The Hazard of Being an English Football League Manager: Empirical Estimates for Three Recent League Seasons

Ray Bachan

Brighton Business School

University of Brighton

Lewes Road

Brighton

BN2 4AT

United Kingdom

Barry Reilly

Department of Economics

University of Sussex

Falmer

Brighton

BN1 9QN

United Kingdom

Robert Witt

Department of Economics

University of Surrey

Guildford

GU2 7XH

United Kingdom

Abstract

This paper uses data drawn from the English Football League to model separate hazard rates for club managers for the 2001/2, 2002/3 and 2003/4 seasons. On average over the three seasons, approximately one-third of managers involuntarily exited employment status with their club. We model the hazard using a standard logistic model exploiting information on the spell at risk rather than the individual. The role of neglected heterogeneity is also examined using random and fixed effects logistic models within a discrete-time setting. League position at the start of the spell at risk is found to be the most important determinant of a manager's involuntary exit. A variety of individual specific human capital and other covariates are found to be unimportant in determining the hazard and no role for unobservable heterogeneity as captured by random effects is detected.

June, 2006

Keywords: hazard rate logistic regression model sports

Introduction

The high turnover rate among managers in the English Football League is well documented in Audas, Dobson and Goddard (1999). Scully (1994) used a long time series of United States data to examine the relationship between managerial efficiency and managerial survival rates in basketball, (American) football and baseball. In spite of this, there has been a modest literature that attempts to explain the employment survival of football managers. Audas, Dobson and Goddard (*op.cit.*) and Hope (2003) provide rare examples. The former authors use continuous time duration modelling and exploit match-level data over a 25-year period to explain involuntary separations of football managers. On the other hand, Hope (2003) uses Operations Research techniques to determine the optimal time to sack a manager. The purpose of this paper is to add to this modest literature through an examination of managerial exit patterns in three recent league seasons. This allows us to examine whether the effects of important covariates are relatively stable over a short time period. Our empirical approach emphasizes the role of intra-season performance indicators as determinants of involuntary exit.

In contrast to the existing empirical work on this issue, we exploit a discrete-time approach in the hazard modelling. The calendar month within the season provides the measure for the spell at risk. This facilitates use of a standard logistic model and allows an easier introduction of time-varying covariates. In addition, we endeavour to capture neglected heterogeneity through exploiting the panel nature of the spell at risk data.

This paper is structured as follows. The next section contains a description of the empirical methodology used and is followed by a section describing the data. The penultimate section discusses the empirical results and a final section offers some conclusions.

Methodology

Hazard modelling has traditionally been undertaken within a continuous-time framework (see Lancaster (1992)). More recently, however, situating the analysis within a discrete-time context has become popular as it allows for a degree of simplification in the modelling. Jenkins (1995), exploiting the work of Allison (1982), suggests a useful way for the estimation of discrete-time hazard models using a binary logistic regression model. The approach requires re-organization of the data away from the individual as the unit of observation to the spell at risk of event occurrence. The individual-level data are thus replaced by spell at risk data stacked by individual. Each individual unit contributes multiple observations to an expanded likelihood function. In the first period an individual unit either stays or exits the state occupied. Those that survive into the second period either remain or exit in that period and so forth into subsequent periods with the numbers surviving, and hence contributing to the likelihood function, diminishing as individuals leave the state in question. The dependent variable in this type of setting could be denoted $y_{i,t}$ and defined as 1 if the i^{th} individual exits the state at time period t , and zero otherwise.

The re-organization of the data allows construction of an unbalanced panel dataset where a maximum of N individuals are observed over T discrete time periods. This could be developed a little further by introducing covariates and separate intercepts for each failure period to capture the baseline hazard. The logistic is a computationally straight-forward transformation and possesses an added advantage of providing a non-proportional hazard. This may be formulated as follows:

$$\text{prob}(y_{i,t}=1) = \frac{\exp[\boldsymbol{\beta}' \mathbf{x}_{i,t} + \boldsymbol{\delta}' \mathbf{D}_{i,t}]}{1 + \exp[\boldsymbol{\beta}' \mathbf{x}_{i,t} + \boldsymbol{\delta}' \mathbf{D}_{i,t}]} \quad [1]$$

The dependent variable represents the probability of individual i exiting in a small interval after period t conditional on having survived to period t , where $\mathbf{x}_{i,t}$ is a vector of

covariates, which may or may not vary over time. The vector $\mathbf{D}_{i,t}$ denotes the baseline hazard, which can be specified by a set of dummy variables for each failure time period (see Sueyoshi (1995)). This provides another advantage in that the estimates for the baseline hazard are derived directly as part of the estimation procedure. This approach has the merit of simplicity and also allows an easier introduction of time-varying covariates than parametric continuous-time hazard models. This latter advantage has some relevance in the current application given the potential for a strong correlation between the hazard of interest and time-varying performance indicators.

The discrete approach, though computationally straightforward, is not free of criticism. The first relates to the inflation of the sample size through the re-organization of the data. Though a legitimate concern, the estimates obtained are still maximum likelihood and retain the asymptotic properties of this class of estimator. The second relates to the treatment of the repeated observations as if they were independent of each other. This is clearly not the case given the data re-organization with correlation across observations allowing the potential for some degree of inefficiency in the estimates and a downward bias in the sampling variance. However, as Robinson (1982) demonstrates with a probit model, ignoring the correlation across time does not affect the consistency of the maximum likelihood estimates. Following Allison (1982), we exploit an approach within the discrete-time framework that provides a solution to this particular problem (see below).

Neglected heterogeneity, a potentially important issue in hazard modelling, could also be addressed by exploiting the panel nature of the data. In particular, the binary logistic model could be re-specified to allow for individual-specific fixed effects as follows:

$$\text{prob}(y_{i,t}=1) = \frac{\exp[\boldsymbol{\beta}' \mathbf{x}_{i,t} + \alpha_i]}{1 + \exp[\boldsymbol{\beta}' \mathbf{x}_{i,t} + \alpha_i]} \quad [2]$$

The construction of the relevant likelihood for this logistic model and its maximization with respect to the β vector and α_i terms yields consistent estimators only if the number of time periods goes to infinity. In empirical applications this is impractical and, with a fixed T , estimates of the fixed effects and the parameter vector are inconsistent. The Monte Carlo experiments reported by Katz (2001) suggest that the scale of the bias may attenuate with $T > 15$. The inconsistency arises because the number of incidental parameters increases without bound as $N \rightarrow \infty$. Chamberlain (1980) suggests maximizing a log-likelihood function conditional on a set of minimal sufficient statistics that help sweep out the fixed effects. A sufficient statistic in the case of the incidental parameters for the logit is $\sum_{t=1}^T y_{i,t}$. The model parameters are only identified through the ‘within’ dimension of the data.

As in the case of the fixed effects linear regression model, only time-varying covariates are permissible with this procedure. The groups of data points used to construct the estimator for β are confined to those that exhibit a change in status. The groups that exhibit no change in status are discarded as they provide no information about the vector β . The use of this procedure requires belief that all the information required for β is contained in the non-discarded data. This is a fairly stringent requirement and may not be appropriate for all applications. A further limitation with this approach is that the fixed effects cannot be retrieved for further interrogation or analysis. A crude way of overcoming this problem would be to estimate a Linear Probability Model (LPM) with fixed effects. However, this procedure is inherently heteroscedastic and will have a strong tendency to predict outside the lower end of the $[0,1]$ boundary given the low hazard rates generally encountered in these applications. This particular approach is thus not pursued here.

The neglected heterogeneity could alternatively be captured through treating the omitted factors as random. This sits more comfortably within the hazard modelling tradition than the fixed effects approach. The random effects specification is expressed as:

$$\text{prob}(y_{i,t}=1)= \frac{\exp[\beta' \mathbf{x}_{i,t} + u_i]}{1 + \exp[\beta' \mathbf{x}_{i,t} + u_i]} \quad [3]$$

with $u_i \sim N(0, \sigma_u^2)$, where $\sigma_u^2 = \frac{\rho}{1-\rho}$ and ρ captures the correlation in unobservables across the time units (see Greene Ch.19.5 for further details). The u_i random term varies across the individuals and is assumed to follow a normal distribution. This could be viewed as more plausible than the Gamma distribution traditionally assumed for such heterogeneity in the early duration literature (for example, see Lancaster (1979)). The random effects logistic model can be estimated by maximum likelihood techniques. The additional merit of this model is that it provides a basis for testing the proposition of no cross-period correlations (i.e., $\rho=0$) and, implicitly, whether σ_u^2 is zero. It thus provides a formal test for the presence of random effects. If $\rho=0$ then the resultant estimates are comparable to those from a pooled (or stacked) logistic model. The estimation for this paper was conducted using the STATA (version 8) software package.

Data

The empirical analysis uses data collected from the 33rd edition of the Rothmans Football Yearbook and various editions of the Sky Sports Football Yearbooks. We focus on three seasons: 2001/2, 2002/3 and 2003/4. The managerial exits examined in each season comprise only involuntary separations. The voluntary separations are excluded from our analysis given that the process determining this type of separation is likely to be distinct from that governing the involuntary type. Once voluntary separations are excluded, the number of managers available at the start of each of the three seasons is 89, 91, and 87 for the 2001/2, 2002/3 and 2003/4 respectively. The discrete approach we adopt uses the month as the measure for the spell at risk and there are nine of these in the football season. The choice of time interval is arbitrary but anything shorter did not prove sensible for the construction of the time-varying performance covariates of most interest to us. The analysis is restricted to modelling the single risk of whether the individual involuntarily exits the state or not.

We model time to failure starting in early August and concluding in late April given our interest in intra-season separations. This implies that if a manager has survived to the end of April we ignore whether he gets sacked in May or later in the summer close season. We take the view that involuntary separations in May (or later) might be taken to reflect factors other than those related to within season performance. Thus, we start with the maximum number of managers that were not subject to voluntary departure from their club (N) and for those that survive to the end of April, we have a maximum of $T=9$ person-specific observations for each individual. The stacked (or panel) data inflate to 626, 736 and 646 spell at risk observations in the 2001/2, 2002/3 and 2003/4 seasons respectively reflecting the loss of individuals due to exits.

The Kaplan-Meier hazard rates for each of the nine months in each of the three seasons are contained in table 1 with asymptotic standard errors reported in parentheses where relevant. The raw data suggest the peak exit month was October in the first two seasons, though the pattern appears more even across months in the final season. There were no exits in the months of August, December or February in the 2002/3 season and this

curtails how we choose to specify the baseline hazard in the pooled (or stacked) logit model in the three seasons (see below).

The use of the discrete-time approach outlined earlier allows for a very simple introduction of time-varying monthly covariates in our analysis. A variety of different measures were initially explored to determine their use as covariates in the hazard function. These included an extensive set of time-varying performance measures and time-invariant personal and club-level characteristics (see below). However, only a small sub-set of these exerted any independent influence on the hazard and for brevity, the reported analysis is restricted to those that recorded some effect in one or other of the models estimated. These comprise two time-varying covariates and one time-invariant covariate described in table 2. The role of other covariates is discussed later.

The position (15th or lower) used to determine whether a team is in the relegation zone or not is broader than conventionally understood. Although this cut-off point is arbitrary it is felt that it adequately captures league under-performance. It is anticipated that this covariate will exert a positive effect on the hazard. In work not reported here we also used separately continuous measures relating to actual league position of the club at the start of the risk month and the number of wins in the previous month. The estimated effects for both were as anticipated and registered well determined effects in all cases but offered little enhancement to the regression models in terms of an improvement in the maximised log-likelihood values. Thus, for the analysis undertaken here, we report estimates based on the binary measure capturing whether the club is in the relegation zone or not at the start of the risk month.

The attendance measure used is defined as attendance expressed as a percentage of a club's ground capacity averaged over the home league games in the month prior to the risk month. We believe that this percentage rate measure is more meaningful in this particular context given the wide variation in absolute attendance figures across all four leagues. It is arguable that what is important to a league club is not absolute attendance but the distance it reflects from capacity. This measure could be taken to provide a

barometer of fan support for the manager. More importantly, however, the attendance rate provides the club board the financial incentive to either retain or dispense with the services of the current incumbent. Given this interpretation, we anticipate that the greater the attendance rate, the lower the hazard is likely to be.

As noted above, these two measures are based on realizations at the start of the risk month. In the case of the August spell at risk the realisations for the two time-varying covariates relate to the league position at the end of the last season and the attendance figure is averaged over the month of April from the previous season. Finally, one time-invariant characteristic found to contain some explanatory power in a number of the estimated logistic regression models is a binary measure that captures whether a manager was capped at senior international level during his professional playing career or not.

Our empirical approach also allows for the introduction of factors that capture unobservables. These unobservable effects could be taken to reflect individual heterogeneity that capture coaching ability, man-management skills or other motivational attributes that are assumed to vary across individuals. These omitted factors might also be club-specific in nature and reflect the attitude, patience, indulgence or otherwise of those that have the responsibility for running the football club. The omitted factors are separately treated as either fixed (see [2] above) or random (see [3] above) in the estimated specifications. The importance of these latter factors is tested empirically.

Empirical Results

The absence of exits in three of the nine months for the middle season (see table 1) implies that a flexible baseline hazard based on intercepts for individual months does not provide a complete replacement for the constant term in the empirical specification. Therefore, the baseline hazard used in this study conflates the months into pairs for the first six months and allows a final intercept to capture the February to April risk period. This treatment allows for a reasonably flexible baseline hazard that fully replaces the constant term.

The maximum likelihood estimates for the hazard rate models for the three separate league seasons are reported in tables 3 to 5. The first column of each table reports estimates using the above-defined baseline hazard specification but excluding any covariates. The estimates give the effect on the log odds ratio of exit. Once transformed into probability effects these provide Kaplan-Meier estimates but, given the conflation discussed above, are not directly comparable to those reported in table 1. The relevant transformation is $[1 + \exp[-\delta_k]]^{-1}$ where δ_k is the logistic parameter corresponding to the k^{th} period at risk. The highest risk period is October/November in the first two seasons though the February to April period has a slight edge over October/November in the final season reviewed here. A possible conjecture for this result in regard to the two earlier seasons is that, from a club management board perspective, this period is far enough into the season to assess managerial competence but sufficiently early to replace an under-performing manager and salvage club fortunes for the season in question.

The introduction of the three covariates (see column two in tables 3 to 5) predictably alters the baseline hazard estimates. However, the October/November period remains the one containing the highest exit risk for the two earlier seasons and the estimated *ceteris paribus* effect for this particular period in the final season is rendered statistically insignificant.

The most important covariate determining the hazard, in terms of both statistical significance and magnitude, is reserved for being in the relegation zone at the start of the

risk month. The estimated effects for this covariate are positive and well determined in all three cases but, on the basis of asymptotic t-tests (not reported here), there is no evidence of a statistical change in this effect over the three seasons to which our analysis relates. The estimated effect for the attendance rate, the other time-varying covariate in the specification, is poorly determined in the 2002/3 season with an asymptotic t-value close to unity but comfortably achieves statistical significance at a conventional level in the specifications corresponding to the earliest and latest seasons. There is again no evidence, on the basis of an asymptotic t-test, of a statistical difference in the effects of this covariate between the 2001/2 and the 2003/4 season. The time-invariant covariate capturing whether the current manager was capped in his playing career registers an estimated effect that is just within the boundary of statistical significance at the 10% level using a two-tailed test for the last two seasons but does not register an independent effect in the earliest 2001/2 season. The result for the later two seasons may be explained by a club managerial selection policy that erroneously infers the presence of managerial and coaching qualities from an individual's playing career.

The estimates reported in the third column of tables 3 to 5 are based on a pooled (or stacked) logit model specification where a constant replaces the baseline hazard. The specifications in columns two and three are tested against each other using a conventional likelihood ratio test. The null hypothesis comprises the model reported in the third column of these three tables. The resultant tests with prob-values in parentheses are computed as 5.7 (0.12), 11.4 (0.01) and 4.8 (0.19) using the log-likelihood values reported in tables 3 to 5 respectively. This suggests that the logistic model with the flexible baseline hazard is more congruent with the data than the model containing a constant term for only the 2002/3 season. However, it is worth noting that the estimated effects for the covariates of interest appear insensitive to whether a constant term or a flexible baseline hazard is actually used in estimation.

The estimates for the fixed effects model are reported in column four of tables 3 to 5. Given the nature of the estimator only time-varying covariates are permissible. The estimator only exploits information for those cases where the status changed – hence the

reduced sample sizes reported for this column in these tables. Once the fixed effects are introduced, the model estimates suggest a role for attendance that is confined to the fixed effects specifications for the first two seasons. However, the fact that a large volume of information is discarded in the construction of this estimator provides pause for some interpretational caution here.

The estimates reported in the final column of tables 3 to 5 are based on a random effects logistic model with a constant term. The estimated coefficients for σ_u and hence ρ are both negligible and poorly determined in a statistical sense in all cases. This suggests no cross-period correlation and thus, implicitly, no evidence of random effects (or neglected heterogeneity) in the estimated regression models. The estimated effects for the covariates are thus identical to those reported for the pooled logit specification containing a constant term (see column three of these tables). It should be noted that replacing the constant term by the flexible baseline hazard in this final specification does not alter the finding in regard to neglected heterogeneity for any of the three seasons.

The role of a set of other variables on the hazard of exit is now briefly investigated. All but one is time-varying in nature. Given the absence of any evidence of random effects, our preferred model for this exercise is the pooled logit with the flexible baseline hazard. The variables of interest are introduced separately into this model to assess their statistical significance using either an asymptotic t-test or a likelihood ratio test, and then subject to a joint statistical test overall. The results are reported in table 6. There is no variation in the hazard rates across the division a league club plays. The age of the manager at the start of the season, his length of managerial service at the club, the number of clubs he has managed prior to the current one, and whether or not he is English exert no independent influence on the hazard. In addition, a good FA cup run is unlikely to provide respite to a manager under-performing in the league, though the evidence for the 2001/2 season could be construed as marginal. And managers who once played for the club they currently manage are not subject to any degree of favouritism in regard to their employment status. The null hypothesis of no joint effects for these variables overall is also upheld by the data for all three seasons.

Conclusions

This paper used data drawn from the English League to model hazard rates for club managers in three separate seasons: 2001/2, 2002/3 and 2003/4. Over one-third of managers involuntarily exited their employment status within these three seasons. We used a suggestion by Allison (1982) and Jenkins (1995) to model the hazard on the basis of the spell at risk, rather than the individual, using a standard logistic model. We defined the spell of risk in terms of months of the football season. We also examined the role of neglected heterogeneity exploiting both random and fixed effects logistic models in this setting.

The raw data suggest that October was generally a high risk month for involuntary departures. There was no evidence that controlling for neglected heterogeneity through random effects was important to understanding the determinants of involuntary exit. Our preferred specification was a pooled logistic model with a flexible baseline hazard, though this was not found to dominate unambiguously in all three seasons. In using this model, we found that the hazard rates were affected by a small set of covariates with the most important a time-varying measure defined around success on the field of play. In this regard, our findings resonate with those of Audas, Dobson and Goddard (*op.cit.*). In particular, league position was the most important determinant of an individual manager's exit. A measure for home attendance designed to proxy for, *inter alia*, the financial well-being of the club also exerted a role over the three seasons reviewed here. It should come as little surprise to English Football League managers that the key to survival is avoiding a lowly position in the league within which they ply their trade. No independent role for managerial age, experience, length of service or ethnicity was detected. The fact that a manager had a playing relationship with the club in the past also counted for little. Those that run English League football clubs appear understandably to be characterised by a distinct lack of sentiment.

On a more general note, we believe that the discrete-time approach adopted in this study appears eminently suitable to modelling hazard rates in applications where fast-changing

performance indicators are potentially important determinants of managerial survival. This would generally be the case in most professional sports labour market applications. In addition, the use of random effects within the discrete-time framework to capture neglected heterogeneity provides an added dimension that is relatively easy to implement with available software (e.g., STATA). However, we are not entirely convinced that the use of Chamberlain's fixed effects estimator to model neglected heterogeneity was all that informative in our application given the amount of information the procedure discarded. Finally, the arbitrary choice of a month to represent the spell at risk could also be relaxed by the interested investigator to allow for a much finer discrete unit based on weeks or even days at risk. The extent to which the foregoing enhances empirical analysis is likely to be application dependent.

ACKNOWLEDGEMENTS

The authors acknowledge the constructive comments provided by two anonymous referees of this journal. However, the usual disclaimer applies

REFERENCES

Allison PD (1982). Discrete-time Methods for the Analysis of Event Histories, Chapter 2 in *Sociological Methodology 1982*, S.Leinhardt (editor), pp.61-97, Jossey-Bass Publishers, San Francisco.

Audas R, Dobson S, and Goddard J (1999). Organizational Performance and Managerial Turnover, *Managerial and Decision Economics* **20**: 305-318.

Chamberlain G (1980). Analysis of Covariance with Qualitative Data, *Review of Economic Studies* **XLVII**: 225-238.

Greene WH (2000). *Econometric Analysis*, Fourth Edition, International Edition, Prentice Hall.

Hope C (2003). When Should You Sack a Football Manager? Results from a Simple Model Applied to the English Premiership, *J Opl Res Soc* **54**: 1167-1176.

Jenkins S (1995). Easy Estimation Methods for Discrete-Time Duration Models, *Oxford Bulletin of Economics and Statistics* **57**: 129-138.

Katz E (2001). Bias in Conditional and Unconditional Fixed Effects Logit Estimation, *Political Analysis* **9**(4): 379-384.

Lancaster T (1979). Econometric Methods for the Duration of Unemployment, *Econometrica* **47**: 939-956.

Lancaster T (1992). *The Econometric Analysis of Transition Data*, Cambridge University Press.

Robinson PM (1982). On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables, *Econometrica* **50**: 27-41.

Rothmans Football Yearbook 2002-2003, 33rd Edition, London: Headline.

Scully GW (1994). Managerial Efficiency and Survivability in Professional Team Sports, *Managerial and Decision Economics* **15**: 403-411.

StataCorp. 2005 *Stata Statistical Software: Release 9*. College Station, TX: StataCorp LP.

Sueyoshi GT (1995). A Class of Binary Response Models for Grouped Duration Data, *Journal of Applied Econometrics* **10**: 411-431.

Sky Sports Football Yearbook, 2003-2004 and 2004-2005 Editions, London: Headline.

Table 1: Kaplan Meier Hazard Rates by Month and League Season

Season	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
2001/2	0.011 (0.011)	0.068 (0.028)	0.146 (0.045)	0.057 (0.029)	0.061 (0.031)	0.048 (0.029)	0.034 (0.025)	0.070 (0.036)	0.075 (0.039)
2002/3	0.000	0.011 (0.011)	0.100 (0.035)	0.025 (0.017)	0.000	0.038 (0.022)	0.000	0.039 (0.023)	0.055 (0.028)
2003/4	0.023 (0.016)	0.071 (0.030)	0.051 (0.026)	0.067 (0.031)	0.028 (0.021)	0.029 (0.021)	0.076 (0.035)	0.098 (0.042)	0.036 (0.026)

Notes to table 1:

- (a) For the 2001/2 season: managers at start = 89; involuntary exits = 40; spells at risk=626.
(b) For the 2002/3 season: managers at start = 91; involuntary exits = 22; spells at risk=736.
(c) For the 2003/4 season: managers at start = 87; involuntary exits = 34; spells at risk=646.

Table 2: The Relevant Covariates

Variable Name	Description of Covariates	2001/2	2002/3	2003/4
Relegation Zone	=1 if the league position at the start of the month was 15 th place or below; =0 otherwise.	0.32	0.34	0.33
Attendance	Home attendance as a percentage of a club's ground capacity over the previous month.	62.5	62.6	65.9
International	=1 if the current manager was capped at senior international level as a player; =0 otherwise	0.36	0.28	0.24

Notes to table 2:

- (a) The sample averages reported are based on the pooled sample of 626 for 2001/2 season, 736 for 2002/3 season and 646 for 2003/4 season.

Table 3: Maximum Likelihood Estimates of the Hazard Rate for 2001/2 Season

Variables	Pooled Logit with Flexible Baseline Hazard	Pooled Logit with Flexible Baseline Hazard and Covariates	Pooled Logit with Constant and Covariates	Fixed Effects Logit with Covariates	Random Effects Logit with Constant and Covariates
Constant	‡	‡	-1.778* (0.521)	‡	-1.778* (0.521)
August/September	-3.190* (0.386)	-2.366* (0.639)	‡	‡	‡
October/November	-2.140* (0.264)	-1.292* (0.571)	‡	‡	‡
December/January	-2.850* (0.389)	-1.922* (0.635)	‡	‡	‡
February to April	-2.766* (0.326)	-1.773* (0.602)	‡	‡	‡
Relegation Zone	†	0.790* (0.346)	0.784* (0.342)	0.379 (0.471)	0.784* (0.342)
International	†	-0.123 (0.379)	-0.145 (0.376)	‡	-0.145 (0.376)
Attendance	†	-0.021* (0.008)	-0.021* (0.008)	-0.050* (0.026)	-0.021* (0.008)
σ_u	†	†	†	†	0.001 (1.010)
ρ	†	†	†	‡	0.0000 (0.0003)
Log(L)	-145.74	-137.33	-140.20	-53.45	-140.20
Sample Size	626	626	626	184	626

Notes to table 3:

(a) † denotes not used in estimation. (b) ‡ not applicable in estimation. (c) * denotes statistical significance at the 0.10 level or better. (d) Log(L) denotes the log-likelihood value. (e) The log-likelihood value for the fixed effects logit model is not comparable to the other log-likelihood values.

Table 4: Maximum Likelihood Estimates of the Hazard Rate for 2002/3 Season

Variables	Pooled Logit with Flexible Baseline Hazard	Pooled Logit with Flexible Baseline Hazard and Covariates	Pooled Logit with Constant and Covariates	Fixed Effects Logit with Covariates	Random Effects Logit with Constant and Covariates
Constant	‡	‡	-3.981* (0.721)	‡	-3.981* (0.724)
August/ September	-5.198* (1.003)	-5.923* (1.248)	‡	‡	‡
October/ November	-2.677* (0.312)	-3.368* (0.762)	‡	‡	‡
December/ January	-3.945* (0.583)	-4.554* (0.917)	‡	‡	‡
February to April	-3.439* (0.384)	-4.032* (0.779)	‡	‡	‡
Relegation Zone	†	1.713* (0.500)	1.671* (0.495)	0.751 (0.671)	1.671* (0.495)
International	†	0.790* (0.477)	0.889* (0.470)	‡	0.889* (0.470)
Attendance	†	-0.009 (0.009)	-0.012 (0.010)	-0.075* (0.043)	-0.012 (0.010)
σ_u	†	†	†	†	0.001 (0.620)
ρ	†	†	†	‡	0.0000 (0.0003)
Log(L)	-93.07	-83.95	-89.63	-30.57	-89.63
Sample Size	736	736	736	294	736

Notes to table 4:

(a) See notes to table 3.

Table 5: Maximum Likelihood Estimates of the Hazard Rate for 2003/4 Season

Variables	Pooled Logit with Flexible Baseline Hazard	Pooled Logit with Flexible Baseline Hazard and Covariates	Pooled Logit with Constant and Covariates	Fixed Effects Logit with Covariates	Random Effects Logit with Constant and Covariates
Constant	‡	‡	-1.727* (0.610)	‡	-1.727* (0.610)
August/ September	-3.020* (0.362)	-1.952* (0.680)	‡	‡	‡
October/ November	-2.780* (0.344)	-1.606* (0.668)	‡	‡	‡
December/ January	-3.512* (0.507)	-2.139* (0.781)	‡	‡	‡
February to April	-2.565* (0.288)	-1.084 (0.687)	‡	‡	‡
Relegation Zone	†	1.247* (0.381)	1.206* (0.378)	0.729 (0.496)	1.206* (0.378)
International	†	0.737* (0.424)	0.712* (0.420)	‡	0.712* (0.420)
Attendance	†	-0.035* (0.010)	-0.033* (0.010)	-0.037 (0.029)	-0.033* (0.010)
σ_u	†	†	†	†	0.001 (0.462)
ρ	†	†	†	‡	0.0000 (0.0002)
Log(L)	-131.58	-116.41	-118.80	-46.44	-118.80
Sample Size	646	646	646	167	646

Notes to table 5:

(a) See notes to table 3.

Table 6: The Set of Unimportant Covariates

Variable Description	Test Statistic	2001/2 Season	2002/3 Season	2003/4 Season
Length of Service (in Months) up to the July prior to the season start.	t-ratio	0.03	-0.38	-1.15
Manager's Age (in Years) at August prior to the season start.	t-ratio	0.43	-1.16	-0.68
Manager is English=1; 0 otherwise.	t-ratio	0.35	0.33	0.06
Manager was a player at the current club=1; 0 otherwise.	t-ratio	-0.08	0.13	-0.02
Number of clubs managed prior to the current club.	t-ratio	-0.70	-0.77	-1.09
FA Cup Run [†]	t-ratio	-1.64	0.94	-1.08
League Division Dummies.	χ^2_3	6.06 (0.11)	1.77 (0.62)	5.65 (0.13)
Joint Significance of all the above Variables	χ^2_9	11.72 (0.23)	3.91 (0.92)	9.32 (0.41)

Notes to table 6:

(a) The t-ratios are asymptotic and the relevant critical value at the 0.05 level is ± 1.96 .

(b) [†] denotes time-varying covariates.

(c) The FA cup run variable is an ordinal variable that captures the highest round achieved at the start of the risk month.

(d) The pooled (or stacked) logit refers to the specification with the flexible baseline hazard reported in column two of tables 3, 4 and 5.

(e) For the chi-squared values the numbers in parentheses refer to the significance level of the test.