.0151</td>
<td>-0.0015</td>
<td>-0.0133</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDWAY (8)</td>
<td>0.2545</td>
<td>0.1047</td>
<td>0.1163</td>
<td>0.0215</td>
<td>-0.0244</td>
<td>0.0095</td>
<td>-0.1118</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERPTSHOMETEAM (9)</td>
<td>-0.4274</td>
<td>0.0167</td>
<td>-0.0071</td>
<td>0.0143</td>
<td>-0.0123</td>
<td>-0.0285</td>
<td>0.6709</td>
<td>-0.1002</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>PERPTSAWAYTEAM (10)</td>
<td>0.3936</td>
<td>0.0761</td>
<td>0.0944</td>
<td>0.0318</td>
<td>-0.0453</td>
<td>-0.0185</td>
<td>-0.1072</td>
<td>0.6812</td>
<td>-0.3345</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
APPENDIX TWO
Table 5.5.1 Second Stage Regression Results (GMM)

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>HTRESULT</th>
<th></th>
<th></th>
<th>Dependent Variable</th>
<th>ATRESULT</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GLSAWAY</td>
<td>-1.509651</td>
<td>-0.50+</td>
<td></td>
<td>Coef</td>
<td>0.72+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHTSONTGTINCGLSAWAY</td>
<td>0.9406478</td>
<td>0.31+</td>
<td>-1.112299</td>
<td>Coef</td>
<td>-1.34+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTSUCPASSAWAY</td>
<td>-0.0022446</td>
<td>-0.93+</td>
<td></td>
<td>Coef</td>
<td>0.06+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TACKWONAWAY</td>
<td>0.0543823</td>
<td>0.76+</td>
<td>-0.5823355</td>
<td>Coef</td>
<td>0.41+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAVESMADEAWAY</td>
<td>-0.3732114</td>
<td>-0.23+</td>
<td></td>
<td>Coef</td>
<td>0.41+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON1</td>
<td>0.020069</td>
<td>0.17+</td>
<td>-0.153833</td>
<td>Coef</td>
<td>-0.12+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON2</td>
<td>0.0104613</td>
<td>0.07+</td>
<td>0.0135492</td>
<td>Coef</td>
<td>0.17+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON3</td>
<td>0.1594889</td>
<td>1.82*</td>
<td>-0.0136899</td>
<td>Coef</td>
<td>-0.13+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDHOME</td>
<td>-0.0067351</td>
<td>-1.67*</td>
<td></td>
<td>Coef</td>
<td>1.37+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDAWAY</td>
<td>0.0039797</td>
<td>0.78+</td>
<td>-0.0090509</td>
<td>Coef</td>
<td>-2.47**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>0.0127409</td>
<td>4.21***</td>
<td></td>
<td>Coef</td>
<td>-3.80***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERPPTSAYWAYTEAM</td>
<td>0.0065147</td>
<td>-2.13**</td>
<td></td>
<td>Coef</td>
<td>4.12***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLSHOME</td>
<td>0.2435308</td>
<td>0.16+</td>
<td>1.233377</td>
<td>Coef</td>
<td>0.86+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHTSONTGTINCGLSHOME</td>
<td>0.3557496</td>
<td>0.23+</td>
<td>-0.6032012</td>
<td>Coef</td>
<td>-0.41+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTSUCPASSHOME</td>
<td>-0.0028338</td>
<td>-3.15+</td>
<td></td>
<td>Coef</td>
<td>-0.88+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TACKWONHOME</td>
<td>-0.0144952</td>
<td>-0.91+</td>
<td></td>
<td>Coef</td>
<td>-0.37+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAVESMADEHOME</td>
<td>-0.9549389</td>
<td>-0.32+</td>
<td></td>
<td>Coef</td>
<td>1.34+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endogenous Variables</td>
<td>Dependent Variable</td>
<td>Coef</td>
<td>z</td>
<td>Dependent Variable</td>
<td>Coef</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------</td>
<td>--------</td>
<td>--------</td>
<td>-------------------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATTTP</td>
<td></td>
<td></td>
<td>DEFTP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coef</td>
<td>-.1366878</td>
<td>-3.04***</td>
<td>Coef</td>
<td>-.0156589</td>
<td>-0.87+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>z</td>
<td>-</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFTP</td>
<td>-.1366878</td>
<td>-3.04***</td>
<td></td>
<td>DEFTP</td>
<td>-.0156589</td>
<td>-0.87+</td>
<td></td>
</tr>
<tr>
<td>ATTTP</td>
<td>-.0156589</td>
<td>-0.87+</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTGLS</td>
<td>1.473206</td>
<td>4.52***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTSHTSONINCGLS</td>
<td>2.638492</td>
<td>14.75***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTASSTS</td>
<td>1.258768</td>
<td>3.55***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTSCUPASS</td>
<td>.5646039</td>
<td>28.12***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTYC</td>
<td>2.113925</td>
<td>4.64***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTRC</td>
<td>-11.79963</td>
<td>-3.11***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATTVDUM</td>
<td>.6499869</td>
<td>1.91*</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFTACKWON</td>
<td>-.8729349</td>
<td>10.74***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFTOTCLE</td>
<td>-.5552967</td>
<td>12.23***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFBLO</td>
<td>-.9418797</td>
<td>7.15***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFINT</td>
<td>.7283325</td>
<td>10.77***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFYC</td>
<td>.290439</td>
<td>0.80+</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFRC</td>
<td>-22.86256</td>
<td>-7.05***</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEFVDUM</td>
<td>-.5143088</td>
<td>-1.82*</td>
<td></td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON 1</td>
<td>-.6329719</td>
<td>-1.29+</td>
<td></td>
<td>-.6170797</td>
<td>-3.82***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON 2</td>
<td>-.6164807</td>
<td>-1.30+</td>
<td></td>
<td>-.1552198</td>
<td>-3.56***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEASON 3</td>
<td>-.6712726</td>
<td>-1.47+</td>
<td></td>
<td>.1405018</td>
<td>0.38+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: + not significant, *p<.10, **p<.05, ***p<.01
### Table 7.5.1 Second Stage Regression Estimates Obtained When Assessing the Impact of the Performance of Positional Units Upon the Performance of the Opposing Team (GMM)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>HTRESULT</th>
<th>HTRESULT</th>
<th>ATRESULT</th>
<th>ATRESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Endogenous Variable</strong></td>
<td>Coef</td>
<td>z</td>
<td>Coef</td>
<td>z</td>
</tr>
<tr>
<td>TPAATT</td>
<td>-.0063653</td>
<td>-2.53**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>-</td>
<td>-.0050237</td>
<td>-3.14***</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHDEF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Independent Variables</strong></td>
<td>Coef</td>
<td>z</td>
<td>Coef</td>
<td>z</td>
</tr>
<tr>
<td>TPHDEF</td>
<td>.0058235</td>
<td>2.60***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPHATT</td>
<td>-</td>
<td>-</td>
<td>.0051841</td>
<td>3.18***</td>
</tr>
<tr>
<td>TPADEF</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TPAATT</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SEASON1</td>
<td>.1021739</td>
<td>1.06+</td>
<td>.0622137</td>
<td>0.64+</td>
</tr>
<tr>
<td>SEASON2</td>
<td>.0780577</td>
<td>0.78+</td>
<td>.0846915</td>
<td>0.86+</td>
</tr>
<tr>
<td>SEASON3</td>
<td>-.0493208</td>
<td>-0.51+</td>
<td>-.0435665</td>
<td>-0.45+</td>
</tr>
<tr>
<td>GDHOME</td>
<td>.0063363</td>
<td>1.74*</td>
<td>.0076641</td>
<td>2.16**</td>
</tr>
<tr>
<td>GDWAY</td>
<td>-.0080281</td>
<td>-2.13**</td>
<td>-.0064301</td>
<td>-1.73*</td>
</tr>
<tr>
<td>PERPTSHOMETEAM</td>
<td>.0234374</td>
<td>9.50***</td>
<td>.0232147</td>
<td>9.36***</td>
</tr>
<tr>
<td>PERPTSAWAYTEAM</td>
<td>-.0154941</td>
<td>-6.24***</td>
<td>-.0167456</td>
<td>-6.74***</td>
</tr>
</tbody>
</table>
APPENDIX THREE

Figure 6.2.3 - Typical 4-4-2 VS 3-5-2 Team Formation Match up
Figure 6.2.4 - Typical 4-4-2 VS 4-3-3 Team Formation Match Up
Figure 6.2.5 - Typical 4-4-2 VS 4-5-1 Team Formation Match Up
Figure 6.2.6 - Typical 4-4-2 VS 5-3-2 Team Formation Match Up
Figure 6.2.7 - Typical 4-4-2 VS 5-4-1 Team Formation Match Up
Figure 6.2.8 - Typical 4-4-2 VS 4-1-4-1 Team Formation Match Up
Figure 6.2.9 - Typical 4-4-2 VS 4-2-3-1 Team Formation Match Up
Figure 6.2.10 - Typical 4-4-2 VS 4-3-2-1 Team Formation Match Up
Figure 6.2.11 - Typical 4-4-2 VS 4-4-1-1 Team Formation Match Up
Figure 6.2.12 - Typical 4-4-2 VS 4-1-2-1-2 Team Formation Match Up
Figure 6.2.13 - Typical 4-5-1 VS 4-3-2-1 Team Formation Match Up
Figure 6.2.14 - Typical 4-1-4-1 VS 4-1-4-1 Team Formation Match Up
Figure 6.2.15 - Typical 4-1-4-1 VS 4-4-1-1 Team Formation Match Up
Figure 6.2.16 - Typical 4-2-3-1 VS 3-4-3 Team Formation Match Up
Figure 6.2.17 - Typical 4-2-3-1 VS 3-5-2 Team Formation Match Up
Figure 6.2.18 - Typical 4-2-3-1 VS 4-3-3 Team Formation Match Up
Figure 6.2.19 - Typical 4-2-3-1 VS 4-5-1 Team Formation Match Up
Figure 6.2.20 - Typical 4-2-3-1 VS 4-1-4-1 Team Formation Match Up
Figure 6.2.21 - Typical 4-2-3-1 VS 4-2-3-1 Team Formation Match Up
Figure 6.2.22 - Typical 4-2-3-1 VS 4-3-2-1 Team Formation Match Up
Figure 6.2.23 - Typical 4-4-1-1 VS 3-4-3 Team Formation Match Up
Figure 6.2.24 - Typical 4-4-1-1 VS 4-2-3-1 Team Formation Match Up
Figure 6.2.25 - Typical 4-4-1-1 VS 4-3-2-1 Team Formation Match Up
Figure 6.2.26 - Typical 4-4-1-1 VS 4-4-1-1 Team Formation Match Up
Figure 6.2.27 - Typical 4-1-2-1-2 VS 3-5-2 Team Formation Match Up
Figure 6.2.28 - Typical 4-1-2-1-2 VS 4-3-3 Team Formation Match Up
Figure 6.2.29 - Typical 4-1-2-1-2 VS 4-5-1 Team Formation Match Up
Figure 6.2.30 - Typical 4-1-2-1-2 VS 5-4-1 Team Formation Match Up
Figure 6.2.31 - Typical 4-1-2-1-2 VS 3-4-2-1 Team Formation Match Up
Figure 6.2.32 - Typical 4-1-2-1-2 VS 4-1-4-1 Team Formation Match Up
Figure 6.2.33 - Typical 4-1-2-1-2 VS 4-2-2-2 Team Formation Match Up
Figure 6.2.34 - Typical 4-1-2-1-2 VS 4-2-3-1 Team Formation Match Up
Figure 6.2.35 - Typical 4-1-2-1-2 VS 4-3-2-1 Team Formation Match Up
Figure 6.2.36 - Typical 4-1-2-1-2 VS 4-4-1-1 Team Formation Match Up
Figure 6.2.37 - Typical 4-1-2-1-2 VS 4-1-2-1-2 Team Formation Match Up