

Football Analytics using Stata

By Dr Babatunde Buraimo
Senior Lecturer in Sports Management and Economics
The Centre for Sports Business
Management School
The University of Liverpool
Liverpool
L69 7ZH

Telephone: +44 (0)151 795 3536
Email: b.buraimo@liverpool.ac.uk

Introduction

The application of statistics and econometrics (analytics) to football has evolved over the past decades. Both the academic community and the football sector invest efforts researching and informing decision making by developing complex and sophisticated empirical models. The successful development of models has the potential to improve decision-making and offer firms and comparative advantage. The benefits are not limited to individual firms, and league administrators can also use football analytics to improve policy which in turn could maximise consumer welfare.

Advances in football analytics have been driven by a number of factors. Firstly, access to data has improved and at a very basic level, comprehensive data can be accessed and acquired with limited constraint although more sophisticated data are costly. Secondly, the popularity of analytics has become fashionable following the success of the film *Moneyball*. Even if the benefits of analytics are disputed (and there are antagonists), the popularity of the film and the eloquence of the book has served to at least open the debate. Thirdly, the wide adoption football analytics means that many conventional and received wisdom can be empirically tested.

In this course, various concepts and theories will be discussed. Their relevance to football management and decision making within football will be established. Having examined their theoretical relevance, students will engage in the process of data management and organisation, as well as the development of empirical models to test various hypotheses and theories. The differences between statistical significance and economic relevance will be explored. The process of empirical modelling will be by means of the data analysis and statistical package, Stata. In all cases, the relevance of various findings to decision makers will be discussed and explored.

Learning outcomes

By the end of this course, students will be able to

- understand critically a range of concepts and theories along with their relevance to management decision making in football.
- organise and manage data using the data analysis and statistical package, Stata.
- replicate and advance empirical models in the published literature.
- employ a range of diagnostics to assess the robustness of models.
- assess the statistical significance and economic relevance of findings and what they mean for decision takers in the football sector.

Programme of study

The benefit from the course, students should engage thoroughly prior to attending and during class. Prior to attendance, students should familiarise themselves with published articles across a range of topics in football. Students must read and familiarise themselves with a nominated article for the session. The nominated articles are denoted with an asterisk (*) in the schedule below. In addition to the nominated article, students

should also read at least one other article. The articles will form the basis of in-class discussion prior to engaging in any empirical modelling.

Day 1: Uncertainty of outcome and stadium attendance demand for football

Synopsis: The uncertainty of outcome hypothesis has been the mainstay of sports and by implication football since its inception by Rottenberg (1956). This seminal article has been cited numerous times and its relevance and contents have not been limited to citations by other academics but have also played important roles in administering sports leagues and judicial rulings in law courts. In this session, the uncertainty of outcome hypothesis is re-examined. The re-examination of the hypothesis will take place in the context of stadium attendance, however, its relevance will be scrutinised across numerous leagues, large and small. We consider the concept and the associated theories as well as the manner in which the concept could be captured using empirical data. Models of stadium attendance demand are developed in order to assess if there is any statistical or economic significant impact on stadium attendees of the concept.

Buraimo, B., Forrest, D., & Simmons, R. (2007). Outcome uncertainty measures: How closely do they predict a close game. *Statistical thinking in sports*, 167-178.

Buraimo, B., & Simmons, R. (2008). Do sports fans really value uncertainty of outcome? Evidence from the English Premier League. *International Journal of Sport Finance*, 3(3), 146.*

Pawlowski, T., Nalbantis, G., & Coates, D. (2018). Perceived game uncertainty, suspense and the demand for sport. *Economic Inquiry*, 56(1), 173-192.

Rottenberg, S. (1956). The baseball players' labor market. *Journal of political economy*, 64(3), 242-258.

Day 2: Television audience demand for football

Synopsis: Interesting in modelling stadium demand arguably belongs to a post-stadium attendance era when considering premier leagues. This is not to suggest that modelling and analysis stadium attendance is irrelevant since this is the main source of revenue for sub-premier football leagues. However, for premier leagues (across Europe), modelling television audience demand is far more important since revenue from broadcasters dominate other sources including stadium revenue. In this session, we consider not just the relevance of uncertainty of outcome but of other factors that affect television audience demand

Buraimo, B., & Simmons, R. (2009). A tale of two audiences: Spectators, television viewers and outcome uncertainty in Spanish football. *Journal of Economics and Business*, 61(4), 326-338.*

Feddersen, A., & Rott, A. (2011). Determinants of demand for televised live football: Features of the German national football team. *Journal of Sports Economics*, 12(3), 352-369.

Forrest, D., Simmons, R., & Buraimo, B. (2005). Outcome uncertainty and the couch potato audience. *Scottish Journal of Political Economy*, 52(4), 641-661.

Schreyer, D., Schmidt, S. L., & Torgler, B. (2016). Game Outcome Uncertainty and Television Audience Demand: New Evidence from German Football. *German Economic Review*.

Day 3: Superstars and television audience demand for football

Synopsis: The economics of superstars has been a feature of the literature and its relevance to sport and football is not surprising. As broadcasting and media proliferation shapes the football broadcast market, the contestation of football as and sporting pursuit or as pure entertainment has policy implications for the industry. In this session, we empirically test these premises: do television audiences prefer sporting contest, as proxied by uncertainty of outcome, or are their interests dominated by superstars?

Brandes, L., Franck, E., & Nüesch, S. (2008). Local heroes and superstars: An empirical analysis of star attraction in German soccer. *Journal of Sports Economics*, 9(3), 266-286.

Buraimo, B., & Simmons, R. (2015). Uncertainty of outcome or star quality? Television audience demand for English Premier League football. *International Journal of the Economics of Business*, 22(3), 449-469.*

Lucifora, C., & Simmons, R. (2003). Superstar effects in sport: Evidence from Italian soccer. *Journal of Sports Economics*, 4(1), 35-55.

Rosen, S. (1981). The economics of superstars. *The American economic review*, 71(5), 845-858.

Day 4: Modelling the impact of corruption on consumer demand for football

Synopsis: The biggest threat to football (and sport) is that of corruption and match-fixing. As the demand for fixes increases, the willingness to supply and the prices at which fixes are supplied provide some insight of the problem. Whilst there is much discussion in the literature on the economics of corruption in sport, and by implication football, very little empirical analyses of the effects of corruption on consumers feature in the literature. In this session, using Italian football as a setting and considering the corruption scandal of 2006, we model how corruption affects consumer demand for football.

Boeri, T., & Severgnini, B. (2011). Match rigging and the career concerns of referees. *Labour Economics*, 18(3), 349-359.

Buraimo, B., Migali, G., & Simmons, R. (2016). An analysis of consumer response to corruption: Italy's Calciopoli scandal. *Oxford Bulletin of Economics and Statistics*, 78(1), 22-41.*

Duggan, M., & Levitt, S. D. (2002). Winning isn't everything: Corruption in sumo wrestling. *American Economic Review*, 92(5), 1594-1605.

Preston, I., & Szymanski, S. (2003). Cheating in contests. *Oxford review of economic policy*, 19(4), 612-624.

Day 5: Impact of salaries on performance in football

Synopsis: In this session, attention switches to a labour market issue and focuses on the men in black (or yellow, red, green, blue...), the referee. The analysis is that of principal-agent theory. For some time, the referees in Europe's premier league were poorly remunerated considering the responsibilities bestowed on them. Whilst the literature is prevalent with the study of referee bias, the focus of this session is to model the performances of referees. The context for the empirical analysis is that of a natural experiment. Up until a specific moment in time, all referees in our setting were amateurs who were paid expenses for their efforts. After a period of time, a selection of referees was employed on a professional basis with the others maintaining their amateur status. In this session, we model the effects of payment to football referees on the performances.

Bryson, A., Buraimo, B., & Simmons, R. (2011). Do salaries improve worker performance?. *Labour Economics*, 18(4), 424-433.*

Dawson, P., & Dobson, S. (2010). The influence of social pressure and nationality on individual decisions: Evidence from the behaviour of referees. *Journal of Economic Psychology*, 31(2), 181-191.

Dohmen, T. J. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic Inquiry*, 46(3), 411-424.

Rocha, B., Sanches, F., Souza, I., & Carlos Domingos da Silva, J. (2013). Does monitoring affect corruption? Career concerns and home bias in football refereeing. *Applied Economics Letters*, 20(8), 728-731.

Other reading

Greene, W. H. (2018). *Econometric analysis*. Pearson Education.

Verbeek, M. (2017). *A guide to modern econometrics*. John Wiley & Sons

Teaching resources

In order to assist students, sample data sets will be provided.

Assessment

The assessment will require groups of students to develop empirical models.

1. Students will be assigned to groups.
2. Groups will be required to compile a data set. Students will be directed to relevant sources or will be provided with data to organise and manage.
3. Groups will be required to develop empirical models along with diagnostics to check their robustness and suitability.
4. Groups will be required to consider the policy implications of their findings.
5. Groups will be required to compile a 1,500 word study reporting their findings. The presentation of the studies should conform to the submission requirements of the journal, *Economics Letters*.