
Is soccer dying? A time series approach

ALEJANDRO GAVIRIA

Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, USA

E-mail: agaviria@weber.ucsd.edu

Received 20 May 1998

A Markov switching regime model is used to investigate structural breaks in the average goals per game in the premier soccer leagues of England, Italy and Spain. The analysis reveals a structural break around 1965 in the three countries.

I. INTRODUCTION

On Sunday, 17 July 1994, the 95 000 spectators gathered in Pasadena, California to watch the final match of the World Cup witnessed an unprecedented event: for the first time ever a World Cup final went scoreless and the soccer world champion had to be decided by penalty kicks. Four years before in Rome, Italy, things had not been very different; the World Cup final was decided then by a lone penalty kick towards the end of a dreary game. Bluntly, not a single field goal has been scored in the two top soccer matches of this decade.

No wonder some sports commentators hastened to foretell the end of soccer (see, for example, Cantor 1996 and Gardner 1996).¹ They argued that the balance between offense and defence – a cornerstone of most sport competitions – had been lost and that defensive tactics were so widespread that soccer had lost much of its appeal. The International Soccer Federation (FIFA) did not turn a deaf ear this time around. Breaking with its proverbial conservatism, FIFA introduced the most drastic modifications to the soccer rules in 70 years; all of them aimed to give offensive soccer a much-needed edge.

A simple Markov switching model in the spirit of Hamilton is used here (1989) to study changes in the annual average goals per game in the premier soccer leagues of England, Italy and Spain. The analysis shows that soccer has indeed become more defensive-minded over time – at least in the narrow sense of fewer goals scored per

game. This change, however, did not happen during the 1990s or even the 1980s. The available evidence unambiguously indicates that average goals per game fell dramatically – and almost coincidentally in the three countries – during the 1960s. No big changes in this variable seem to have occurred ever since.

II. DATA

In this paper the Historical Domestic Results chapter of the Soccer Statistics Foundation (RSSSF) archive is used to compute the annual average goals per game in the premier soccer leagues of England, Italy and Spain.² These three leagues have a long – albeit discontinuous – history dating back to the beginning of the century. Both the English and Italian leagues were suspended during World War II – the English from 1940 to 1946 and the Italian in 1944 and 1945. The Spanish league, for its part, was suspended from 1936 to 1939 during the Spanish Civil War. The post war years are focused upon so as to avoid the ‘war’ moratoriums.

All three leagues have a similar structure; round robin tournaments comprising around 20 teams. Every year the worst teams are relegated and replaced by the best teams of the second tier league and hence the roster of teams changes from one year to the next. The number of teams have also changed slightly over time. The Spanish league gradually increased the number of teams from 16 in the 1940s to 22 in 1996, and the Italian league decreased the

¹ Gardner's (1996, p. 223) reaction to the scoreless final of World Cup '94 is illustrative: 'There was nothing new here. Just caution. More of a chess game than a soccer game, ending in stalemate. A dreadful anticlimax. A tactical imposed 0–0 tie.'

² The raw data can be downloaded at <http://www.risc.uni-linz.ac.at/non-official/rsssf/intclub.html>.

number of teams from 18 to 16 during the 1960s and 1970s and increased it back to 18 during the 1990s. The English league, for its part, has included 22 teams for most of the post war years with a few exceptions in the late 1980s.

Ideally, one should control for changes in the structure of the tournaments. For example, a rise in the number of teams may bring in a few lousy teams, which in turn may cause an increase the average goals per game.³ This observation notwithstanding, few would argue that one can explain the stark decline in the average goals per game described below by appealing to small changes in the regulations and the mechanics of the tournaments (more on this later).

III. EMPIRICAL MODEL AND ESTIMATION

We assume first that the annual average goals per game in each league (y_t) are generated by the following state-dependent process in the spirit of Hamilton (1992):

$$y_t = \mu_{s_t} + \varepsilon_t \quad (1)$$

where μ is the mean of y_t conditional on the state s_t , and $\varepsilon_t \sim$ i.i.d. $N(0, \sigma_{s_t}^2)$. We also assume that there are only two states of nature ($s_t = 1$ and $s_t = 2$), and that s_t is the unobserved outcome of a Markov chain with transitional probabilities given by

$$P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix} \quad (2)$$

Finally, we assume that s_t and ε_t (the two sources of randomness in the model) are uncorrelated for all t and τ .

The process described above is completely characterized by the vector α of population parameters, where $\alpha = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p_{11}, p_{22})$. Here I estimate α by Maximum Likelihood using the algorithms developed by Hamilton (1994) and Kim (1994).⁴ The estimation algorithm has three steps. First, given an initial guess for α , we compute the conditional probabilities $\xi_{t/t}$, where $\xi_{t/t}$ is defined as the probability that observation t th was generated by state 1 conditional on the information available through date t . Then, we use the conditional probabilities to compute the smoothed probabilities $\xi_{t/T}$ – defined as before but now conditional on the whole sample. Last, we use the smoothed probabilities to update α . As usual, we repeat these three steps until some convergence is achieved.

The estimation algorithm yields two distinct outputs: the population parameters characterizing each state of nature

and the smoothed probabilities showing our best inference about the value of the state of nature at each t . With this information we can compute the ‘estimated’ values of y_t at each date according to

$$\hat{y}_t = \xi_{t/T} \hat{\mu}_1 + (1 - \xi_{t/T}) \hat{\mu}_2 \quad (3)$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the estimates of the mean goals per game in states 1 and 2, and $\xi_{t/T}$ is the smoothed probability at t . As discussed below, we can use the evolution of \hat{y}_t over time to identify the presence of regime shifts in y_t .

IV. RESULTS

Figure 1 displays the yearly average goals per game along with the corresponding estimated values (\hat{y}_t) for the English, Italian and Spanish leagues. A first fact is evident: all three leagues experienced a shift from a high scoring to a low scoring state during the 1960s.⁵ The shift in the English league lagged the shifts in the Italian and Spanish leagues by approximately four years. As shown, the mean of \hat{y}_t decreased by more than 0.7 goals per game in the English and Italian leagues and by more than 1.1 goals per game in the Spanish league. The variance decreased even more dramatically – more than 60% in general and as much as 85% in the English league. Soccer no doubt became more stingy and more predictable during the prodigal and unruly 1960s.

The estimates of the transition probabilities suggests that movements across states are rare and often irreversible. Indeed, for the English and Spanish leagues the low scoring state is an absorbing state (i.e., $\hat{p}_{22} = 1$). For the Italian league, on the other hand, there is evidence of a recent transition back to the pre-war scoring averages, which surely reflects the influx of scoring talent that took place in Italy during most of the 1990s.

V. DISCUSSION

A question is in order now: can we identify the forces underlying the shift from high to low scoring described above? First, we know that the regime shift was not brought about by a modification in the rules; such a thing just did not happen during the 1960s. Second, explanations cast in terms of coordination problems are also dubious. We may argue, for example, that the regime shift was caused by a jump from a high-scoring to a low-scoring equilibrium. The presence of multiple equilibria, however, needs the unpalatable assumption of strategic

³ Baseball is a case in point. According to Lockwood (1998), ‘If baseball history tells us anything, it is that expansion seasons produce off-the-chart numbers.’ Indeed, ‘baseball has expanded five times up to 1998 – 1961, 1962, 1969, 1969, 1977 and 1993 – and every time except 1962 home run totals increased dramatically. In 1977, they jumped 50%.’ The reason: the overall dilution of pitching talent.

⁴ These two algorithms are special cases of the EM principle developed by Dempster *et al.* (1977). See Hamilton (1994) for details.

⁵ Appendix 1 shows that average goals per game in the World Cup has followed a similar pattern.

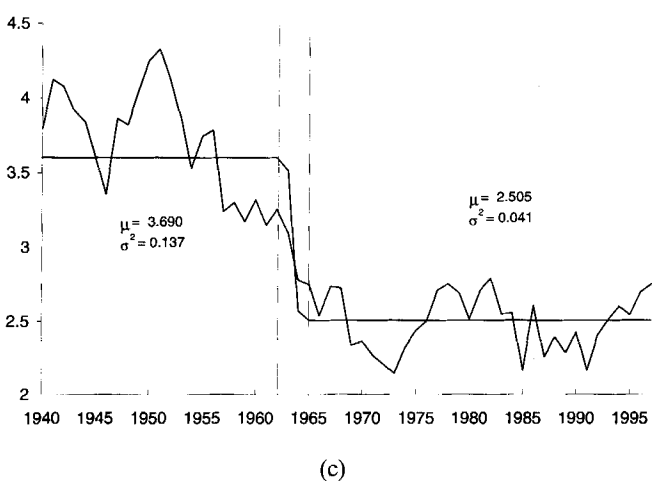
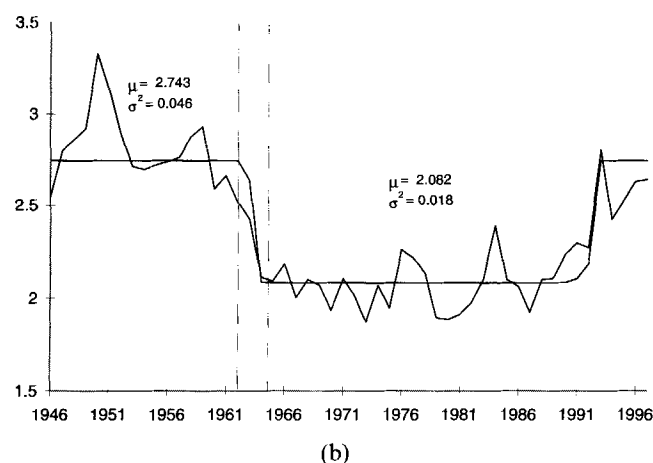
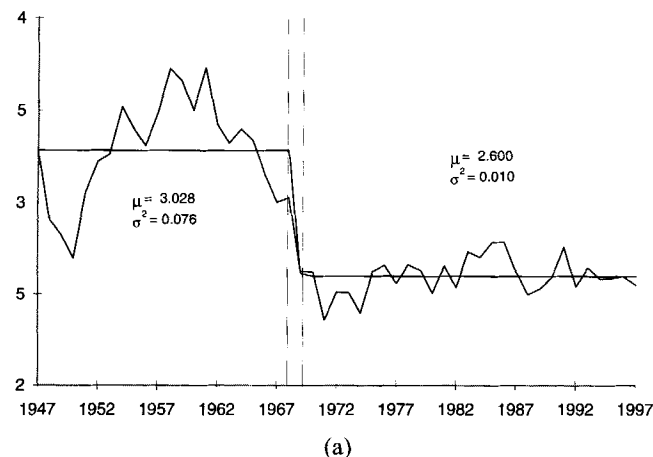


Fig. 1. Goals per game (annual averages): (a) England 1947–97; (b) Italy 1946–97; (c) Spain 1940–97

complementarities (i.e., offensive play is optimal when our opponent plays offensively, and defensive play is optimal otherwise).

The regime shift, I think, was mainly driven by a few tactical innovations (and their subsequent diffusion). In the early 1960s, the Inter Milan broke new ground by winning twice the prestigious European Cup while using an innovative defensive system. This system, known as *catenaccio* and devised by Inter's coach Helenio Herrera, used at least five defenders and only one striker. Although the *catenaccio*'s formation was not widely adopted outside Italy, most of its main tactics were widely emulated all over the world. As a result, the off-side trap, the *libero* ('sweeper') and the man-to-man marking became the bread and butter of soccer everywhere.

The new defensive tactics proved to be quite successful. England, for example, won its first World Cup ever in 1966 by relying mainly on defensive tactics to counter its more talented opponents (Rote, 1978). Likewise, two mediocre Argentinean teams, Racing Club and Estudiantes de la Plata, won the World Club Championship in the late 1960s by relying on a combination of aggressive tackling and ultra-defensive play (Rote, 1978). In short, the new tactics made it possible to win it all with limited talent. This brought about a true revolution to soccer; suddenly, it was not the players who won games. That was the coach's job.

Needless to say, the emergence of the new tactics elevated soccer coaches to a pre-eminent position. Suddenly, they came to be perceived as the key to success. Of course, they also came to be perceived as the guilty party when success was elusive. This put unprecedented pressure on the coaches. Now more than ever, they were afraid of defeat. This fear made them more conservative, which in soccer means more defensive-minded. Gardner (1996) puts it plainly, 'the orthodox view was that coaches who concentrated on building strong defenses were the ones who won titles, and the ones who kept their jobs'.

We can now tell a coherent story of how the defensive tactics got locked in. The story is simple: the new emphasis on tactics gave coaches a prominent role, and coaches in prominent role had no choice but to emphasize tactics. The vicious circle was then complete. Suddenly, soccer became more like chess and less like ballet. Moreover, the defensive trend was reinforced by the consolidation of the top continental cups (the European Cup and the Libertadores Cup in South America).⁶ Arguably, winning at any cost became then more important than ever because only winners have access to the continental cups and their large bounties.⁷

The answer to the question posed in the title should be obvious by now. Soccer is not dying. Either it has been

⁶ The European Champions Cup began in 1956, the UEFA cup (formerly called the Fairs Cup) began in 1958, and the Libertadores Cup began in 1960.

⁷ The continental cups may have also played a crucial role by speeding up the diffusion of new tactics across national borders

dead for a while or it is well alive. It is, of course, entirely up to the eye of the beholder to provide a definitive answer.

ACKNOWLEDGEMENTS

The Alfred P. Sloan foundation provided generous financial support. Clive Granger, Pierre A. Chiappori and Steve Raphael provided helpful suggestions.

REFERENCES

- Cantor, A. (1996) *Goal: A Celebration of Soccer*. Simon & Schuster, New York.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society Series B*, **39**, 1–38.
- Gardner, P. (1996) *The Simplest Game*, Macmillan, New York.
- Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–84.
- Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.
- Kim, C. (1994) Dynamic linear models with Markov-switching, *Journal of Econometrics*, **60**, 1–22.

Lockwood, W. (1998) Expansionist theory: Einstein know old records in serious danger of blowing up, *San Diego Union Tribune*, 6 April.

Rote Jr., K. (1978) *Kyle Rote Jr.'s Complete Book of Soccer*, Simon & Schuster, New York.

APPENDIX: AVERAGE GOALS PER GAME IN THE WORLD CUP TOURNAMENTS (1930–1994)

1930	3.89
1934	4.11
1938	4.66
1950	4.00
1954	5.38
1958	3.60
1962	2.78
1966	2.78
1970	2.96
1974	2.55
1978	2.68
1982	2.80
1986	2.54
1990	2.21
1994	2.71
