The Economics of Football

Stephen Dobson

School of Management and Economics,
The Queen's University of Belfast

and John Goddard

Department of Economics,
University of Wales Swansea

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Stephen Dobson and John Goddard 2001

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface Plantin MT 10/13 pt *System* QuarkXPress™ [SE]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Dobson, Stephen.
The economics of football / Stephen Dobson, John Goddard.
p. cm.
Include bibliographical references and index.
ISBN 0 521 66158 7
1. Soccer – Economic aspects – Great Britain. 2. Soccer – Great Britain – France.
I. Goddard, John. II. Title.
GV943.3.D63 2001
338.4′3796334′0941–dc21 2001018128

ISBN 0 521 66158 7 hardback

Contents

List of figures *page* ix
List of tables xi
Preface xv
Acknowledgements xviii

1 Introduction 1
1.1 The economics of team sports: three seminal contributions and the subsequent development of the subject 3
1.2 Outline of this volume 11
1.3 The use of econometrics in this volume 16

2 Professional football: historical development and economic structure 22
2.1 The commercial structure of Europe's 'big five' football leagues 24
2.2 The English football league: competitive structure and team performance 38
2.3 Match attendances 56
2.4 Admission prices and gate revenues 71
2.5 English football and the broadcasting media 80
2.6 Other sources of revenue and cross-subsidy 87
2.7 Football's labour market: players' wages and the transfer system 90
2.8 The historical development of professional football: international comparisons 101
2.9 Convergence and divergence in the league gate revenues of English football clubs 109
2.10 Conclusion 123

3 Competitive balance and uncertainty of outcome 125
3.1 Economic theory of competitive balance in sports leagues 126
3.2 Patterns in football league match results 150
3.3 Modelling and forecasting match results 167
3.4 Conclusion 187

4 The labour and transfer markets 189
4.1 Player mobility, migration and career structure in the English league 191
4.2 Patterns of international migration for professional footballers 203
4.3 The earnings of professional footballers 209
4.4 Determinants of players' compensation: empirical evidence 221
4.5 The football transfer market 227
4.6 Conclusion 236

5 The contribution of the football manager 240
5.1 The role of the football manager 242
5.2 Measuring the managerial contribution: the production frontier approach 245
5.3 Measuring managerial efficiency in English football 252
5.4 Conclusion 262

6 Managerial change and team performance 264
6.1 Patterns of managerial change in English football 265
6.2 Determinants of managerial change: evidence from North American team sports 280
6.3 Specification of job-departure hazard functions 284
6.4 Hazard functions for English managers: duration measured in seasons 286
6.5 Hazard functions for English managers: duration measured in matches 296
6.6 The effect of managerial change on team performance: evidence from North American team sports 303
6.7 Managerial succession effects in English football: performance measured by season 306
6.8 Managerial succession effects in English football: performance measured by match 309
6.9 Conclusion 314

7 The demand for football attendance 317
7.1 Econometric analysis of football attendances 318
7.2 Case studies in the sociology of football attendance: FC Barcelona, and Celtic and Rangers 335
7.3 Modelling the demand for attendance at English league football, 1947–1997 342
7.4 Conclusion 371

8 Information transmission and efficiency: share prices and fixed-odds betting 373
8.1 Ownership, finance and English football's stock market boom 375
8.2 Football team performance and share price movements: an event study 381
8.3 Efficiency of prices in the fixed-odds betting market 397
8.4 Conclusion 416

9 Professional football: current issues and future prospects 418
9.1 Sources of competitive imbalance in English football 419
9.2 The future of professional football 425
9.3 Conclusion 439

List of references 440
Index 452

Figures

2.1 Average home league attendances, first-division clubs, five European leagues, 2000 season *page* 25
2.2 Nationalities of all squad players of first-division clubs, five European leagues, end of 1999 season 33
2.3 Percentage shares in aggregate performance of clubs in Groups 1 to 5 54
2.4 Percentage shares in aggregate attendance of clubs in Groups 1 to 5 60
2.5 Trends in average home-league attendances, selected clubs, four European leagues 61
2.6 Percentage shares in aggregate gate revenue of clubs in Groups 1 to 5 78
2.7 Cross-sectional relationships between revenue and growth and implications for the time-path of revenue 113
2.8 Time-path of revenue under conditional beta convergence 114
2.9 Calculation of Gini coefficient using football club revenue data 116
3.1 Competitive balance under reserve clause and under free agency 135
3.2 Implications for competitive balance of a salary cap 138
3.3 Implications for competitive balance of revenue sharing 139
3.4 Derivation of teams' iso-profit curves, open model 142
3.5 Competitive balance under profit and revenue maximising assumptions, open model 144
3.6 Standard normal probability density function 173
3.7 Implications of shifts in the mean of the standard normal probability density function 175
4.1 Optimum investment in playing talent, rank-order tournament model 219
5.1 Deterministic production frontier 247
5.2 Stochastic production frontiers 249
6.1 Distribution of complete managerial spells, by duration, 1973–1999, all departures 277

6.2 Distribution of complete managerial spells, by duration, 1973–1999, involuntary departures 278
6.3 Distribution of complete managerial spells, by duration, 1973–1999, voluntary departures 279
6.4 Estimated stochastic production frontiers, 1973 and 1999 seasons 288
7.1 Trends in league attendance and performance, selected English clubs 356
8.1 Trends in index value of logarithmic share price: Bolton Wanderers 395
8.2 Trends in index value of logarithmic share price: Charlton Athletic 396
8.3 Trends in index value of logarithmic share price: Manchester United 398
8.4 Trends in index value of logarithmic share price: Newcastle United 399

Tables

2.1 Nationalities of squad players at first-division clubs, five European leagues, summer 1999 *page* 32
2.2 Highest-paid players, five European leagues, 1999 36
2.3 World record transfers and 12 highest transfers, summer 2000 37
2.4 Most successful English league teams in league and cup competition, 1920–1999 44
2.5 Most successful teams in league competition, European comparisons, 1946–1999 46
2.6 League championship dominance, European comparisons, 1940s–1999 50
2.7 Group definitions 52
2.8 English league attendances, aggregate and by division, 1922–1999 57
2.9 Aggregate revenues and costs, Premier League and Football League clubs, 1999 season 71
2.10 Financial and employment data, top six Premier League clubs and divisional averages, 1999 season 72
2.11 English league admission prices, 1926–1999 74
2.12 English league gate revenues, 1926–1999 76
2.13 Shirt sponsorship contracts, Premier League clubs, 2000 season 88
2.14 Gross expenditure on transfers between English league member clubs, 1973–1999 94
2.15 Revenue, wages and salaries and net transfer expenditure, 1994–1999 97
2.16 Net transfer expenditure within the English league, 1973–1999 100
2.17 Average league attendances, French First Division, 1981–2000 104
2.18 Convergence in gate revenues from English league matches, 1926–1997 117

3.1 Percentages of home wins, draws and away wins and average number of goals scored by home and away teams per season, English league, 1973–1999 151
3.2 Percentages of home wins, draws and away wins and average number of goals scored by home and away teams per season, European comparisons, 1973–1999 152
3.3 Joint and marginal percentage distributions of goals scored by home and away teams, English league, 1973–1981 154
3.4 Joint and marginal percentage distributions of goals scored by home and away teams, English league, 1982–1990 154
3.5 Joint and marginal percentage distributions of goals scored by home and away teams, English league, 1991–1999 155
3.6 Percentage distributions of goals scored by home team, conditional on the number of goals scored by away team, English league, 1991–1999 156
3.7 Percentage distributions of goals scored by away team, conditional on the number of goals scored by home team, English league, 1991–1999 157
3.8 Long runs of consecutive results, English league, 1973–1999 159
3.9 Probabilities of home win conditional on recent results 160
3.10 Probabilities of home loss conditional on recent results 161
3.11 Probabilities of away win conditional on recent results 162
3.12 Probabilities of away loss conditional on recent results 163
3.13 Actual and expected numbers of unbeaten and win-less spells and spells of consecutive wins and losses, of various durations 168
3.14 Ordered probit estimation of match results forecasting model, 1973–1999 and 1989–1998 180
3.15 Match results forecasting model's predicted results and actual results 183
3.16 24-month playing records of Middlesbrough and Arsenal up to 24 April 1999 184
3.17 Part-calculation of $\hat{y}^{\star}_{i,j}$ for Middlesbrough v. Arsenal (1) 184
3.18 Part-calculation of $\hat{y}^{\star}_{i,j}$ for Middlesbrough v. Arsenal (2) 185
3.19 Premier League table on the morning of 24 April 1999 186
4.1 Percentage distribution of all players, by age, 31 May 1979 or 1999 192
4.2 Average age of players, by birthplace and division, 31 May 1979 or 1999 193
4.3 Percentage distribution of all players, by month of birth, 1979 or 1999 194

4.4 Percentage distribution of all players, by birthplace, 1979 or 1999 194
4.5 Percentage distribution of players born in England and Wales, by birthplace, and population comparisons, 1979 or 1999 195
4.6 Percentage distribution of players' first clubs, by region, 1979 or 1999 196
4.7 Most frequent first clubs: top 20, 1979 or 1999 197
4.8 Percentages of players whose first club and whose current club is in their region of birth, 1979 or 1999 198
4.9 Distribution of players, by birthplace, selected first clubs, 1979 or 1999 199
4.10 Percentage distribution of players, by duration of present spell at present club, 1979 or 1999 200
4.11 Percentage distribution of players, by number of clubs (past and present) in England and Wales, 1979 or 1999 201
4.12 Distribution of players who joined present club by transfer from another club in England and Wales, by division of current club (in May 1979 or 1999) and division of previous club at time of departure 202
4.13 Analysis of home countries and countries of employment, Euro 2000 squad players 207
4.14 Permanent and temporary transfers, 1973–1996 233
4.15 Variable definitions 235
4.16 Summary statistics for selected variables 236
4.17 Estimated model for determinants of transfer fees 237
5.1 Fixed effects, random effects and maximum likelihood estimates of production frontier, Premier League, 1993–1998 257
5.2 Summary statistics for manager efficiency scores 258
5.3 Managers with highest efficiency scores in production frontier estimations 260
6.1 Managerial spells ranked by duration in matches, top 50 266
6.2 Managerial spells ranked by win ratio, top 50 268
6.3 Managerial spells ranked by adjusted win ratio, top 50 270
6.4 Managerial departures, by season, 1973–1999 274
6.5 Managerial departures, by month and by division, 1973–1999 275
6.6 Complete managerial spells, by duration (matches), summary statistics 276
6.7 Playing seasons ranked by efficiency score obtained from estimation of (6.3): top 50 and bottom 10 290

6.8 Distribution of complete and right-censored managerial spells, by duration (in seasons) and crude estimates of the survivor, distribution, probability and hazard functions 293
6.9 Involuntary and voluntary manager job-departure hazard functions: duration in seasons 294
6.10 Distribution of complete and right-censored managerial spells, by duration (in matches) and crude estimates of the survivor, distribution, probability and hazard functions 299
6.11 Involuntary and voluntary managerial job-departure hazard functions: duration in matches 300
6.12 Relationship between managerial succession and performance, measured by season 308
6.13 Relationship between managerial succession and performance, measured by match 312
7.1 Average capacity utilisation, 1974–1999 322
7.2 Attendance model, first-stage estimation results 348
7.3 Attendance model, second-stage estimation results 369
8.1 Date and method of flotation and share price returns from flotation to December 1999, floated English clubs 378
8.2 Estimation results for determinants of share price returns, pooled model 390
8.3 Estimation results for determinants of share price returns, individual models (1) 392
8.4 Estimation results for determinants of share price returns, individual models (2) 406
8.5 Bookmakers' implicit match result probabilities, descriptive statistics 406
8.6 Match results forecasting model's evaluated match result probabilities, descriptive statistics 408
8.7 Average returns from £1 bets, long, medium and short odds 411
8.8 Average returns from £1 bets placed on forecasting model's prediction for each match 412
8.9 Average returns from £1 bets based on the outcome of each match with the highest expected return, according to the forecasting model (bets ranked in descending order of expected return) 412
8.10 Average returns from every possible £1 bet (bets ranked in descending order of expected return according to forecasting model) 412
8.11 Fixed-odds betting on Premier League fixtures: 24 April 1999 415

1 Introduction

Academic interest in the economics of professional team sports dates back as far as the mid-1950s. Since then, many books and journal articles have been written on the subject. Much of the academic literature originates in the USA. In common with trends that are evident throughout the subject discipline of economics, research on the economics of sport has become increasingly sophisticated, both theoretically and in its use of econometric methodology, especially in recent times. Papers on the economics of sport now appear regularly in many of the leading economics journals, and most economists would agree that in view of its social, cultural and economic importance, professional sport is a legitimate area of interest for both theoretical and empirical researchers. Indeed, many would argue that the unique configurations of individual and team incentives, and the interactions between co-operative and competitive modes of behaviour that professional team sports tend to generate, make this particularly fertile territory in which to explore the perennial questions about incentives, effort, risk and reward which lie at the heart of all areas of economic inquiry.

This volume makes a contribution to the burgeoning literature on the economics of team sports by providing the first comprehensive survey of research that is focused on professional football at club level. A survey of the economics of professional football seems appropriate at the present time, if not long overdue. The spectacular recent increase in the size of football's audience is, of course, a strong motivating factor. Such a survey will recognise and reflect not only football's global popularity at the start of the twenty-first century, but also the special historical significance of England as the original birthplace of the sport. Club football played in the English Premier League and Football League provides the laboratory for most of the original, empirical research that is reported in this volume.

Each chapter of this volume concentrates on a particular aspect of the economics of professional football. The previous theoretical and empirical literature that is relevant to each topic is reviewed, and new and original empirical analyses are presented. The sections that describe the

existing literature aim to convey an impression of the breadth and depth of previous academic research into the economics of professional team sports. Much has been written already about football, and very much more has been written about other sports, especially in the USA where attention naturally tends to focus on sports such as baseball, basketball and (American) football. Though football is the main subject of this volume, due attention and emphasis is devoted to insights that have been obtained from research into other sports, wherever these turn out to be of wider relevance.

As already emphasised, anyone who reads the academic literature on the economics of sport cannot fail to be struck by the sheer volume of column inches that has been devoted to this topic. Of all the articles that have been published, however, there are perhaps a few that have had a particularly important influence in shaping the research agenda for the economics of team sports in general, and for the economics of football in particular. Many of the ideas contained in these articles are as relevant to researchers today as they were when the articles concerned were originally published. In this introductory chapter, three such articles are highlighted and reviewed in some detail. There is, of course, an element of subjectivity in choosing such a small number of articles out of the many excellent ones that have been published. Even so, a consensus seems to have evolved that regards the articles by Simon Rottenberg (1956) and Walter Neale (1964) as fundamental to the subsequent development of research on team sports in general. Both articles address various economic implications of the structural features of the markets within which professional sports teams operate. The article by Peter Sloane (1971) has also had a major influence on the developing the research agenda, especially in respect of the economics of football. In section 1.1, the principal contributions of each of these articles are reviewed in turn, and the subsequent development of the economics of team sports as an academic discipline is outlined. The aim of section 1.1 is to place the research that is reported in the rest of this volume into its proper context. This is followed in section 1.2 by a summary of the contents of each of the following chapters, and in section 1.3 by an outline of the use of econometric methods in the empirical investigations that are reported in this volume.

1.1 The economics of team sports: three seminal contributions and the subsequent development of the subject

Rottenberg: 'The baseball players' labor market', Journal of Political Economy, *1956*

Rottenberg is widely credited with writing the first academic analysis of the economics of professional team sports. The paper was written at a time when US professional baseball players' contracts included a reserve clause. Once he signed his first one-year contract with a team in organised baseball, a player ceased to be a free agent. On expiry of his present contract, his team retained the option to renew his contract for another year.[1] This served to limit players' freedom of movement, by binding them to their present employers. Effectively, the reserve clause created a monopsony in the players' labour market: each contracted player could negotiate with only one potential buyer of his services. The baseball authorities defended the reserve clause on the grounds that it was necessary to ensure an equal distribution of playing talent among opposing teams. Without the reserve clause, the rich teams (with the largest potential markets) would outbid the poorer ones for the best available players. This would tend to reduce uncertainty of outcome and spectator interest in the league competition as a whole, and depress the attendances and revenues of all teams.

Rottenberg's contribution was to argue that free agency in the players' labour market would not necessarily lead to a concentration of the best players in the richest teams. In other words, a reserve clause was not a necessary condition to ensure competitive balance. Professional team sports are intrinsically different from other businesses, in which a firm is likely to prosper if it can eliminate competition and establish a position as a monopoly supplier. In sports, it does not pay a rich team to accumulate star players to the extent that (sporting) competition is greatly diminished, because of the joint nature of 'production' in sports. Consequently, a team that attempts to accumulate all of the best available playing talent will find at some stage that diminishing returns begin to set in: 'In baseball no team can be successful unless its competitors also survive and prosper sufficiently so that the differences in the quality of play among teams are not "too great." . . . At some point, therefore, a first star player is worth more to poor team B than, say, a third star to rich team A. At this point, B is in a position to bid players away from A in the market' (Rottenberg, 1956, pp. 254, 255).

[1] This was subject only to a rule that prevented the wage being cut by more than 25 per cent in any one year if the team was in one of the major leagues.

If teams are rational profit maximisers, the distribution of playing talent among the teams should be reasonably even. Neither a reserve clause nor explicit collusion is necessary in order to bring about this result. It is in each team's self-interest to ensure that it does not become too strong relative to its competitors: 'It follows that players will be distributed among teams so that they are put to their most "productive" use; each will play for the team that is able to get the highest return from his services. But this is exactly the result which would be yielded by a free market' (Rottenberg, 1956, p. 256).

A reserve clause will therefore deliver roughly the same distribution of playing talent among teams as free agency. Whether players are free agents or not, the distribution of playing talent is determined by the incentive to maximise the capitalised value of the services supplied by individual players. If there is another team for which this capitalised value would be higher than it is for the player's present team, then there is a price at which it is advantageous for both teams to trade the player's contract.

Rottenberg also discusses the implications of the reserve clause, and the monopsony power it confers on teams as buyers of playing services, for players' salaries. Each player's reservation wage (the minimum wage he would accept to play baseball) is determined primarily by the next highest wage he could earn outside baseball adjusted to reflect his valuation of the non-pecuniary costs and benefits of playing baseball. Although theoretically the team has the contractual power to impose the reservation wage on all players, Rottenberg notes that in practice this does not seem to happen. Many players earn far more from baseball than they could in alternative employment. This is attributed to the fact that players as well as teams have bargaining power in wage negotiations: in an extreme case, a player can simply threaten to withdraw his services. If a player's reservation wage is $10,000, but he is worth $20,000 to his team, then a wage anywhere between $10,000 and $20,000 is possible, depending on the 'shrewdness and guile of the parties in devising their bargaining strategies' (Rottenberg, 1956, p. 253). Competition among sellers, however, imposes limitations on players' bargaining power. A star player worth $40,000 to his team cannot extract a wage beyond $30,000 if a lesser player worth $20,000 is willing to accept a wage of $10,000 to fill the same position.

The main effect of the reserve clause is that players receive salaries below their value to the team that employs them. In other words, it tends to direct rents away from players and towards teams. The reserve clause does not achieve its stated aim of influencing the allocation of playing talent between teams. Rottenberg concludes by considering several

alternative regimes that might produce a more or less equal distribution of playing talent between teams with free agency in the players' labour market. These include:

- *Revenue pooling.* If all revenues are shared equally, teams have no pecuniary incentive to spend on players to enhance their own performance and revenue. An equal distribution of mediocre playing talent is the most likely outcome.
- *Salary capping.* The effect of a salary cap on the distribution of talent will depend on the extent to which teams can circumvent it by offering players non-pecuniary rewards.
- *Allocating multiple team franchises in large cities.* If reasonable equality between each team's potential market size can be achieved, this is expected to create a more equal distribution of playing talent.

Neale: 'The peculiar economics of professional sports', Quarterly Journal of Economics, *1964*

Neale's analysis begins by emphasising the joint nature of production in professional sports. Heavyweight boxing is used as an example to introduce what Neale calls the 'Louis Schmelling Paradox'. World champion Joe Louis' earnings were higher if there was an evenly matched contender available for him to fight than if the nearest contender was relatively weak. The same point applies also in baseball. 'Suppose the Yankees used their wealth to buy up not only all the good players but also all of the teams in the American League: no games, no receipts, no Yankees. When, for a brief period in the late fifties, the Yankees lost the championship and opened the possibility of a non-Yankee World Series they found themselves – anomalously – facing sporting disgrace and bigger crowds' (Neale, 1964, p. 2).

Does this imply that professional sport is an industry in which monopoly is less profitable than competition, contradicting what we teach students and what they can read in any Principles text? Neale addresses this paradox by distinguishing between 'sporting' and 'economic' competition. Sporting competition is more profitable than sporting monopoly for the reasons outlined above, but sporting competition is not the same as economic competition. Similarly, although in law the sports team is a firm (which may be motivated by profit), it is not a firm in the economist's sense. A single team cannot supply the entire market; if it did it would have no one to play. Teams must co-operate with each other to produce individual matches and a viable league competition, so there is joint production. The league's organising body exerts strict controls over a wide range of matters including competition rules and schedules, player

mobility and the entry and exit of clubs. In short, the league rather than the individual team is the 'firm' in the economist's sense. A sports league should be regarded as analogous to a multi-plant firm, in which the individual teams are 'plants', subject to decisions which are taken and implemented collectively at league level.

If the sports league is the 'firm' in the economic sense, this raises the question as to why it is unusual to observe direct competition between rival leagues operating within the same sport. Although the National League and American League do operate simultaneously in baseball, analytically they should be regarded as one larger 'multi-league' firm, since they come together at the end of each season to produce the World Series. Geographical division is a more common form of segmentation though, according to Neale, one that is inherently unstable. Where direct competition is prevented by geography, profit incentives tend to promote enlargement and the elimination of geographical boundaries.[2] Competition between different sports is more common than competition between rival leagues within the same sport, though segmentation based on nation, region, season of the year or even social class are still common.[3] All such forms of segmentation tend to inhibit direct head-to-head competition.

Neale suggests that the general lack of competition between sports leagues arises because the cost and demand characteristics of the market for professional team sports tend to create conditions of natural monopoly, making it efficient for a single league to supply the entire market. On the cost side, Neale suggests that the long-run average cost curve is probably horizontal. Although an increase in the scale of production might entail the use of less efficient playing inputs, raising average costs, this tends to be offset by an 'enthusiasm effect'. If the sport operates on a larger scale, public enthusiasm encourages more people to take up playing, eventually raising the supply of players at the highest level. To some extent, the enthusiasm effect makes supply and demand interdependent: if more people play the sport, more will also want to attend matches at the professional level. Finally, the existence of rival leagues would effectively break the monopsony power of teams as buyers of

[2] This observation seems to be borne out by recent changes in the structure of competition in European football.

[3] Traditionally, English cricket and rugby union were upper- or middle-class sports, while football and rugby league were working-class sports. More recently, the class divisions may have become more blurred, but have not disappeared altogether. Cricket is played only in summer; rugby league switched from a winter to a summer calendar in the late 1990s; football and rugby union are still played in winter. Rugby league still has a strong regional identity centred on Lancashire and Yorkshire. Cricket is popular throughout England, but very much a minority sport in Scotland.

playing services, enhancing players' bargaining power in wage negotiations. This would tend to make costs higher than they are when one league operates as a monopoly supplier.

On the demand side, Neale suggests that baseball teams produce a number of streams of utility: directly for spectators who buy tickets for seats in the stadium and for television viewers who watch the match at home; and indirectly for everyone who enjoys following the championship race as it unfolds. The closer the competition, the greater is the indirect effect. For newspapers and television companies in particular, the indirect effect is a marketable commodity that helps sell more of their product. The size of the indirect effect depends on the scale and universality of the championship, and is therefore maximised when the league is a monopoly supplier. Overall 'it is clear that professional sports are a natural monopoly, marked by definite peculiarities both in the structure and in the functioning of their markets' (Neale, 1964, p. 14).

An important implication is that the peculiar economic characteristics of professional sports leagues and their constituent teams should be recognised by legislatures, by the courts and by the general public, whenever practices such as collective decision making or other (apparently) anti-competitive modes of behaviour come under scrutiny.

Sloane: 'The economics of professional football: the football club as a utility maximiser', Scottish Journal of Political Economy, *1971*

Sloane's paper questions Neale's conclusion that the league rather than the individual team or club is the relevant 'firm' (or decision making unit) in professional team sports. In the case of English football, for example, the sport's governing bodies merely set the rules within which clubs can freely operate. Most economic decisions, such as how much money to spend on stadium development and how many players to employ, are made by the clubs. Although the total quantity of 'output' (the number of matches played by each club) is regulated, this clearly reflects the clubs' common interest. In cartels, it is not unusual for firms to reach joint decisions concerning price or production, but this does not imply that the cartel should be elevated to the theoretical status of a 'firm'. In short, Sloane suggests that Neale's argument tends to overemphasise mutual interdependence. 'The fact that clubs together produce a joint product is neither a necessary nor a sufficient condition for analysing the industry as though the league was a firm' (Sloane, 1971, p. 128).

Having argued that the club is the relevant economic decision maker, Sloane goes on to raise a number of important questions concerning the objectives of sports clubs. Implicit in the reasoning of both Rottenberg

and Neale is an assumption of profit maximisation. Despite the 'peculiarities' of sports economics elucidated by Neale, the behaviour of professional sports teams is analysed within a very conventional analytical framework. While this may be reasonable in the case of US professional team sports, where many teams do have an established track record of profitability, Sloane suggests that it may not be universally applicable. Through most of the history of English football, profit making clubs have been very much the exception and not the rule. Chairmen and directors with a controlling interest in football clubs are usually individuals who have achieved success in business in other fields. Their motives for investing may include a desire for power or prestige, or simple sporting enthusiasm: a wish to see the local club succeed on the field of play. In many cases, profit or pecuniary gain seems unlikely to be a significant motivating factor. If so, it may be sensible to view the objective of the football club as one of utility maximisation subject to a financial solvency constraint. The financial solvency constraint recognises that the benevolence of any chairman or director must reach its limit at some point.

Non-profit maximising models of the firm had received considerable attention in the economics literature during the decade prior to the publication of Sloane's article. 'A major drawback to the general introduction of the utility maximisation assumption in the theory of the firm is that it may be rationalised so that it is consistent with almost any type of behaviour and therefore tends to lack operational significance' (Sloane, 1971, p. 133).

In the case of football clubs, however, it is not too difficult to identify several plausible and easily quantifiable objectives. Sloane suggests the following:

- *Profit*. Sloane's expectation that profit is not the sole or even the most important objective does not preclude its inclusion as one of a number of arguments in the utility function.
- *Security*. Simple survival may be a major objective for many clubs. Decisions (concerning, for example, sales of players) may aim more at ensuring security than at maximising playing success.
- *Attendance or revenue*. A capacity crowd enhances atmosphere and a sense of occasion, and may in itself be seen as a measure of success. Recently, an increasing willingness to charge whatever ticket prices the market will bear suggests that revenue (or profit) carry a heavier weight in the utility function than in earlier periods, when it was usual to charge the same price for all matches, irrespective of the level of demand.
- *Playing success*. This is probably the most important objective of all, and one to which chairmen, directors, managers, players and spectators can all subscribe.

- *Health of the league*. This enters the utility function in recognition of clubs' mutual interdependence.

Formally, the club's objective is to maximise:

$$U = u(P,A,X,\pi_R - \pi_0 - T) \qquad \text{subject to } \pi_R \geq \pi_0 + T \qquad (1.1)$$

where P = playing success; A = average attendance; X = health of the league; π_R = recorded profit; π_0 = minimum acceptable after-tax profit; and T = taxes.

It is important to note that the utility maximising model has implications very different to those which follow from the profit maximising assumptions of Rottenberg and Neale. In particular, if the weighting of P in the utility function is heavy relative to that of X and π_R, the argument that diminishing returns would prevent the accumulation of playing talent in the hands of a small number of rich clubs does not necessarily hold, unless there are binding financial constraints preventing expenditure on new players. The idea that profit incentives should help maintain a reasonably even allocation of playing talent between richer and poorer teams therefore breaks down. The case for regulation to over-ride the 'free market' outcome, whether in the form of a reserve clause, revenue sharing or the taxation of transfer fees, therefore seems to be enhanced if clubs are pursuing non-profit rather than profit objectives.

The economics of team sports since the 1970s

Since the appearance of the pioneering work of Rottenberg, Neale and Sloane, there has been a proliferation of published academic research on the economics of team sports, in the form of both journal articles and books. Many of these contributions will be reviewed in their proper place later in this volume. At this point, however, we shall highlight just a few of the main themes in the development of the economics of team sports literature over the last three decades, by identifying a handful of journal articles that seem to us to have been particularly influential. The ideas that were originally developed in these articles will figure prominently, and will be considered in greater detail, throughout this volume.

While Rottenberg and Neale both adopted a discursive style for the presentation of their insights, an article by El-Hodiri and Quirk (1971) demonstrated that these and other ideas could also be represented and developed within a more formal, mathematical framework. The kind of model developed by El-Hodiri and Quirk is now the standard apparatus for analysing questions about the implications for competitive balance within a sports league of institutional or policy changes, such as the introduction of free agency, salary caps or revenue sharing arrangements.

Nearly a quarter of a century later, and with the benefit of hindsight after witnessing a number of attempts at introducing such changes by the North American sports governing bodies in the interim, Fort and Quirk (1995) and Vrooman (1995) provided a compelling demonstration of the usefulness of this framework for policy analysis in team sports.

Sports economists have often claimed that an abundance of easily accessible data on the characteristics and performance of individual personnel makes professional team sports an ideal laboratory for the empirical scrutiny of economic theories and hypotheses that might prove difficult to test elsewhere. An article by Scully (1974), which investigated the relationship between the performance of individual players in US Major League Baseball and their compensation, provides a classic example of a contribution of this kind. Scully's motivation was to demonstrate that severe restrictions on the rights of the players to sell their services in the labour market (in the form of the reserve clause) led to exploitation perpetrated by the teams as employers, in the sense that individual players' wages were systematically below their contributions to their teams' revenue-earning capability. But the applicability of Scully's methods and the relevance of his findings went much further than this. Identifying the links between pay, productivity and human capital is a task of central importance towards achieving a much broader understanding of how labour markets work, as well as insights into specific issues such as inequality and discrimination in opportunity and compensation.

Developments elsewhere in the field of economics are also capable of providing valuable insights into the economics of sports. Two such examples are articles by Rosen (1981) on the economics of superstars, and by Lazear and Rosen (1981) on the earnings of chief executives in large corporations. Both provide insights into an issue of major controversy in the public debate about the economics of sports: the massively inflated earnings of the leading stars. Rosen demonstrated how in certain professions the technological conditions of supply create markets in which the top performers can service very large audiences at little or no incremental cost to themselves as the audience size increases. The outcome is that the extra compensation of the top performers (in comparison with slightly less talented rivals) is hugely disproportionate to their extra talent. Film stars and musicians are obvious examples, but the model is relevant to sports stars as well. Lazear and Rosen likened the high earnings of chief executives to a lavish first prize in a tournament. The possibility of emerging as a future prize winner provides incentives for others to strive for success early in their careers, with obvious benefits for the productivity of the workforce as a whole. The internal labour market of a sports team, with its squad of aspiring youngsters competing for first-team places that

may eventually enable them to acquire superstar status, may resemble that of a corporation in this respect.

For most of the 1970s and 1980s, research into the economics of football specifically (as opposed to team sports in general) focused mainly on identifying the demand for attendance at football matches. An article by Hart, Hutton and Sharot (1975) was an early forerunner of a number of other UK demand studies that appeared over the course of the next two decades. For a broader perspective, one could also consult books and other works published by social historians (such as Walvin, 1975), sociologists (Dunning, Murphy and Williams, 1988; Guilianotti, 1999), or geographers (Bale, 1992). During the 1990s, however, the range of football-related issues attracting the interest of British and other European economists increased rapidly. Recent years have seen the publication of articles on topics as diverse as the settlement of transfer fees for football players moving between clubs; the implications of the creation of a European league for competitive balance in domestic and European football; the performance and job security of football managers; and discrimination against black players in football's labour market. If the present volume makes even a small contribution towards disseminating knowledge and stimulating interest in this relatively new, exciting and rapidly expanding area of theoretical and empirical research, then it will have achieved one of its most important objectives.

1.2 Outline of this volume

This section contains a brief outline of the contents of each of the following chapters of this volume. Chapter 2 provides an overview of the historical development of English football as a business, and looks in some detail at football's present-day commercial structure. The chapter also draws comparisons between the historical development and commercial structure of club football in England and in a number of other countries, including several in mainland Europe as well as Japan. Recent changes in the economic and financial structure of professional football in England and elsewhere have raised widespread concern about the implications of growing economic disparities between clubs operating at different levels within the league. As seen above, economists studying professional team sports have for many years been aware of the link between the distribution of resources among the members of sports leagues, and the degree of competitive balance. Explicit or implicit types of cross-subsidy and regulations restricting free economic competition in both the product and players' labour markets have been a pervasive feature of the history of English football. The retain-and-transfer system, the maximum wage, the

minimum admission price and gate and television revenue sharing are some of the most prominent measures of this kind, all of which are described in chapter 2. The chapter concludes by reporting an empirical investigation in which a number of standard economic measures of convergence, divergence and inequality are applied to an English club-level league gate revenues data set.

Competitive balance and uncertainty of outcome are the main themes of chapter 3. Following Rottenberg, a number of US economists have developed theoretical models of competitive balance in sports league competition, which formalise some of Rottenberg's original insights, and permit exploration of a number of other policy issues. Naturally, the US literature on this topic has developed primarily with North American professional team sports in mind. Chapter 3 begins with a description of the North American model for the organisation and regulation of professional team sports. A number of features which differ significantly from their counterparts in the British or European model are highlighted. The chapter then reviews the main findings of the US literature, and considers their relevance for the case of English football. Two key assumptions of the US literature in particular seem questionable if these models are applied directly to European football. Following Sloane, the first is the assumption that team owners are motivated solely by profit. The second is the assumption that playing talent is drawn from a pool whose total size is fixed. Especially since the Bosman ruling in 1995 (see chapter 2), European football labour markets have been open rather than closed, with players frequently hired from outside the domestic league. In chapter 3 the competitive balance model is adapted for English football by introducing behavioural assumptions other than profit maximisation, and by treating the labour market as open.

Closely balanced competition is important because it maximises uncertainty of outcome or unpredictability: an essential characteristic of any sports contest. The empirical sections of chapter 3 examine the extent to which individual match results in English football are predictable or unpredictable. Patterns and trends in English league match results and goal scoring records since the 1970s are identified and compared with equivalent data from a number of other European leagues. The main long-term trend in the data is a progressive improvement in the performance of away teams, and a corresponding decline in the importance of home advantage. Analysis of sequences of match results reveals evidence of a negative 'persistence' effect: a recent run of good results appears to create either pressure or complacency, increasing the risk that the next result will be bad. Chapter 3 concludes by developing a statistical model that processes the patterns in sequences of past match results,

in order to assess probable outcomes and provide forecasts for future results.

Professional football players are the main focus of attention in chapter 4. The chapter starts by presenting a descriptive profile of various personal and career characteristics of the players who turned out most regularly for their clubs in English league matches played during the 1979 and 1999 seasons. The analysis helps identify a number of significant changes to the composition of English football's regular labour force that have taken place over the last 20 years. This is followed by a more general investigation of patterns of international migration among professional footballers, which draws on the findings of previous sociological research.

Chapter 4 then goes on to consider the determinants of players' compensation. While the earnings of some superstar players are now a cause of major controversy, above-inflation increases in players' wages, especially at the highest level, have been a permanent feature of English football since the abolition of the maximum wage in the early 1960s. Chapter 4 considers some theoretical explanations for the extremely high earnings of the top superstars in sports and in similar fields such as film, music and publishing. Scarcity of supply of the highest talent, together with the very large audience reach of the top performers, are important factors which help to explain highly skewed earnings distributions. Another important aspect of footballers' compensation is the fact that wage structures are often extremely hierarchical, and players who are perhaps only a small fraction better than others frequently earn several times as much. By regarding the massive salaries of the top players as equivalent to a generous first prize in a tournament, encouraging all players to contribute maximum effort to the team's cause in an attempt to become the next prize winner, the rank-order tournament model provides one plausible explanation for observed wage differentials. Finally, chapter 4 concludes with an empirical investigation of the English football transfer market. The trading of players for cash is a feature of professional football worldwide, though one that is currently subject to the intense scrutiny of European legislators. The empirical model presented in chapter 4 shows how transfer fees are influenced by the characteristics of the players concerned, and of the buying and selling clubs.

Rottenberg considered briefly the nature of the production process in team sports, suggesting that the output of the club could be measured in terms of its attendance. Other researchers have taken up his idea of a team production function, but have typically employed team performance measures (such as win ratio), rather than attendance, as a measure of output. There have been numerous team production function studies based on North American sports such as baseball and basketball, in

which many dimensions of playing skill are quantifiable, making it easy to measure individual productivity. But in American football and (association) football, there is greater interaction among the player inputs, so the measurement of productivity is more problematic. Chapter 5 includes a review of the empirical literature on team production functions, and describes the task of modelling team production in football.

An interesting application of the team production function approach is in quantifying the contribution of the manager to the team's performance. Typically, the managerial input is measured as part of a residual that is obtained after controlling for the players' contribution to performance. Having controlled for team quality, the variation in performance that remains unexplained is attributed either to the manager or to luck and other random factors. Chapter 5 concludes by describing some recent research which estimates team production functions using English football data, and uses them to obtain efficiency scores for football managers.

The measurement of managerial efficiency forms only one part of the research agenda on the role of managers in team sports. Sociologists and economists have also examined the link between team performance and the security of the manager's job tenure, and the effect of a change of manager on the subsequent performance of the team. This literature, which is reviewed in chapter 6, contributes to a broader research programme that is concerned with the two-way link between organisational performance and the managerial contribution.

Chapter 6 goes on to present some new empirical results for English football managers. Chronic job insecurity is undoubtedly one of the defining characteristics of the football manager's position. An important aim of the empirical analysis in chapter 6 is to quantify the factors that are most likely to trigger a change of manager, including match results over the short term, longer-term team performance measures and various personal and human capital attributes of the manager himself. Of course, changing the manager is a rational decision for club owners to take if it is likely to lead to an improvement in fortune under a new incumbent. Opinion as to the average effect on team performance of a change of manager seems to be quite mixed. The empirical results for English football presented in chapter 6 suggest that while a change of manager often has a disruptive effect that makes matters worse in the short term, in the longer term teams that change managers seem to recover faster from a run of poor results than teams that retain their manager following a poor run.

Until quite recently, the economic analysis of football was concerned almost exclusively with the determinants of demand, or with the statistical modelling of match and season attendances. Even in the present-day

world of highly lucrative television contracts, merchandising and sponsorship, gate revenues still constitute a significant proportion of the total revenues of most football clubs. Furthermore, spectators who attend matches in person are not just passive consumers. Their presence contributes in a fundamental way to the quality of the product, by generating atmosphere and a sense of occasion. Most spectators also seek to influence match outcomes though the effects of vocal encouragement and criticism of players, managers and match officials. For these reasons, as well as the 'enthusiasm effect' suggested by Neale, supply and demand are really interdependent.

Identifying the causes of variations in attendances, from season to season and from match to match, is therefore the main objective of chapter 7. The chapter reviews the empirical literature on the demand for football attendance, highlighting some of the main problems of model specification, estimation and interpretation encountered by researchers in this area. Sociological case studies from Spain and Scotland are also used to illustrate the impact of broader social or cultural influences on patterns of demand for football attendance. Chapter 7 concludes by presenting an empirical analysis of English club-level attendances during the post-Second World War period. The analysis shows the effects on attendances of factors such as short-term loyalty, team performance and admission prices. It also investigates the relationship between various socio-economic, demographic and football-related characteristics of each club and its home town, and average attendances over the long term.

Neoclassical economics predicts that unco-ordinated interactions between large numbers of self-interested buyers and sellers operating in free, unregulated markets can normally be relied upon to deliver an efficient allocation of scarce resources. A key assumption is that all relevant information about the prices and characteristics of goods and services is transmitted rapidly and accurately among all participants in the relevant market. Chapter 8 presents two case studies, which focus on the efficiency with which relevant information about the performance of football teams on the field of play is impounded into the prices at which trade takes place in particular markets.

The first case study concerns the secondary market for trade in the shares of football clubs quoted on the London Stock Exchange (LSE). Following a description of the historical evolution of present-day structures of ownership and control in English club football, the empirical investigation seeks to quantify the extent to which major events that take place on the field of play, such as European qualification and elimination, or domestic promotion and relegation, carry implications for the share prices of football clubs that are floated on the LSE. If new information

about team performance is transmitted and absorbed by the markets rapidly, a direct link between fluctuations in fortunes on the field of play and variations in share prices should be discernible empirically. The second case study in chapter 8 concerns the market for fixed-odds betting on the results of football matches. The main objective here is to establish whether a high-street bookmaker's published odds reflect all information that is relevant to predicting the match outcome.

This volume concludes in chapter 9 with a discussion of some of the major policy issues concerning football's future development as a sport and as a business that are currently facing the sport's governing bodies, club owners, broadcasters, sponsors and legislators. Among the policy issues considered, with the aid of the economic analysis presented in previous chapters of this volume, are the proposals for the formation of a European Super League; rocketing inflation in players' wages; plans to reform or even abolish the players' transfer market; the prospects for the survival of the smaller clubs; and the future of football's relationship with the broadcasting media. In one way or another, decisions concerning all such matters impinge upon what is surely the most fundamental issue in the economics of team sports: the need to maintain competitive balance. Forty-five years on, this is one issue that is still as relevant today as it was at the time Rottenberg's original ground-breaking article was published.

1.3 The use of econometrics in this volume

As indicated above, chapters 2–8 of this volume report the results of new empirical investigations of a number of the key topics that are covered in each chapter. Regression analysis, in various guises, is the main empirical technique used in these investigations. For readers who are familiar with econometrics, in the empirical sections of this volume we have attempted to strike a reasonable balance between the objectives of technical rigour and accessibility. With accessibility in mind, we have chosen not to burden the text with lengthy technical discussions about model specification or estimation issues. We have also chosen not to extend the tables of empirical results (many of which are already large) by including extensive listings of diagnostic test statistics. Readers who are interested or concerned about such matters should consult the relevant journal articles cited in the text, where the various technical issues are given proper emphasis and attention.

For the benefit of readers who are unfamiliar with econometrics, the presentation of the empirical results in this volume has been confined to specific sections within each chapter. These sections can be skipped

without prejudicing understanding of the material contained elsewhere. Readers who skip the empirical sections will find a concise and completely non-technical summary of the main empirical findings in the concluding sections of each chapter. For readers without a background in statistics or econometrics who do wish to read the empirical sections, the rest of this sub-section provides a brief, non-technical description of the basics of regression analysis, and introduces some of the main terminology and jargon. Readers with an econometrics background are invited to skip the rest of section 1.3 and proceed directly to chapter 2.

Most econometric studies seek either to *explain* observed facts about the real world, or to construct a model that can *predict* the outcome of events that have not yet occurred. This volume includes examples of both types of econometric investigation. Typically, a regression model takes the following form:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + u_i \qquad (1.2)$$

The definitions of variables and symbols used in (1.2) are as follows:

y_i	is the dependent variable
$x_{2i}, x_{3i}, \ldots, x_{ki}$	are the independent or explanatory variables
u_i	is the error term
$\beta_1, \beta_2, \beta_3, \ldots, \beta_k$	are the regression coefficients.

The dependent variable y_i is the variable whose behaviour the model seeks either to explain or to predict. The independent variables $x_{2i}, x_{3i}, \ldots, x_{ki}$ are other variables thought to influence or determine y_i. The coefficients $\beta_2, \beta_3, \ldots, \beta_k$ identify the impact on y_i of small changes in the values of each of $x_{2i}, x_{3i}, \ldots, x_{ki}$ respectively. For example, if x_{2i} increases by one unit while all other variables in (1.2) are held constant, the numerical adjustment to the value of y_i is given by the coefficient β_2. The one remaining coefficient, β_1, allows for the scaling of the dependent and independent variables; in most (but not all) regression models β_1 does not have any important interpretation.[4] Finally the error term u_i allows for any variation in the dependent variable that is not accounted for by corresponding variation in the independent variables. In most regression models, the value taken by the error term is assumed to be purely random. For purposes of statistical inference (see below) it is often useful

[4] Occasionally, other notational conventions are used in an expression such as (1.2). Different symbols may be used for both variables and coefficients. It is sometimes convenient to write an expression such as (1.2) more concisely using matrix notation, i.e. $y_i = \beta' x_i + u_i$, where $\beta' = (\beta_1\ \beta_2 \ldots \beta_k)$ and $x'_i = (1\ x_{2i} \ldots x_{ki})$. Alternatively it is sometimes convenient to use variable names rather than *y*s and *x*s, i.e. $DEPVAR = \beta_1 + \beta_2\ INDVAR2 + \ldots + \beta_k INDVARk + ERROR$, where the names actually used describe the variables.

to assume that the error term follows some specific probability distribution, such as the *normal distribution*.

The '*i*'-subscripts on the variables y_i, x_{2i}, x_{3i}, . . . , x_{ki} and u_i in (1.2) indicate that the relationship between the independent variables and the dependent variable holds over a number of *observations* contained within a sample of data. The observations are indexed over $i = 1 \ldots N$, where N is the total number of observations.[5] The sample therefore comprises N complete sets of observed numerical values of x_{2i}, x_{3i}, . . . , x_{ki} and y_i. The numerical values of the coefficients $\beta_1, \beta_2, \beta_3, \ldots, \beta_k$ are all unknown, and the purpose of the regression analysis is to process the data in order to obtain the best possible numerical estimates of these coefficients. The choice of estimation method depends on the specification of the model. Estimation methods used in this volume (and referred to in the text) include Ordinary Least Squares (OLS), Generalised Least Squares (GLS), Weighted Least Squares (WLS), Fixed Effects and Random Effects (FE and RE), Maximum Likelihood (ML) and Seemingly Unrelated Regressions estimation (SUR). Estimated regression coefficients are usually denoted $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, . . . , $\hat{\beta}_k$, to distinguish them from the corresponding true (but unknown) coefficients β_1, β_2, β_3, . . . , β_k.

Once a regression model has been estimated, the signs and numerical magnitudes of the estimated coefficients convey information about the direction and strength of the relationship between each independent variable and the dependent variable. Because the estimated coefficients are based on a limited sample of data, there is always a suspicion of imprecision or unreliability attached to them. For this reason, estimation methods for regression coefficients provide a *standard error* with each coefficient estimate. The standard error reflects the reliability of the estimate: the smaller the standard error, the greater the reliability. It can also be used to construct a *confidence interval* for the true value of the coefficient. A confidence interval is simply a range of likely values for the true coefficient located on either side of the estimated coefficient. We can claim that the true coefficient should lie somewhere within this range, subject only to a small (and specified) probability of being wrong.

The ratio of an estimated coefficient to its standard error, usually known as a *z-statistic* or *t-statistic* (depending on which type of regression model is being used), provides a convenient method for assessing whether the estimation has succeeded in identifying a relationship that is reliable

[5] In cases where data are observed over a number of different time periods, '*t*'-subscripts are used instead of '*i*'-subscripts in order to denote time. For data with both a cross-sectional and a time series dimension, variables may be indexed using both '*i*'- and '*t*'-subscripts.

in a statistical sense. This is one of the main tasks of *statistical inference*. The z-statistic (or t-statistic) for the coefficient estimate $\hat{\beta}_2$ is calculated as $z = \hat{\beta}_2 / \text{se}(\hat{\beta}_2)$ where $\text{se}(\hat{\beta}_2)$ is $\hat{\beta}_2$'s standard error. If the absolute value of the z-statistic exceeds a certain critical value, we say that $\hat{\beta}_2$ is *statistically significant* or *significantly different from zero*. This means that we have sufficient statistical evidence to reject the *null hypothesis* that the true value of the coefficient β_2 is zero, in favour of the *alternative hypothesis* that β_2 is non-zero. In this case the null and alternative hypotheses would be written $H_0: \beta_2 = 0$ and $H_1: \beta_2 \neq 0$. This is an example of a *two-tail* hypothesis test. Another possibility is a *one-tail* test, in which the null and alternative hypotheses might take the form $H_0: \beta_2 \leq 0$ and $H_1: \beta_2 > 0$, respectively. Whenever a *hypothesis test* is carried out, it is usual to quote an accompanying *significance level*, such as 1, 5 or 10 per cent. This expresses the probability that the test may cause us to draw a wrong inference by rejecting a null hypothesis that is actually true. With a significance level of 5 per cent, if β_2 really were zero the test would run a 5 per cent risk of incorrectly concluding that β_2 is non-zero. The smaller the significance level the greater the confidence we can have in any inference that an estimated coefficient is statistically significant, and that the true coefficient is therefore different from zero.

If the purpose of a regression analysis is explanation, then by following the procedures outlined above, we may be able to infer (with only a small risk of making a mistake) that β_2 is non-zero. If so, then according to (1.2) we have shown that the independent variable x_{2i} does have an effect on the dependent variable y_i. In other words, some of the variation in y_i has been explained by corresponding variation in x_{2i}. The estimated coefficients, standard errors and z- or t-statistics are therefore of considerable interest: they show which independent variables are important in explaining the behaviour of the dependent variable. Consequently, much of the discussion of the empirical results in this volume concentrates on this aspect of the estimated regression models.

If the purpose of the regression analysis is prediction, the estimated coefficients, standard errors and z- or t-statistics are important, because they enable us to determine which independent variables should be included in the model that is to be used to generate the best possible prediction. In order to obtain a prediction, we need a set of values for all of the independent variables for an out-of-sample observation. These are substituted into the estimated model (already estimated using the in-sample data), so as to produce a prediction for the out-of-sample value of the dependent variable.

The *linear regression model*, also known as the *multiple regression model*, is probably the best known and most widely used type of regression model.

It is used in most situations where the dependent variable is continuous, and can take any numerical value. The linear regression model is employed a number of times in this volume: in chapter 2, to estimate models of convergence in football club gate revenues; in chapter 4, to estimate a model that explains transfer fees in terms of player and club characteristics; in chapter 6, to estimate the effect of a change of manager on team performance during the season after the change takes place; in chapter 7, to estimate models that explain variations in match attendances; and in chapter 8, to estimate models that explain movements in football club share prices.

Not all data, however, are amenable to analysis using the linear regression model. A model capable of predicting match results, for example, has a discrete or qualitative dependent variable, which can take only one of three non-numerical values: 'win', 'draw' or 'lose'. The *ordered probit model*, one of a broader class of *discrete choice models*, is the appropriate regression model to use in this case. This type of model is estimated for the first time in chapter 3, and is employed again in chapters 6 and 8. A second example arises when the relationship between measurable team inputs (playing talent or financial resources) and team performance is considered. In estimating this relationship, it is useful to be able to decompose the error term (the variation in team performance that cannot be explained by variations in the measurable inputs) into two parts: one reflecting the manager's contribution to performance, and the other reflecting random influences or luck. In chapters 5 and 6 *stochastic frontier analysis* is used to specify a regression model that has a special error term with the required structure. A final example arises when we seek to identify the factors most important in triggering the decision to terminate a manager's appointment. In this case the dependent variable in the regression model is an indicator variable, which takes the form $y_i = 0$ if the manager was retained and $y_i = 1$ if the manager was removed after match i was played. In chapter 6 yet another type of regression model is needed to estimate what are known as job-departure *hazard functions* for football managers.

In the space available in section 1.3, it has been possible to provide only a brief and highly selective outline of some of the basics of regression analysis. Readers who wish to discover more are strongly encouraged to explore one or more of the excellent econometrics texts that are currently available. In teaching econometrics to undergraduates and postgraduates for a number of years, the authors have found the texts by Gujarati (1995), Kennedy (1998), Pindyck and Rubinfeld (1998) and Greene (1999) to be especially helpful. Gujarati and Pindyck and Rubinfeld both provide a detailed and comprehensive beginner's guide

to econometrics, while Kennedy explains the terminology of econometrics in a clear, succinct and non-technical manner. Greene, meanwhile, offers a more advanced treatment, including full technical details of a large number of more highly specialised econometric models and estimation techniques.