

When should you sack the manager? Results from a simple model applied to the English Premiership.

Chris Hope
Judge Institute of Management
April 2002

Abstract

By the end of October 2001, less than three months into the 2001-02 football season, 21 managers from the 92 English Premiership and Nationwide League clubs had lost their jobs. Many commentators thought that the clubs were too quick to act. So what strategy should a football club adopt when deciding whether to sack its manager?

This paper introduces a simple model assuming that a club's objective is to maximize the number of league points that it scores per season. The club's strategy consists of three choices:

- the length of the honeymoon period during which it will not consider sacking a new manager,
- the level of the trapdoor, the average number of points scored per game; if the manager's performance falls below this, he will get the sack,
- the weight that it will give to more recent games compared to earlier ones.

Data from the last six seasons of the Premiership are used to calibrate the model. The best strategy is to have only a short honeymoon period of 8 games, to set the trapdoor at 0.74 points per game, and to put 47% of the weight on the last five games. A club adopting this strategy would obtain on average 56.8 points per season, compared to a Premiership average of 51.8 points. It would employ an average of 5.7 managers every ten seasons, against the Premiership average of 4.5 managers. It would have sacked John Gregory at Aston Villa and Walter Smith at Everton, but not Ruud Gullit or Gianluca Vialli at Chelsea, nor Joe Royle at Everton, just.

Priorities for further work include developing the model to include the distinction between home and away games, the quality of the opposition, the importance of avoiding relegation, and the different aspirations of different clubs, and applying it to other leagues than the Premiership.

Introduction

The 2001 – 2002 season was an uncomfortable one for the managers of the 92 clubs in English Premiership and Nationwide League football. By the 26th of October 2001, 18 managers had lost their jobs since the start of the season in August, and Walter Smith, then Manager of Everton, was lamenting

“I feel managers are being sacked too quickly. I think circumstances placed upon directors in the financial sense means some are getting panicked into changing their managers too quickly. It means clubs never get the stability which I believe is so vital to them.”

http://news.bbc.co.uk/sport/hi/english/football/newsid_1621000/1621182.stm

Smith’s lament went unheeded, and at the end of October, just five days later, it was reported that

“Oldham Athletic have sacked manager Andy Ritchie - the 21st manager to leave his club since the start of August.”

http://news.bbc.co.uk/sport/hi/english/football/teams/o/oldham_athletic/newsid_1629000/1629700.stm

Smith himself was to lose his job in March 2002.

So were the clubs too quick to act? What can we say about the strategy that a football club should follow in deciding whether to sack its manager?

This paper introduces a simple model of this decision problem. The club is assumed to want to maximize the number of league points that it scores per season. To do this it would like to employ the best manager it can. There are five *types* of manager: poor, fair, good, excellent and world class. But the club cannot observe the quality of its manager directly. Instead it looks at results on the pitch.

The club has three choices to make:

- the length of the *honeymoon* period during which it will not consider sacking a new manager,
- the level of the *trapdoor*, the average number of points scored per game; if the manager’s record falls below this, he will get the sack,
- the weight that it will give to more recent games compared to earlier ones. The club increases this weight by increasing the amount of *smoothing* it applies when keeping track of the manager’s results.

The choice is complicated by the likelihood that a change of manager will initially *inspire* the team and get a *boost* in performance, and then require some time to *rebuild*, during which the team’s performance will *drop*. And even the best manager will eventually see his performance drop somewhat with *age*, and even more as his skills and relationship with the club *decay*.

Whenever a new manager is appointed, he will demand a *contract* for a number of seasons, and a hefty *salary*. If he is sacked before his contract expires, the club will have to pay it up, using money that could otherwise have been used to *buy success* on the pitch, for example through buying new players.

So the essence of the club's dilemma is this: every time it sacks a manager it may get a short-lived boost in performance, but it incurs a substantial cost and a subsequent period of rebuilding, both of which cost points. But if it doesn't sack a mediocre manager, it will continue to perform badly.

If the club sets the honeymoon period too short, it risks wasting money by sacking a lot of managers, some of whom might have turned out to be excellent or world class, but unlucky in their first few games. If it sets the honeymoon period too long, it will keep even poor managers for longer than their performance would merit.

If the club sets the trapdoor too high, it will sack a lot of managers, some of whom could be superb managers going through a sticky spell. Too low, and even mediocre managers will never get the sack.

If the club relies too much on the most recent results, it will sack a lot of managers, as even a short bad patch will lead to dismissal. If the club uses too little smoothing, it will take a long while to sack even those managers whose performance has aged and decayed.

The next section describes a model built using the terms in italics in the above description. Then, results from the English Premiership are used to calibrate the model, and the records of some well-known managers are examined. The model is used to explore a range of strategies that a club could adopt, and some general guidelines are found. Then an optimal strategy is calculated, its implications described, and it is applied to some real managers. Finally the simplifications of the model are discussed and some proposals for further research are made.

The model

In this section the model is described in enough detail to allow it to be reproduced. Anyone not interested in the details of the model can skip to the next section to see the data and results for the Premiership.

Points per game

In the English Premiership and Nationwide football (soccer) leagues, games are scored at 3 points for a win, 1 point for a draw, and 0 points for a loss. Each team plays every other team twice during a season, once at home and once away.

Assume that, for each game, a club has a probability of winning of $p(win)$, of drawing $p(draw)$, and of losing $1 - p(win) - p(draw)$. The mean number of points scored per game, g , will be given by

$$g = 3p(win) + p(draw)$$

The league

Averaged over a whole season and all the teams in the league, the number of wins must equal the number of losses, so, using capital letters to represent the average over the season for the league as a whole,

$$P(win) = 1 - P(win) - P(draw), \text{ or}$$

$$P(draw) = 1 - 2P(win)$$

So

$$\begin{aligned} G &= 3P(win) + P(draw) \\ &= 3P(win) + 1 - 2P(win) \\ &= 1 + P(win) \end{aligned}$$

At the two extremes, if all games in the league are drawn, $P(win) = 0$, and $G = 1$; if all games are either won or lost $P(win) = 0.5$ and $G = 1.5$. $P(win)$ is an input to the model.

The managers

Table 1 shows the five types of manager that are assumed to be available, their quality measured by $g(normal)$, the mean points per game they will obtain, and the probability of obtaining them each time a new manager is hired.

Table 1 The five types of manager

Type	$g(normal)$	prob
World class	$G + 2d$	$P(2d)$
Excellent	$G + d$	$P(d)$
Good	G	$1 - 2(P(2d) + P(d))$
Fair	$G - d$	$P(d)$
Poor	$G - 2d$	$P(2d)$

Table 1 shows that a good manager will equal the mean performance in the league as a whole, while the excellent and world class managers will do better than this, with fair and poor managers their mirror image in both quality and probability. If the population of managers in the league is drawn randomly from this table, then the mean number of points per game over the season will be G , as required. The model inputs are d , $P(2d)$ and $P(d)$. G is derived from $P(win)$.

Time profile for a manager

Figure 1 shows how a manager’s performance, measured by g , is assumed to vary over time. For the first few games (*inspire*), the team are inspired by the new appointment and obtain a boost in performance. Then there is a period of rebuilding (*rebuild*) and the team’s performance drops below the long term average. Then there is the normal period (*age – rebuild – inspire*), where the team performs at the level $g(normal)$ described in table 1. Then, if the manager has not been sacked, there is a period (*decay*) where his performance drops by one type from table 1, and finally by two types, but never below ‘poor’. The model inputs are *inspire*, *boost*, *rebuild*, *drop*, *age*, and *decay*.

Figure 1 Manager performance by number of games

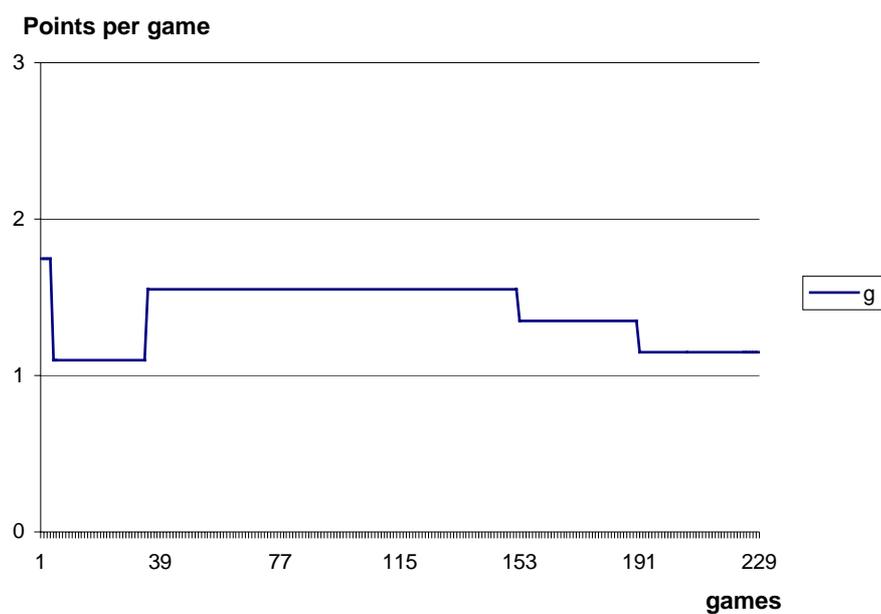


Figure 1 shows the profile for an excellent manager, with $P(win) = 0.35$ (so $G = 1.35$ points per game), $d = 0.2$ points per game, $inspire = 4$ games, $boost = 0.2$ points per game, $rebuild = 30$ games, $drop = 0.45$ points per game, $age = 152$ games (4 Premiership seasons), and $decay = 38$ games (1 Premiership season).

Formally, at any stage of a manager's career with a club, g is described by the relationship in table 2

Table 2 Value of g versus games played

Game From	to	g
1	<i>inspire</i>	$g(normal) + boost$
<i>inspire</i> + 1	<i>inspire</i> + <i>rebuild</i>	$g(normal) - drop$
<i>inspire</i> + <i>rebuild</i> + 1	<i>age</i> - 1	$g(normal)$
<i>age</i>	<i>age</i> + <i>decay</i> - 1	$\max(g(normal) - d, G - 2d)$
<i>age</i> + <i>decay</i>	infinity	$\max(g(normal) - 2d, G - 2d)$

If $inspire + rebuild > age - 1$, there is no period when the manager is performing 'normally'. However, this is not a problem with any reasonable set of input values.

Generating the results of games

It is assumed that all types of managers, at all stages of their career, have the same proportion of draws. Data from the English Premiership, shown later, support this assumption.

$$p(draw) = P(draw) = 1 - 2P(win).$$

Since

$$g = 3p(win) + p(draw)$$

we have

$$p(win) = (g - P(draw))/3$$

$$p(lose) = 1 - P(draw) - (g - P(draw))/3$$

$$= (3 - 2P(draw) - g)/3$$

Clearly this assumption will cease to make sense if $p(win) < 0$ or $p(lose) < 0$, so the model checks to ensure that these conditions are not broken, for any type of manager, in any period of his career.

For each game, a uniform random number between 0 and 1 is generated, and if the number is less than $p(lose)$, 0 points are scored, if the number is between $p(lose)$ and

$p(\textit{lose}) + p(\textit{draw})$, 1 point is scored, and if the number is above $p(\textit{lose}) + p(\textit{draw})$, 3 points are scored. Formally, where m represents the number of the game,

$$\textit{result}(m) = \begin{array}{l} 0 \text{ if the game was lost} \\ 1 \text{ if the game was drawn} \\ 3 \text{ if the game was won} \end{array}$$

The performance of a manager

After his first game the performance of the manager is given by

$$\textit{perf}(1) = \textit{result}(1)$$

After any subsequent game his performance is given by an exponentially weighted average of his results to date:

$$\textit{perf}(m) = \textit{smooth} * \textit{result}(m) + (1 - \textit{smooth}) * \textit{perf}(m-1)$$

where \textit{smooth} is between 0 and 1, and is a choice variable in the model. The higher the value of \textit{smooth} , the more weight is given to the most recent results. At the extreme, if $\textit{smooth} = 1$, only the result of the most recent game is considered. This form of smoothing is commonly used in quality control in industry, and has the great advantage that only the most recent result and the most recent previous value of performance need to be considered (Eppen, Gould and Schmidt, 1990).

Sacking the manager

The manager is assumed to be given a honeymoon period. If $m < \textit{honeymoon}$, the manager will not be sacked however badly he performs. After every game beyond the honeymoon period, the manager is sacked if his performance drops below a certain value. Formally

If $(m \geq \textit{honeymoon})$ and $\textit{perf}(m) < \textit{trapdoor}$, then the manager is sacked.

$\textit{Honeymoon}$ and $\textit{trapdoor}$ are choice variables in the model.

Once the manager has been sacked, a new manager is drawn at random from the types in table 1, and m is reset to 1.

Performance of the club

The aim of the club is to score as many points per season as possible. A running tally of the points scored is kept for 380 games (10 Premiership seasons). This is then adjusted in two ways.

1. Every time a manager is sacked before his contract expires, points are deducted from the club's total. The number of points deducted is found by multiplying the

amount of time left on the contract by the manager's salary, and by the number of points that each £million pound's investment can buy.

Formally, if the manager is sacked after *msack* games, then points are deducted if (*msack* < *contract*). The number of points deducted is

$$(contract - msack) * salary * buy_success/1000$$

where

contract is the length of the contract in games,

salary is the manager's salary in £000 per game,

and

buy_success is the number of points that can be bought for £1 million.

contract, *salary* and *buy_success* are model inputs.

Given the progression of the Premiership towards a pure business proposition, there was a temptation to cast the whole model in terms of discounted cash flow, but, on reflection, I opted to stay with the traditional league points as the measure of success.

2. At the end of the 380 games, the club's points are adjusted upwards by an amount equal to the points that would be lost through rebuilding and paying off the managers' contracts by a club that had employed an average number of managers over the period. This adjustment has no effect on the operation of the model, but is necessary to allow the number of points scored by the club to be compared fairly with the premiership average.

Formally if the average number of managers is *manave*, the average number of sackings will be *manave* - 1, and the average tenure of the managers will be

$$avegame = 380/manave \text{ games.}$$

Then the number of points lost for each sacking is

$$Aveloss = rebuild*drop - inspire* boost + \max(0,((contract - avegame) * salary * buy_success/1000))$$

and the number of points added to the club's total is

$$(manave - 1) * aveloss.$$

manave is an input to the model.

The total points scored by the club over the 380 games including the two adjustments is divided by 10, to give an average number of points per season. It is this value of points per season that the club is trying to maximize.

The English Premiership

Since the 1995-96 season, the English Premiership has been made up of 20 clubs (24 clubs before that date), so a season has 38 games. The six seasons 1995-96 to 2000-01 make up the database of complete seasons in the present format. Mables-tables.com provides results of each game played in the Premiership in these seasons, in a spreadsheet format. Over the six seasons, the proportion of wins is 0.363, and so the average number of points per game is 1.363.

At the end of each season, the bottom 3 clubs are relegated and replaced by clubs from the Nationwide League division 1. The top 3 clubs (4 from the 2001-2002 season) qualify for the European Champions League.

Points per game and types of manager

Table 3 shows the mean number of points scored per season and per game by clubs in the top 3 places, the next 4 places, the middle 6 places, the next 4 places, and the bottom 3 places.

Table 3 Mean points by finishing position

English Premiership 1995-96 to 2000-01

Finishing position	Points	
	per season	per game
1 – 3	74.78	1.968
4 – 7	60.96	1.604
8 – 13	50.31	1.324
14 – 17	41.38	1.089
18 – 20	33.56	0.883

Source: mables-tables.com

A good approximation to these mean points is given by model inputs of $P(\text{win}) = 0.37$, $d = 0.27$, $P(d) = 0.2$, $P(2d) = 0.15$. They imply mean points per game, $g(\text{normal})$, of 0.83, 1.1, 1.37, 1.64 and 1.91 for poor (bottom 3), fair (next 4), good (mid-table 6), excellent (next 4) and world class (top 3) managers respectively. These values are all within 0.05 of the actual Premiership values.

The proportion of drawn games in the five categories was 0.25, 0.26, 0.29, 0.29, and 0.26 respectively. This near constancy inspired the model assumption that all types of manager have the same proportion of drawn games. A $P(\text{win})$ value of 0.37 implies $P(\text{draw}) = 0.26$.

Time profile of manager performance

It is not so simple to find reasonable values of *inspire*, *boost*, *rebuild*, *drop*, *age*, and *decay* to describe the time profile of a manager's performance.

Most commentators seem confident that there is an initial period during which the team's performance is boosted:

"If every incoming manager has a honeymoon [inspired in our terminology] period where they inspire good form and fortune, then Yugoslavia will be hoping that their recently appointed management trio can have triple the effect of a normal new arrival." Beale, 2001

and a subsequent period of rebuilding:

"Marseille look to Blanc to rebuild the club" County Life, 2001

"Affable and enthusiastic as he is, does anyone really believe Keegan is the type to rebuild a club like City?" Kelly, 2001.

But the statistical evidence, particularly on the initial inspired period, is not so strong. It could be that commentators are comparing the performance of the new manager with the last few games of the outgoing manager, which are likely to have been rather poor to earn him the sack. However what is needed for the model is a comparison of the new manager's first few games with his subsequent normal performance.

Figures 2 below shows the performance of five high profile managers from their first game until their sacking (or until the end of the 2000 – 01 season in the case of Wenger).

Figure 2 The performance of five Premiership managers

Figure 2a Arsene Wenger at Arsenal

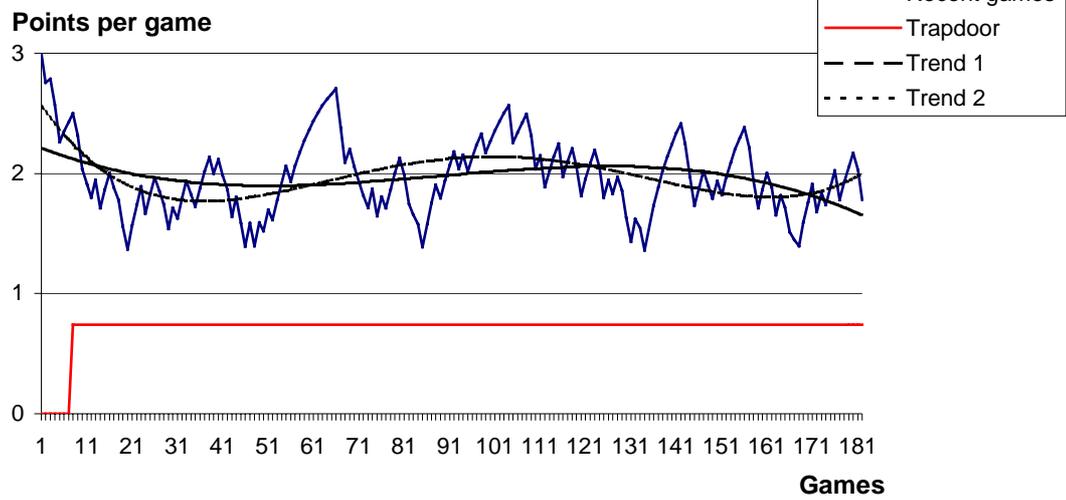


Figure 2b John Gregory at Aston Villa

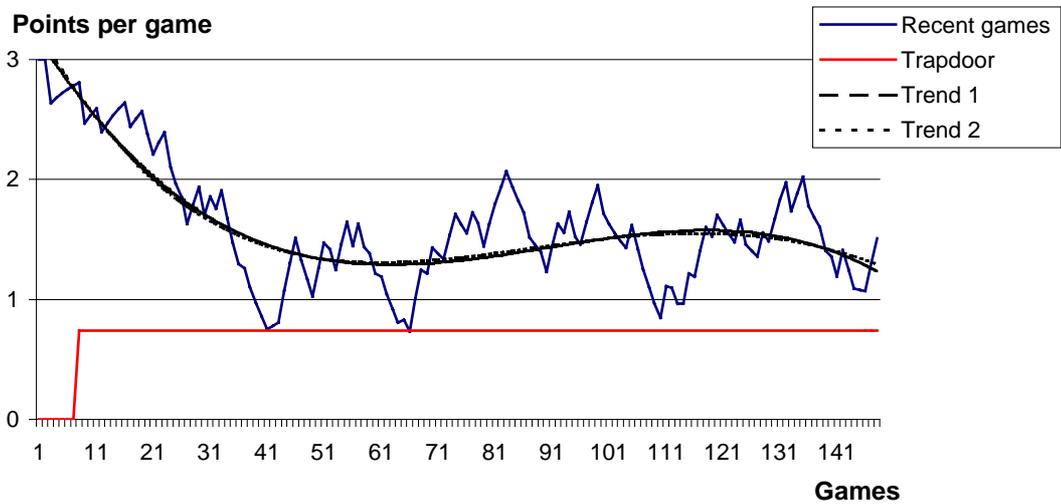


Figure 2c Ruud Gullit at Chelsea

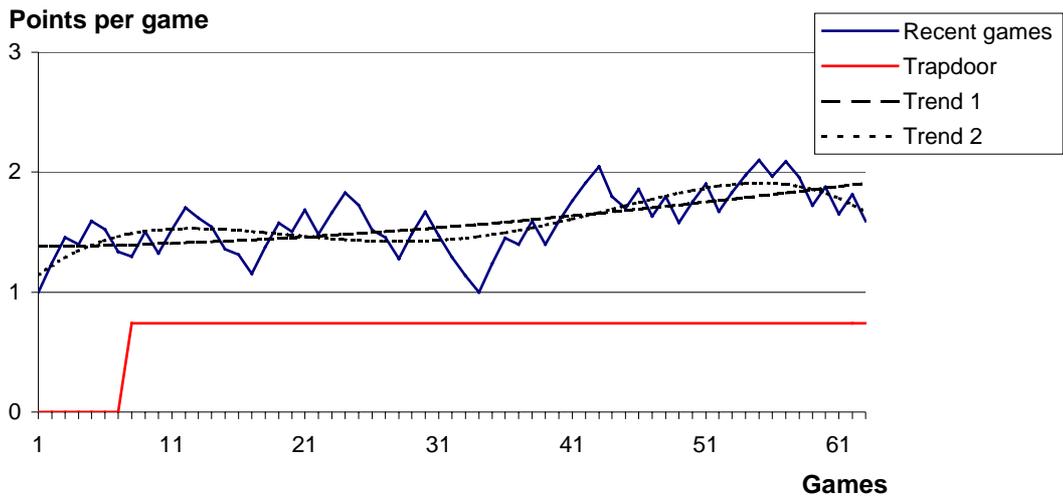


Figure 2d Gianluca Vialli at Chelsea

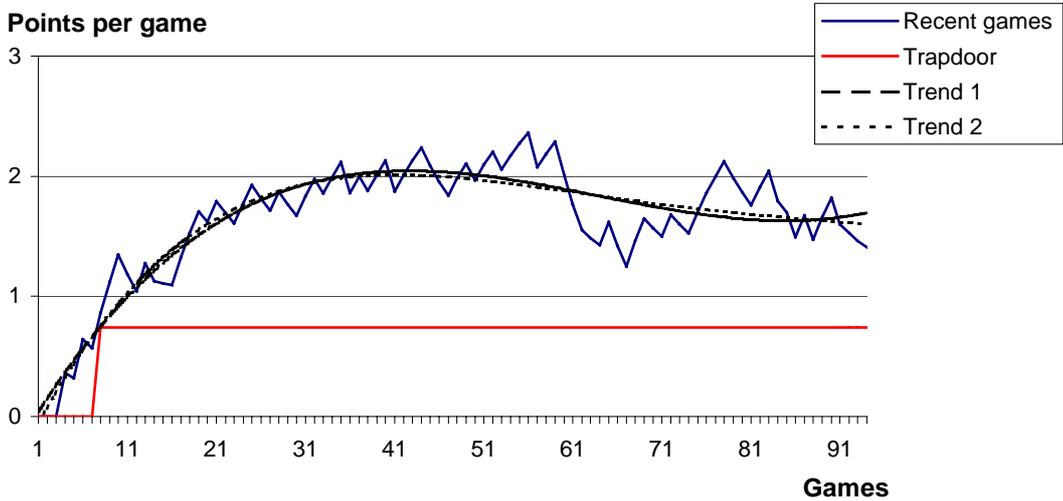
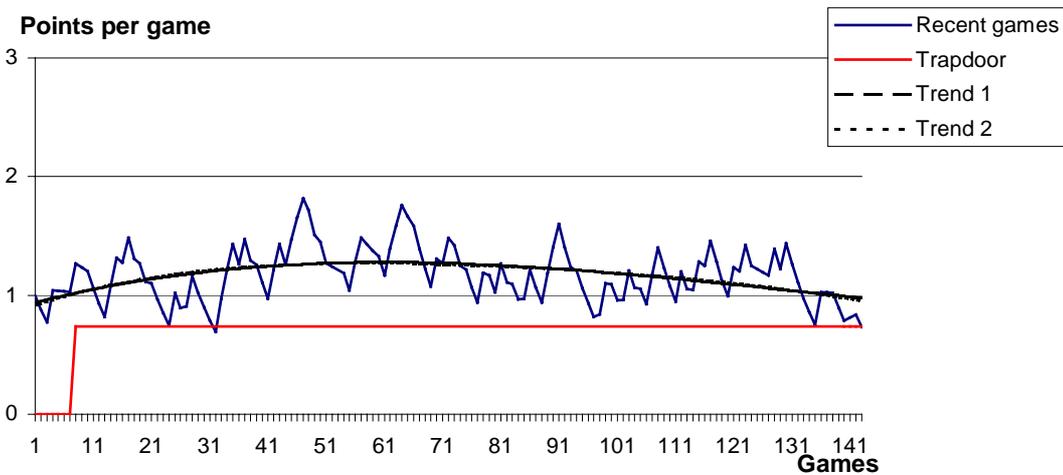


Figure 2e Walter Smith at Everton



Source: mables-tables.com, extended manually using www.soccerbase.com for Gregory and Smith

The five charts show the exponentially weighted average points per game using a smoothing value of 0.121 (the jagged line labelled ‘recent games’), and two polynomial approximations to the recent games line. Trend 1 is a 3rd order polynomial approximation (the lowest order that can detect an initial boost, a drop, a normal period, then another drop), and trend 2 a 4th order polynomial. Ignore the trapdoor line for now; it shows an output from the optimal strategy found in a later section.

Visual inspection of the five charts shows that only Wenger and Gregory show any sign of the initial boost in performance. However all five show some evidence of a drop in performance during a rebuilding period, and at least a suspicion of a late decline, which of course does not have to come from physical aging, but could just be due to personality clashes or the desire for a fresh start.

From visual inspection of these graphs, plus commonsense, the input values shown in table 4 are used in the model (see figure 1 and the accompanying explanation for a definition of the variables).

Table 4 Input value distributions for a manager’s time profile

	min	max	mean	sd	Distribution
<i>inspire</i>	0	12	6	1.73	Truncated Binomial (12,0.5)
<i>boost</i>	0.15	0.45	0.3	0.074	Truncated Normal (0.3,0.1)
<i>rebuild</i>	10	50	30	3.9	Truncated Binomial (60,0.5)
<i>drop</i>	0.1	0.4	0.25	0.074	Truncated Normal (0.4,0.1)
<i>age</i>	76	228	152	37.5	Truncated Normal (152,50)
<i>decay</i>	19	76	38	16.7	Truncated Normal (38,19)

As these values do not come from a statistical analysis of the underlying data, they are obviously open to criticism, which is one reason for presenting the raw data of the charts in figure 2. Anyone wishing to suggest alternative values can use these charts as a starting place for their own investigations.

The mean values of *inspire*, *boost*, *rebuild* and *drop* imply that the sacking of a manager leads on average to the loss of 10.2 points (excluding any loss due to paying up the manager’s contract) while the new manager rebuilds the team.

Contract length, salaries, buying success and number of managers

The normal length for a contract in the Premiership seems to be about three years. This is the reported length of Alex Ferguson’s contract extension at Manchester United (McClaren, 2002).

Some contracts are shorter:

“Graham Taylor returned to Aston Villa on Tuesday, signing a 21/2-year contract”

while a few are longer:

“McClaren would prove very difficult to dislodge from Middlesbrough as he signed a lucrative contract five-year contract last summer”

<http://www.myboro.co.uk/news.asp?page=1&sort=2&storyid=650>

“David [O’Leary] has a six-year contract” Fitton, 2002

As the details of contracts are not published, it is necessary to rely upon intelligent speculation to deduce salaries. At the top of the Premiership, we have

“the salary on offer more than doubles Arsene Wenger’s existing salary, and a figure of three million pounds a year plus bonuses is being talked about” Parry, 2000.

and

“Ferguson, who is getting a reported salary of two million pounds this season, would expect to receive a sizable pay increase that would vault him past Arsenal manager Arsene Wenger’s reported 2.5 million salary.”

www.canoe.ca/Slam020205/soc_fer-ap.html

Away from the top of the table, we see

“despite the insurance of a £1.2 million salary, McClaren is finding the Premiership a cruel environment” Fitton, 2001.

and

“Actual salary from their respective clubs

Bryan Robson	1.0
David O’Leary	1.6
Claudio Ranieri	1.55”

www.sportinglife.com/totaljobs/player_salary/

When it comes to buying success, there is one piece of solid statistical evidence. The Deloitte & Touche Annual Review Of Football Finance for 1997/98 reported that league position was correlated with spending so that a rise of 16 league positions was related to an extra spend of £22M (Deloitte & Touche, 1998).

16 league positions in 1997-98 involved a difference of about 37 points (2nd place 77 points, 18th place 40 points). This would imply a value for *buy_success* of 1.68 points per £million.

Putting these together, we have the values shown in table 5 for a manager’s contract

Table 5 Input value distributions for manager's contract and buying success

	min	max	mean	sd	Distribution
<i>contract</i>	76	152	114	16.7	Truncated Normal (114,19)
<i>salary</i>	26	80	53	9.7	Truncated Normal (53,10)
<i>buy_success</i>	0	3	1.5	0.49	Truncated Normal (1.5,0.5)

The mean *contract* of 114 games equates to 3 Premiership seasons, while the mean *salary* of 53 000 per game equates to £2 million per year. The value for *buy_success* is very uncertain, with a mean of 1.5 points per £million, slightly below the figure implied in the Deloitte and Touche report.

The final piece of information needed is the average number of managers employed by a club over the ten seasons that the model looks at. The numbers of managers employed in the ten seasons to May 2001, excluding caretaker managers, and counting joint managers as one, for the 20 clubs making up the Premiership in 2000-01 are shown in table 6.

Table 6 Managers in 10 seasons by club

<i>Premiership</i>	<i>1991 -2001</i>
Man Utd	1
Arsenal	4
Liverpool	3
Leeds	3
Ipswich	2
Chelsea	7
Sunderland	5
Aston Villa	4
Charlton	2
Southampton	8
Newcastle	5
Tottenham	8
Leicester	5
Middlesbrough	4
West Ham	2
Everton	6
Derby	3
Man City	5
Coventry	6
Bradford	7
Mean	4.5

Source: www.soccerbase.com

The average value in table 6, 4.5, is used as the value of *manave* in the model. With this value of *manave*, the average tenure of a manager is 84.4 games, and the mean number of points lost from paying up a sacked manager's contract is

$$(114 - 84.4) * 53 * 1.5/1000 = 2.35 \text{ points.}$$

Adding this to the 10.2 mean points lost through disruption, and multiplying by the average number of sackings, 3.5, shows that an average club will lose 44 points, or 4.4 per season through the turnover of managers.

Exploratory analysis

The three choice variables in the model are *honeymoon*, *trapdoor* and *smooth*.

Many commentators imply that new managers have a honeymoon period during which they will not be sacked:

“Alex Miller was the ideal appointment to stabilise a club which lacked leadership and direction. Now, almost twelve months on, the honeymoon is over and Miller too will soon be under pressure unless the situation alters radically.” Gordon, 1998.

The shortest tenure (excluding caretakers) in the Premiership in the last season is 17 games (98 days) for Colin Todd at Derby in October 2001 to January 2002, but only 15 of the games were Premiership games. In 2000-01, Bradford were in the Premiership and sacked their manager Chris Hutchings in November 2000 after just 12 league games in charge (Austin, 2002).

In 1974, in the precursor to the Premiership, Brian Clough was in charge of Leeds United for only 44 days and 9 division 1 (old) games (<http://www.soccerbase.com/footballlive>).

So one aspect of the investigation with the model concerns the appropriate length of any honeymoon period.

The *trapdoor* is the level of performance below which a manager will get the sack. Over the last six seasons, if a club averaged fewer than 39.67 points in a season, or 1.04 points per game, it could expect to be relegated. So a first guess at an appropriate level for *trapdoor* might be just over 1.

But the fluctuating nature of results means that even an excellent or world class manager might average below 1 point a game for short periods, particularly during their rebuilding period. So the interaction of *trapdoor* with *smooth* and *honeymoon* needs to be considered.

Any value of *smooth* above about 0.3 implies that the most recent games will have far more weight than earlier ones. Table 7 illustrates this.

**Table 7 Weight on last 5 games
by value of *smooth***

<i>smooth</i>	<i>percent</i> weight on last 5 games
0.1	41
0.2	67
0.3	83
0.4	92
0.5	97
0.6	99

With a value of *smooth* above about 0.3, any manager with a run of as few as five games without a win would see their performance drop close to or below 1, and so be sacked if the value of *trapdoor* was as high as this.

To quantify this effect, it is necessary to run the model. Clearly it is not enough to just run the model once for ten seasons. Many of the model inputs are uncertain, and even if they were not, the results of individual games certainly are. So the model is run 5000 times for ten seasons with each combination of choice variables, sampling different values from the uncertain inputs, recording the sackings and results of each game, and calculating the mean number of points per season for the club. The risk analysis software @RISK from Palisade Corporation is used to perform the calculations.

Table 8 shows 27 results of this exploratory analysis, for three values of each of the three choice variables. The three values for each choice variable are chosen to cover the range of plausible values. Each entry in the table is the mean points per season for one combination of the choice variables. So the first entry shows that a combination of *honeymoon* = 1, *trapdoor* = 0.1 and *smooth* = 0.1 gives a mean result of 46.5 points per season.

**Table 8 Mean points per season
by *honeymoon*, *trapdoor*
and *smooth***

***honeymoon* = 1**

	<i>smooth</i>:		
	0.1	0.3	0.5
<i>trapdoor</i>:			
0.1	46.5	51.8	55.8
0.5	52.9	53.3	27.6
1.0	52.5	13.4	-13.0

***honeymoon* = 8**

	<i>smooth</i>:		
	0.1	0.3	0.5
<i>trapdoor</i>:			
0.1	45.4	51.0	55.7
0.5	52.7	53.6	36.7
1.0	55.4	33.8	27.9

***honeymoon* = 40**

	<i>smooth</i>:		
	0.1	0.3	0.5
<i>trapdoor</i>:			
0.1	45.3	49.6	53.9
0.5	51.5	53.5	47.7
1.0	53.9	46.3	43.7

Source: 5000 model calculations

The values in table 8 are only shown to 1 decimal place, as the mean number of points per season over the ten seasons has a standard deviation of about 7. So even with 5000 runs, the standard error of the mean is about $7/\sqrt{5000} = 0.1$ points per season, giving a 95% confidence interval for each of the results in table 8 of about plus or minus 0.2 points per season.

With this in mind, there are still several insights that can be gained from table 8.

The first is that a combination of a short honeymoon, a high trapdoor and a lot of weight on recent games is not a good idea. At the extreme of a *honeymoon* of 1, a *trapdoor* of 1.0 and *smooth* of 0.5, it even leads to an obviously unreal negative mean number of points per season, as an average of 90 managers are tried, fall below the trapdoor of 1.0 points per game, and are sacked and paid off during the ten seasons. But even the less extreme combinations in this bottom right hand area of the three blocks produce poor results.

The second insight is that the combination of a low *trapdoor* of 0.1 and a low *smooth* of 0.1 produces only about 46 points per season, a poor result, whatever the honeymoon period. The problem here is the opposite one, with the first manager employed nearly always being retained for practically the whole of the ten seasons, whatever his quality.

The third insight is that a long honeymoon period of 40 games generally does not produce the best results. It leads to poor managers being retained for longer than their performance would suggest.

Finally, the two best results in the table, with mean points per season of 55.8 and 55.7, both have a low *trapdoor* of 0.1 and a high *smooth* of 0.5, and honeymoons of 1 and 8 games respectively. However, there is another result with a honeymoon of 8 games, a high *trapdoor* of 1.0 and a low *smooth* of 0.1 that gives nearly the same number of points per season, 55.4.

Looking in more detail at the two best solutions with honeymoons of 8 games shows that they are of a different character, despite giving a similar result. The solution with a low trapdoor and high smoothing employs on average fewer than 6 managers in the ten seasons, and their average quality is almost mid-way between good and excellent (strictly speaking, the type of manager is an ordinal measure, and so does not have a mean, but by specifying the symmetric distribution of manager types shown in table 1, it becomes an interval measure whose mean can be taken). The solution with a high trapdoor and low smoothing employs more managers, just over 9 in the ten seasons on average, but their average quality is higher, 4 times closer to excellent than good. The only reason it doesn't end up with the higher result is that the consequence of sacking so many managers is a big loss of points through rebuilding and paying off contracts.

This is about as far as the exploratory analysis can take us. There is nothing to guarantee, or even suggest, that one of the 27 combinations of choice variables in table 8 is the true optimal combination, the best that is available. Indeed, having two such different strategies coming so close to giving the best result of the 27 suggests that finding the optimal strategy will be a challenge.

The optimal strategy

Traditional optimisation methods either perform poorly or cannot be applied when a problem has several local maxima, and a lot of uncertainty. Both of these conditions apply here. The only method that holds out some hope of finding the optimal strategy is RISKOptimizer, from Palisade Corporation, which combines a genetic algorithm-based optimiser (to avoid getting stuck on the local maxima), with Latin Hypercube Monte Carlo simulation (to handle the uncertain inputs) (Palisade Corporation, 2000).

Running RISKOptimizer with 5000 iterations per simulation, and all other settings at their standard values allows about 700 simulations to be performed overnight on a 1.5 GHz Pentium 4 desktop computer. Three overnight runs were performed and gave the three optimal strategies shown in table 9. All of them give a mean result of about 57 points per season.

Table 9 Optimal strategies found by RISKOptimizer

	<i>honeymoon</i>	<i>trapdoor</i>	<i>smooth</i>	Mean points per season
Run 1	7	0.84	0.094	57.0
Run 2	8	0.74	0.126	56.8
Run 3	8	0.74	0.121	56.9

Source: Overnight RISKOptimizer runs

Since each of these results has a 95% confidence interval of about 0.2 points per season, we can say that they are all significantly better than the best of the 27 strategies used in the exploratory analysis in table 8, which gave a mean result of under 56 points per season.

However, they are not distinguishable from each other with any degree of confidence (two mean results from 5000 iterations need to differ by 0.28 points per season to be 95% sure that they are actually different). To attempt to distinguish them, an @RISK run of 38000 iterations was performed with each of the three candidate optimal strategies from table 9 (38000 iterations gives a 95% confidence interval of 0.07 points per season for each result, or 0.1 for the difference between results). The three candidates from table 9 scored 56.66, 56.78 and **56.81** points per season respectively. We can say with 95% confidence that the candidate strategy from run 1 is worse than the two very similar candidate strategies from runs 2 and 3.

So the optimal strategy for a club would seem to be to allow a manager a honeymoon period of 8 games, and then sack him if his weighted average performance with a smoothing value of 0.121 (putting 47% of the weight on the last five games) falls below 0.74 points per game.

If a club follows this strategy it can expect to obtain 56.81 points per season. It can expect to employ 5.7 managers on average over the ten seasons, and the average quality of those managers will be 3/5ths of the way between good and excellent. To put this into perspective, 56.81 points would have secured a finishing position of 10th, 8th, 8th, 6th, 8th and 8th in the six seasons so far of a 20-club Premiership.

Applying the optimal strategy to real managers

Looking again at figure 2 for the five managers, the red line shows the *trapdoor* and *honeymoon* implied by the optimal strategy. The managers' performance in the charts is already weighted using the optimal value of *smooth* (that is why the rather unusual smoothing value of 0.121 was used to draw the charts), so the charts can be used to see directly whether a club using the optimal strategy would have sacked any of the five managers. If at any point, the blue, weighted performance, line falls below the red, trapdoor, line, the club would have sacked the manager at that point.

Arsene Wenger's weighted performance has never dropped remotely close to the trapdoor. This is reassuring, but hardly a stern test of the model, as there would be severe doubts about the utility of any model that recommended the sacking of Wenger!

John Gregory's weighted performance at Aston Villa dropped to 0.75, perilously close to the trapdoor of 0.74, after his 41st Premiership game in charge, the 0-3 defeat at home to Chelsea on 21st March 1999, and fell to 0.73, just below the trapdoor, after his 66th game in charge, the 0-1 home defeat by Newcastle on 4th December 1999. A club employing the optimal strategy in the model would have sacked him at that point. His weighted performance did not dip close to the trapdoor again, and at the time he left the club (not sacked; citing pressure and a clash of personalities) it was comfortably above 1 point per game.

Ruud Gullit's weighted performance at Chelsea never dropped close to the trapdoor. In fact it only once dropped as low as 1 point per game, after his 34th game in charge, the 3-1 defeat away at Newcastle on 16th April 1997. At the time of his sacking in February 1998, his weighted performance was at a very creditable 1.6 points per game.

His successor, Gianluca Vialli, flirted with the sack immediately by losing five of his first seven Premiership games. But he was saved by the honeymoon period of 8 games, and by his eighth game had pulled his weighted performance up to 0.86 points per game, just keeping clear of the trapdoor. From that point on, his weighted performance was comfortably clear of the trapdoor, remaining above 1 point per game. At the time of his sacking in September 2000, it was standing at 1.4 points per game, and his sacking was greeted with a lot of criticism.

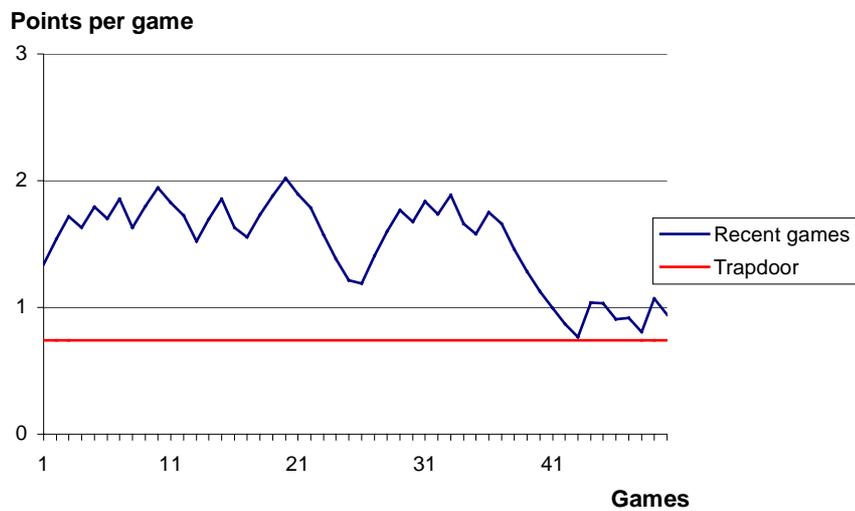
"The email response has been massive, and the majority believe Blues chairman Ken Bates has made a mistake." BBC Sport Online, 2000.

Walter Smith's weighted performance at Everton dropped to 0.75 points per game after his 24th Premiership game, the 2-1 away defeat at Derby County on 7th February 1999, and to 0.69, below the trapdoor of 0.74, after his 32nd game, the 1-2 home defeat by Sheffield Wednesday on 5th April 1999. A club using the optimal strategy in the model would have sacked him at that point. Although his performance dropped below 1 point per game several times after that, the next time it fell below the

trapdoor was after his 143rd Premiership game, the last one before he was actually sacked (although in reality there was also a cup defeat after this last league game).

One final comparison with a real manager is instructive. Figure 3 shows the end of the Everton career of Joe Royle, one of Walter Smith's predecessors. This was not shown in figure 2 as the early part of his Everton career was during the 24-club Premiership. His weighted performance dropped to 0.76 points per game after the 4-1 away defeat at Newcastle on 29th January 1997. This was just above the trapdoor of 0.74 points per game. So the model would just not have sacked him, although in reality he was dismissed in March 1997.

Figure 3 Joe Royle at Everton (end)



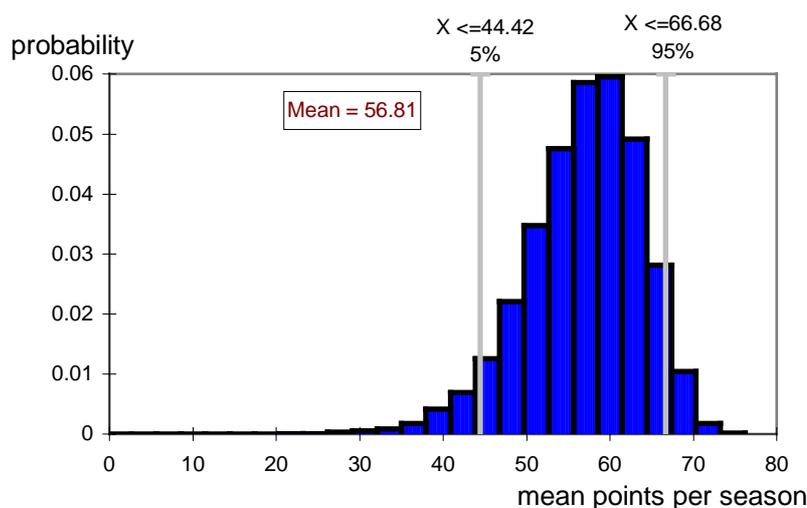
Source: mables-tables.com

Discussion and suggestions for further research

Variation in the results

The results presented in the previous section sounded reassuringly precise, and the comparison with the record of six real managers gives further confidence that the model is reasonable, but figure 4 shows how much the results can vary even with this optimal strategy.

Figure 4 Distribution of mean points per season using the optimal strategy



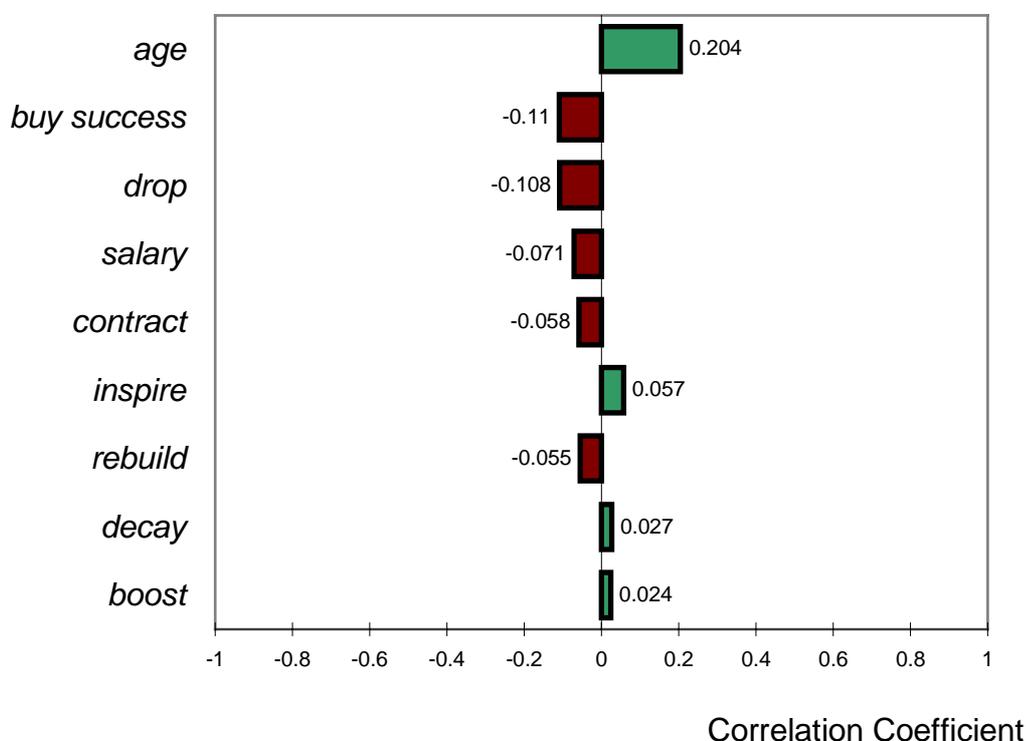
Although the mean number of points per season over ten seasons is 56.81, for 5% of the ten season periods the mean number of points per season falls below 44.42, and for 5% of them it rises above 66.68.

The former would see the club on the fringes of relegation, as the top club relegated in the six seasons of a 20-club Premiership so far has scored 38, 40, 40, 36, 33 and 34 points. Football is an uncertain game, and following the best strategy gives no guarantee of success, even over ten seasons. Perhaps this is a good thing for the long-term health of the sport.

The latter would see the club regularly challenging for a European place, as 66.68 points would have secured a finishing position of 4th, 5th, 3rd, 5th, 5th and 5th in the six seasons so far of a 20-club Premiership.

Figure 5 shows the uncertain inputs to the model and their correlation with the number of points per season. All of the correlations are in the expected direction (the longer the time to age, the more points per season, the more points that can be bought for £1 million, the lower the points per season etc), but none of the correlations are particularly strong. The main lesson from figure 5 is that it would probably be worthwhile making some effort to find a better estimate of the time it takes for a manager's performance to begin to decline (*age*); the value used in the model is little more than an educated guess.

Figure 5 Correlation of inputs with mean points per season using the optimal strategy



Is the optimal strategy really optimal?

The previous section used 38000 @RISK iterations to show that the very similar optimal strategies from runs 2 and 3 were better than the candidate from run 1.

However, in each overnight run, 25 or so of the top strategies gave results within about 0.28 points per season of the optimal strategy. It needs a difference of 0.28 points per season to be able to dismiss a strategy as sub-optimal. Table 10 shows these strategies for run 3. Checking each of these with 38000 @RISK runs would be too time consuming, so it remains an open question whether any of these are actually the optimal strategy.

Although they do differ, they all have some features in common:

- The honeymoon periods are short, ranging from a minimum of three games to a maximum of eight.
- The trapdoors lie in a fairly small range, between 0.69 and 0.92 points per game.

- The smoothing coefficients are fairly gentle, between 0.09 and 0.14. If the trapdoor values are towards the top of the range, the smoothing coefficients are towards the bottom, and vice versa.

Table 10 Best 26 Strategies found by RISKOptimiser in run 3

Rank	<i>honeymoon</i>	<i>trapdoor</i>	<i>smooth</i>	Mean points per season
1	8	0.74	0.12	56.90
2	8	0.74	0.13	56.87
3	8	0.75	0.13	56.87
4	8	0.74	0.13	56.86
5	8	0.71	0.13	56.82
6	8	0.76	0.13	56.80
7	8	0.75	0.13	56.80
8	7	0.91	0.09	56.78
9	7	0.75	0.13	56.77
10	6	0.75	0.13	56.77
11	7	0.82	0.09	56.77
12	9	0.75	0.13	56.75
13	8	0.69	0.14	56.75
14	7	0.91	0.09	56.72
15	7	0.91	0.09	56.71
16	7	0.85	0.09	56.70
17	7	0.85	0.09	56.69
18	8	0.82	0.13	56.67
19	8	0.72	0.13	56.66
20	7	0.92	0.09	56.66
21	7	0.91	0.10	56.66
22	7	0.91	0.09	56.66
23	3	0.91	0.09	56.65
24	8	0.91	0.09	56.65
25	8	0.79	0.13	56.64
26	4	0.91	0.09	56.63

Source: 719 RISKOptimiser runs

What table 10 also illustrates is that there are many strategies that give almost the same results as the optimal strategy. All the strategies in table 10 appear to give a mean result within 0.3 points per season, or 3 points over 10 seasons, of the optimal strategy. 3 points is the difference between losing and winning one game in ten seasons – perhaps a single penalty miss by your captain and a flash of brilliance from an opposition striker.

What's missing from the model

The model described here is clearly only a first attempt at applying management science techniques to the hiring and firing of football managers. The more glaring simplifications include:

- No consideration of whether games are played at home or away. In the last six seasons of the Premiership, clubs averaged just under 9 home wins, but only just under 5 away wins per season. In reality a club might be less willing to forgive a poor run of results at home, than a poor run away from home.
- No consideration of the quality of the opposition. In reality a club might be more willing to forgive a poor run of results against the top teams in the league, than against more modest opposition.
- No consideration of the critical importance of avoiding relegation at the end of each season. In reality a club might be more willing to sack its manager if it is in a relegation place, but hopes to use the initial boost in performance of a new manager to draw clear of danger.
- No consideration of games other than Premiership ones. In reality a club would be expected to take notice particularly of good performances in European competitions, and poor performances against lower league opponents in the domestic cup competitions.
- No consideration of the financial and other costs of sacking a manager other than paying off his contract. In reality, the rebuilding period often includes other changes in the playing and management staff, which can have severe financial implications, and frequent changes of manager can have an unsettling effect on players and fans alike. On the other hand, the arrival of a new manager does sometimes lead to at least a temporary increase in attendance at games, bringing some financial benefits.
- No consideration of the different aspirations of different clubs. Implicitly the model is calibrated against the average club in the Premiership. In reality it is clear that there is a group of clubs at the top of the Premiership whose objectives are focussed on success in European competition. This is perhaps most clearly seen in the sackings of Gullit and Vialli at Chelsea when the model would suggest they were performing well above the trapdoor level. Clubs at the top of the Premiership, with more money to spend, and a realistic chance of winning trophies, might be expected to have access to a better than average pool of managers.
- No consideration of other choice variables than the triplet of *honeymoon*, *trapdoor* and *smooth*. In theory the clubs should perhaps be performing more complicated calculations using Bayesian statistics to update the posterior probabilities of manager quality after every game. In practice, they may use simple rules of thumb, including pressure from fans, rather than calculation when deciding it is time for a manager to go. Clubs also have the chance to observe the manager's performance in other areas of the job, such as coaching

and transfer dealings, and probably take this into account when deciding whether to sack him or not.

Despite all these shortcomings, and no doubt others that are obvious to the reader if not the author, the model does seem to give reasonable insights, and the calculations required to apply the optimal strategy (as opposed to those required to discover it) are simple. Any club considering the future of its manager might find it worthwhile at least to take a few minutes to check how close his performance is to the trapdoor before pulling the lever that consigns him to the sack.

Further research

The first priority is to update the data in the model once the complete information for the 2001-02 season becomes available. This will allow the optimal strategy to be recalculated using the most recent data. It will also enable the performance of all the current Premiership managers to be monitored and compared to the trapdoor.

The second task is to adapt the model so that it can be applied to the Nationwide League clubs in England, where the seasons consist of 46 games, and the financial circumstances are very different to those in the Premiership. Looking at other soccer leagues, in Spain and Italy particularly, or other sports, would also seem to be possible.

Development of the model to address the more important shortcomings identified above would seem to be worthwhile. The best candidates for inclusion are incorporating home and away games, the quality of the opposition, the importance of avoiding relegation, and the different aspirations of different clubs. Quantitative information exists for all of these.

It would also be worth trying to improve the accuracy of the inputs to simulate the actual Premiership more closely. In particular, the restriction to five types of manager, symmetrically arranged around the 'good' manager who has an average performance, is not a perfect representation of the Premiership. Inspection of table 3 shows that the 'good' managers actually do slightly worse than this on average, while both the 'poor' and 'world class' managers do rather better than the model assumes. Figure 5 shows that more systematic investigation to estimate the time profile of a manager's performance, particularly any drop in performance after several seasons in the job, would also be worthwhile.

A final observation is probably worth making. When I started this investigation, I expected to be able to draw on a body of previous similar research. I was thinking that there would be existing models that I could adapt to include uncertain inputs, and the new optimisation techniques that perform well under these conditions. But searching for football and manager on the Web of Science (<http://wos.mimas.ac.uk/>) did not turn up any academic research on the subject. It seems that there are hundreds of quantitative studies of the fitness and skill of players, but very little on the management aspects of the sport.

The foremost football research centres in the UK at Leicester University (www.le.ac.uk/snccfr), Birkbeck College (www.football-research.org/) and Liverpool University (www.liv.ac.uk/footballindustry/) concentrate on the social and governance aspects of the game. Deloitte & Touche Sport produces excellent annual publications on football finance. But quantitative modelling so far seems to have been absent. For a £1 billion per year industry (Deloitte & Touche, 2000), this seems an odd omission, and surely an opportunity to be grasped.

References

- Austin S, 2002, Quick-fire managers,
news.bbc.co.uk/sport/hi/english/football/newsid_1761000/1761595.stm
- BBC Sport online, 2000, Fans stand behind Vialli,
http://news.bbc.co.uk/sport/hi/english/football/teams/c/chelsea/newsid_924000/924617.stm
- Beale J, 2001, Incoming Trio Enthusiastic,
<http://www.megasoccer.com/story/coun/you/home/6136.html>
- Country Life, 2001,
http://www.countrylife.co.uk/worldsoccer/newsapril01/news_blanc.html
- Deloitte & Touche, 2000, English Football Revenues Set to Double,
<http://www.deloitte.com/vc/0,1029,sid=1018&cid=2934,00.html>
- Deloitte & Touche, 1998, Annual Review Of Football Finance 1997/98, p41.
- Eppen G , Gould F and Schmidt C, 1990, Quantitative Concepts for Management, 3rd Ed, Prentice Hall.
- Fitton P, 2001, Mac on the rack already,
www.soccernet.com/england/news/2001/0826/20010826efcmfcrep.html
- Fitton P, 2002, O'Leary's a winner, says Ridsdale,
<http://www.soccernet.com/england/news/2002/0224/20020224lufcridsdale.html>
- Gordon R, 1998, <http://www.scottishfootball.com/features/rg7.htm>
- Kelly D, 2001,
<http://www.megasoccer.com/page/genpage.php?nt=world&nv=World&ns=home&page=columns%2Fkelly10.html>
- McClaren S, 2002, Fergie quit? He's the last of a dying breed, Observer, 3 March 2002, <http://www.observer.co.uk/sport/story/0,6903,660885,00.html>
- Palisade Corporation, 2000, Guide to RISKOptimizer, Newfield, NY, USA.
- Parry C, 2000, <http://www.arsenal-world.net/news/1100/051100b.htm>

Appendix: Cautionary lessons from an alternative formulation

An earlier version of the model assumed that all managers would keep the same proportion of wins to draws as in the league as a whole. So

$$\begin{aligned} p(\text{win})/p(\text{draw}) &= P(\text{win})/P(\text{draw}) \\ &= P(\text{win})/(1-2P(\text{win})) \end{aligned}$$

$$\begin{aligned} g &= 3p(\text{win}) + p(\text{draw}) \\ &= 3p(\text{win}) + p(\text{win})(1-2P(\text{win}))/P(\text{win}) \\ &= p(\text{win})*(3 + (1-2P(\text{win}))/P(\text{win})) \\ &= p(\text{win})*(1+P(\text{win}))/P(\text{win}) \end{aligned}$$

so, for any type of manager,

$$\begin{aligned} p(\text{win}) &= g * P(\text{win})/(1 + P(\text{win})) \\ p(\text{draw}) &= g * (1 - 2P(\text{win}))/ (1 + P(\text{win})) \\ p(\text{lose}) &= 1 - g * (1 - P(\text{win}))/ (1 + P(\text{win})) \end{aligned}$$

Clearly this assumption would cease to make sense if $p(\text{lose}) < 0$, ie if

$g > (1 + P(\text{win}))/ (1 - P(\text{win}))$, so the model checked to ensure that this condition was not broken.

However, this formulation gave too few draws for the worst managers, and too many for the best, and so was abandoned in favour of the formulation described in the text.

It also produced a quirky result from the optimisation procedure, as during the inspired period at the start of a manager's tenure, the probability of losing a game would be very small, often below 2%, for the world class managers. So the model would find it optimal to sack the manager if he lost his first game, since it was almost impossible to sack a world class manager this way. This was clearly unrealistic behaviour that was an artefact of the model.