Score-based rating systems in sports statistics

Michal Černý and Vladimír Holý

Department of Econometrics
Faculty of Informatics and Statistics
Prague University of Economics and Business
Winston Churchill Square 4
CZ13067 Prague, Czech Republic
cernym@vse.cz

Abstract

In many sports disciplines, players are assigned a rating that is defined as a certain function of the results of matches the player has played in the past. When a player plays another match, it is assumed that the increase/decrease of their rating after a win/loss depends on the *difference between their rating and that of the opponent*. The Elo rating system (Elo rating, ER) is defined as follows: when two players A, B with ratings $r_t^{(A)}, r_t^{(B)}$ in time t play a game with outcome $Y_t^{(A)} \in \{0, 1\}$ (here 1 stands for the win of A), their ratings are updated as $r_{t+1}^{(A)} = r_t^{(A)} + k(Y_t^{(A)} - \Lambda(r_t^{(A)} - r_t^{(B)}))$, $r_{t+1}^{(B)} = r_t^{(B)} + k((1 - Y_t^{(A)}) - \Lambda(r_t^{(B)} - r_t^{(A)}))$, where $k > 0$ is a constant and Λ is the logistic function. If the game outcomes $Y_t^{(A)}$ are treated as random variables, ER generates sequences of players' ratings which can be viewed as time series.

First we will derive some basic statistical properties. For example, we will show what happens if the true probability of winning differs from the probability implied by the player's rating. It turns out that there exists a certain weak form of mean reversion. We also illustrate some undesirable properties, in particular a certain form of asymptotic inconsistency, which in the long run unfairly favors strong players and penalizes the weak.

We further show that ER with the logistic link function is equivalent to a special case of a GAS (Generalized Autoregressive Score) time-series model containing the autoregression term and the so-called score term (here, *score* stands for the gradient of log-likelihood of the outcome generating distribution). This result has interesting consequences: it makes it possible to construct generalized ER systems even for games where the match outcome may be more complex than simple 0-1 win/loss information. For example, these generalizations, called *score-based rating systems*, include outcomes such as "winning by five goals," "finishing second in a race of six runners," "finishing second with a five-second gap behind the winner," or a match whose outcome is a general permutation of the set of players.

References

[1] D. Creal, S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications, *Journal of Applied Econometrics* **28**(5), 777–795.

[2] V. Holý and J. Zouhar (2022). Modelling time-varying rankings with autoregressive and score-driven dynamics, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **71**(5), 1427–1450.

[3] M. Černý and V. Holý (2025). Score-based ranking systems in sports, Submitted to: *The American Statistician*.