



Munich Personal RePEc Archive

Previsão da eficácia ofensiva do futebol profissional: Um caso Português

Jorge Caiado and Aníbal Vieira and Ana Bonito and Carlos Reis and Francisco Fernandes

2006

Online at <http://mpa.ub.uni-muenchen.de/2185/>
MPRA Paper No. 2185, posted 12. March 2007

Previsão da Eficácia Ofensiva do Futebol Profissional: Um Caso Português

Jorge Caiado^{a,b,*}, Aníbal Vieira^a, Ana Bonito^a, Carlos Reis^a e Francisco Fernandes^a

^a *Escola Superior de Ciências Empresariais, Instituto Politécnico de Setúbal, Campus do IPS, Estefanilha, 2914-503 Setúbal*

^b *Centro de Matemática Aplicada à Previsão e à Decisão Económica, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa*

Resumo

A previsão desempenha um papel importante no planeamento, tomada de decisão e controlo em qualquer domínio de actividade, incluindo o fenómeno desportivo do futebol. A experiência tem mostrado que os modelos extrapolativos ou não causais (modelos univariados), que se baseiam no conhecimento exclusivo dos seus valores passados para prever o futuro, são muitas vezes mais eficientes do que os modelos causais ou multivariados. Pretende-se com este artigo elaborar um exercício de modelação e previsão da eficácia ofensiva da equipa de futebol do Sport Lisboa e Benfica na Liga Portuguesa de Futebol. Para modelar e prever a série anual do número de golos, são utilizados métodos de previsão determinísticos (tendência linear, médias móveis, alisamento exponencial, holt, naïve) e estocásticos (modelos ARMA, passeio aleatório). No processo de selecção dos melhores modelos, são considerados as funções do erro quadrático médio, erro absoluto médio e erro percentual absoluto médio de previsão a um passo sobre as últimas três observações conhecidas da série observada.

Palavras chave: Alisamento exponencial, futebol, médias móveis, modelo ARMA, previsão.

Abstract

The forecast plays an important role in the planning, the decision-making and control in any domain of activity, including the sportive phenomenon of the soccer. The experience has shown that the extrapolative or not casual models (univariate models), that use only the information of its past values to forecast the future, can often predict future with more accuracy than causal or multivariate models. In this paper, we model and forecast the offensive effectiveness of the soccer team Sport Lisbon and Benfica, in Portuguese soccer league, by using deterministic methods (linear trend, moving average, exponential smoothing, holt, naïve) and stochastic models (ARMA models, random walk). The model selection criteria used in our study were the mean squared error, the mean absolute error and the mean absolute percentage error based in a one-step forecast of the last three observations.

Keywords: Exponential smoothing, soccer, moving average, ARMA model, forecast.

* Contacto por e-mail: jcaiado@esce.ips.pt (Jorge Caiado).

1 – Introdução

O desporto tem vindo a adquirir uma magnitude crescente na actualidade, onde se inclui o futebol, inquestionavelmente o desporto mais popular do mundo. Enquanto que o baseball, o basketbol, o futebol americano e o hóquei no gelo são considerados os maiores passatempos desportivos nos Estados Unidos, o futebol é o desporto nacional na maioria dos países da Europa, América Latina, África e na maior parte dos países da Ásia. Apesar da sua popularidade mundial, o futebol tem merecido pouca atenção por parte da comunidade científica quando comparada com a vasta literatura existente sobre os mais importantes desportos americanos.

Dos poucos estudos científicos sobre futebol, destacam-se alguns ligados ao futebol europeu. Hart, Hutton e Sharot (1975) estudaram a assistência de espectadores aos jogos de 4 equipas de topo da liga inglesa de futebol entre 1969 e 1972, e em particular, analisaram as variações no número de espectadores em função de factores geográficos, demográficos e da atractividade dos jogos em si. Num outro trabalho, Bird (1982) usou dados de séries temporais entre 1948 e 1975 para estimar a assistência de espectadores no futebol inglês com base no preço dos bilhetes (incluindo os custos de viagem), o rendimento nacional, e outras variáveis ligadas ao sucesso da Inglaterra no campeonato do mundo, ao estado do tempo, ao golos marcados e ao fenómeno do hooliganismo.

Dobson e Goddard (1992) analisaram a procura de bilhetes de lugares sentados e lugares em pé com base numa amostra de 795 jogos da liga inglesa nas épocas de 1989-1990 e 1990-1991. Concluíram que o número de espectadores que assistem aos jogos de pé aumenta significativamente nos jogos entre as equipas rivais e nos jogos entre as equipas que tiveram bom desempenho em jogos recentes, ao passo que as vendas de bilhetes para lugares sentados dependem mais da performance das equipas ao longo das épocas passadas.

Num estudo sobre o desempenho desportivo das equipas de futebol da liga italiana, Lucifora e Simmons (2003) verificaram que o número de golos, o número de minutos e as convocatórias para selecção nacional são factores preditivos do aumento do nível salarial do jogadores. Na liga inglesa, como os salários individuais dos seus jogadores são geralmente desconhecidos para o público, Szymanski e Smith (1997) basearam-se no orçamento total da equipa para medir a sua performance desportiva, e

concluíram que existe uma forte correlação entre a percentagem de vitórias e o orçamento da equipa.

Anderson, Ekman e Edman (2003) analisaram a performance de previsões feitas por especialistas e não especialistas em futebol. Para o efeito, questionaram 250 indivíduos com diferentes níveis de conhecimentos sobre futebol acerca dos seus prognósticos para os resultados da primeira ronda do campeonato do mundo de futebol de 2002. Das verificações empíricas obtidas, estes autores concluíram que os especialistas em futebol (jornalistas, adeptos e treinadores de futebol) não conseguem obter melhores previsões do que os não especialistas, que se baseiam em juízos subjectivos face à ausência de informação.

A modelação e previsão de resultados desportivos e a sua aplicação no desenvolvimento de estratégias de apostas tem surgido bastante na literatura científica. São exemplos os estudos sobre a liga de futebol americano (Craig e Hall 1994, Gray e Gray 1997, Glickman e Stern 1998, Vergin 2001), o futebol inglês (Lee 1997, Crowder, Dixon, Ledford e Robinson, 2002), o basebol nos Estados Unidos (Albert 1994, Gandar, Zuber e Lamb 2001, Schall e Smith, 2000), e o futebol australiano (Stefani e Clarke 1992, Clarke 1993 e Brailsford, Easton, Gray e Gray, 1995).

O papel da previsão através de julgamentos e especulações de analistas baseados nas suas experiências e conhecimentos passados tem sido extensivamente investigado. Alguns estudos têm mostrado que os analistas experientes sobrestimam a sua habilidade para prever com precisão devido ao seu excesso de confiança (Ayton 1992). Para além disso, existem enviesamentos e limitações que caracterizam o julgamento humano quando se pretende minimizar as consequências negativas da previsão (Caiado 2002).

Este artigo tem como objectivo central a modelação e previsão da série temporal do número de golos marcados pelo maior e mais popular clube português de futebol profissional - o Sport Lisboa e Benfica. Para o estudo empírico, foram utilizados métodos de previsão determinísticos (tendência linear, médias móveis, alisamento exponencial, naïve) e estocásticos (modelo ARMA de Box-Jenkins, modelo de passeio aleatório). No processo de escolha dos melhores modelos, tomou-se em consideração os erros absolutos de previsão a um passo sobre as últimas três observações conhecidas da série observada.

2 - Metodologia de Análise

A previsão desempenha um papel importante no planeamento, tomada de decisão e controlo em qualquer domínio de actividade, incluindo o fenómeno desportivo do futebol. Uma previsão consiste na tentativa de extrapolar o comportamento futuro a partir das condições actuais, utilizando dados históricos do passado. Neste processo, o analista pressupõe que as condições no presente, mesmo as que se ignoram, se vão manter inalteradas no futuro. A experiência tem mostrado que os modelos extrapolativos ou não causais (modelos univariados), que se baseiam no conhecimento exclusivo dos seus valores passados para prever o futuro, são muitas vezes mais eficientes do que os modelos causais ou multivariados.

Em seguida faz-se uma breve exposição dos métodos determinísticos e estocásticos de previsão utilizados no estudo empírico.

a) Modelos de tendência linear

Estes modelos pressupõe um comportamento linear de tendência da série. Quando numa série temporal se não detectam movimentos de carácter oscilatório nem movimentos sazonais e se admite a existência de variação não explicada, pode adoptar-se o seguinte modelo determinístico para a tendência,

$$Y_t = b_0 + b_1t + \varepsilon_t,$$

que descreve a série Y_t como função da variável tempo (t) e da componente aleatória ε_t que representa aquilo que não consegue explicar-se. A estimação dos parâmetros b_0 e b_1 é feita pelo método dos mínimos quadrados.

b) Modelos de médias móveis

Estes métodos são adequados para avaliar a tendência de uma série e são muito sensíveis às alterações recentes do seu comportamento. Utilizam a informação relativa ao termos da série numa vizinhança de alguns períodos para estimar o seu nível corrente. Para séries com tendência localmente estacionária, o método de médias móveis simples obtém previsões constantes para todo o horizonte da previsão,

$$\hat{Y}_{T+m} = \frac{1}{k} \sum_{t=T-k+1}^T Y_t = \frac{Y_T + Y_{T-1} + \dots + Y_{T-k+1}}{k}, \quad m = 1, 2, \dots,$$

onde \hat{Y}_{T+m} é a previsão da série Y no momento $T+m$ (com T a origem da previsão e m o horizonte da previsão) e $Y_T, Y_{T-1}, \dots, Y_{T-k+1}$ são as k observações incluídas na média móvel. Outros métodos de médias móveis atribuem uma maior importância à informação passada mais recente relativamente à mais antiga.

c) Modelos de alisamento exponencial

Este métodos de previsão utilizam ponderadores exponencialmente decrescentes com a antiguidade das observações. São de vários tipos os modelos de alisamento exponencial:

- Modelos de *alisamento exponencial simples* (séries com tendência localmente estacionária e sem movimentos sazonais),

$$M_t = \alpha Y_t + (1 - \alpha)M_{t-1}, \text{ com } 0 < \alpha < 1,$$

onde M_t é a estimativa do nível da série no momento t e α é a constante de alisamento.

- Modelos de *alisamento exponencial duplo* (séries com tendência linear e sem movimentos sazonais),

$$M_t = \alpha Y_t + (1 - \alpha)M_{t-1},$$

$$D_t = \alpha M_t + (1 - \alpha)D_{t-1},$$

$$\hat{Y}_{T+m} = \hat{a}(T) + \hat{b}(T)m, \quad m = 1, 2, \dots,$$

onde M_t é a série de alisamento exponencial simples, D_t é a série de alisamento exponencial duplo, $a(t) = 2M_t - D_t$ e $b(t) = [\alpha/(1 - \alpha)](M_t - D_t)$ são as estimativas do nível e do declive da série, respectivamente e \hat{Y}_{T+m} a função de previsão a m passos.

- Modelo de *Holt-Winters sem sazonalidade* (séries com tendência linear e sem movimentos sazonais),

$$a(t) = \alpha Y_t + (1 - \alpha)[a(t-1) + b(t-1)], \quad 0 < \alpha < 1,$$

$$b(t) = \beta[a(t) - a(t-1)] + (1 - \beta)b(t-1), \quad 0 < \beta < 1,$$

$$\hat{Y}_{T+m} = \hat{a}(T) + \hat{b}(T)m, \quad m = 1, 2, \dots,$$

onde $a(t)$ e $b(t)$ são as equações de actualização do nível e do declive de tendência da série, respectivamente e \hat{Y}_{T+m} a função de previsão a m passos. Para mais detalhes veja-se Makridakis, Wheelwright e Hyndman (1998).

d) Previsões naïve

Nos mercados onde a informação está muito dispersa e que são influenciados por um elevado número de intervenientes, as previsões dos analistas feitas para o dia, mês, trimestre ou ano seguinte correspondem aos valores observados no actual dia, mês, trimestre ou ano, $Y_t = Y_{t-1}$. São por isso chamadas previsões *naïve*.

e) Modelo de passeio aleatório

No modelo de passeio aleatório (*random walk*), o valor da série no momento t é igual ao seu valor passado mais um choque aleatório (geralmente conhecido por ruído branco),

$$Y_t = Y_{t-1} + \varepsilon_t .$$

Este modelo é muito utilizado na modelização e previsão de séries económicas e financeiras não estacionárias. Os “passeios aleatórios” caracterizam-se por movimentos de tendência crescente ou decrescente em períodos longos seguidos de mudanças bruscas de sentido imprevisíveis.

f) Modelos ARMA

Modelos lineares introduzidos por Box e Jenkins (1970) que permitem modelar e prever séries cronológicas estacionárias e não estacionárias através de uma representação autoregressiva e de médias móveis (ARMA). Sob certas condições de estacionaridade e invertibilidade (veja-se Box e Jenkins 1970), o modelo genérico ARMA tem a representação,

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} .$$

Este modelo descreve a série Y_t como função dos seus valores passados e como combinação linear de uma sucessão de choques aleatórios. O processo integrado autoregressivo e médias móveis ou ARIMA(p,d,q) assume para o modelo comum ARIMA(1,1,1) a forma,

$$Z_t = \phi_1 Z_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \text{ onde } Z_t = Y_t - Y_{t-1} .$$

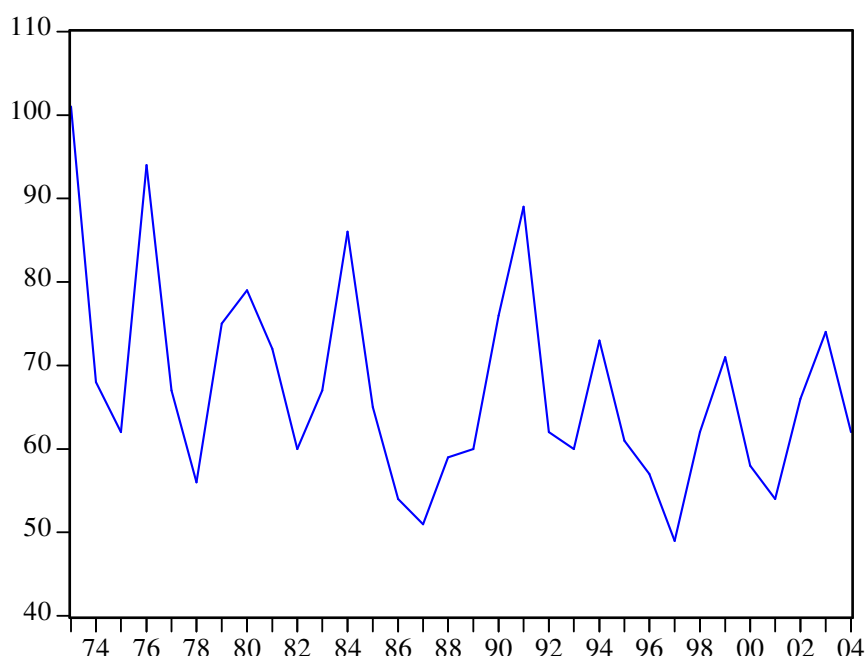
Box e Jenkins (1970) propuseram uma metodologia de modelização ARIMA baseada nas etapas de identificação, estimação, avaliação do diagnóstico, selecção de modelos e previsão, que é actualmente ainda muito utilizada por especialistas em séries temporais e métodos de previsão. Desenvolvimentos mais recentes podem ver-se em Box, Jenkins e Reinsel (1994) e Wei (1990).

3 - Estudo Empírico

A série em estudo refere-se ao número de golos marcados pelo Sport Lisboa e Benfica na Liga Portuguesa de Futebol, nas últimas 32 épocas de 1972/73 a 2003/04 (fonte: www.zerozero.pt). Na Figura 1 pode observar-se o cronograma da série. Esta parece apresentar uma tendência ligeiramente decrescente ao longo do período em análise, acompanhada de oscilações de carácter aleatório.

Figura 1

Série do número de golos marcados pelo Sport Lisboa e Benfica entre 1972/73 e 2003/04



A escolha do melhor modelo de previsão será determinada com base no modelo que produzir menores erros de previsão a um passo sobre as últimas 3 observações conhecidas do número de golos marcados. Os critérios dos erros de previsão utilizados são os seguintes: Erro quadrático médio, $EQM = (1/m) \sum_{t=1}^m e_t^2$ (onde $e_t = Y_t - P_t$ é o desvio entre o valor observado e o valor previsto da série no momento t); Erro absoluto médio, $EAM = (1/m) \sum_{t=1}^m |e_t|$; Erro percentual absoluto médio, $EPAM = (1/m) \sum_{t=1}^m |e_t/Y_t| \times 100$.

No Quadro 1 apresentam-se as previsões a um passo à frente do número de golos marcados nas épocas de 2001/02, 2002/03 e 2003/04 e os respectivos valores do erro quadrático médio, erro absoluto médio e erro percentual absoluto médio. Na última linha do Quadro 1 encontram-se as previsões *out-of-sample* para a época de 2004/05 com base nos diferentes modelos de previsão. Os Quadros 2 e 3 mostram os erros de previsão absolutos e percentuais absolutos.

Quadro 1

Erros de previsão do número de golos do Benfica entre 2001/02 e 2003/04

Épocas	Golos	Modelos determinísticos									Modelos estocásticos		
		MTL	MM2	MM3	MM4	MM8	AES	AED	HOLT	NAIVE	RW	ARIMA (1,1,2)	ARIMA (0,1,2)
01/02	66	57	56	61	61	61	64	56	56	54	52	57	61
02/03	74	58	60	59	62	60	64	58	59	66	63	74	70
03/04	62	59	70	65	63	61	65	60	65	74	71	62	69
EQM		115,3	120,0	86,3	56,7	74,0	37,7	120,0	111,3	117,3	132,7	27,0	30,0
EAM		9,3	10,7	7,7	6,0	6,7	5,0	9,3	9,3	10,7	11,3	3,0	5,3
EPAM		13,4	15,7	10,9	8,5	9,4	7,1	13,3	13,4	16,1	16,9	4,5	8,1
Previsão 04/05		59	68	67	64	62	65	60	63	62	59	66	61

Notas: MTL é o modelos de tendência linear; MM2, MM3, MM4 e MM8 são os modelo de médias móveis de ordens 2, 3, 4 e 8, respectivamente; AES e AED são os modelo de alisamento exponencial simples e duplo com pesquisa óptima das constantes de alisamento α pelo software Eviews 5; HOLT é o modelo de Holt-Winters sem sazonalidade com pesquisa óptima das constantes de alisamento α e β ; NAIVE é o modelo de previsão naïve; RW é o modelo de passeio aleatório (*random walk*); ARIMA(1,1,2) é um modelo integrado autoregressivo de ordem 1 e de média móveis de ordem 2 com a restrição $\theta_2=0$; ARIMA(0,1,2) é um modelo integrado de médias móveis de ordem 2. Assinala-se a sombreado os menores erros de previsão.

Quadro 2

Erros de previsão absolutos do número de golos do Benfica entre 2001/02 e 2003/04

Épocas	Modelos determinísticos									Modelos estocásticos		
	MTL	MM2	MM3	MM4	MM8	AES	AED	HOLT	NAIVE	RW	ARIMA (1,1,2)	ARIMA (0,1,2)
2001/02	9	10	5	5	5	2	10	10	12	14	9	5
2002/03	16	14	15	12	14	10	16	15	8	11	0	4
2003/04	3	8	3	1	1	3	2	3	12	9	0	7

Quadro 3

Erros de previsão percentuais do número de golos do Benfica entre 2001/02 e 2003/04

Épocas	Modelos determinísticos								Modelos estocásticos			
	MTL	MM2	MM3	MM4	MM8	AES	AED	HOLT	NAIVE	RW	ARIMA (1,1,2)	ARIMA (0,1,2)
2001/02	13,6	15,2	7,6	7,6	7,6	3,0	15,2	15,2	18,2	21,2	13,6	7,6
2002/03	21,6	18,9	20,3	16,2	18,9	13,5	21,6	20,3	10,8	14,9	0,0	5,4
2003/04	4,8	12,9	4,8	1,6	1,6	4,8	3,2	4,8	19,4	14,5	0,0	11,3

O modelo ARIMA(1,1,2) com uma parâmetro autoregressivo de ordm 1 e um parâmetro de médias móveis de ordem 2 e uma diferenciação simples revelou-se o melhor método para prever a performance ofensiva do futebol do Benfica, a avaliar pelas funções EQM (27,0), EAM (3,0) e EPAM (4,5%) sobre as últimas três observações disponíveis. Em contraste, o modelo que apresenta piores previsões foi o modelo de passeio aleatório.

Dos modelos determinísticos utilizados na previsão, destaca-se o modelo de alisamento exponencial simples com um EQM de 37,7, um EAM de 5,0 e um EPAM de 7,1%, que conseguiu obter a melhor previsão para a época de 2001/02 (erro absoluto de 2 golos).

4 – Conclusões

Muitos analistas e investigadores na área do desporto têm a preocupação de modelar e prever o comportamento futuro de alguns importantes e imprevisíveis resultados desportivos. O fenómeno do futebol tem merecido um enorme e crescente interesse não só por parte dos agentes desportivos e adeptos como também por investidores institucionais e individuais.

Neste artigo forma utilizados métodos de previsão determinísticos e estocásticos para prever a eficácia ofensiva do mais popular clube português – o Sport Lisboa e Benfica. Os resultdos empíricos sugerem que os modelos autoregressivos e de médias móveis (ARMA) de Box-Jenkins são aqueles que conseguem acompanhar melhor as oscilações da série do número de golos marcados e produzir melhores previsões. Os modelos de alisamento exponencial simples e de médias móveis de ordem 4 revelaram também boas qualidades preditivas.

5 – Referências

- Albert, J. (1994), “Exploring baseball hitting data: What about those breakdown statistics?”, *Journal of the American Statistical Association*, 89, 1066-1074.
- Anderson, P., Ekman, M. e Edman, J. (2003), “Forecasting the fast and frugal way: A study of performance and information-processing strategies of experts and non-experts when predicting the World Cup 2002 in soccer”, Working Paper Series in Business Administration No 2003:9.
- Ayton, P. (1992). “On the competence and incompetence of experts”. In: F. Bolger e G. Wright (Eds.), *Expertise and Decision Support*, Plenum Press, 77-105.
- Bird, P. (1982), “The demand for league football”, *Applied Economics*, 14, 637-649.
- Box, G. E. P. e Jenkins, G. (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- Box, G. E. P., Jenkins, G. e Reinsel, G. C. (1994), *Time Series Analysis, Forecasting and Control*, 3.^a ed., Prentice-Hall, New York.
- Brailsford, T. J., Easton, S. A., Gray, P. K. e Gray, S. F. (1995), “The efficiency of Australian football betting markets”, *Australian Journal of Management*, 20, 167-195.
- Caiado, J. (2002), “Métodos de previsão em economia e gestão: desenvolvimentos recentes”, *Inforbanca*, XIV, n.º 53, 18-19 e 44-47.
- Clarke, S. R. (1993), “Computer forecasting of Australian rules football for a daily newspaper”, *Journal of the Operational Research Society*, 44, 753-759.
- Craig, L. A. e Hall, A. R. (1994), “Trying out for the team: Do exhibitions matter? Evidence from the National Football League”, *Journal of the American Statistical Association*, 89, 1091-1099.
- Crowder, M., Dixon, M., Ledford, A. e Robinson, M. (2002), “Dynamic modeling and prediction of English Football League matches for betting”, *Journal of the Royal Statistical Society Series D (The Statistician)*, 51, 157-168.

- Dobson, S. e Goddard, J. (1992), “The demand for standing and seated viewing accommodation in the English Football League”, *Applied Economics*, 24, 1155-1163.
- Gandar, J. M., Zuber, R. A. e Lamb, R. P. (2001), “The home field advantage revisited: a search for the bias in other sports betting markets”, *Journal of Economics and Business*, 53, 439-453.
- Glickman, M. E. e Stern, H. S. (1998), “A state-space model for National Football League scores”, *Journal of the American Statistical Association*, 93, 25-35.
- Gray, P. K. e Gray, S. F. (1997), “Testing market efficiency: Evidence from the NFL sports betting market”, *Journal of Finance*, 52, 1725-1737.
- Hart, S., Hutton, J. e Sharot, T. (1975), “A statistical analysis of association football attendances”, *Journal of Royal Statistical Society Series C*, 24, 17-27.
- Lee, A. J. (1997), “Modelling scores in the premier league: Is Manchester United really the best?”, *Chance*, 10, 15-19.
- Lucifora, C. e Simmons, R. (2003), “Superstar effects in sport: Evidence from Italian soccer”, *Journal of Sports Economics*, 3, 169-203.
- Makridakis, S., Wheelwright, S. e Hyndman, R. (1998), *Forecasting: Methods and Applications*, 3.^a ed., John Wiley & Sons, New York.
- Schall, T. e Smith, G. (2000), “Do baseball players regress toward the mean?”, *American Statistician*, 54, 231-235.
- Stefani, R. T. e Clarke, S. R. (1992), “Predictions and home advantage for Australian Rules football”, *Journal of Applied Statistics*, 9, 251-261.
- Szymanski, S. e Smith, R. (1997), “The English football industry: performance, profit and industrial structure”, *International Review of Applied Economics*, 11, 135-154.
- Vergin, R. C. (2001), “Overreaction in the NFL point spread market”, *Applied Financial Economics*, 11, 497-509.
- Wei, W. W. S. (1990), *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley Publishing Company, New York.