Should He Stay or Should He Go? Estimating the Effect of Firing the Manager in Soccer

Rolf Poulsen
Department of Statistics and Operations Research
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Oe Denmark
rolf@math.ku.dk

November 8, 1999

1 Introduction

Professional sport teams often fire their managers when the team performs poorly. But does it help? Let's find out. We use soccer (or football as it is known in most of the world and in the rest of this article) data to address this question.

2 Collecting Data

To do such an analysis, we need data about the firing of managers and about the performance of football teams. We use data from the top two English soccer leagues. Based on anecdotal evidence, the English league is not the "most dangerous" for managers, there are other leagues where sackings are more frequent. It is, however, the only league where the amount of data accessible to us makes it possible to do "statistical analysis" rather than "case studies".

Results of football matches. These are easy to quantify. The results (and dates) of all matches can be obtained either from Rothmans Football Yearbook (RFY), bought in computer readable form, or found (for recent seasons) on the Internet. Directing your browser to the home page of The Association of Football Statisticians (`http://www.innotts.co.uk/~soccerstats/`) is a good way to start.
A match can result in a home win, an away win or a draw (around 25% of all matches end in draws). 3 points are given for a win, 1 point for a draw and 0 points for a loss. League tables (and hence championships, promotions and relegations) are based on how many points a team collects over the season (a round robin tournament with home and away games). Therefore the number of points a team gains from a match is the central measure of performance.

Firing of managers. This information is not registered in structured form, at least not if by "structured" we mean "a large computer readable file with precise dates". This matter can be solved by cross-referencing the "Milestones diary" and the "English League Managers" sections in a volume of RFY (yielding information for one season).(Since this source uses the word "sack" rather that "fire" we will do the same for the rest the article.) In this way we get exact dates and are sure that we have not missed any sackings. The data used in the analysis can be found on the web-page `http://www.math.ku.dk/∼rolf/`. The board of directors or the chairman of the

Season(s)	PL	D1
1993-94	6 (22)	9(24)
1994-95	8 (22)	13 (24)
1995-96	3*(20)	8*(24)
1996-97	7 (20)	9 (24)
1997-98	4 (20)	13 (24)
1993-98	28	52

Table 1: Number of regular season sackings in the Premier League and Division One. (*: Crude estimates, exact dates not available) Numbers in parentheses indicate the total number of teams in the division. Note that it is possible for a club to sack more than one manager in a season.

club decides if the manager is to be sacked. The source may say that the manager "has been sacked", "has resigned", "leaves after mutual consent", or use some more poetic term. It is all the same to us and we will refer to it as a "sacking". In our analysis we use the time of sacking, not the time at which a new manager is appointed. We could make up a story that justified this, but the primary reason is that it is the easiest thing to observe. Many changes on the managerial front take place during the summer break (mid-May to mid-August). Several of these are in effect sackings but we are only able to analyze regular season sackings. The number of regular season manager sackings in the top two English divisions for the 1993-94 to 1997-98 seasons are given in Table 1. Judging from this, Premier League managers have 25% risk of being sacked, while Division One managers have a sacking risk of over 40%. But as one would probably expect, this risk is not "evenly spread", clubs from top halves of tables rarely sack their managers.
Of course we would like more observations of sackings, but for now we will have to make do with the data in Table 1.

3 Estimating the "Sacking Effect"

Sackings rarely come as bolt from the blue; typically managers are sacked after a spell of poor results. In other words the "mechanism" that produces our observations (the sackings) is not independent of "the underlying random process" of interest (the results). This means that we have to be careful when we want to draw inference otherwise we might get out spurious results ("regression towards the mean") due to selection anomalies ("self-selection"). To illustrate, suppose a team has, purely by chance, performed very poorly and that its manager is sacked. Then we are likely to see an increase in performance after the sacking when compared to the "before" results. We try to circumvent the problems by comparing the "after sacking" performance (the "treatment group") to that of teams which had an equally poor run of form/luck but did not sack the manager (the "control group"). It also seems natural not just to decide what "poor" performance is based on actual results, but to measure performance relative to some "expected performance" of the team. So we need an estimate of what "expected performance" is. Of course for a "sacking" team such an estimate can be affected by the (unlucky) events that led to the sacking, but we still believe that more is gained than lost even if we use a relatively simple model to measure expected performance; at least when we use all matches in the season and all teams in the league to find the estimates.

We use two methods to investigate the effects of sacking the manager. The first is a simple method that looks only at wins/draws/loses and actual points. This is robust in the sense that it does not rely on a formal model. However, this method is not very efficient. The strength of the opposition is not taken into account. In the long run this "noise" will probably "average out", but try telling that to a newly appointed manager whose first game is away against a top team. The second method builds on a model suggested by Alan Lee (Chance, Winter 1997) that estimates the home-advantage and relative strength of all the teams. This can be used to make an adjusted measure of team performance. We then compare the "before"and "after" performance of "sacking"and "non-sacking" teams based on this measure.

3.1 A Simple Comparison

To set up the data for analysis we use the following recipe.

- Find all sackings and their exact dates.
- Write a piece of software that, for any given sacking, returns the number of points the team got in the n_B games immediately before the sacking and in the n_A matches after. Now choose specific n_B and n_A values, run the program and

BEFORE	AFTER							
# points "before"	No sacking			Sacking			Same mean?	
	#obs.	mean	std. dev.	#obs.	mean	std. dev.	t-val.	p-val.
[0; 4]	451	3.691	2.21	22	4.455	2.28	1.64	0.101
[5; 7]	890	3.884	2.14	23	4.174	2.80	0.635	0.526
[8; 13]	1079	4.121	2.24	14	3.281	2.62	-1.39	0.165

Table 2: Robust analysis with "before" performance measure based on 6 matches ($n_B = 6$) and the "after" performance on 3 ($n_A = 3$). The test probabilities (p-val.) are calculated using a 2-sided t-test based on equal variances. A high p-value indicates that the means are equal, while a low p-value (traditionally less than 0.05) leads us to reject that hypothesis.

store the results.

- Divide the sacking teams into (a reasonable number of) classes based on "before" performance (in n_B games), where "performance" means "points". For example: Put the teams that got 0 to 4 points in 6 "before" games in one class (which is then called the [0; 4]-class). Now calculate the mean and standard deviation of the "after" points for the teams in the different classes.

- Write another piece of software that, for any given performance class, finds instances where a team has a performance that puts it in this class but the manager is *not* sacked. Calculate the mean and standard deviation of the "after" points for such "non-sacking" teams in the different classes.

- Compare "after" performance of "sacking" and "non-sacking" teams in the same class. In practice this amounts to using a statistical test to check if the "after points" means are the same.

We choose $n_B = 6$ and $n_A = 3$ in the reported analysis, but, within reason, the results are insensitive to this choice. It is common to look at a period of 6 matches when discussing the "form" of teams (as people reading the gambling sections of newspapers will know). The length of the "after" period is a compromise; on one hand we want a short period so that "all other things are equal", but on the other hand we want a longer period because that gives us more data. It would be possible to use a more "continuous" weighting scheme, which would mean allowing a sacking

BEFORE	AFTER							
Excess points	No Sacking			Sacking			Same mean?	
"before"	#obs.	mean	std. dev.	#obs.	mean	std. dev.	t-val.	p-val.
$]-6;-3]$	154	0.1281	2.32	19	1.3402	2.48	2.125	0.035
$]-3;-1]$	235	0.1030	2.33	20	-0.3575	1.76	-0.860	0.391
$]-1;6]$	686	-0.1311	2.42	20	-0.2425	2.08	0.203	0.839

Table 3: Sophisticated analysis with $n_B = 6$ and $n_A = 3$. The test probabilities are calculated as in Table 2.

effect to gradually decay (or increase, for that matter) over time, but we will not do that. Regarding the division into performance classes we also have to compromise. With a high number of classes we feel more sure that homogeneity within each class is reasonable, but a high number of classes "thins out" our – already somewhat scarce – data, so it becomes hard to achieve statistically significant results. We use 3 classes in the reported analysis and have chosen the point intervals such that there are roughly the same number of "sacking" teams in each class; in other we treat the best, middle and worst thirds of "sacking" teams separately.

Table 2 gives results for the robust method, i.e. when performance measurement and division into classes is based solely on actual points. The observed number of sackings in this table is smaller than the number of sackings in Table 1 since we have had to discard observations when sackings took place less than 6 games into the season or less than 3 games before the end of the season. We note a monotonicity of differences of means; the worst performing teams have the largest positive effects, and for the best performing sacking teams there is actually an indication of a negative effect (the t-value is negative). But none of the means are statistically significantly different. So: The signs (on t-values) are "right" but the std. deviations are too large, which we take as a strong encouragement to perform a more efficient/sophisticated analysis.

3.2 A More Sophisticated Comparison

We use the (quite reasonable) model described in the SIDEBAR to capture the home advantage effect and the fact that some teams *are* better than others. Once the model has been estimated (which can been done with standard statistical software), the probability distribution of the outcome of any match can be determined. In particular we can find the number of points a team can expect from a match. Now imagine that all points are substituted by "excess points", that is by "actual points" - "expected points". We can then perform the same analysis as before. The results are shown in Table 3.

We see that for the sacking teams with the worst runs of form there is in fact a significant (at the 5% level; the p-value is 0.035) increase in performance compared to teams that do not sack their manager. But if teams are doing well, or just "not too bad" then there appears to be no gain from sacking the manager. If we use 6 performance classes (of roughly equal size) then mean differences are negative for the "lower" classes, but the data has been thinned too much for significance. If we use only 1 performance class then we find a test probability of around 40% of "no sacking effect". This indicates that sacking the manager is not something you should do in general to boost performance. To use an analogy, sacking the manager is like taking medicine: It helps if you are sick, but you don't get similar positive effects if you are well.

4 Conclusion

Overall, we can give no definitive answer to the question posed in the title of the paper. If we could, we would have used a different title. But if you are the chairman of a club that is performing really poorly, then sacking the manager is probably not a bad idea. But of course a manager typically has some compensation scheme. The effects of this on the club economy has not taken into account. Despite its relevance, doing this seems an almost impossible task. What is possible, however, is to gather more data on sackings in order to further validate the results in this article (or the opposite). We are constantly doing this. Another interesting thing would be to see if the results are robust across sports, in particular we would like to look at sports where results tend to be more predictable (such as American football or basketball) than in soccer. Finally, let us stress that we have quite deliberately avoided any discussion as to why there is (or is not) an effect from sacking the manager. We leave that to sports psychologists or to the true experts – the fans.

References

Dixon, M. J. & Coles, S. G. (1997), 'Modelling Association Football Scores and Inefficiencies in the Football Betting Market', *Applied Statistics* **46**(2), 265–280.

Lee, A. J. (1997), 'Modeling Scores in the Premier League: Is Manchester United Really the Best?', *Chance* **10**(1), 15–19.

$\boxed{\textbf{SIDEBAR}}$

Alan Lee (Chance, Winter 1997) suggests modeling the number of goals scored, say X, by a particular team in a particular match as a Poisson(λ)-distributed random variable, which means that

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2 \ldots.$$

Further it is assumed that i) matches are independent, ii) the number of home and away goals are independent (this may be questionable, Dixon & Coles (1997) suggest an improvement, but it is not totally unreasonable), and iii) home and away scoring intensities in a particular match satisfy

$$\begin{aligned} \ln \lambda_{HOME} &= \beta + \beta_{HOME} + \beta_{OFF}(HOME\_TEAM) + \beta_{DEF}(AWAY\_TEAM), \\ \ln \lambda_{AWAY} &= \beta + \beta_{OFF}(AWAY\_TEAM) + \beta_{DEF}(HOME\_TEAM). \end{aligned}$$

To identify and interpret parameters we impose the restrictions

$$\sum \beta_{OFF} = 0, \qquad \sum \beta_{DEF} = 0.$$

This generalized linear model can easily be estimated using maximum likelihood, for example by the procedure `glim` in `Splus`. When this has been done it is easy to find the win/draw/loss probabilities for any match in the season, and from this we can find the number of points a team can expect from any given match.