2009

Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success

Joel Oberstone
University of San Francisco, joel@usfca.edu

Recommended Citation
Joel Oberstone (2009) "Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success," Journal of Quantitative Analysis in Sports: Vol. 5 : Iss. 3, Article 10.

Journal of Quantitative Analysis in Sports

Volume 5, *Issue* 3 2009 *Article* 10

Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success

Joel Oberstone, *University of San Francisco*

Recommended Citation:
Joel Oberstone (2009) "Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success," *Journal of Quantitative Analysis in Sports*: Vol. 5 : Iss. 3, Article 10.
Available at: http://www.bepress.com/jqas/vol5/iss3/10
DOI: 10.2202/1559-0410.1183

Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success

Joel Oberstone

Abstract

This paper develops a robust, statistically significant, six independent variable multiple regression model that accounts for the relative success of English Premier League football clubs based on an array of twenty-four pitch actions collected during the 2007-2008 season ($p<0.0000$). Additionally, one-way ANOVA is used to identify those specific pitch actions that statistically separate the top 4 clubs from (1) the dozen clubs forming the middle of the pack and, (2) by a greater contrast, the bottom 4 clubs. Thirteen pitch actions yield statistically significant differences among these three tiers of clubs using this second method of analysis.

KEYWORDS: English Premier League, football, team performance analysis, multiple regression, ANOVA, retrodiction

INTRODUCTION

Unlike professional American sports, football in England—as well as in many leagues in Europe and South America—has grave outcomes for poorly performing clubs. At the end of each English Premier League (EPL) 38-game season, the top football clubs get rich(er), while the bottom three are *relegated*—unceremoniously booted out of the league: demoted to a lower division euphemistically called the Coca-Cola Championship.[1] In addition to being downgraded to a lower league status, a relegated EPL club will experience an accompanying dramatic decline in income resulting from the loss of league television revenue sharing, a drop in club merchandise sales, and dramatically lower attendance. Further, there will be accompanying losses of star personnel—players whose salaries the club could no longer support.[2]

The Deloitte Annual Review of Football Finance 2008 report estimates that English Premier League teams will have generated revenues of approximately £1.9 billion for the 2007-2008 season.[3] Television alone provides approximately £900 million that are proportionally divided among the Premiership teams according to their place in the final league standings. First place Manchester United received £49.3m as compared to relegated and last place Derby County's £29.1m. However, this season, Derby County will not enjoy this sizable financial influx. Instead, they will have to cope with the dramatically smaller television revenue pot of the Championship. Teams relegated, including two parachute payments of £11.9 million over the following two seasons, will still experience a net loss of £30 million (James [2007]).

In addition to the English Premier League's associated rewards, a club's final place in the league table determines their eligibility to play in prestigious and extremely lucrative cup matches such as the UEFA Cup and the Champions League, e.g., only the top four teams are invited to play in the UEFA Champions League games—the most celebrated of cup challenges. The Champions League, for example, pays clubs between €2 million and €15 million of fixed sums, and up

---

[1] Relegation is identical to what would happen if, say, the three worst teams in Major League Baseball—the ones with the lowest win-loss percentages during each season—were demoted to Triple A baseball.

[2] Even if players on relegated clubs were offered the same salaries usually supported only in the Premiership, many would not want to play at anything except the highest, most prestigious level possible. Championship football is, often, not acceptable for players with options.

[3] http://www.deloitte.com/dtt/article/0,1002,sid=70402%2526cid=56148,00.html

to €40 million in bonus reward based on considerations such as the value of the national market and number of games played.[4]

Given these serious performance implications, how does a team identify the kinds of performance that most dramatically affect their success during the English Premier League season? To be blunt, what differentiates a top-notch team from an also ran? What, exactly, do the better teams do on the pitch that differentiates them from the others? Are there specific offensive and defensive pitch actions that are more influential than others in a team's achievement over the season? The following section describes the unique problems associated with analyzing football data, how they have historically affected the body of analysis on the subject as compared to other sports, and the data sources used to make these analyses possible.

The Challenge of Analyzing Football Performance

Although football analysis has made important contributions to understanding the sport's performance, the vast body of analysis has been applied to the primary professional sports in the United States—baseball, basketball, and American football (Anderson, Edman and Ekman [2005], Bennett and Flueck [1983], Hirotsu and Wright [2002], Boulier and Steckler [2003], Ammar and Wright [2004/2005], Lewis [2004]). In large measure, these sports are constructed with separable, concise actions and numerous, built-in breaks, that are relatively easy to analyze. Conversely, it is the unique qualities of football that makes it extremely difficult to dissect because it is a game intended to be ongoing, a continuous flow of action with virtually no down time. Timeouts are not part of the game. There are no "breaks" between plays, no manager conferences, and the game actions only stop at the conclusion of each half or for a serious player injury. Only three substitutions are allowed per game, no opportunity for the players to huddle and discuss strategies between plays (or to catch their breath), and no changeover between the instantaneous change of possession and transition between offense and defense. Football is, in some ways, a metaphor for life: it simply keeps on going until it's over. In fact, one of the commercial problems that football presents is how to unobtrusively insert advertising during each half of play without any scheduled breaks.[5] Unmistakably, football performance analysis

[4] http://www.uefa.com/UEFA/index.html

[5] Many football clubs use continuously running, animated advertisements projected on the special, electronic facing surface of the field-level stadium stands.

DOI: 10.2202/1559-0410.1183

poses significant challenges: it is the will-o'-the-wisp, the moving target, the thumb-to-the-nose of sports analysis.

DATA SOURCES FOR EPL FIXTURES

To date, comprehensive performance data documenting detailed game actions of each player on the pitch are available, albeit primarily for commercial use. Several of the top organizations that record and dissect post-match data include Opta Sportsdata, PA Sport-Actim Index, and Prozone.[6] Each of these groups meticulously videotape and analyze the thousands of actions that take place during a game and then market these data to the groups described in the following section. Assessments of the relative effectiveness of these organizations in analyzing player performance includes Barros and Leach [2006], Bradley, et al. [2007], Espitia-Escuer et al. [2006], McHale [2007], McHale and Scarf [2007], O'Donoghue [2002], Reed and Hughes [2006], Szczepański [2008], Valter et al. [2006].

Conversely, the analysis of *team* performance is relatively sparse when contrasted with *player* performance. This preference is exemplified by the well-known Opta and Actim Indexes that "grade" player performance and the Prozone Manager that concentrates primarily, but not exclusively, on the player movement (work rate) and field location throughout the game. Even though measures of team performance are collected, the primary focus is on individual player accomplishments. Studies that have examined team performance include (Barros and Leach [2006], Carmichael and Ward [2000], Crowder et al. [2002], Hirotsu and Write [2002, 2003], Hope [2003], Fitt, and Kabelka [2006], Jones and Mellalieu [2004]).

The work of Carmichael and Ward is an especially engaging exception: it examines English Premier League *team* performance based on the relationship between input (the skills and other characteristics of the team members) and output (a victory or loss) in head-to-head competitions between clubs. An example of their performance function is shown in the following linear equation:

$$Y_k = \alpha H + \sum_{i=1}^{n} \beta x_i + \sum_{j=1}^{20} \mu F_j + e \qquad (1)$$

---

[6] PA Sport-Actim Index, http://www.pa-sport.com/en/actim/actim-index-premier.html; Opta Index, http://www.optasportsdata.com; ProZone, http://www.prozonesports.com/index.htm

where Y_k is the match score for the k^{th} match in terms of the final score goal differential, dichotomous variables, H, indicating whether the observed team is playing at home, play differences between the two teams, x_j, on 20 independent, dichotomous variables, F_j, such as shots on target, blocked shots, and percentage of successful passes. In general, the measures use score differentials rather than ratios. However, no paper uses summative season long performance comparisons in an attempt to identify specific team performance criteria—pitch actions—that might be used to either (1) better understand the factors associated with a team's success over a season or (2) separate the top clubs from the others based on significantly different pitch action performance.

Data collected by Opta Sportsdata for the complete 2007-2008 English Premier League season is used in this paper to uncover these pivotal pitch actions. The analysis follows.

METHOD OF ANALYSIS

Other than the obvious outcome of winning more games than an opponent, little attention has been given to identifying what the better football clubs do differently from the average or poorly performing clubs over a seasons performance. Is it possible to pinpoint a set of specific actions that can be used to retrodict the number of games an EPL team wins? Are there specific actions that individually pose statistically significant differences between the best, middle, and worst tier of English Premier League teams?

Out of the hundreds of different pitch actions, football team performance can be generally organized into the five families illustrated in Figure 1: (1) goal attempts, (2) passing, (3) defending, (4) crossing, and (5) discipline. Because these action groups are described in less-than-explicit terms, each will be broken into finer, detailed actions. An illustration of the expanded set of twenty-four (24) actions is shown in Figure 2.

In order to be able to analyze the different Premiership teams, performance data from league games needed to be gathered. There are several, well-established, prestigious organizations that are renowned for their post-match data collection and analysis. These include (1) PA Sport-Actim, (2) Prozone, and (3) Opta Sportsdata. Each of these organizations meticulously videotape and analyze English Premier League games. These data have significant commercial value in betting and odds making and are purchased by: (1) Newspapers, e.g., News International, Mirror Group Newspapers, Daily Mail, and The Guardian; (2) Television and Internet media, e.g., ESPN, Yahoo!, Setanta Sports, Sky Italia, and

DOI: 10.2202/1559-0410.1183

Figure 1. Team Pitch Action Groups.

BSkyB, (3) Mobile phone carriers; and (4) English Premier League football clubs that use the information to "grade" their team and player performances and those of their opponents.

Premiership clubs will have preferences for which of these organizations to employ, however all three are widely viewed by them as legitimate, prestigious sources of football data. Although their specific data sources vary, all provide post-match analysis by recording the hundreds of different pitch actions that occur during the match, including passes, tackles, assists, fouls, shot on goals, blocks, offside, clearances, and saves. Even greater detail is recorded including such measures as the pitch location of the action, whether the player involved used his left or right foot, and, if the action is a pass, the direction and distance of the play. These data are organized and tabulated into relevant statistics that can be used to primarily profile player productivity and, secondarily, team performances. There is no shortage of data, however complications arise in trying to make sense out of this daunting volume of numbers. The next section describes several analysis methods that provide insight to club performance. More specifically,

- A linear multiple regression analysis is conducted to see if it is possible to identify a retrodictive set of pitch performance actions that form a statistically significant model of a club's final league standings (Groebner [2008], Levine et al. [2008], Schleifer and Bell [1995]).

- Separate from a regression model, individual pitch actions will also be examined to see which ones, if any, suggest important differences between the best teams (top four), worst teams (bottom four), and the best of the rest (middle 12 teams) using analysis of variance (ANOVA).[7]

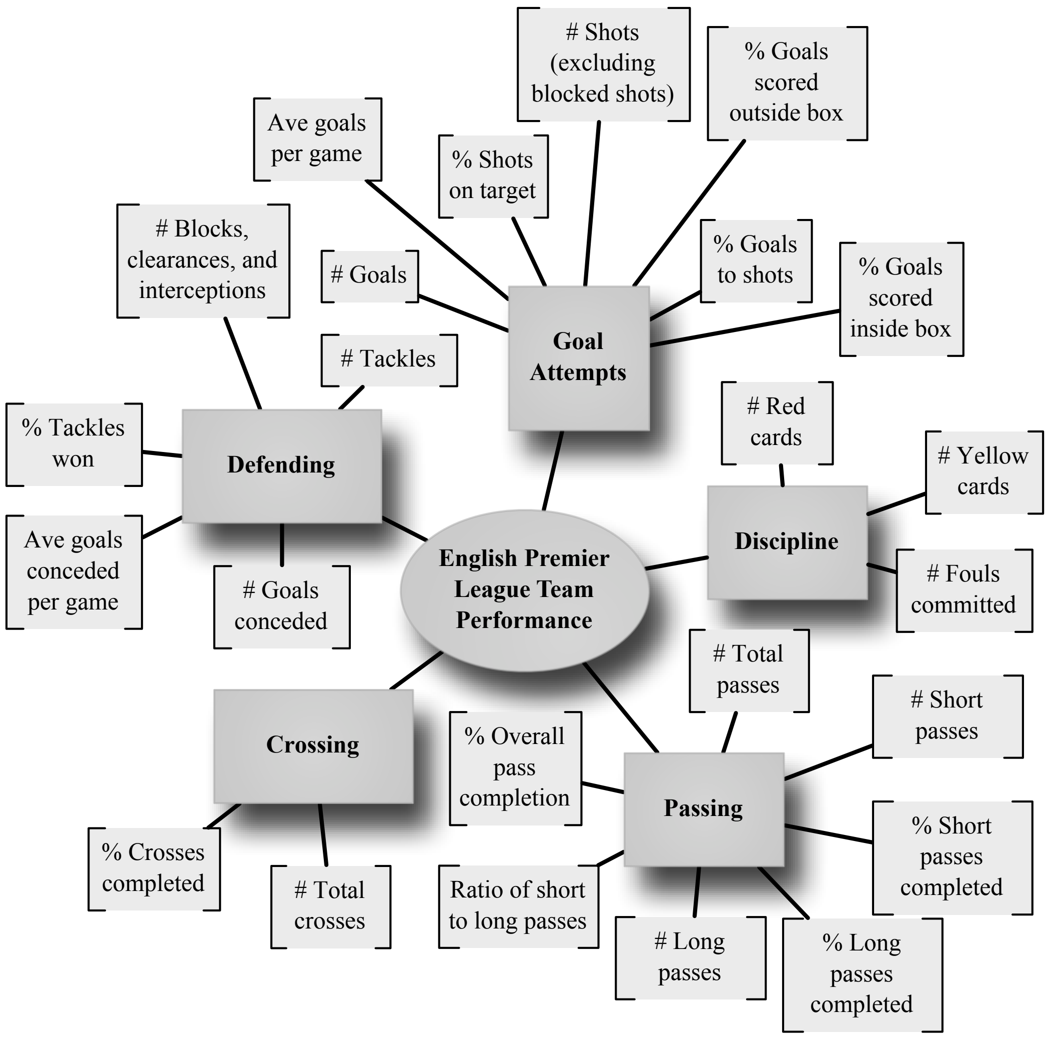


Figure 2. OPTA Team Performance Actions.

---

[7] The division of the 20 teams into three tiers is arbitrary but with a purpose. The top four teams automatically qualify for the single most prestigious cup competition: the UEFA Champions League. The bottom four teams include the three that will be relegated to the lower division Championship, leaving the rest as the widely varying middle group.

DOI: 10.2202/1559-0410.1183

BUILDING A MULTIPLE REGRESSION MODEL

Because we are primarily interested in being able to understand reasons for the differences between varying levels of club achievement in terms of pitch performance, an obvious choice for the dependent variable (degree of success), Y, would be the league points earned during the 38 game season. After that comes the task of identifying what group of specific pitch actions (independent variables), if any, contribute statistically significant associations in explaining the league points earned over the season.

A careful review of the original 24 pitch actions of the five groups of measures shown in Figure 2 discloses pairs of overlapping or redundant measures, e.g., since every team plays a 38-game season, the number of goals scored (X_1) and average goals per game (X_2) provide the same information. The same can be said of the percent of goals scored inside the box (X_6) and the percent of goals scored outside the box (X_7). After eliminating other similar features, the original cluster of 24 variables is cautiously reduced to the 17 independent variables shown in Table 1.[8] The multiple regression results for the 17-variable model are presented in Table 2.[9]

Although the model is statistically significant (p=0.0177), an examination of the t-values suggests that only one independent variable makes a statistically significant contribution to the overall model: average goals conceded per game, X_{18} (p=0.022). This leads to the suspicion that multicollinearity issues are present. The correlation matrix for the independent variable set supports this reservation as evidenced by the numerous performance factors with r-values>0.75 shown in bolded text (Table 3). We still need to refine the analysis by identifying the weakest independent based on the highest p-value>0.05. The technique is called backward elimination in stepwise regression [Groebner, et al., 2008].

With each successive removal, the model is increasingly refined until it is comprised of only statistically significant variables. Care is taken at each step to insure that multicollinearity is not an issue as well. After approximately one dozen of these refinements, the set of 17 pitch measures is pared down to a set of 6 statistically significant pitch actions: (1) X_5 (% goals to shots); (2) X_7 (% goals scored outside of box); (3) X_{11} (ratio of short/long passes); (4) X_{15} (total crosses); (5) X_{18} (average goals conceded per game); and (6) X_{23} (yellow cards).

---

[8] The original set of 24 pitch action variables is shown in Appendix A.

[9] Since the model sample size is limited to the twenty (n=20) English Premier League teams, the maximum number of k-independent variables (pitch actions) we can use in this study is n-k-1 ≥ 2 or k=17.

	Points	Ave goals per game	# Shots on goal	% Shots on Target	% Goals to Shots	% of goals scored outside box	# Total passes	Ratio of short/long passes	Overall Pass Completion %	# Total Crosses	Cross Completion %	Ave goals conceded per game	# Tackles	% Tackles Won	# Blocks, Clearances & Interceptions	# Fouls	# Yellow Cards	# Red Cards
Football Club	**Y**	**X2**	**X3**	**X4**	**X5**	**X7**	**X10**	**X11**	**X12**	**X15**	**X16**	**X18**	**X19**	**X20**	**X21**	**X22**	**X23**	**X24**
Arsenal	83	1.95	473	43.97%	15.64%	10.81%	18,831	8.68	83.18%	883	24.01%	0.82	878	74.72%	2,438	407	55	3
Aston Villa	60	1.87	405	44.94%	17.53%	16.90%	11,438	4.77	69.76%	926	25.59%	1.34	885	79.77%	2,826	559	54	4
Birmingham City	35	1.21	314	45.54%	14.65%	21.74%	10,630	3.79	63.74%	769	22.50%	1.63	935	76.58%	2,809	546	70	3
Blackburn	58	1.32	421	43.94%	11.88%	20.00%	13,650	4.98	74.03%	874	22.43%	1.26	833	74.67%	2,736	583	72	6
Bolton	37	0.95	340	43.82%	10.59%	11.11%	11,855	4.48	67.85%	901	23.53%	1.42	866	76.44%	2,859	547	76	0
Chelsea	85	1.71	455	43.74%	14.29%	18.46%	17,250	7.01	81.16%	937	20.70%	0.68	917	76.23%	2,302	478	63	5
Derby County	11	0.53	288	40.28%	6.94%	20.00%	11,733	4.42	67.30%	798	21.30%	2.34	898	72.72%	2,685	548	63	1
Everton	65	1.45	361	46.81%	15.24%	9.09%	12,397	4.79	69.32%	868	23.27%	0.87	816	77.21%	2,773	494	40	3
Fulham	36	1.00	367	41.96%	10.35%	15.79%	12,574	4.44	70.77%	842	22.09%	1.58	851	76.62%	2,661	493	55	6
Liverpool	76	1.76	535	40.00%	12.52%	19.40%	16,360	4.93	77.06%	922	25.27%	0.74	1,089	73.74%	2,651	465	45	1
Manchester City	55	1.18	323	44.27%	13.93%	17.78%	15,147	5.62	76.23%	691	21.27%	1.39	916	77.62%	2,492	427	50	4
Manchester United	87	2.11	547	47.53%	14.63%	17.50%	17,417	6.21	80.50%	930	24.41%	0.58	978	76.18%	2,574	425	51	2
Middlesbrough	42	1.13	352	43.18%	12.22%	13.95%	12,199	4.21	71.01%	946	23.68%	1.39	835	75.45%	2,805	562	86	2
Newcastle United	43	1.18	358	42.74%	12.57%	17.78%	13,947	4.76	72.64%	957	19.85%	1.71	876	76.14%	2,516	483	63	1
Portsmouth	57	1.26	411	38.93%	11.68%	20.83%	12,793	4.62	73.50%	654	21.87%	1.05	963	75.91%	2,733	541	55	3
Reading	36	1.08	332	39.76%	12.35%	12.20%	10,431	3.74	64.35%	1072	22.67%	1.74	800	74.38%	2,369	492	59	5
Sunderland	39	0.95	358	38.27%	10.06%	11.11%	12,727	4.67	70.07%	882	24.94%	1.55	892	74.33%	2,831	517	65	4
Tottenham	46	1.74	424	45.99%	15.57%	9.09%	15,464	5.37	77.43%	987	22.19%	1.61	960	74.58%	2,708	410	51	1
West Ham United	49	1.11	384	40.10%	10.94%	11.90%	13,762	4.82	74.11%	904	24.23%	1.32	848	74.76%	2,546	562	63	1
Wigan Athletic	40	0.89	363	38.84%	9.37%	14.71%	11,672	4.42	68.87%	756	25.00%	1.34	883	78.94%	2,504	527	59	4

Table 1. Opta Sportsdata EPL Team Pitch Actions for 2007-2008 Season.

DOI: 10.2202/1559-0410.1183

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.999
R Square	0.998
Adjusted R Squa	0.980
Standard Error	2.787
Observations	20.000

ANOVA	*df*	*SS*	*MS*	*F*	*Significant F*
Regression	17	7327.910	1046.844	55.935	0.0177
Residual	2	72.090	6.008		
Total	19	7400			

	Coeff	*Std Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-278.270	207.320	1.342	0.312	-1,170.400	613.830
X2	-139.660	106.920	1.306	0.322	-599.740	320.410
X3	0.362	0.327	1.108	0.383	-1.044	1.769
X4	79.583	101.690	0.783	0.516	-357.990	517.160
X5	1,226.000	847.550	1.446	0.285	-2,421.100	4,873.000
X7	273.080	135.650	2.013	0.182	-310.630	856.790
X10	-7.963	7.621	1.045	0.406	-40.755	24.829
X11	12.692	8.186	1.550	0.261	-22.534	47.918
X12	282.660	196.160	1.441	0.286	-561.410	1,126.700
X15	0.103	0.049	2.106	0.170	-0.107	0.312
X15	333.010	209.490	1.590	0.253	-568.440	1,234.500
X18	-35.609	5.327	6.684	0.022	-58.533	-12.685
X19	0.015	0.010	1.520	0.268	-0.027	0.057
X20	-147.350	118.590	1.243	0.340	-657.620	362.930
X21	11.935	10.729	1.112	0.382	-34.234	58.103
X22	-0.065	0.078	0.833	0.493	-0.400	0.270
X23	-0.582	0.255	2.277	0.151	-1.680	0.517
X24	-1.201	1.210	0.992	0.426	-6.409	4.008

Table 2. Multiple Regression Model for Set of 17 Team Performance Actions (k=17 independent variables)

	X2	*X3*	*X4*	*X5*	*X7*	*X10*	*X11*	*X12*	*X15*	*X16*	*X18*	*X19*	*X20*	*X21*	*X21*	*X23*	*X24*
X2	1.000																
X3	**0.839**	1.000															
X4	0.581	0.222	1.000														
X5	**0.839**	0.420	0.727	1.000													
X7	-0.032	0.076	-0.075	-0.116	1.000												
X10	0.698	**0.776**	0.296	0.377	-0.024	1.000											
X11	0.648	0.601	0.332	0.437	-0.133	**0.889**	1.000										
X12	0.703	**0.782**	0.269	0.386	-0.014	**0.971**	**0.870**	1.000									
X15	0.315	0.257	0.179	0.254	-0.450	0.129	0.062	0.081	1.000								
X16	0.262	0.347	-0.113	0.103	-0.322	-0.032	-0.015	-0.027	0.137	1.000							
X18	**-0.757**	**-0.809**	-0.308	-0.508	0.026	-0.658	-0.595	-0.677	-0.036	-0.334	1.000						
X19	0.431	0.589	-0.005	0.147	0.426	0.470	0.190	0.423	-0.195	0.134	-0.349	1.000					
X20	0.158	-0.098	0.317	0.393	0.039	-0.213	-0.059	-0.131	-0.276	0.162	-0.207	-0.149	1.000				
X21	-0.177	-0.224	0.132	-0.060	-0.024	-0.478	-0.492	-0.440	-0.179	0.312	0.223	0.025	0.061	1.000			
X21	-0.552	-0.460	-0.335	-0.439	0.253	-0.726	-0.635	-0.637	-0.168	0.127	0.356	-0.387	0.077	0.500	1.000		
X23	-0.455	-0.392	-0.140	-0.370	0.095	-0.353	-0.276	-0.323	0.103	-0.100	0.367	-0.381	-0.153	0.248	0.613	1.000	
X24	0.002	-0.051	-0.069	0.090	0.169	-0.112	0.048	-0.050	-0.123	-0.077	-0.091	-0.350	0.286	-0.239	0.074	-0.060	1.000

Table 3. Correlation Matrix Indicating Possible Multicollinearity Problems Between Pairs of Independent Variables for 2007-2008 EPL Season (17 Pitch Factors).

The coefficients of the 6-variable regression model produced in Table 4 can now be used to estimate English Premier League team league points earned, $\hat{Y}$. The model equation is:

$$\begin{aligned}\hat{Y} &= 30.999 + 99.231X_5 + 80.159X_7 + 4.471X_{11} \\ &\quad + 0.029X_{15} - 31.708X_{18} - 0.161X_{23}\end{aligned} \tag{2}$$

The coefficient of any single independent variable indicates the average rate of change in the team's success (points earned) *when all other variables remain constant.* Consequently, a team that (1) increases the percent goals to shots accuracy, X_5, by 3% would have an accompanying gain of about 3 table points (0.03 x 99.231); (2) increases the percentage of goals scored outside of the box, X_7, by 5% will improve its table position by about 4 points (0.05 x 80.159); (3) decreases the ratio of short/long passes, X_{11}, from, say, 6 to 4 (a two-point drop) will result in a loss of approximately 9 table points (2 x 4.471); (4) increases the number of crosses during the season, X_{15}, by 100, would generate a gain of about

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.995					
R Square	0.990					
Adjusted R Square	0.985					
Standard Error	2.407					
Observations	20.000					
ANOVA	*df*	*SS*	*MS*	*F*	*Sig F*	
Regression	6	7324.696	1220.783	210.748	**0.0000**	
Residual	13	75.304	5.793			
Total	19	7400.000				
	Coeff	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**30.999**	8.992	3.447	**0.004**	11.573	50.424
X5	**99.231**	27.573	3.599	**0.003**	39.663	158.799
X7	**80.159**	15.503	5.171	**0.000**	46.667	113.651
X11	**4.471**	0.618	7.230	**0.000**	3.135	5.806
X15	**0.029**	0.007	4.329	**0.001**	0.014	0.043
X18	**-31.708**	1.739	-18.235	**0.000**	-35.465	-27.951
X23	**-0.161**	0.058	-2.771	**0.016**	-0.287	-0.036

Table 4. Multiple Regression Analysis of Final 6-Variable Model.

DOI: 10.2202/1559-0410.1183

of 2.9 points in the table (100 x 0.029); (5) allows a 0.50 increase in the average goals conceded per game, X_{18}, loses almost 16 points (0.50 x -31.708); (6) gives up 10 less yellow cards, X_{23}, over the season would gain an average of 1.61 points in the table (10 x -0.161). The final regression equation establishes a strong overall model (p<0.0000), a statistically significant set of contributions by the independent variables (p<< 0.05), and a correlation matrix and variation inflation factor (VIF) values that indicates that multicollinearity is not evident, i.e., r < 0.60 (Table 5).[10]

As a test of the regression model accuracy, the 6 factor values for the 20 English Premier League teams of the 2007-2008 season are substituted into the model equation and the retrodiction of the points earned are calculated in Table 6. The fit of the retrodicted data for the actual 2007-2008 English Premier League season performance delivers an R^2=0.990 and p<0.0000, and serves as compelling evidence of the model accuracy. These results are also plotted in a scatter diagram (Figure 3) and a complimentary radar chart (Figure 4). It is of special importance that the extremely low error held across the entire range of team performance—including the historically poor performance of last place Derby County's 11 points.[11] The radar chart of Figure 4 captures another view of the strong model fit for the actual versus retrodicted values of season league points for each of the 20 Premiership teams illustrating nearly perfectly overlapping plots. The residual plot for these data shown in Figure 5 visually supports the suitability of the linear model, however a more formal test is also performed that does, indeed, support the assumption of uniform error variance in Appendix B (Newbold, et al., 2006]). The influence that these 6 pitch actions play in competitions is sometimes subtle but always influential. Contributions that these actions can typically make in a fixture follows.

	X5	*X7*	*X11*	*X15*	*X18*	*X23*	*VIF*
X5	1.000						1.635
X7	-0.116	1.000					1.319
X11	0.437	-0.133	1.000				1.643
X15	0.254	-0.450	0.062	1.000			1.436
X18	-0.508	0.026	-0.595	-0.036	1.000		1.837
X23	-0.370	0.095	-0.276	0.103	0.367	1.000	1.311

Table 5. Correlation Matrix and VIF Values for Final 6-Variable Regression Model.

[10] It is generally accepted that correlation values between independent variables of ≤ |0.80| and VIF values ≤5 indicate that multicollinearity is not a problem.

[11] The 11 point total was the fewest points earned in the history of the Premiership.

	Actual Points	% Goals to Shots	% of goals scored outside box	Ratio of short/long passes	% cross completion	Ave goals conceded per game	# Yellow Cards	Predicted Points
Football Club	***Y***	**X5**	**X7**	**X11**	**X15**	**X18**	**X23**	$\hat{Y}$
Arsenal	**83**	15.64%	10.81%	8.68	883	0.82	55	**84.516**
Aston Villa	**60**	17.53%	16.90%	4.77	926	1.34	54	**58.605**
Birmingham City	**35**	14.65%	21.74%	3.79	769	1.63	70	**38.500**
Blackburn	**58**	11.88%	20.00%	4.98	874	1.26	72	**54.958**
Bolton	**37**	10.59%	11.11%	4.48	901	1.42	76	**38.477**
Chelsea	**85**	14.29%	18.46%	7.01	937	0.68	63	**86.374**
Derby County	**11**	6.94%	20.00%	4.42	798	2.34	63	**11.979**
Everton	**65**	15.24%	9.09%	4.79	868	0.87	40	**66.236**
Fulham	**36**	10.35%	15.79%	4.44	842	1.58	55	**39.273**
Liverpool	**76**	12.52%	19.40%	4.93	922	0.74	45	**75.832**
Manchester City	**55**	13.93%	17.78%	5.62	691	1.39	50	**51.649**
Manchester United	**87**	14.63%	17.50%	6.21	930	0.58	51	**87.122**
Middlesbrough	**42**	12.22%	13.95%	4.21	946	1.39	86	**41.909**
Newcastle United	**43**	12.57%	17.78%	4.76	957	1.71	63	**42.065**
Portsmouth	**57**	11.68%	20.83%	4.62	654	1.05	55	**56.212**
Reading	**36**	12.35%	12.20%	3.74	1072	1.74	59	**36.249**
Sunderland	**39**	10.06%	11.11%	4.67	882	1.55	65	**35.822**
Tottenham	**46**	15.57%	9.09%	5.37	987	1.61	51	**45.883**
West Ham United	**49**	10.94%	11.90%	4.82	904	1.32	63	**46.953**
Wigan Athletic	**40**	9.37%	14.71%	4.42	756	1.34	59	**41.389**
Coefficients	***30.999***	***99.231***	***80.159***	***4.471***	***0.029***	***-31.708***	***-0.161***	

Table 6. Regression Model Retrodiction of EPL League Points Earned.

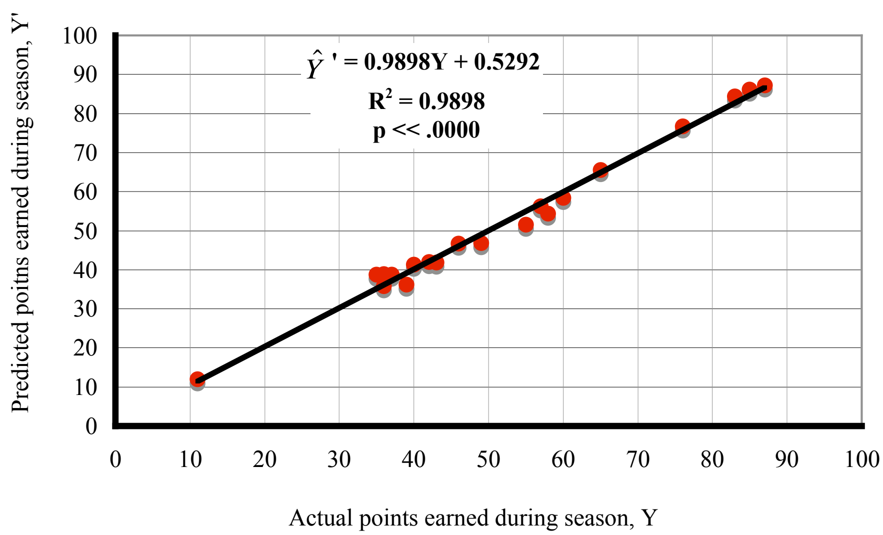


Figure 3. Scatter Diagram of 6-Variable EPL Multiple Regression Model.

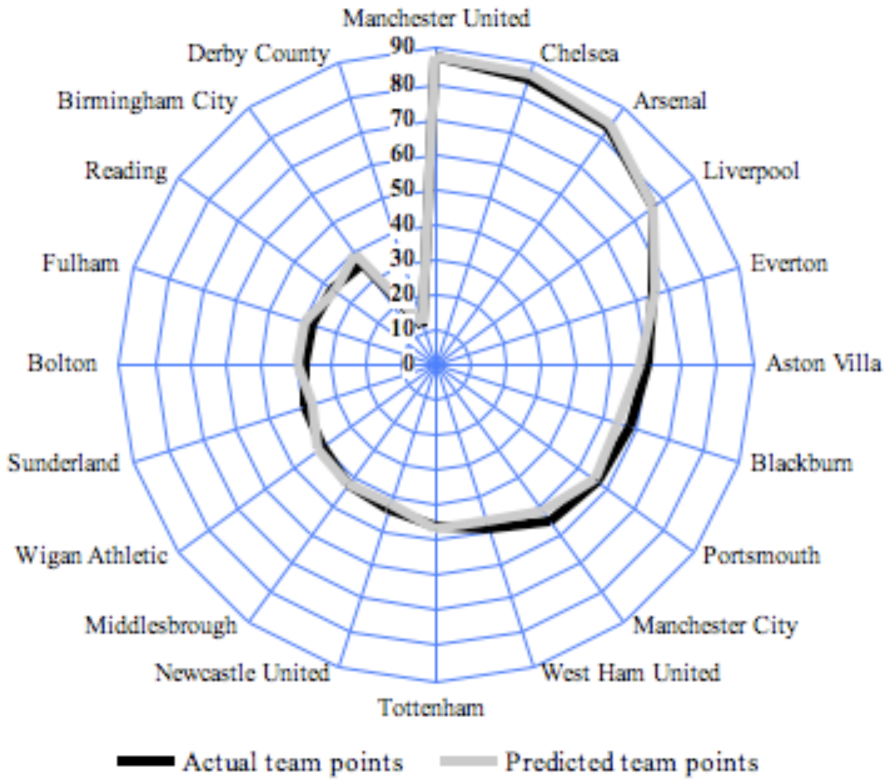


Figure 4. Comparison of EPL Team Performance for 6-Variable Multiple Regression Model.

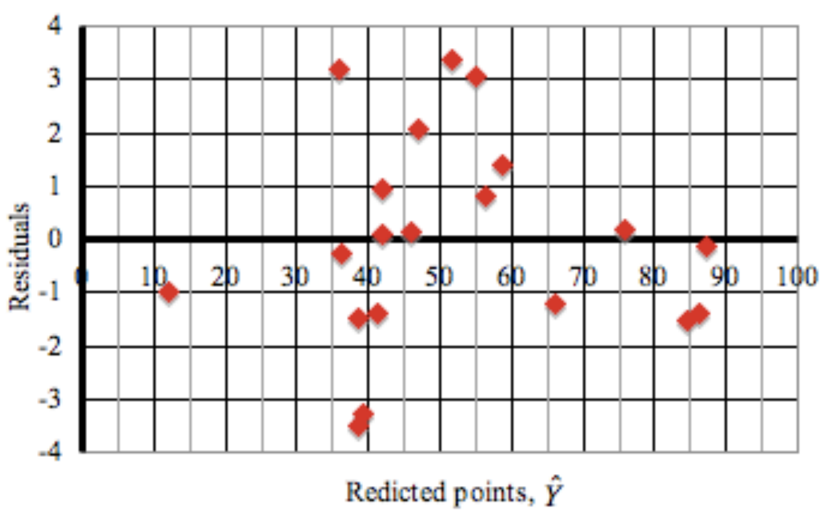


Figure 5. Residual Plot for Retrodicted EPL 2007-2008 Season Points

Shooting accuracy—the percentage of goals scored to shots on goal, X_5—establishes the ability of the better teams to take advantage of the typically rare scoring opportunities. It is not unusual for even the winning team to get less than a handful of goal attempts during any given fixture, i.e., two of the most common scores of an EPL game is 1-0 and 2-1.

When teams can score from distance as well as from close range, the complexity of defending them increases dramatically. *Goals coming from outside the box*, $\underline{X}_7$, force defenders to play in a slightly more advanced position—to defend further away from the goal—because of this varied threat. This can, in turn, create more space inside the box for "poaching" or scoring easy goals by teammates. However, it is important to note that scoring a large proportion of shots from the outside is not a panacea if the team is not scoring a sufficient number of goals, i.e., the most unsuccessful team during the 2007-2008 EPL season, Derby County, had one of the highest percentages of goals scored from outside the box. The problem was that they simply did not score many goals.

Teams that use higher ratios of *short-to-long passes*, X_{11}, employ a more reliable way of advancing the ball than teams that move the ball with a higher proportion of less accurate, long passes. An additional benefit of the preference for the shorter passing game is that it typically keeps possession of the ball for longer periods of time—another important ingredient for both scoring goals while, at the same time, denying the opponent possession and scoring opportunities.[12] However, in order to effectively implement the short passing game, a more skilled, creative player with excellent conditioning and pace to support the constant movement and overlapping runs required in the "short passing game" is needed.

Teams that generate a greater *number of crosses*, X_{15}—typically intended for teammates in advanced positions—can create more immediate scoring opportunities and can improve the quality of shots taken by the *suddenness* of the advancement if the cross is successful. It is important to keep in mind, however, that no more than approximately 25 percent of all crosses are controlled by the offensive team.

Arguably, one of the most important features of a successful team is the ability to defend, as evidenced by a very low value for number of goals allowed per game, X_{18}. With the likelihood that most games will be low scoring affairs, this factor is crucial.

Lastly, when a player commits a foul that results in a *yellow card,* X_{23}, the team not only loses possession of the ball: the opponent is awarded a free kick that may create a potential scoring opportunity. Further, the yellow card can also result in a player ejection (red-carded) if it is the second card infraction received during the fixture. Having to play "one man down"is often a kiss of death in a football competition.[13]

USING ANOVA TO DIFFERENTIATE PITCH ACTIONS AMONG THE TOP, MIDDLE, AND BOTTOM TIER EPL CLUBS

Analysis of variance (ANOVA) is used to explore the original set of 24 pitch actions to see if statistically significant *differences* exist between the EPL clubs

[12] A moderate combination of four or five short, one or two-touch passes will maintain possession considerably longer than a single, long pass.

[13] Even if a player receives a single yellow card in a competition, his accumulation of yellow cards across a number of games may result in him being ineligible for one or more games in the future.

DOI: 10.2202/1559-0410.1183

forming the top, bottom, and middle of the final league table. The composition of these three subdivisions is:

- The *top four* teams qualify for the most prestigious cup competition in the world, i.e., UEFA Champions League (Manchester United, Chelsea, Arsenal, and Liverpool)
- The *bottom four* include the three teams that will be relegated to the lower division Championship plus the fourth worst team that is typically only one or two points above the relegation zone (Fulham, Reading, Birmingham City, Derby County)
- The *middle dozen* (12) teams form the “rest of the pack” (Everton, Aston Villa, Blackburn, Portsmouth, Manchester City, West Ham United, Tottenham, Newcastle United, Middlesbrough, Wigan Athletic, Sunderland, and Bolton)

ANOVA provides the opportunity to identify pitch actions that might not be a statistically significant retrodictor variable in the regression model, but are pitch factors that establish statistically significant differences between the three EPL subdivisions (Levine et al. [2008], Winston, [2004]). The ANOVA results are summarized in Table 7 and show that thirteen (13) pitch measures are significantly different among these three team clusters. The likely influences of these measures—discussed within each pitch factor family—include:

> Within the *goal attempts cluster actions*—ignoring the obvious duplicate information provided by the total goals scored (X_1) and the average goals scored per game (X_2)—the only other statistically significant difference found among the three clusters of clubs action was the number of shots taken, X_3 ($p<0.0000$). Conversely, the actions reflecting (1) the accuracy of the goal attempts such as the percent of shots on target (X_4) and percent of goals to shots ratio (X_5) and (2) the general distance of the goals scored—neither the percent of goals scored inside (X_6) or outside (X_7) the box presented statistically significant differences. *Better teams simply create more opportunities.*
>
> Conversely, almost the entire collection of *passing cluster actions* was significant. The only exception was the number of long passes, X_9. All other actions—short passes, X_8 ($p=0.042$), total passes, X_{10} ($p<0.0000$), ratio of short-to-long, X_{11} ($p=0.0003$), overall pass completion percentage, X_{12} ($p<0.0000$), percent short passes completed, X_{13} ($p=0.0018$), and percent long passes completed, X_{14} ($p<0.0000$)—exhibited statistically significant differences between the three club clusters. *It is likely that the more successful teams make significantly more passes and are more accurate with their passing than the lesser teams, keep possession of the ball more consistently, and create more scoring opportunities that the other groups. Further, the less accurate selection of long passing in advancing the ball by the weaker teams could be a likely sign of desperation—something that the better teams have a lesser need to do.*
>
> No significant differences were found between the three groups among the *crossing cluster actions* consisting of the total number of crosses (X_{15}) and the accuracy of the crosses (X_{16}). *EPL teams have very similar crossing skills: neither action is a defining feature of club success*.

		Mean ActionValues				
	Action	Top 4 clubs	Middle 12 clubs	Bottom 4 clubs	*F-val*	*p-val*
Goal Attempts	**X1**	71.500	47.583	36.250	11.411	**0.0007**
	X2	1.882	1.252	0.954	11.411	**0.0007**
	X3	502.5	375.0	325.3	27.536	**0.0000**
	X4	0.438	0.427	0.419	0.453	0.6433
	X5	0.143	0.126	0.111	1.668	0.2181
	X6	0.835	0.855	0.826	0.911	0.4210
	X7	0.165	0.145	0.174	0.911	0.4210
Passing	**X8**	15,146	10,830	9,119	3.840	**0.0421**
	X9	2,319	2,257	2,223	0.285	0.7557
	X10	17,465	13,088	11,342	28.013	**0.0000**
	X11	6.707	4.792	4.099	13.772	**0.0003**
	X12	0.805	0.721	0.665	22.107	**0.0000**
	X13	0.837	0.774	0.770	9.340	**0.0018**
	X14	0.600	0.466	0.424	27.619	**0.0000**
Crossing	X15	918.0	862.2	870.3	0.442	0.6502
	X16	0.236	0.232	0.221	0.882	0.4322
Defending	**X17**	26.750	51.500	69.250	22.124	**0.0000**
	X18	0.704	1.355	1.822	22.124	**0.0000**
	X19	965.5	881.1	871.0	3.470	0.0545
	X20	0.752	0.763	0.751	1.149	0.3404
	X21	2,491	2,694	2,631	2.693	0.0964
Discipline	**X22**	443.8	517.7	519.8	3.840	**0.0421**
	X23	53.5	61.2	61.8	0.818	0.4578
	X24	2.750	2.750	3.750	0.471	0.6323

Table 7. One-Way ANOVA of Pitch Actions for 3-Tier Group of EPL Football Clubs.

In contrast, the *defending and discipline* pitch action clusters both yield highly different results between the three tiers. Pitch factors X_{17} (p<0.0000) and X_{18} (p<0.0000) are, essentially the same measurement, i.e., goals conceded during the season and goals conceded per game. The number of tackles, X_{19} (p=0.0545), and the total fouls, X_{22} (p=0.0421) are significantly different.[14] A possible reason for these differences is that both pitch actions are often associated with either controlling or keeping possession of the ball. *The better clubs are significantly more successful at (1) gaining possession of the ball by making more tackles and (2) maintaining possession by fouling less than the middle and bottom tier clubs.*

[14] Even though the number of tackles, X_{19}, is slightly over the desired .05 level of significance, it is, nevertheless viewed as having *practical* significance.

DOI: 10.2202/1559-0410.1183

An illustration of how one of these significantly different pitch actions differentiates the three clusters of teams is shown in Appendix C using the overall pass completion percentage, X_{12}.

CONCLUSIONS

A retrodictive, linear multiple regression model is used to define 6 statistically significant team pitch factors that are keys to a club's ultimate success as measured by points earned over the 2007-2008 English Premier League season (Figure 6). The influence of marginal changes in the factor coefficients on table

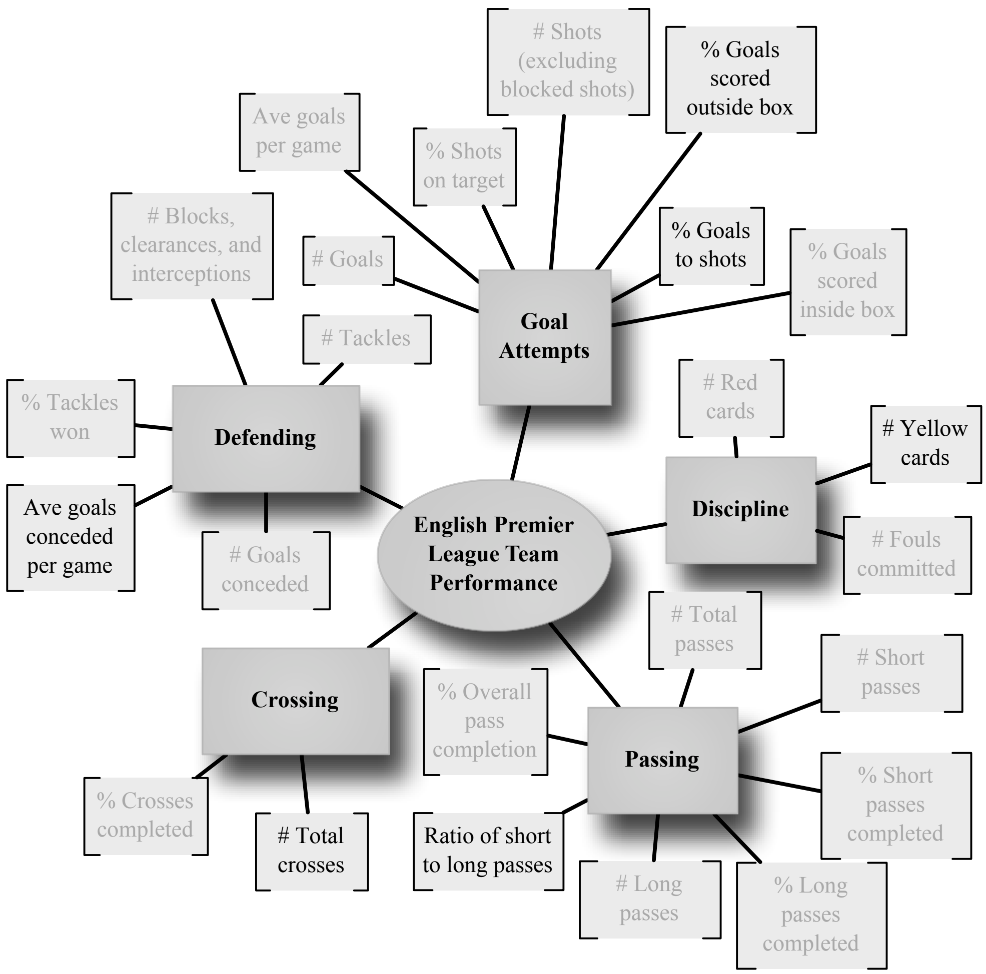


Figure 6. Six (6) Key EPL Pitch Actions of Multiple Regression Model.

placement is discussed along with the common role that these 6 factors play in a fixture.

A second, complementary statistical analysis using ANOVA uncovers a set of 13 pitch factors that exhibit statistically significant differences when compared between the three tiers of teams that form the top, middle, and bottom clusters of the EPL. Again, a practical explanation of how these pitch factors influence team performance is presented. The significant specific pitch actions that were identified using the ANOVA assessment are shown in Figure 7.

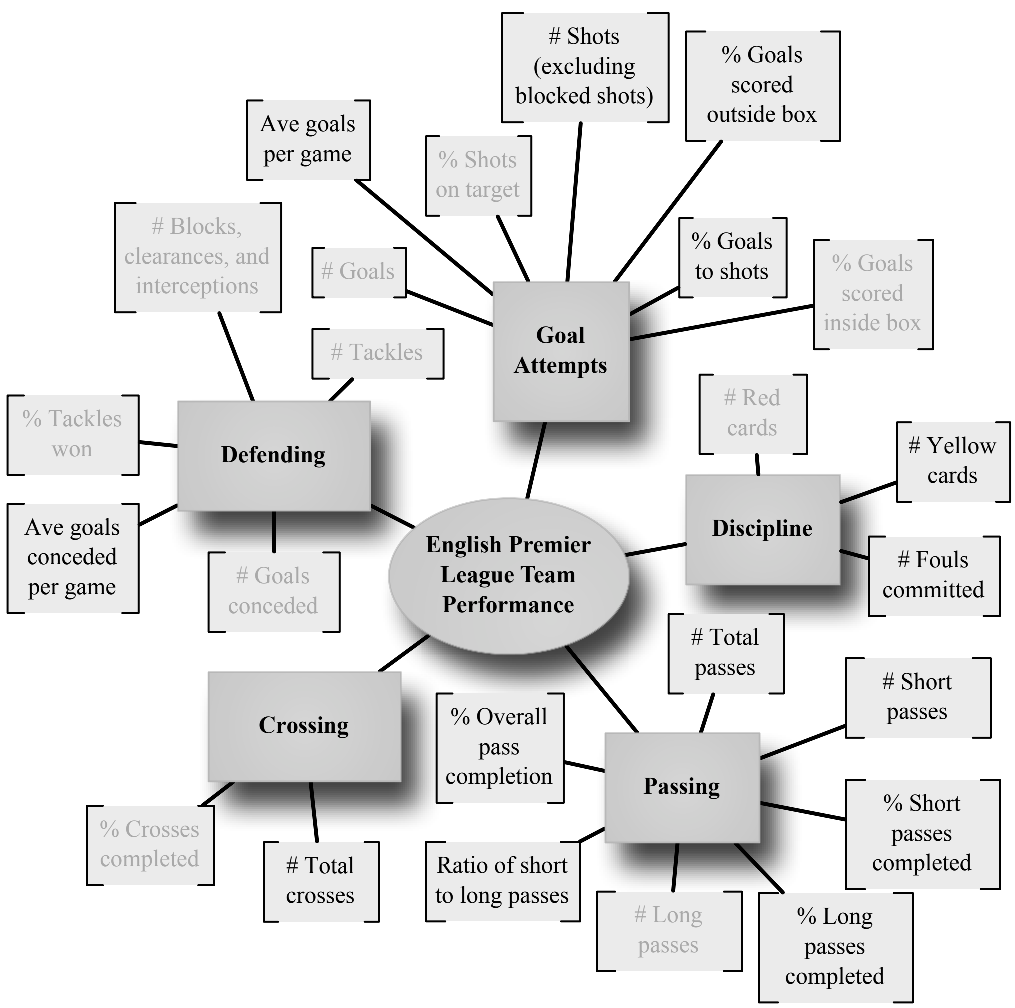


Figure 7. Thirteen (13) Significant Pitch Factors Differentiating 3-Tier Groups of EPL Clubs Using ANOVA.

Analysis genre, other than a multiple regression model, may also offer insight into assessing team success. Because there is no single standard of excellence, no comparable measure equivalent to the league points that a team earns, individual player performance ratings are more subjective in nature. Both the Actim Index and the Opta Index grade individual players according to a different, complex set of weighted points awarded or deducted for outstanding or poor performance associated with numerous pitch actions during each game. These measures of composite player performance definitely beg a more subjective measure of goodness as apposed to a club's league points won.

Clearly, the success of a team will most likely be indicated by the collective performance of its players throughout the season. Accordingly, this paper looks at the team performance in a way similar to the assessments of player performance without the notion of creating an artificial index by identifying the key pitch actions that differentiate the relative success of the 20 EPL football clubs. Follow-up work that examines the opportunity to generalize the model findings might include (1) studying multiple EPL seasons and (2) applying the model to other, prestigious football leagues, e.g., La Liga (Spain), Serie A (Italy), Eredivisie (Netherlands), Ligue 1 (France), and Bundesliga (Germany).

Appendix A-1. Opta Sportsdata Team Pitch Actions for 2007-2008 English Premier League Season (Part 1 of 3).

		Goal Attempts						
	Points	Goals	Ave Goals per game	Shots (excl blocked shots)	% Shots on Target	% Goals to Shots	% of goals scored inside box	% of goals scored outside box
Football Club	**Y**	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
Manchester United	87	80	2.11	547	47.53%	14.63%	82.50%	17.50%
Chelsea	85	65	1.71	455	43.74%	14.29%	81.54%	18.46%
Arsenal	83	74	1.95	473	43.97%	15.64%	89.19%	10.81%
Liverpool	76	67	1.76	535	40.00%	12.52%	80.60%	19.40%
Everton	65	55	1.45	361	46.81%	15.24%	90.91%	9.09%
Aston Villa	60	71	1.87	405	44.94%	17.53%	83.10%	16.90%
Blackburn	58	50	1.32	421	43.94%	11.88%	80.00%	20.00%
Portsmouth	57	48	1.26	411	38.93%	11.68%	79.17%	20.83%
Manchester City	55	45	1.18	323	44.27%	13.93%	82.22%	17.78%
West Ham United	49	42	1.11	384	40.10%	10.94%	88.10%	11.90%
Tottenham	46	66	1.74	424	45.99%	15.57%	90.91%	9.09%
Newcastle United	43	45	1.18	358	42.74%	12.57%	82.22%	17.78%
Middlesbrough	42	43	1.13	352	43.18%	12.22%	86.05%	13.95%
Wigan Athletic	40	34	0.89	363	38.84%	9.37%	85.29%	14.71%
Sunderland	39	36	0.95	358	38.27%	10.06%	88.89%	11.11%
Bolton	37	36	0.95	340	43.82%	10.59%	88.89%	11.11%
Fulham	36	38	1.00	367	41.96%	10.35%	84.21%	15.79%
Reading	36	41	1.08	332	39.76%	12.35%	87.80%	12.20%
Birmingham City	35	46	1.21	314	45.54%	14.65%	78.26%	21.74%
Derby County	11	20	0.53	288	40.28%	6.94%	80.00%	20.00%

Table A-1. 2007-2008 EPL Opta Sportsdata for Goal Attempt Actions.

DOI: 10.2202/1559-0410.1183

Appendix A-2. Opta Sportsdata Team Pitch Actions for 2007-2008 English Premier League Season (Part 2 of 3).

	Passing						
	Short Passes	Long Passes	Total passes	Ratio of short/long passes	Overall Pass Completion %	% short passes completed	% long passes completed
Football Club	**X8**	**X9**	**X10**	**X11**	**X12**	**X13**	**X14**
Manchester United	15,003	2,414	17,417	6.21	80.50%	83.86%	59.65%
Chelsea	15,097	2,153	17,250	7.01	81.16%	84.47%	58.10%
Arsenal	16,885	1,946	18,831	8.68	83.18%	85.48%	63.36%
Liverpool	13,599	2,761	16,360	4.93	77.06%	80.79%	58.75%
Everton	10,255	2,142	12,397	4.79	69.32%	75.40%	40.34%
Aston Villa	9,456	1,982	11,438	4.77	69.76%	74.55%	47.02%
Blackburn	11,369	2,281	13,650	4.98	74.03%	78.92%	49.85%
Portsmouth	10,515	2,278	12,793	4.62	73.50%	79.08%	47.89%
Manchester City	12,858	2,289	15,147	5.62	76.23%	80.82%	50.46%
West Ham United	11,396	2,366	13,762	4.82	74.11%	79.27%	49.28%
Tottenham	13,036	2,428	15,464	5.37	77.43%	82.05%	52.68%
Newcastle United	11,527	2,420	13,947	4.76	72.64%	78.96%	42.69%
Middlesbrough	9,858	2,341	12,199	4.21	71.01%	77.42%	44.08%
Wigan Athletic	9,519	2,153	11,672	4.42	68.87%	74.42%	44.36%
Sunderland	10,483	2,244	12,727	4.67	70.07%	75.13%	46.52%
Bolton	9,692	2,163	11,855	4.48	67.85%	73.17%	43.97%
Fulham	10,264	2,310	12,574	4.44	70.77%	75.40%	47.49%
Reading	8,232	2,199	10,431	3.74	64.35%	74.55%	44.34%
Birmingham City	8,411	2,219	10,630	3.79	63.74%	78.92%	37.18%
Derby County	9,568	2,165	11,733	4.42	67.30%	79.08%	40.69%

Table A-2. 2007-2008 EPL Opta Sportsdata for Passing Actions.

Appendix A-3. Opta Sportsdata Team Pitch Actions for 2007-2008 English Premier League Season (Part 3 of 3).

	Crossing		Defending					Discipline		
	Total Crosses	Cross Completion %	Goals Conceded	Ave goals conceded per game	Tackles	Tackles Won %	Blocks, Clearances & Interceptions	Fouls	Yellow Cards	Red Cards
Football Club	**X15**	**X16**	**X17**	**X18**	**X19**	**X20**	**X21**	**X22**	**X23**	**X24**
Manchester United	930	24.41%	22	0.58	978	76.18%	2,574	425	51	2
Chelsea	937	20.70%	26	0.68	917	76.23%	2,302	478	63	5
Arsenal	883	24.01%	31	0.82	878	74.72%	2,438	407	55	3
Liverpool	922	25.27%	28	0.74	1,089	73.74%	2,651	465	45	1
Everton	868	23.27%	33	0.87	816	77.21%	2,773	494	40	3
Aston Villa	926	25.59%	51	1.34	885	79.77%	2,826	559	54	4
Blackburn	874	22.43%	48	1.26	833	74.67%	2,736	583	72	6
Portsmouth	654	21.87%	40	1.05	963	75.91%	2,733	541	55	3
Manchester City	691	21.27%	53	1.39	916	77.62%	2,492	427	50	4
West Ham United	904	24.23%	50	1.32	848	74.76%	2,546	562	63	1
Tottenham	987	22.19%	61	1.61	960	74.58%	2,708	410	51	1
Newcastle United	957	19.85%	65	1.71	876	76.14%	2,516	483	63	1
Middlesbrough	946	23.68%	53	1.39	835	75.45%	2,805	562	86	2
Wigan Athletic	756	25.00%	51	1.34	883	78.94%	2,504	527	59	4
Sunderland	882	24.94%	59	1.55	892	74.33%	2,831	517	65	4
Bolton	901	23.53%	54	1.42	866	76.44%	2,859	547	76	0
Fulham	842	22.09%	60	1.58	851	76.62%	2,661	493	55	6
Reading	1072	22.67%	66	1.74	800	74.38%	2,369	492	59	5
Birmingham City	769	22.50%	62	1.63	935	76.58%	2,809	546	70	3
Derby County	798	21.30%	89	2.34	898	72.72%	2,685	548	63	1

Table A-3. 2007-2008 EPL Opta Sportsdata for Crossing, Defending, and Discipline Actions.

DOI: 10.2202/1559-0410.1183

Appendix B. Uniform Variance Test for Heteroscedasticity

This procedure tests the error variance uniformity, e^2, of a regression model with the dependent variable, $\hat{Y}$. For our model, let R^2 be the coefficient of determination and n be the number of observations. Reject the null hypothesis of equal error variance if nR^2 is larger than a critical value of the χ^2 random variable with 1 degree of freedom and α=0.05. Then, for our conditions of n=20 observations, α =0.05, df=1, and $\chi^2_{crit} = \chi^2_{df=1,\alpha=0.05} = 3.84$, we find R^2=0.055, and nR^2 =1.10 (Table B). Since nR^2 =1.10 < 3.84, the null hypothesis is not rejected and we may assume that the assumption of uniform variance holds for our model.

Predicted Y	*e2*
84.516094	2.29854
58.604623	1.94708
38.500399	12.2528
54.958351	9.25163
38.47663	2.18044
86.373729	1.88713
11.97856	0.95758
66.235693	1.52694
39.273221	10.714
75.831722	0.02832
51.648541	11.2323
87.121895	0.01486
41.909089	0.00826
42.064515	0.87513
56.211785	0.62128
36.248771	0.06189
35.821943	10.1
45.882581	0.01379
46.953132	4.18967
41.388726	1.92856
	72.0902

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.2352					
R Square	0.0553					
Adjusted R Square	0.0028					
Standard Error	19.611					
Observations	20					
ANOVA						
	df	*SS*	*MS*	*F*	*Signif F*	
Regression	1	405.28	405.279	1.053793	0.3182	
Residual	18	6922.6	384.591			
Total	19	7327.9				
	Coeff	*Std Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	55.814	5.7477	9.71069	1.4E-08	43.739	67.8899
e2	-1.0582	1.0309	-1.02654	0.318233	-3.224	1.10754

Table B. Regression Test of Heteroscedasticity (Uniform Variance).

Appendix C. ANOVA Pitch Action Cluster Sample of Overall Pass

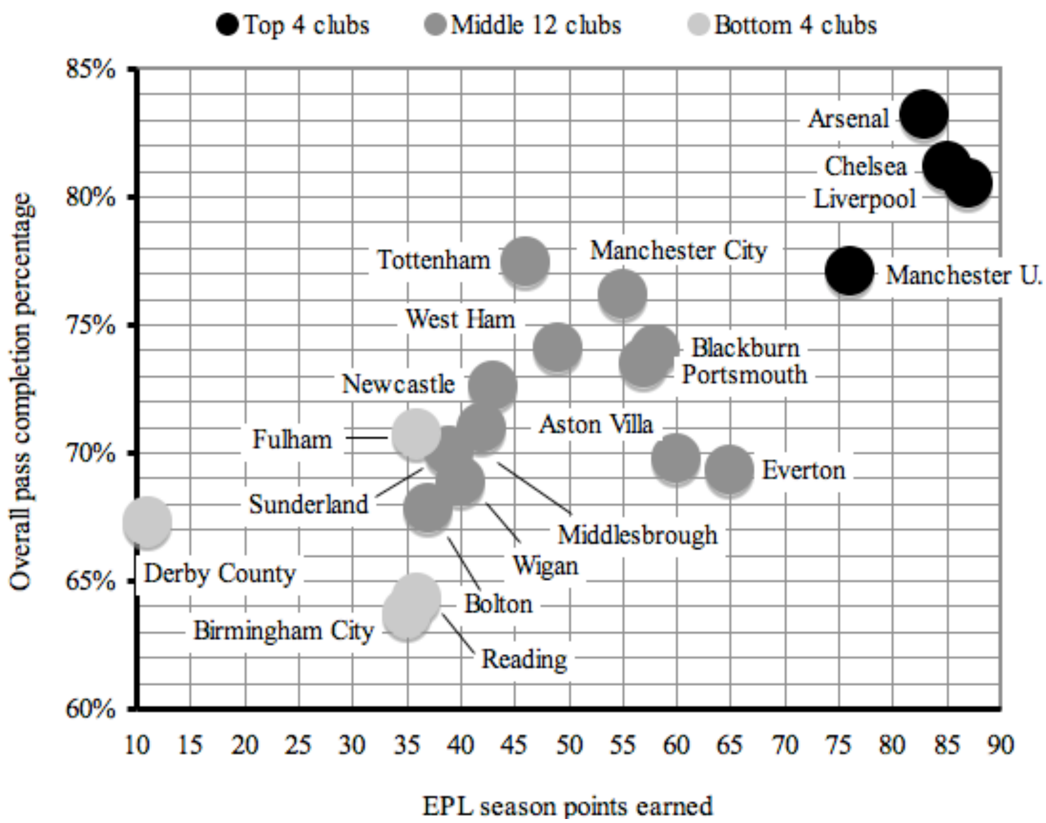
Top 4	Middle 12	Bottom 4
80.50%	69.32%	70.77%
81.16%	69.76%	64.35%
83.18%	74.03%	63.74%
77.06%	73.50%	67.30%
	76.23%	
	74.11%	
	77.43%	
	72.64%	
	71.01%	
	68.87%	
	70.07%	
	67.85%	

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Top 4	4	3	0.805	0.001
Middle 12	12	9	0.721	0.001
Bottom 4	4	3	0.665	0.001

ANOVA

Source of Variatior	*SS*	*df*	*MS*	*F*	*P-value*
Between Groups	0.0398	2	0.0199	22.107	**0.00002**
Within Groups	0.0153	17	0.0009		
Total	0.0551	19			



DOI: 10.2202/1559-0410.1183

REFERENCES

1. Ammar, Salwa and Wright, Ronald, "Comparing the impact of star rookies Carmelo Anthony and Lebron James: An example on simulating team performances in the National Basketball Association League," INFORMS transactions in Education, vol: 5, 2004/05.

2. Andersson, Patric, Edman, Jan and Ekman, Mattias, "Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts," IJF, vol: 21, Jul/Sep 2005, 3, page(s): 565-576.

3. Barros, Carlos Pestana and Leach, Stephanie, "Performance evaluation of the English Premier Football League with data envelopment analysis," Applied Economics, Volume 38, Number 12, 10 July 2006 , pp. 1449-1458(10).

4. Bennett JM, Flueck JA. 1983. An evaluation of Major League Baseball offensive performance models. The American Statistician 37: 76-82.

5. Boulier, Bryan L. and Stekler, H.O., "Predicting the outcomes of National Football League games," IJF, vol: 19, Apr/Jun 2003, 2, page(s): 257-270.

6. Bradley, Paul; O'Donoghue, Peter; Wooster, Blake; Tordoff, Phil, The reliability of ProZone MatchViewer: a video-based technical performance analysis system, International Journal of Performance Analysis in Sport, Volume 7, Number 3, October 2007 , pp. 117-129(13), University of Wales Institute, Cardiff.

7. Carmichael F., Thomas D., Ward R., Team Performance: The Case of English Premiership Football, Managerial and Decision Economics, Vol. 21, No. 1 (Jan-Feb 2000), pp. 31-45.

8. Crowder, Martin, Dixon, Mark, Ledford, Anthony and Robinson, Mike, "Dynamic modelling and prediction of English football league matches for betting," Journal of the Royal Statistical Society, series D (The Statistician), vol: 51, Apr 2002, 2, page(s): 157-168.

9. Espitia-Escuer, Manuel; García-Cebrián, Lucía Isabel, "Performance in sports teams: Results and potential in the professional soccer league in Spain," Management Decision, Volume 44, Number 8, 2006 , pp. 1020-1030(11).

10. Fitt, A.D., Kabelka, M. and Howls, C.J., "Valuation of soccer spread bets," JORS, vol: 57, Aug 2006, 8, page(s): 975-985.

11. Groebner, David; Shannon, Patrick; Fry, Phillip; Smith, Kent, "Business Statistics: A Decision-Making Approach," Prentice-Hall, 7th Edition (2008), Upper Saddle River, New Jersey, 07458, pp. 726-730.

12. Hirotsu N.; Wright M., "Determining the best strategy for changing the configuration of a football team," Journal of the Operational Research Society, Volume 54, Number 8, August 2003 , pp. 878-887(10).

13. Hirotsu, N. and Wright, M., "Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions," Journal of the Operational Research Society, vol: 53, Jan 2002, 1, page(s): 88-96

14. Hope, Chris, "When should you sack a football manager? Results from a simple model applied to the English Premiership," Journal of the Operational Research Society vol: 54, Nov 2003, 11, page(s): 1167-1176 (10).

15. James, Stuart, “Thirty Million Reasons Not to Get Relegated This Season,” The Guardian, Saturday 3 May 2007.

16. Jones P; James, N.; Mellalieu, S; “Possession as a performance indicator in soccer,” International Journal of Performance Analysis in Sport, Volume 4, Number 1, August 2004 , pp. 98-102(5)

17. Levine, David; Stephan, David; Krehbiel, Timothy; Berenson, Mark, “Statistics for Management Using Microsoft Excel,” Prentice-Hall, 5th Edition (2008), Upper Saddle River, New Jersey, 07458, pp. 628-629.

18. Lewis, Michael, Moneyball: The Art of Winning an Unfair Game, W.W. Norton & Company (2004).

19. McHale, Ian, and Scarf, Phil, (2007). "Modelling soccer matches using bivariate discrete distributions". Statistica Neerlandica, 61, 4, 432-445.

20.McHale, I.G. (2007). "Quantitative studies in sport". The Operational Research Society, Keynote Papers. ISBN: 0 903440 39 3.

21.Newbold, Paul; Carlson, William; Thorne, Betty, “Statistics for Business and Economics,” Prentice-Hall, 6th Edition (2006), Upper Saddle River, New Jersey 07458, pp. 564-568.

22. O'Donoghue P.G., "Time-motion analysis of work-rate in English FA Premier League soccer ," International Journal of Performance Analysis in Sport, Volume 2, Number 1, 15 August 2002 , pp. 36-43(8).

23. Reed, David; Hughes, Mike, “An Exploration of Team Sport as a Dynamical System,” International Journal of Performance Analysis in Sport, Volume 6, Number 2, November 2006 , pp. 114-125(12).

24. Schleiffer, Arthur and Bell, David, “Data Analysis, Regression, and Forecasting,” Managerial Decision Analysis Series, Course Technology, Inc., One Main Street, Cambridge, MA 02142 (1995).

25. Szczepański, Łkasz, “Measuring the effectiveness of strategies and quantifying players' performance in football,” International Journal of Performance Analysis in Sport, Volume 8, Number 2, July 2008 , pp. 55-66(12).

26. Valter, Di Salvo; Adam, Collins; Barry, McNeill; Marco, Cardinale, “Validation of Prozone ®: A new video-based performance analysis system,” International Journal of Performance Analysis in Sport, Volume 6, Number 1, June 2006 , pp. 108-119(12).

27. Winston, Wayne, “Microsoft Excel Data Analysis and Business Modeling,” Microsoft Press