

The University of San Francisco
**USF Scholarship: a digital repository @ Gleeson Library |
Geschke Center**

Business Analytics and Information Systems

School of Management

2010

Comparing English Premier League Goalkeepers: Identifying the Pitch Actions that Differentiate the Best from the Rest

Joel Oberstone

University of San Francisco, joel@usfca.edu

Follow this and additional works at: <http://repository.usfca.edu/at>

Recommended Citation

Joel Oberstone (2010) "Comparing English Premier League Goalkeepers: Identifying the Pitch Actions that Differentiate the Best from the Rest," *Journal of Quantitative Analysis in Sports*: Vol. 6 : Iss. 1, Article 9.

This Article is brought to you for free and open access by the School of Management at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Business Analytics and Information Systems by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.

*Journal of Quantitative Analysis in
Sports*

Volume 6, Issue 1

2010

Article 9

Comparing English Premier League
Goalkeepers: Identifying the Pitch Actions that
Differentiate the Best from the Rest

Joel Oberstone, *University of San Francisco*

Recommended Citation:

Joel Oberstone (2010) "Comparing English Premier League Goalkeepers: Identifying the Pitch Actions that Differentiate the Best from the Rest," *Journal of Quantitative Analysis in Sports*: Vol. 6 : Iss. 1, Article 9.

Available at: <http://www.bepress.com/jqas/vol6/iss1/9>

DOI: 10.2202/1559-0410.1221

©2010 American Statistical Association. All rights reserved.

Comparing English Premier League Goalkeepers: Identifying the Pitch Actions that Differentiate the Best from the Rest

Joel Oberstone

Abstract

The Opta Index is a prestigious performance measure used to assess English Premier League (EPL) football players. Although the Opta model is proprietary, the general structure uses a multiattribute collection of subjectively weighted pitch measures that either rewards or penalizes a player with a potential range of points based on the quality of his game performance. In addition, the specific set of measures used depends upon player position: forwards, midfielders, defenders, and goalkeepers each have their own unique set of measures even though there might be some overlap. Although the player's Opta Index is calculated for each game, it is the cumulative "grade card"—the final Opta Index calculated at the end of the thirty-eight game EPL season in May—that is of particular importance. The index, along with the large array of player pitch data, is commercially distributed to the EPL clubs and appears in a wide variety of television and print media outlets. This paper proposes an alternative to using the full set of Opta data by identifying those specific pitch actions that form a statistically significant retrodictive linear regression model for the 2007-2008 EPL season. Additionally, the importance of evaluating pitch actions historically assumed to be clearly pertinent measures—such as goals allowed per game for the goalkeeper—will be not only be appraised from a statistical viewpoint, but also from a practical perspective.

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

INTRODUCTION

The value of world-class football players is often reflected by field performance measures using a variety of models (Andersson, Edman, and Ekman, [2005]; Barros and Leach [2006]; Crowder, Dixon, Ledford, Robinson [2005]). Other efforts have focused on the use of performance indices such as those generated by Opta Sportsdata, PA Sports (Actim), and Castrol.¹ Considerable research has been done recently using the index models in the evaluation of team and player performance (McHale and Scarf [2007]; Oberstone [2009]).

The index values embrace a proprietary array of relatively weighted, pitch performance criteria assigned by the analysts of each data collection and analysis organization. Broadly speaking, these models use a variety of statistical and multicriteria methods that results in a composite measure of goodness or “grade card” for each player. One simple form of this type of model is:

$$R_{ij} = \sum_{k=1}^{K_j} w_{jk} r_{ijk} \quad [\text{Eq. 1}]$$

for each i - j combination of player-position
where

R_{ij} = composite rating/index score of i^{th} player at the j^{th} position

w_{jk} = weight of importance of the k^{th} performance factor at the j^{th} position (independent of specific player; reflection of team preferred “style of play”).

x_{ijk} = performance *value* of the i^{th} player at the j^{th} position on the k^{th} performance factor, e.g., average number of shots inside the box per 90 minutes, goals scored per 90 minutes (or minutes per goal), etc.

r_{ijk} = performance *rating* of the i^{th} player at the j^{th} position on the k^{th} performance factor, $f(x_{ijk})$, i.e., transformed/converted value of x_{ijk} using factor performance chart

K_j = total number of performance factors associated with j^{th} position

However, the subjectivity of all of these index measures is inescapable, contrary to a measure of goodness based on the league points earned during the season used to measure team quality (Oberstone [2009]). In spite of the subjectivity of

¹ An example of the Actim Index is given in Appendix A.

player index measures, they are, nevertheless, widely accepted indicators of his season-long performance effort.²

METHOD OF ANALYSIS

As an alternative to the multiattribute models used for index values, this paper proposes the use of a multiple regression model that retrodictively identifies the specific pitch variables that make statistically significant contributions. The Opta Index for the 2007-2008 English Premier League season will be used in this paper. An illustration of six (6) basic pitch actions typically used to assess goalkeeper value is shown in Figure 1.

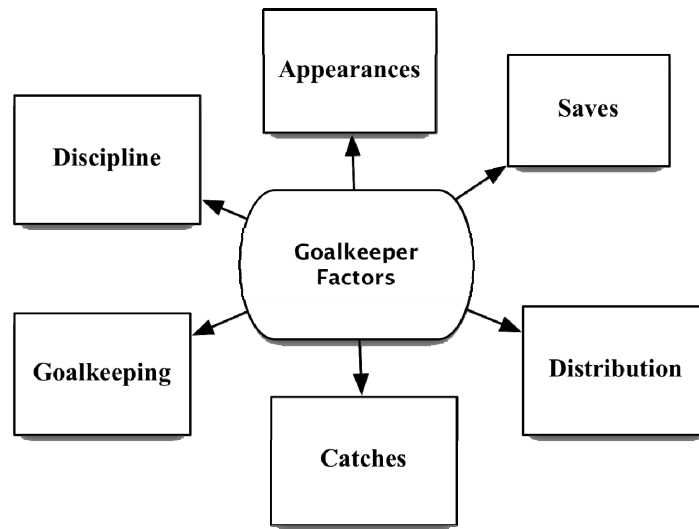


Figure 1. Opta Index Goalkeeper Pitch Action Groups

A more detailed breakdown of the basic pitch actions is provided in the set of Opta Index measures shown in Figure 2 (distribution, discipline, and saves) and Figure 3 (catches, appearances, goalkeeping).

The thirty-four (34) actions in this set have considerable redundancy, e.g., goals conceded per game versus minutes per goal conceded; percent successful distribution versus percent unsuccessful distribution. Since the goalkeepers have a broad range of minutes played, the goals scored total is obviously influenced by the amount of game exposure. Because of this conflict in action measures, the original set can be carefully reduced to 24 actions that have been transformed to a “per 90 minutes of exposure.” The original Opta data is shown in Table 1 and

² The indices also comprise a commercially lucrative industry that closely guards its financial records much the same way it protects the detailed structure of its models.

Table 2.³ The new, normalized, smaller set pitch actions that eliminates redundant measures, is shown in Table 3 and Table 4.

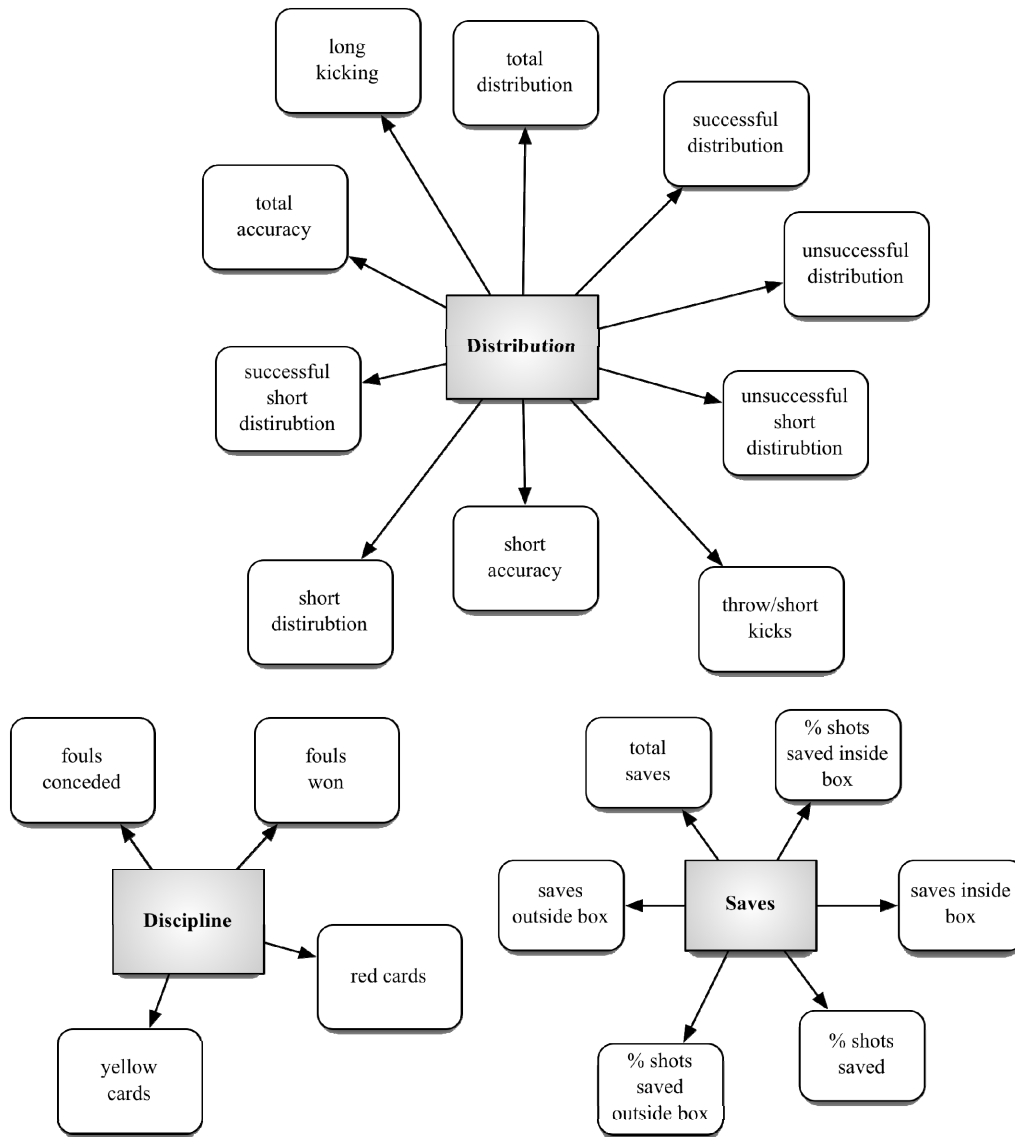


Figure 2. Opta Index Goalkeeper Distribution, Discipline, and Saves Pitch Actions (Part 1 of 2)

³ Goalkeepers with less than 270 minutes of game exposure were eliminated from the original set of data.

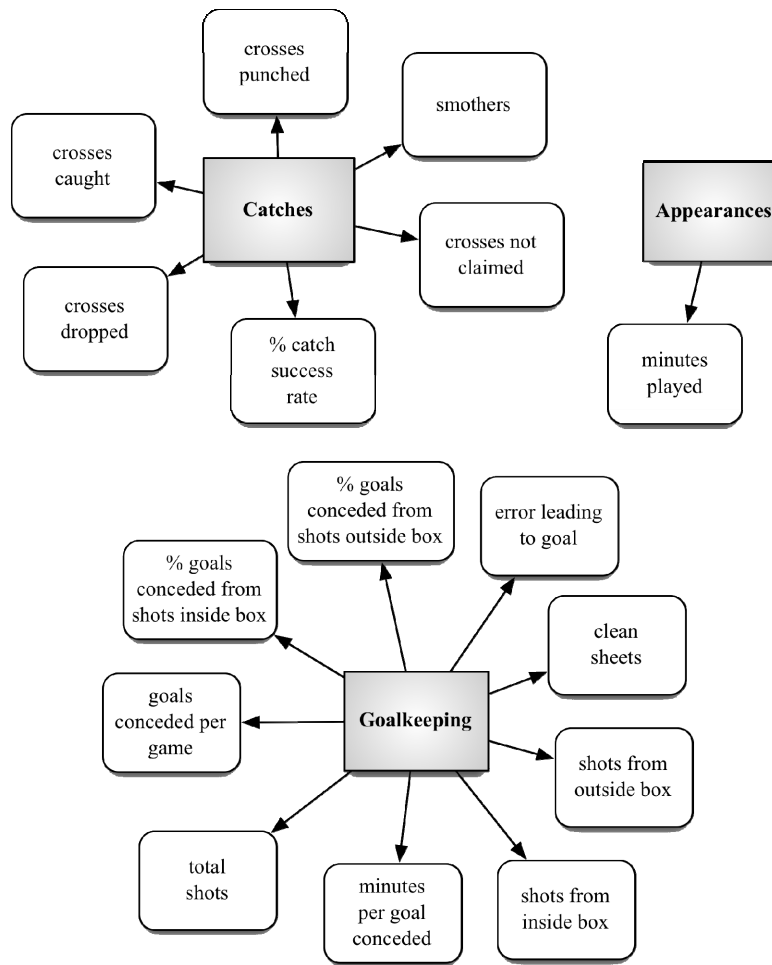


Figure 3. Opta Index Goalkeeper Catches, Appearances, and Goalkeeping Pitch Actions (Part 2 of 2)

Oberstone: Comparing English Premier League Goalkeepers

Player Surname	Average OPTA Score per 90	Time Played	Total Fouls Conceded	Total Fouls Won	Yellow Cards	Red Cards	Goals Conceded	% Goals Conceded Inside Box	% Goals Conceded Outside Box	Total shots	Shots from Inside Box	Shots from Outside Box	Saves Made	Saves Made from Inside Box	Saves Made from Outside Box	Saves to Shots Ratio	Saves to Shots ratio Inside Box	Saves to Shots ratio Outside Box
<i>Name</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>X3</i>	<i>X4</i>	<i>X5</i>	<i>X6</i>	<i>X7</i>	<i>X8</i>	<i>X9</i>	<i>X10</i>	<i>X11</i>	<i>X12</i>	<i>X13</i>	<i>X14</i>	<i>X15</i>	<i>X16</i>	<i>X17</i>
Friedel	853	3420	0	5	0	0	48	0.854	0.146	171	111	60	123	70	53	0.719	0.631	0.883
Reina	749	3420	0	8	2	0	28	0.929	0.071	86	61	25	58	35	23	0.674	0.574	0.920
Hahnmann	913	3420	1	7	3	0	66	0.879	0.121	218	142	76	152	84	68	0.697	0.592	0.895
Green	831	3420	1	2	2	0	50	0.840	0.160	166	108	58	116	66	50	0.699	0.611	0.862
Kirkland	754	3330	1	3	0	0	49	0.837	0.163	150	106	44	101	65	36	0.673	0.613	0.818
Howard	784	3240	0	5	1	0	30	0.900	0.100	119	75	44	89	48	41	0.748	0.640	0.932
James	1169	3150	1	7	2	0	36	0.861	0.139	174	109	65	138	78	60	0.793	0.716	0.923
Carson	598	3126	2	3	1	1	45	0.778	0.222	126	80	46	81	45	36	0.643	0.563	0.783
Taylor	857	3060	1	5	1	0	54	0.963	0.037	188	124	64	134	72	62	0.713	0.581	0.969
Schwarzer	703	3060	0	10	2	0	47	0.766	0.234	135	81	54	88	45	43	0.652	0.556	0.796
Gordon	564	3060	0	4	1	0	55	0.855	0.145	138	92	46	83	45	38	0.601	0.489	0.826
Almunia	928	2610	1	3	0	0	24	0.792	0.208	95	49	46	71	30	41	0.747	0.612	0.891
Van der Sar	971	2565	0	8	2	0	18	0.944	0.056	91	51	40	73	34	39	0.802	0.667	0.975
Jaaskelainen	671	2520	1	9	1	0	42	0.833	0.167	117	78	39	75	43	32	0.641	0.551	0.821
Hart	1065	2340	1	3	0	0	34	0.971	0.029	141	88	53	107	55	52	0.759	0.625	0.981
Cech	945	2314	0	4	0	0	17	0.941	0.059	78	53	25	61	37	24	0.782	0.698	0.960
Robinson	575	2250	0	7	0	0	47	0.745	0.255	108	68	40	61	33	28	0.565	0.485	0.700
Niemi	800	1980	0	2	0	0	39	0.897	0.103	124	90	34	85	55	30	0.685	0.611	0.882
Harper	925	1791	0	6	0	0	28	0.821	0.179	100	65	35	72	42	30	0.720	0.646	0.857
Given	526	1629	0	3	2	0	37	0.730	0.270	86	54	32	49	27	22	0.570	0.500	0.688
Bywater	677	1620	0	2	1	0	41	0.780	0.220	103	71	32	62	39	23	0.602	0.549	0.719
Carroll	792	1260	1	5	0	0	38	0.868	0.132	101	74	27	63	41	22	0.624	0.554	0.815
Keller	880	1170	0	4	0	0	16	0.813	0.188	55	31	24	39	18	21	0.709	0.581	0.875
Cerny	916	1170	0	2	1	0	14	0.643	0.357	58	38	20	44	29	15	0.759	0.763	0.750
Al-Habsi	986	900	0	2	0	0	12	0.833	0.167	50	34	16	38	24	14	0.760	0.706	0.875
Cudicini	926	900	0	3	0	0	7	0.857	0.143	30	18	12	23	12	11	0.767	0.667	0.917
Kuszczak	1089	765	0	2	0	0	4	1.000	0.000	26	14	12	22	10	12	0.846	0.714	1.000
Schmeichel	1181	630	0	2	0	0	5	0.800	0.200	27	19	8	22	15	7	0.815	0.789	0.875
Lehmann	935	561	0	2	1	0	5	0.600	0.400	22	11	11	17	8	9	0.773	0.727	0.818
Price	589	540	0	1	0	0	10	0.800	0.200	26	17	9	16	9	7	0.615	0.529	0.778
Isaksson	475	450	0	0	0	0	14	0.857	0.143	26	21	5	12	9	3	0.462	0.429	0.600
Taylor	856	293	0	1	0	0	6	0.667	0.333	17	13	4	11	9	2	0.647	0.692	0.500
Doyle	496	270	1	2	1	0	6	0.833	0.167	15	11	4	9	6	3	0.600	0.545	0.750
Warner	1045	270	0	3	1	0	5	1.000	0.000	21	18	3	16	13	3	0.762	0.722	1.000
Turnbull	1114	270	0	1	0	0	3	1.000	0.000	16	11	5	13	8	5	0.813	0.727	1.000
Ashdown	763	270	0	0	0	0	4	1.000	0.000	14	12	2	10	8	2	0.714	0.667	1.000
Ward	809	270	0	1	0	0	3	0.667	0.333	12	7	5	9	5	4	0.750	0.714	0.800

Table 1. English Premier League 2007-2008 Season Pitch Actions for Goalkeeper Position (Part 1 of 2).

Player Surname	Average OPTA Score per 90	Goalkeeper Smother	Catches	Punches	Drops	Crosses not Claimed	Catch Success Rate	GK Distribution	GK Successful Distribution	GK Unsuccessful Distribution	GK Accuracy	GK Short Distribution	GK Successful Short Distribution	GK Unsuccessful Short Distribution	GK Short Accuracy	Clean Sheets	Error leading to Goal	Goals per 90
Name	Y	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	X31	X32	X33	X34
Friedel	853	1	56	27	3	1	0.949	464	301	163	0.649	115	109	6	0.948	8	2	1.26
Reina	749	0	38	22	4	4	0.905	485	359	126	0.740	226	218	8	0.965	18	0	0.74
Hahnenmann	913	4	55	28	6	2	0.859	462	251	211	0.543	58	54	4	0.931	8	1	1.74
Green	831	1	71	19	6	8	0.910	566	364	202	0.643	162	161	1	0.994	8	1	1.32
Kirkland	754	0	59	14	6	1	0.908	396	227	169	0.573	43	40	3	0.930	12	2	1.32
Howard	784	0	50	17	2	7	0.943	560	301	259	0.538	139	137	2	0.986	14	1	0.83
James	1169	3	69	27	5	2	0.920	594	346	248	0.582	164	158	6	0.963	16	3	1.03
Carson	598	1	32	11	5	2	0.865	420	210	210	0.500	76	67	9	0.882	9	2	1.30
Taylor	857	1	62	9	7	6	0.899	457	232	225	0.508	53	52	1	0.981	3	0	1.59
Schwarzer	703	0	58	31	4	2	0.935	416	220	196	0.529	66	63	3	0.955	8	1	1.38
Gordon	564	0	35	21	1	1	0.972	410	257	153	0.627	80	78	2	0.975	6	1	1.62
Almunia	928	0	31	19	1	0	0.969	456	330	126	0.724	177	175	2	0.989	11	0	0.83
Van der Sar	971	1	33	19	2	2	0.917	406	292	114	0.719	186	183	3	0.984	14	1	0.63
Jaaskelainen	671	1	25	17	3	3	0.862	287	153	134	0.533	40	40	0	1.000	8	1	1.50
Hart	1065	1	28	31	3	4	0.903	420	228	192	0.543	136	131	5	0.963	7	1	1.31
Cech	945	0	41	13	4	3	0.911	350	227	123	0.649	104	103	1	0.990	14	3	0.66
Robinson	575	3	24	16	2	2	0.923	273	168	105	0.615	62	62	0	1.000	5	5	1.88
Niemi	800	0	23	13	4	1	0.852	300	148	152	0.493	42	40	2	0.952	2	1	1.77
Harper	925	0	19	12	3	4	0.864	279	158	121	0.566	45	42	3	0.933	6	0	1.41
Given	526	2	13	8	0	0	1.000	206	116	90	0.563	41	41	0	1.000	2	1	2.04
Bywater	677	0	22	14	1	0	0.957	230	140	90	0.609	25	22	3	0.880	2	2	2.28
Carroll	792	1	28	8	1	2	0.966	178	81	97	0.455	25	23	2	0.920	1	3	2.71
Keller	880	0	16	13	2	1	0.889	190	106	84	0.558	34	29	5	0.853	5	0	1.23
Cerny	916	0	17	5	3	1	0.850	198	123	75	0.621	48	47	1	0.979	4	0	1.08
Al-Habsi	986	0	18	3	1	0	0.947	119	56	63	0.471	14	13	1	0.929	4	0	1.20
Cudicini	926	0	14	2	1	1	0.933	116	80	36	0.690	44	42	2	0.955	6	0	0.70
Kuszczyk	1089	0	12	4	0	1	1.000	127	82	45	0.646	45	44	1	0.978	5	0	0.47
Schmeichel	1181	0	5	10	2	4	0.714	105	64	41	0.610	33	33	0	1.000	4	0	0.71
Lehmann	935	0	11	3	0	0	1.000	103	76	27	0.738	39	39	0	1.000	2	2	0.80
Price	589	0	16	2	0	0	1.000	71	22	49	0.310	7	2	5	0.286	0	0	1.67
Isaksson	475	0	7	2	0	0	1.000	73	54	19	0.740	35	34	1	0.971	0	0	2.80
Taylor	856	1	0	3	3	0	0.000	51	29	22	0.569	16	15	1	0.938	0	0	1.84
Doyle	496	0	7	1	0	1	1.000	43	22	21	0.512	2	2	0	1.000	0	2	2.00
Warner	1045	1	8	1	1	0	0.889	33	14	19	0.424	3	3	0	1.000	0	2	1.67
Turnbull	1114	0	4	4	0	0	1.000	35	15	20	0.429	3	3	0	1.000	0	0	1.00
Ashdown	763	0	3	1	0	1	1.000	39	28	11	0.718	14	14	0	1.000	0	0	1.33
Ward	809	0	3	4	1	1	0.750	37	22	15	0.595	6	5	1	0.833	1	0	1.00

Table 2. English Premier League 2007-2008 Season Pitch Actions for Goalkeeper Position (Part 2 of 2).

Oberstone: Comparing English Premier League Goalkeepers

Player Surname	Average OPTA Score per 90	Equip 90 min games	Fouls Conceded	Fouls Won	Yellow Cards	Red Cards	% Goals Conceded Inside Box	Shots from Inside Box	Shots from Outside Box	Saves to Shots Ratio	Saves to Shots ratio Inside Box	Saves to Shots ratio Outside Box
<i>Name</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>X3</i>	<i>X4</i>	<i>X5</i>	<i>X7</i>	<i>X10</i>	<i>X11</i>	<i>X15</i>	<i>X16</i>	<i>X17</i>
Friedel	853	38.00	0.000	0.132	0.000	0.000	0.854	2.921	1.579	0.719	0.631	0.883
Reina	749	38.00	0.000	0.211	0.053	0.000	0.929	1.605	0.658	0.674	0.574	0.920
Hahnmann	913	38.00	0.026	0.184	0.079	0.000	0.879	3.737	2.000	0.697	0.592	0.895
Green	831	38.00	0.026	0.053	0.053	0.000	0.840	2.842	1.526	0.699	0.611	0.862
Kirkland	754	37.00	0.026	0.081	0.000	0.000	0.837	2.865	1.189	0.673	0.613	0.818
Howard	784	36.00	0.000	0.139	0.028	0.000	0.900	2.083	1.222	0.748	0.640	0.932
James	1169	35.00	0.026	0.200	0.057	0.000	0.861	3.114	1.857	0.793	0.716	0.923
Carson	598	34.73	0.053	0.086	0.029	0.029	0.778	2.303	1.324	0.643	0.563	0.783
Taylor	857	34.00	0.026	0.147	0.029	0.000	0.963	3.647	1.882	0.713	0.581	0.969
Schwarzer	703	34.00	0.000	0.294	0.059	0.000	0.766	2.382	1.588	0.652	0.556	0.796
Gordon	564	34.00	0.000	0.118	0.029	0.000	0.855	2.706	1.353	0.601	0.489	0.826
Almunia	928	29.00	0.026	0.103	0.000	0.000	0.792	1.690	1.586	0.747	0.612	0.891
Van der Sar	971	28.50	0.000	0.281	0.070	0.000	0.944	1.789	1.404	0.802	0.667	0.975
Jaaskelainen	671	28.00	0.026	0.321	0.036	0.000	0.833	2.786	1.393	0.641	0.551	0.821
Hart	1065	26.00	0.026	0.115	0.000	0.000	0.971	3.385	2.038	0.759	0.625	0.981
Cech	945	25.71	0.000	0.156	0.000	0.000	0.941	2.061	0.972	0.782	0.698	0.960
Robinson	575	25.00	0.000	0.280	0.000	0.000	0.745	2.720	1.600	0.565	0.485	0.700
Niemi	800	22.00	0.000	0.091	0.000	0.000	0.897	4.091	1.545	0.685	0.611	0.882
Harper	925	19.90	0.000	0.302	0.000	0.000	0.821	3.266	1.759	0.720	0.646	0.857
Given	526	18.10	0.000	0.166	0.110	0.000	0.730	2.983	1.768	0.570	0.500	0.688
Bywater	677	18.00	0.000	0.111	0.056	0.000	0.780	3.944	1.778	0.602	0.549	0.719
Carroll	792	14.00	0.026	0.357	0.000	0.000	0.868	5.286	1.929	0.624	0.554	0.815
Keller	880	13.00	0.000	0.308	0.000	0.000	0.813	2.385	1.846	0.709	0.581	0.875
Cerny	916	13.00	0.000	0.154	0.077	0.000	0.643	2.923	1.538	0.759	0.763	0.750
Al-Habsi	986	10.00	0.000	0.200	0.000	0.000	0.833	3.400	1.600	0.760	0.706	0.875
Cudicini	926	10.00	0.000	0.300	0.000	0.000	0.857	1.800	1.200	0.767	0.667	0.917
Kuszczyk	1089	8.50	0.000	0.235	0.000	0.000	1.000	1.647	1.412	0.846	0.714	1.000
Schmeichel	1181	7.00	0.000	0.286	0.000	0.000	0.800	2.714	1.143	0.815	0.789	0.875
Lehmann	935	6.23	0.000	0.321	0.160	0.000	0.600	1.765	1.765	0.773	0.727	0.818
Price	589	6.00	0.000	0.167	0.000	0.000	0.800	2.833	1.500	0.615	0.529	0.778
Isaksson	475	5.00	0.000	0.000	0.000	0.000	0.857	4.200	1.000	0.462	0.429	0.600
Taylor	856	3.26	0.000	0.307	0.000	0.000	0.667	3.993	1.229	0.647	0.692	0.500
Doyle	496	3.00	0.026	0.667	0.333	0.000	0.833	3.667	1.333	0.600	0.545	0.750
Warner	1045	3.00	0.000	1.000	0.333	0.000	1.000	6.000	1.000	0.762	0.722	1.000
Turnbull	1114	3.00	0.000	0.333	0.000	0.000	1.000	3.667	1.667	0.813	0.727	1.000
Ashdown	763	3.00	0.000	0.000	0.000	0.000	1.000	4.000	0.667	0.714	0.667	1.000
Ward	809	3.00	0.000	0.333	0.000	0.000	0.667	2.333	1.667	0.750	0.714	0.800

Table 3. Revised English Premier League 2007-2008 Season 34 Pitch Actions for Goalkeeper Position: “Per 90-Minutes” Format (Part 1 of 2)

Player Surname	Average OPTA Score per 90	Goalkeeper Smother	Catches	Punches	Crosses not Claimed	Catch Success Rate	GK Distribution	GK Accuracy	GK Short Distribution	GK Short Accuracy	Clean Sheets	Error leading to Goal	Goals per 90
<i>Name</i>	<i>OPTA</i>	<i>X18</i>	<i>X19</i>	<i>X20</i>	<i>X22</i>	<i>X23</i>	<i>X24</i>	<i>X27</i>	<i>X28</i>	<i>X31</i>	<i>X32</i>	<i>X33</i>	<i>X34</i>
Friedel	853	0.0263	1.474	0.711	0.0263	0.949	12.211	0.649	3.026	0.948	0.211	0.053	1.26
Reina	749	0.0000	1.000	0.579	0.1053	0.905	12.763	0.740	5.947	0.965	0.474	0.000	0.74
Hahnmann	913	0.1053	1.447	0.737	0.0526	0.859	12.158	0.543	1.526	0.931	0.211	0.026	1.74
Green	831	0.0263	1.868	0.500	0.2105	0.910	14.895	0.643	4.263	0.994	0.211	0.026	1.32
Kirkland	754	0.0000	1.595	0.378	0.0270	0.908	10.703	0.573	1.162	0.930	0.324	0.054	1.32
Howard	784	0.0000	1.389	0.472	0.1944	0.943	15.556	0.538	3.861	0.986	0.389	0.028	0.83
James	1169	0.0857	1.971	0.771	0.0571	0.920	16.971	0.582	4.686	0.963	0.457	0.086	1.03
Carson	598	0.0288	0.921	0.317	0.0576	0.865	12.092	0.500	2.188	0.882	0.259	0.058	1.30
Taylor	857	0.0294	1.824	0.265	0.1765	0.899	13.441	0.508	1.559	0.981	0.088	0.000	1.59
Schwarzer	703	0.0000	1.706	0.912	0.0588	0.935	12.235	0.529	1.941	0.955	0.235	0.029	1.38
Gordon	564	0.0000	1.029	0.618	0.0294	0.972	12.059	0.627	2.353	0.975	0.176	0.029	1.62
Almunia	928	0.0000	1.069	0.655	0.0000	0.969	15.724	0.724	6.103	0.989	0.379	0.000	0.83
Van der Sar	971	0.0351	1.158	0.667	0.0702	0.917	14.246	0.719	6.526	0.984	0.491	0.035	0.63
Jaaskelainen	671	0.0357	0.893	0.607	0.1071	0.862	10.250	0.533	1.429	1.000	0.286	0.036	1.50
Hart	1065	0.0385	1.077	1.192	0.1538	0.903	16.154	0.543	5.231	0.963	0.269	0.038	1.31
Cech	945	0.0000	1.595	0.506	0.1167	0.911	13.613	0.649	4.045	0.990	0.545	0.117	0.66
Robinson	575	0.1200	0.960	0.640	0.0800	0.923	10.920	0.615	2.480	1.000	0.200	0.200	1.88
Niemi	800	0.0000	1.045	0.591	0.0455	0.852	13.636	0.493	1.909	0.952	0.091	0.045	1.77
Harper	925	0.0000	0.955	0.603	0.2010	0.864	14.020	0.566	2.261	0.933	0.302	0.000	1.41
Given	526	0.1105	0.718	0.442	0.0000	1.000	11.381	0.563	2.265	1.000	0.110	0.055	2.04
Bywater	677	0.0000	1.222	0.778	0.0000	0.957	12.778	0.609	1.389	0.880	0.111	0.111	2.28
Carroll	792	0.0714	2.000	0.571	0.1429	0.966	12.714	0.455	1.786	0.920	0.071	0.214	2.71
Keller	880	0.0000	1.231	1.000	0.0769	0.889	14.615	0.558	2.615	0.853	0.385	0.000	1.23
Cerny	916	0.0000	1.308	0.385	0.0769	0.850	15.231	0.621	3.692	0.979	0.308	0.000	1.08
Al-Habsi	986	0.0000	1.800	0.300	0.0000	0.947	11.900	0.471	1.400	0.929	0.400	0.000	1.20
Cudicini	926	0.0000	1.400	0.200	0.1000	0.933	11.600	0.690	4.400	0.955	0.600	0.000	0.70
Kuszczak	1089	0.0000	1.412	0.471	0.1176	1.000	14.941	0.646	5.294	0.978	0.588	0.000	0.47
Schmeichel	1181	0.0000	0.714	1.429	0.5714	0.714	15.000	0.610	4.714	1.000	0.571	0.000	0.71
Lehmann	935	0.0000	1.765	0.481	0.0000	1.000	16.524	0.738	6.257	1.000	0.321	0.321	0.80
Price	589	0.0000	2.667	0.333	0.0000	1.000	11.833	0.310	1.167	0.286	0.000	0.000	1.67
Isaksson	475	0.0000	1.400	0.400	0.0000	1.000	14.600	0.740	7.000	0.971	0.000	0.000	2.80
Taylor	856	0.3072	0.000	0.922	0.0000	0.000	15.666	0.569	4.915	0.938	0.000	0.000	1.84
Doyle	496	0.0000	2.333	0.333	0.3333	1.000	14.333	0.512	0.667	1.000	0.000	0.667	2.00
Warner	1045	0.3333	2.667	0.333	0.0000	0.889	11.000	0.424	1.000	1.000	0.000	0.667	1.67
Turnbull	1114	0.0000	1.333	1.333	0.0000	1.000	11.667	0.429	1.000	1.000	0.000	0.000	1.00
Ashdown	763	0.0000	1.000	0.333	0.3333	1.000	13.000	0.718	4.667	1.000	0.000	0.000	1.33
Ward	809	0.0000	1.000	1.333	0.3333	0.750	12.333	0.595	2.000	0.833	0.333	0.000	1.00

Table 4. Revised English Premier League 2007-2008 Season 34 Pitch Actions for Goalkeeper Position: “Per 90-Minutes” Format (Part 2 of 2)

Oberstone: Comparing English Premier League Goalkeepers

The multiple regression analysis using the revised set of 24 pitch actions is shown in Table 5.

<i>Regression Statistics</i>	
<i>Sum-of-squares</i>	53,610
<i>SD of residuals</i>	66.8390
<i>R squared</i>	0.9577
<i>Adjusted R squared</i>	0.8730
<i>Multiple R</i>	0.9786
<i>F</i>	11.3128
<i>Significance F</i>	4.74E-19

<i>Variable</i>	<i>Coefficient</i>	<i>SE</i>	<i>t ratio</i>	<i>P value</i>	<i>Signif?</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
<i>(constant)</i>	-1,740.800	761.58	2.286	0.041	Yes	-3400.3	-81.344
<i>X1</i>	-0.032	0.12	0.261	0.798	No	-0.2947	0.2316
<i>X2</i>	-11.828	35.78	0.331	0.747	No	-89.789	66.133
<i>X3</i>	3.818	9.26	0.413	0.687	No	-16.349	23.984
<i>X4</i>	-42.969	22.50	1.910	0.080	No	-91.993	6.056
<i>X5</i>	42.519	125.55	0.339	0.741	No	-231.05	316.09
<i>X7</i>	21.757	333.45	0.065	0.949	No	-704.84	748.35
<i>X9</i>	-3.668	7,061,775	0.000	1.000	No	-1.54E+07	1.54E+07
<i>X10</i>	4.077	7,061,775	0.000	1.000	No	-1.54E+07	1.54E+07
<i>X11</i>	-2.970	7,061,775	0.000	1.000	No	-1.54E+07	1.54E+07
<i>X15</i>	4,035.600	1,445.40	2.792	0.016	Yes	886.14	7185.1
<i>X16</i>	-561.910	868.57	0.647	0.530	No	-2454.5	1330.7
<i>X17</i>	-575.040	475.46	1.209	0.250	No	-1611.1	460.99
<i>X18</i>	55.132	30.49	1.808	0.096	No	-11.312	121.58
<i>X19</i>	1.342	2.50	0.537	0.601	No	-4.106	6.79
<i>X20</i>	7.257	3.61	2.009	0.068	No	-0.6159	15.131
<i>X21</i>	20.200	17.32	1.166	0.266	No	-17.544	57.944
<i>X22</i>	-4.737	9.06	0.523	0.610	No	-24.468	14.994
<i>X23</i>	166.990	161.49	1.034	0.322	No	-184.9	518.88
<i>X24</i>	0.276	0.67	0.415	0.686	No	-1.175	1.727
<i>X27</i>	40.208	241.15	0.167	0.870	No	-485.25	565.66
<i>X28</i>	0.181	1.00	0.181	0.859	No	-1.993	2.354
<i>X31</i>	109.730	177.19	0.619	0.547	No	-276.36	495.82
<i>X33</i>	-29.751	17.73	1.678	0.119	No	-68.386	8.884
<i>X34</i>	210.420	140.91	1.493	0.161	No	-96.628	517.46

Table 5. Multiple Regression Analysis of EPL 2007-2008 Season Using 24 Pitch Actions (InStat 3.1).

Although the overall regression model is statistically significant ($p < 0.0000$), almost all independent variables are not. Only X_{15} had a p -value < 0.05 . The wide confidence intervals among the independent variables also suggest the

presence of multicollinearity—the intercorrelation of pairs of independent variables that artificially inflate the “goodness” of the model.⁴

BACKWARD ELIMINATION IN STEPWISE REGRESSION

We will refine the model using an iterative process of removing, one at a time, the independent variable that presents the highest p -values first—the least significant contributor and most likely culprit. This method is referred to as the backward elimination in stepwise regression [Groebner, et al., 2008] and is repeated until we are left with a statistically significant set of independent variables.⁵ Further, although p -values $\leq .05$ are traditionally desired for the final set of independent variables, the stepwise removal process results in progressively smaller improvements.⁶ Consequently, as long as the set of remaining independent variables have p -values $\leq .15$, they will have practical value in predicting our dependent variable (Winston [2004]).

After a dozen iterations of removing weak independent variables, the model is pared down to a set of six (6) statistically significant goalkeeper pitch actions per 90 minutes of play: (1) X_{10} (shots from inside the box); (2) X_{11} (shots from outside the box); (3) X_{20} (punches); (4) X_{28} (short distribution); (5) X_{32} (clean sheets); (6) X_{34} (goals allowed). The multiple regression analysis for the final is shown in Table 6. The six-variable model offers considerable improvement over the original set of 34 pitch actions: the standard residual error is reduced by about 36 percent (from 66.84 to 42.85 points), the adjusted R^2 increases from .873 to .948 and the significance of the independent variables improves from a setting in which only 1 of the 34 is significant to one in which all six are highly significant (p -values $\ll .05$). The regression coefficients of this model can now be used to estimate the Opta Index, \hat{Y} , of the EPL goalkeepers

$$\hat{Y} = 207.201 + 204.443X_{10} + 196.251X_{11} + 65.339X_{20} + 26.315X_{28} + 354.661X_{32} - 363.735X_{34} \quad [\text{Eq. 2}]$$

⁴ Each R^2 quantifies how well that X variable is predicted from the other X variables; the larger this value, the less unique the information provided by and the greater the “overlap” between each variable.

⁵ It is also important to insure that each removal strengthens the model as typically indicated by a decrease in the model standard error and an increase in the adjusted R^2 .

⁶ Among the most important indications of improvement in the model is a decrease in the standard error of the residuals along with a decrease as is the difference between the adjusted R squared and R squared values.

Each variable coefficient indicates the average rate of change in the goalkeeper’s Opta Index score when all other variables are held constant. A brief interpretation of these six pitch action coefficients follows.

Regression	
Sum-of-squares	54,670
SD of residuals	42.853
R squared	0.9568
Adjusted R squared	0.9482
Multiple R	0.9782
F	110.8386
Significance F	2.0472E-25

Variable	Coefficient	SE	t ratio	P value	Signif?	Lower	Upper
Y	207.201	63.037	3.287	0.0026	-	78.462	335.940
X10	204.443	11.691	17.487	0.0000	Yes	180.567	228.319
X11	196.251	24.082	8.149	0.0000	Yes	147.069	245.433
X20	65.339	25.078	2.605	0.0141	Yes	14.123	116.554
X28	26.315	4.577	5.749	0.0000	Yes	16.966	35.663
X32	354.661	67.981	5.217	0.0000	Yes	215.825	493.496
X34	-363.735	23.041	-15.787	0.0000	Yes	-410.790	-316.679

Table 6. Multiple Regression Analysis of EPL 2007-2008 Season Using 6 Pitch Actions (InStat 3.1).

If the goalkeeper:

- (1) faces an average of one more shot from inside the box, X_{10} , per game, his Opta Index score will increase by approximately 204 points—the keeper is facing more shots while maintaining, among other criteria, the same average goals allowed per game;
- (2) manages an average of one less shot from outside the box per game, X_{11} , he will lose 196 Opta Index points—managing less shots exposes the keeper to lower risk and, theoretically, should lower goals-per-game allowed;
- (3) punches away an average of one more shot on goal per game, X_{20} , he will gain 65 points on his index score;
- (4) increases his short distribution by one additional pass or kick per game, X_{28} , it will increase his score by 26 points;
- (5) experiences a decrease of 10% in the proportion of clean sheets, X_{32} , he will lose about 36 points;
- (6) experiences an additional 0.50 goals allowed per game, X_{34} , will lose 0.50 x 363.735 or 182 points.

The retrodicted data for the 2007-2008 English Premier League season goalkeeper performance delivers an $R^2=0.957$ and $p < 0.0000$, and serves as evidence of the model strength. These results are also plotted in a scatter diagram (Figure 4) and a complimentary radar chart (Figure 5). Although the distribution

of the residual error about the retrodicted Opta Index value in Figure 6 visually supports the suitability of the linear model, a more formal test is also performed that does, indeed, support the assumption of uniform error variance (Newbold, et al., 2006]). These results are shown in Table 7.

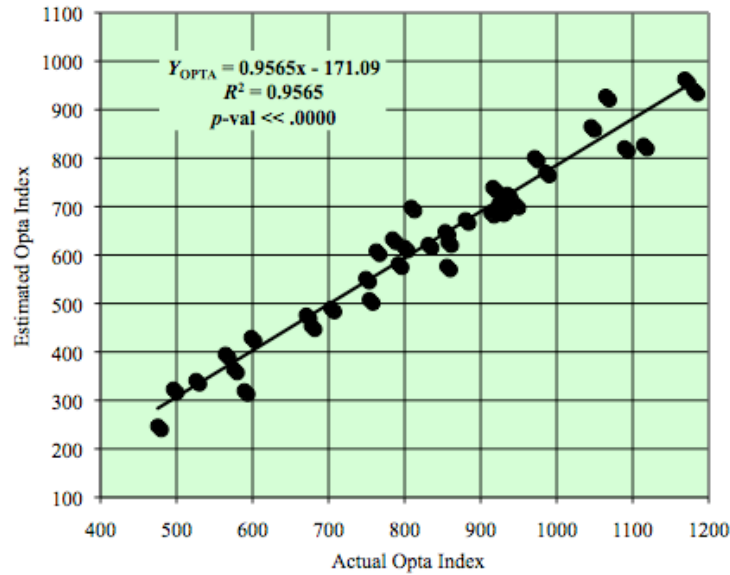


Figure 4. Scatter Diagram of Opta Index Retrodiction for EPL Goalkeepers: 2007-2008 Season

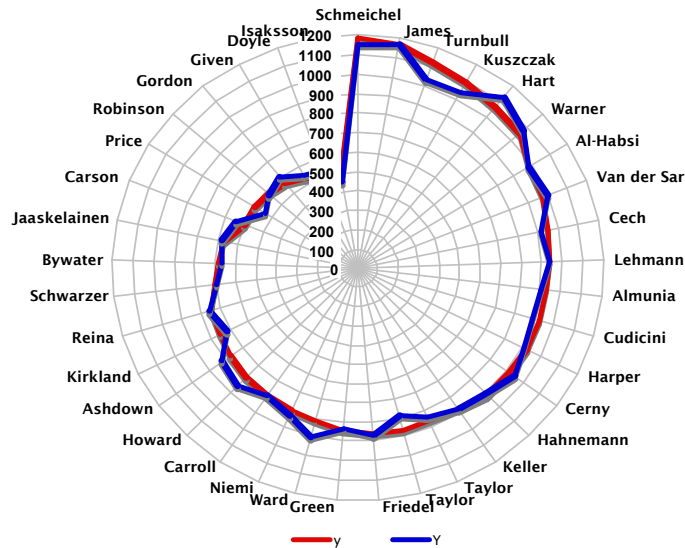


Figure 5. Comparison of Goalkeeper Retrodicted and Actual Opta Index Performance for 6-Variable Multiple Regression Model.

Oberstone: Comparing English Premier League Goalkeepers

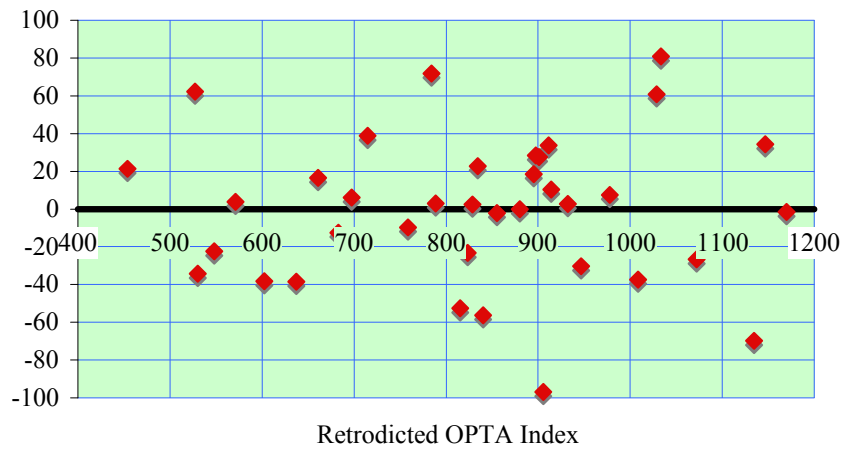


Figure 6. Residual Error of Goalkeeper Retrodiction Values for Opta Index.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.174278042
R Square	0.030372836
Adjusted R Square	0.002669203
Standard Error	183.1586111
Observations	37

ANOVA

	df	SS	MS	F	Significance F
Regression	1	36,779	36,779	1.096	0.302
Residual	35	1,174,148	33,547		
Total	36	1,210,927			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	807.547	36.971	21.843	0.000	732.492	882.603
e2	0.015	0.014	1.047	0.302	-0.014	0.044

Table 7. Regression Test of Heteroscedasticity (Uniform Variance) of EPL Goalkeepers

EXAMINING THE POSSIBLE CONNECTION BETWEEN GOALKEEPER OPTA INDEX SCORE AND AFFILIATED TEAM POINTS DURING EPL SEASON

In addition to the multiple regression model, concern regarding the possible association between the goalkeeper Opta Index score and his team’s success—as measured by the EPL points earned—was examined. The findings shown in Table 8 and Figure 7 suggest that when the 20 team league was segmented into

	FC	Goalkeeper	OPTA Index	Team Points
Top 4	Manchester United	Van der Sar	971	87
	Manchester United	Kuszczak	1089	87
	Chelsea	Cech	945	85
	Chelsea	Cudicini	926	85
	Arsenal	Almunia	928	83
	Arsenal	Lehmann	935	83
	Liverpool	Reina	749	76
Middle 12	Everton	Howard	784	65
	Aston Villa	Carson	598	60
	Aston Villa	Taylor	856	60
	Blackburn Rovers	Friedel	853	58
	Portsmouth	James	1169	57
	Portsmouth	Ashdown	763	57
	Manchester City	Hart	1065	55
	Manchester City	Schmeichel	1181	55
	Manchester City	Isaksson	475	55
	West Ham United	Green	831	49
	Tottenham Hotspur	Robinson	575	46
	Tottenham Hotspur	Cerny	916	46
	Newcastle United	Harper	925	43
	Newcastle United	Given	526	43
	Middlesbrough	Schwarzer	703	42
	Middlesbrough	Turnbull	1114	42
	Wigan Athletic	Kirkland	754	40
	Sunderland	Gordon	564	39
	Sunderland	Ward	809	39
	Bolton Wanderers	Jaaskelainen	671	37
Bolton Wanderers	Al-Habsi	986	37	
Bottom 4	Fulham	Niemi	800	36
	Fulham	Keller	880	36
	Fulham	Warner	1045	36
	Reading	Hahnemann	913	36
	Birmingham City	Taylor	857	35
	Birmingham City	Doyle	496	35
	Derby County	Bywater	677	11
	Derby County	Carroll	792	11
Derby County	Price	589	11	

ANOVA: SINGLE FACTOR

SUMMARY

Groups	Count	Sum	Average	Variance
Top 4	7	6,544	934.869	9,985.0
Mid 12	21	17,117	815.093	43,634.7
Bottom 4	9	7,049	783.258	29,011.8

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	101312.3	2	50,656.15	1.479	0.242	3.276
Within Groups	1164698.2	34	34,255.83			
Total	1266010.5	36				

Table 8. Examination of Possible Influence between Team Points Earned and Goalkeeper Opta Index for Three-Tier ANOVA Grouping.

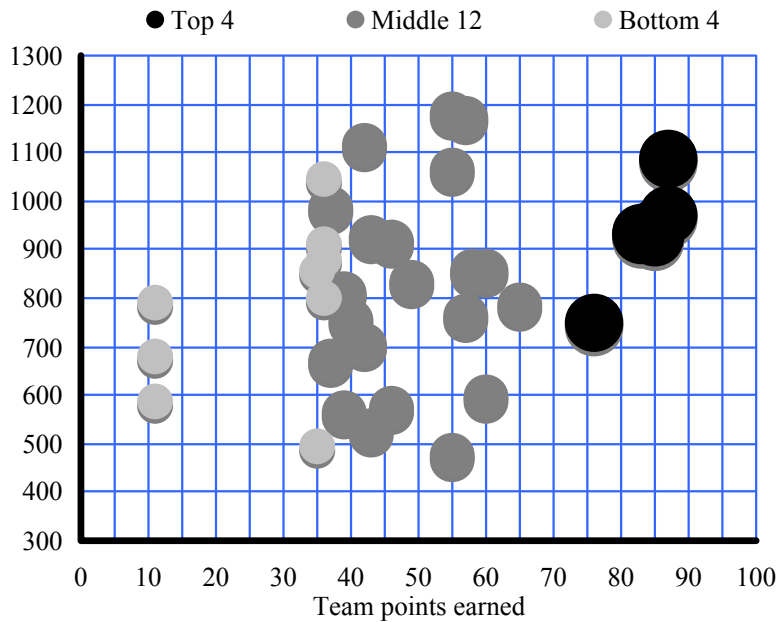


Figure 7. Charting Influence Between Team Points Earned and Goalkeeper Opta Index for Three-Tier ANOVA Grouping.

three groups comprised of the top four clubs, middle dozen, and bottom four, the difference between the tiers based on a single factor ANOVA test was not significant (p -value=.242). Although not conclusive evidence, the likely lack of significance between team success and the goalkeeper’s Opta Index lends a sense of independence to the player’s rating and, therefore, eases concern of an undesired “halo effect” that might favor goalkeepers that play for better teams over those that do not.

CONCLUSION

A linear, multiple regression model that can accurately retrodict the Opta Index of English Premier League goalkeepers during the 2007-2008 EPL season has been developed. The results show that only a small number of pitch actions (independent variables) are needed to accurately define the key performance criteria that are statistically significant indicators of the goalkeeper Opta Index. Care has been taken to insure that the assumption of model linearity has been met.⁷

⁷ The need for determining the model resiliency over several seasons will be examined in a future paper as the new data becomes available.

Although the regression analysis delivers a statistically significant model, use of an index measure as a singular, sufficient gauge of player quality should be viewed with prudence. The precise details describing *how* these index values are determined are proprietary—recipes that are carefully guarded by the three primary organizations that commercially market these esteemed standards—Opta Sportsdata, Actim-PA Sports, and Castrol. Without access to the inner workings of these models, such as the identification of the specific pitch action sets used for each position and how each action is weighted, index values can only be viewed as a “black box” product.

Even though the Opta, Actim, and Castrol index measures provide an interesting meter of player performance and wonderful fodder for heated pub dialog, it is the raw database of pitch actions and the potentially creative ways of employing this information that holds the greatest value for serious sports analysts. Consequently, future research will include the structure of an open player valuation model that will allow the user to customize—to pick and choose—the specific pitch action-weight combinations for all positions: defender, midfielder, forward, and goalkeeper. In addition, vital, non-pitch action considerations that also impact the value of a player will be included, e.g., age, injury history, durability, team fit, emotional-psychological makeup, etc.

Appendix A. Actim “Top 100” Player Rankings for Barclay’s English Premier League 2007-2008 Season.

Actim Player Rankings For The 2007/2008 Barclays English Premier League Season												Last generated on 13/05/2008	
Name	Team	Index Score	Name	Team	Index Score	Name	Team	Index Score	Name	Team	Index Score		
1 Cristiano Ronaldo	Man Utd	838	31 Alexander Hleb	Arsenal	442	51 Christopher Samba	Blackburn	391	71 Luke Young	Middlesbrough	348		
2 El Adebayor	Arsenal	730	32 Simon Davies	Fulham	440	52 Michael Carrick	Man Utd	389	72 Steven Taylor	Newcastle	348		
3 Fernando Torres	Liverpool	682	33 Patrice Evra	Man Utd	439	53 Bacary Sagna	Arsenal	389	73 Blumer Elano	Man City	346		
4 Roque Santa Cruz	Blackburn	652	34 Ayegebeni Yakubu	Everton	437	54 Manuel Almunia	Arsenal	385	74 Mark Schwarzer	Middlesbrough	345		
5 Ashley Young	Aston Villa	640	35 William Gallas	Arsenal	436	55 Nigel Reo-Coker	Aston Villa	384	75 Juliano Belletti	Chelsea	342		
6 Carlos Tevez	Man Utd	598	36 Nicky Shorey	Reading	436	56 James Harper	Reading	384	76 Vedran Corluka	Man City	341		
7 Gareth Barry	Aston Villa	575	37 Mathieu Flamini	Arsenal	429	57 Sebastian Larsson	Birmingham	381	77 Jermaine Jenas	Tottenham	341		
8 Francesc Fabregas	Arsenal	573	38 Phil Neville	Everton	425	58 Kolo Toure	Arsenal	377	78 Ryan Giggs	Man Utd	339		
9 Steven Gerrard	Liverpool	568	39 Stephen Warnock	Blackburn	422	59 John Carew	Aston Villa	375	79 Kenwyne Jones	Sunderland	339		
10 Gael Clichy	Arsenal	565	40 Edwin Van der Sar	Man Utd	419	60 Sol Campbell	Portsmouth	373	80 Lucas Neill	West Ham	336		
11 Wayne Rooney	Man Utd	561	41 Wilfred Bouma	Aston Villa	418	61 Pascal Chimbonda	Tottenham	371	81 Joseph Yobo	Everton	335		
12 David Bentley	Blackburn	555	42 Richard Dunne	Man City	411	62 Phil Jagielka	Everton	371	82 Zat Knight	Aston Villa	332		
13 Joleon Lescott	Everton	549	43 Scott Carson	Aston Villa	410	63 Sylvain Distin	Portsmouth	370	83 Maik Taylor	Birmingham	331		
14 Rio Ferdinand	Man Utd	540	44 Mikel Arteta	Everton	407	64 Chris Kirkland	Wigan	369	84 El-Hadji Diouf	Bolton	326		
15 Gabriel Agbonlahor	Aston Villa	538	45 Stewart Downing	Middlesbrough	405	65 Michael Essien	Chelsea	364	85 Ricardo Carvalho	Chelsea	324		
16 Dimitar Berbatov	Tottenham	519	46 Robert Green	West Ham	399	66 Ashley Cole	Chelsea	363	86 Stephen Ireland	Man City	324		
17 Wes Brown	Man Utd	518	47 Olof Mellberg	Aston Villa	397	67 Stephen Hunt	Reading	359	87 H Hreidarsson	Portsmouth	323		
18 Benjamin Mwaruwari	Man City	517	48 Nemanja Vidic	Man Utd	397	68 Petr Cech	Chelsea	359	88 Andrew Johnson	Everton	321		
19 Robbie Keane	Tottenham	511	49 Glen Johnson	Portsmouth	394	69 Michael Owen	Newcastle	351	89 Alvaro Arbeloa	Liverpool	321		
20 Salomon Kalou	Chelsea	497	50 M Hahnmann	Reading	394	70 George McCartney	West Ham	349	90 Antonio Valencia	Wigan	319		
21 Jose Reina	Liverpool	492							91 Nicky Butt	Newcastle	318		
22 Nicolas Anelka	Chelsea	479							92 John Terry	Chelsea	316		
23 Joe Cole	Chelsea	475							93 Jeremie Aladiere	Middlesbrough	315		
24 Brad Friedel	Blackburn	468							94 Paul Schamer	Wigan	315		
25 Martin Petrov	Man City	467							95 Les Cansley	Everton	312		
26 Martin Laursen	Aston Villa	466							96 Craig Gordon	Sunderland	312		
27 David James	Portsmouth	456							97 Aaron Lennon	Tottenham	312		
28 Jamie Carragher	Liverpool	455							98 Steed Malbranque	Tottenham	311		
29 Tim Howard	Everton	449							99 Liam Ridgeway	Birmingham	310		
30 Frank Lampard	Chelsea	448							100 David Wheeler	Middlesbrough	309		



REFERENCES

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.

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).

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.

Clark, Ephraim A. and Tunaru, Radu and Viney, H. (2005) An option pricing framework for the valuation of football players. *Review of financial economics*, 14 (3). pp. 281-295. ISSN 1058-3300

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.

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

Oberstone, Joel (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*. DOI: 10.2202/1559-0410.1183

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).

Winston, Wayne, "Microsoft Excel Data Analysis and Business Modeling," Microsoft Press (2004), pp. 360-361.