# QUANTIFYING THE EVOLUTION OF FIRST-CLASS RUGBY IN NEW ZEALAND

PRESENTED BY ANKIT PATEL
DATA SCIENTIST, DOT LOVES DATA
PHD. CANDIDATE, VICTORIA UNIVERSITY OF WELLINGTON

# Structural changes to a competition affects the competitiveness of the competition

**DOT Sport** delivers innovative, relatable, data-driven insights to create winning strategies.

Apply cutting edge machine learning to minimise human bias and maximise use of big data.

We have helped a range of sporting organisations win with data over the last 20 years ranging from: *National Governing Bodies*, *Sport Franchises*, to *Academic* and *Commercial* entities.

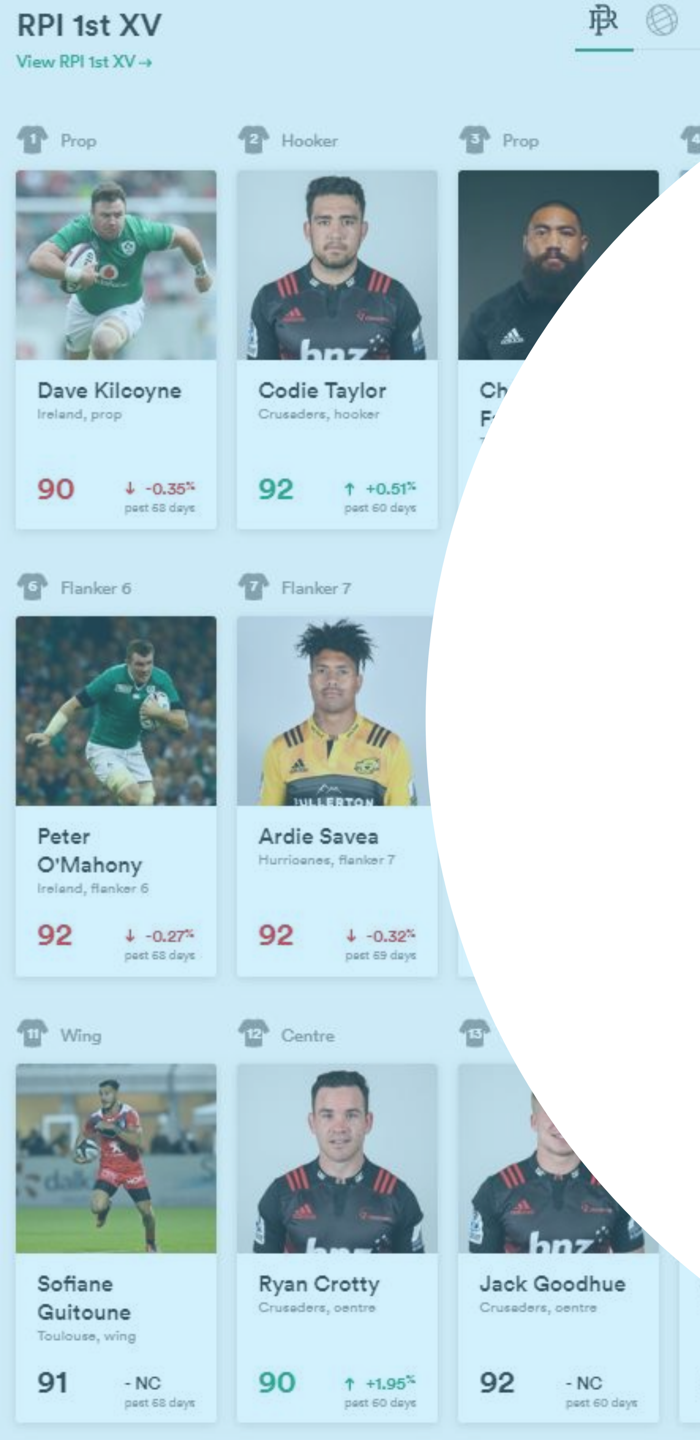
We have explored numerous sports, ranging from *Cricket* and *Rugby* through to *Golf* and Y*achting*.

**BUT,** the intent is always the same – *how to create winning performances*.

- **Rugby union**, commonly known in most of the world simply as **rugby**, is a contact team sport which originated in England in the first half of the 19th century.

- It is based on running with the ball in hand. A game is between two teams of 15 players using an oval-shaped ball on a rectangular field with H-shaped goalposts at each end.


https://www.youtube.com/watch?v=_43F2RsHHL0&t=101s




# WHAT IS RUGBYPASS?

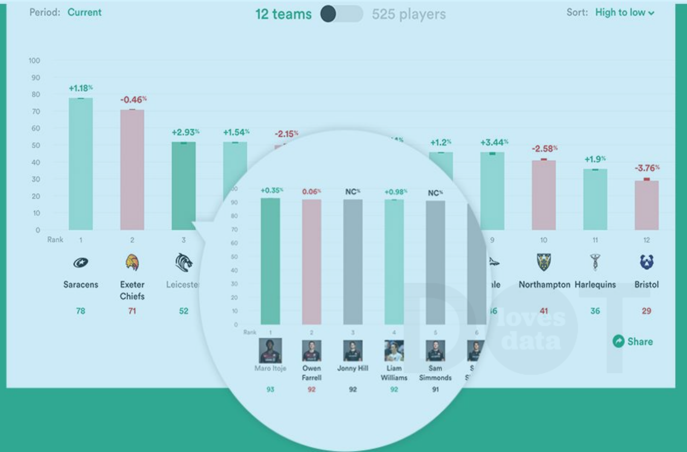
[https://index.rugbypass.com/](https://index.rugbypass.com/)

DOT built a revolutionary rugby rating system based upon individual skill executed in real time

What makes it different? The approach is all about winning, and the contribution to winning.

Winning a game is about winning moments

Parsimonious models align with perception and results are readily interpretable

# New Zealand Rugby

**New Zealand's first-class domestic rugby has gone through many structural and divisional changes over the last 30 years**

- Since 1976 there have been 5 major competition changes and 3 divisional changes.
- New Zealand rugby ushered in professionalism in 1995.

***Structural changes alongside socio-demographic changes such as urban drift have changed the competitiveness of these competitions over time*.**

# QUANTIFY HOW STRUCTURAL CHANGES TO A SPORTS COMPETITION AFFECT COMPETITIVNESS

# <u>HYPOTHESIS</u>

In balanced competitions, the Elo update function needs to be a larger value to update the ratings more quickly to maintain predictivity of the rating system

# WHAT IS *K*?

- Parameter $k$ is a learning rate and can be tuned to achieve optimal predictive performance
- Measures how a team or players rating is impacted by a winning or losing result
- Controls the sensitivity of change in ratings to new match results
- Changes in $k$ across seasons and competitions can be interpreted to understand the underlying mechanics of domestic first-class rugby
- Use the update function, $k$, as a descriptive measure to indicate the impact of latent structural changes in top-flight domestic first-class rugby in New Zealand
- As we hypothesise that $k$ can be used to determine structural changes in a competition, we need to detect statistically significant changes

# SECONDARY OBJECTIVES

1

**Optimise** ***k to produce the predictive ratings of match outcome***

2

***Introduce a ratings deviation to the Elo framework***

## ELO

$$r_i \rightarrow r_i + k\left[y_i - L\left(\frac{r_i - r_j \pm h}{B}\right)\right]$$

$$L(x) = \frac{1}{1+e^{-x}} \text{ and } B = \frac{400}{\ln(10)} \approx 173.7$$

- Lack of a reliability measure for the point estimate and update parameter *k*
- Measures how a winning or losing result impacts the ratings
- Rating remains unchanged for as long as a player remains inactive
- Simply, easy to calculate formula

## GLICKO

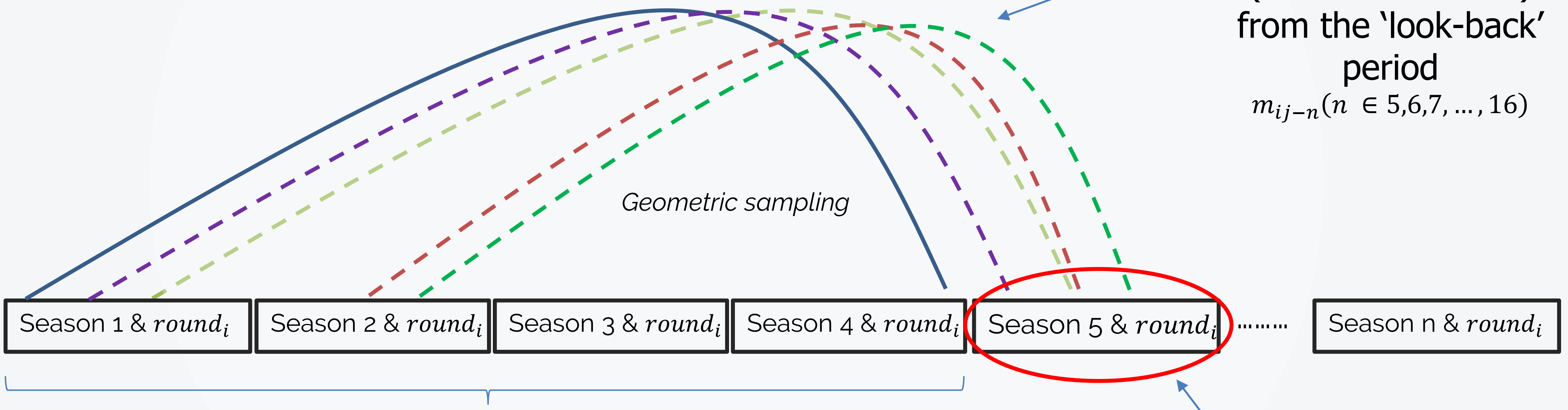
$$r' = r + \frac{1}{\frac{1}{RD^2} + \frac{1}{D^2}} \sum_{j=1}^{m} g(RD_j)\left(s_j - E\left(s|r, r_j, RD_j\right)\right)$$

$$RD = min\left(\sqrt{RD_0^2 + c^2 t}, 350\right)$$

- Glicko ratings incorporates a reliability measure to evaluate the confidence for its predictions
- Indicates a player's expected performance with a 95% confidence
- No *"park the bus"* approach – RD decreases after a game, but slowly increases over-time of inactivity
- Complex calculation requires a computer program

# MODIFIED ELO FRAMEWORK

Time-series geometric sampling (set of $x$ matches) from the 'look-back' period
$m_{ij-n}(n \in 5,6,7,\dots,16)$



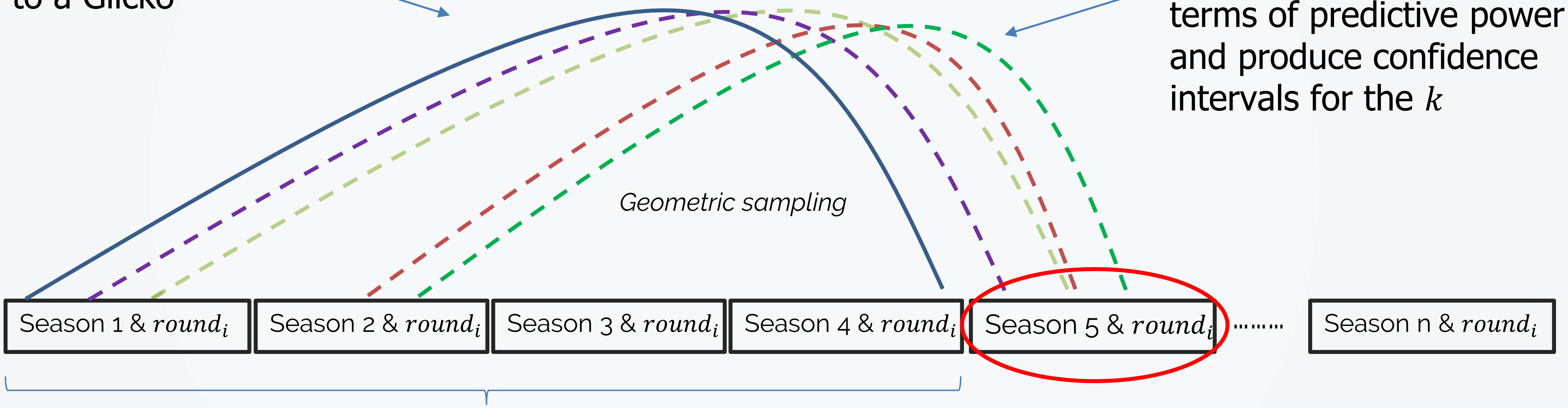

$m_{ij}$ - season-round

Randomly sample without replacement a range of $k$ parameters from a uniform distribution; $k \sim Uni[15, 200]$

# MODIFIED ELO FRAMEWORK cont….

Compare the predictive power of the Elo method to a Glicko model

Bootstrap the top 50 $k's$ in terms of predictive power and produce confidence intervals for the $k$




- Optimise $k$ using log-loss evaluation metric
- $K$ – identify the parameter that produces the greatest predictive power for the sampled ‘look-back period

# INITIAL RATINGS AND HOME GROUND ADVANTAGE

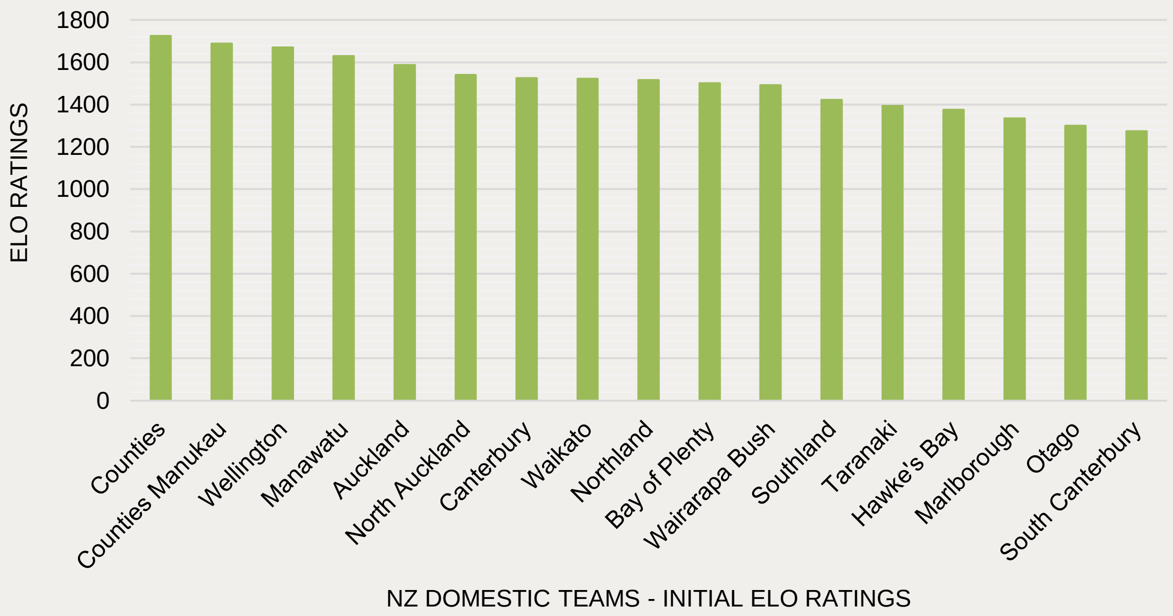
- Initial ratings are set to a mean value of 1500 before evolving the ratings over some training set of matches, eventually reaching equilibrium.
- Set the initial rating equal to a set of static initial ratings $\left\{r_i^{(0)}\right\}$ minimising a log-loss function.
- This is a constrained optimisation problem, where the mean rating is constrained to 1500

$$\min_i \left[ y - logistic\left(\frac{r_h - r_A \pm h}{B}\right)\right]^2 \qquad s.t. \frac{\sum_i (r_i)}{n} = 1500$$

- The constant *h* is optimised simultaneously with the initial ratings.

$$SS\left(\left\{r_i^{(0)}\right\}; h\right) = \sum_R \left[ y_i - L\left(\frac{r_i^{(0)} - r_j^{(0)} \pm h}{B}\right)\right]^2 \qquad s.t. \frac{\sum_i (r_i)}{n} = 1500$$

# HOME GROUND ADVANTAGE ≈ 42



# MODEL VALIDATION AND RESULTS

# VALIDATION AND APPLICATION: WHAT WE FOUND?

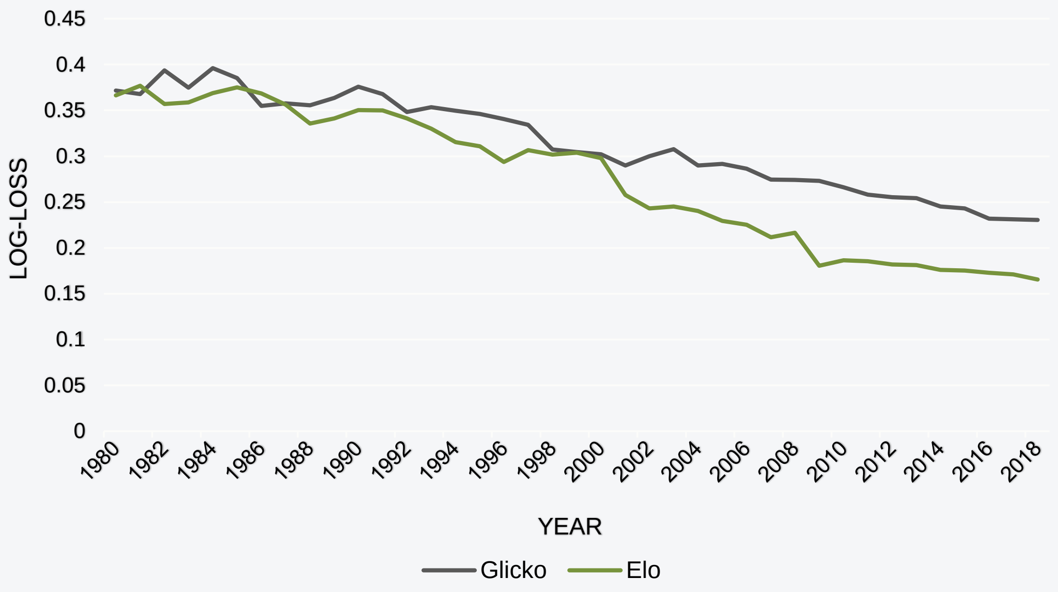



*Figure 1: Log-loss Elo vs. Glicko*

| PROPORTION OF CORRECT FORECASTS | | |
|---|---|---|
| **YEAR** | **ELO** | **GLICKO** |
| 1980- 1984 | 0.842 | 0.807 |
| 1985- 1989 | 0.858 | 0.808 |
| 1990- 1994 | 0.883 | 0.819 |
| 1995- 1999 | 0.869 | 0.829 |
| 2000- 2004 | 0.89 | 0.828 |
| 2005- 2009 | 0.899 | 0.837 |
| 2010- 2014 | 0.897 | 0.84 |
| 2015- 2018 | 0.895 | 0.858 |

*Figure 2: Proportion of correct forecasts*

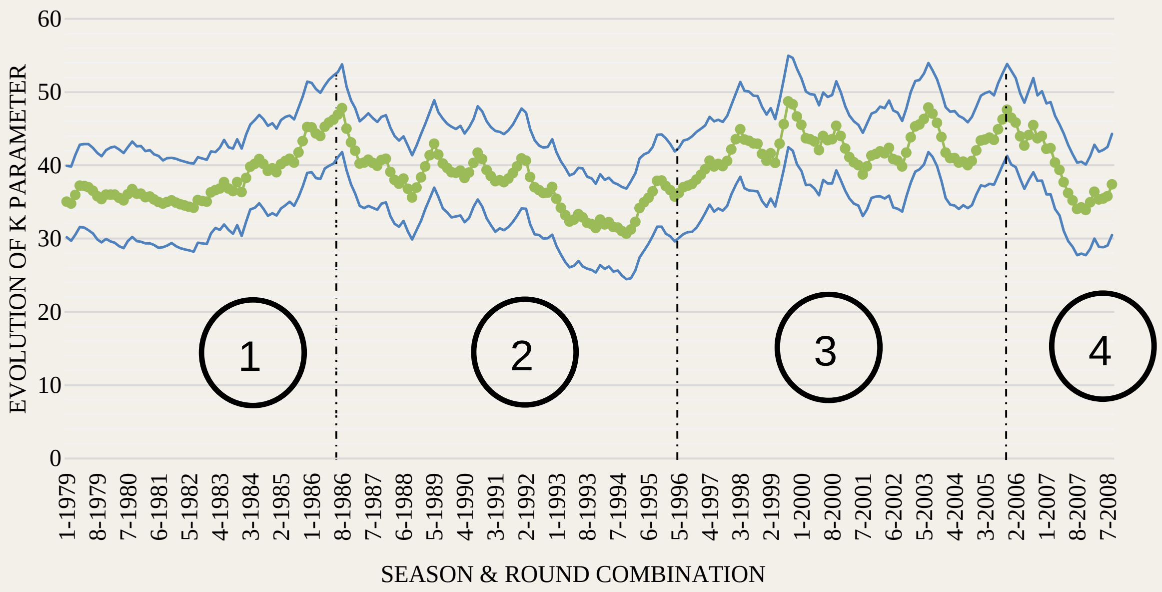



*Figure 3: Evolution of k*

# QUANTIFYING FIRST-CLASS RUGBY EVOLUTION

The optimal $k$ appears quite noisy. Consequently, applying an EWMA (weight = 0.2) removes some noise. This identifies the latent structures of the competition.

There are several structural events influencing the evolution of top 1st class Rugby in New Zealand.

A substantial shift in the nature of the competition, with $k$ becoming increasingly large from 1982-1985.

In 1985, the second tier which was an intra-Island competition was restructured into a 2nd and 3rd division, with promo/ releg.

Between 1985-1996, $k$ tends to decrease. Interesting period for NZ rugby following the 1st world cup and ushering in professionalism.

Interesting, $k$ began to increase, following the introduction of professionalism.

# DISCUSSION AND CONCLUSION

- Several structural events influenced the evolution of top tier first class in New Zealand
- $k$ decreases following any major restructure of the competition
- $k$ describes a competitions structural change and helps determine shifts in competition competitiveness
- Useful for monitoring tiered competitions where either grading, promotion or relegation is involved
- Evolution of $k$, 1982-1985, coincides with the 1$^{st}$ major structural change
- Statistically significant shifts in $k$ following the introduction of professionalism in 1996
- Modified Elo > [out-of-box] Glicko model
- Avg. validation period (7 rounds) for obtaining the optimal $k$

# WHAT NEXT AND THE IMPLICATIONS?

Tune and optimise Glicko parameters

Implement framework to domestic RugbyPass competitions

Are there any practically significant difference between predictive power?

New Zealand Rugby – affect of structural changes on competition

Following any major restructure of the competition, $k$ decreases

Implement as an extension to the RugbyPass system

# REFERENCES


- Abstract, T. (2018). Tennis Abstract: ATP Elo ratings. tennisabstract.com/reports/atp_elo_ratings.html.
- Akhtar, S., Scarf, P., & Rasool, Z. (2015). Rating players in test match cricket. Journal of the Operational Research Society, 66(4), 684-695.
- Aldous, D. (2017). Elo ratings and the sports model: A neglected topic in applied probability? Statist. Sci., 32(4):616–629.
- Allsopp, P. E., & Clarke, S. R. (2004). Rating teams and analysing outcomes in one-day and test cricket. Journal of the Royal Statistical Society: Series A (Statistics in Society), 167(4), 657-667.
- Asif, M., & McHale, I. G. (2016). In-play forecasting of win probability in One-Day International cricket: A dynamic logistic regression model. International Journal of Forecasting, 32(1), 34-43.
- Bracewell, P. J., Forbes, D. G., Jowett, C. A., & Kitson, H. I. (2009). Determining the evenness of domestic sporting competition using a generic rating engine. Journal of Quantitative Analysis in Sports, 5(1).
- Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. Biometrika, 39(3/4):324–345.
- Broadie, M., & Rendleman, R. J. (2013). Are the official world golf rankings biased? a. Journal of Quantitative Analysis in Sports, 9(2), 127-140.
- Broadie, M. (2012). Assessing golfer performance on the PGA TOUR. Interfaces, 42(2), 146-165.
- Christoph Letiner, A. Z. and Hornik, K. (2009). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the euro 2008. International Journal of Forecasting, 26(3):471–481.
- Clarke, S. R. (1988). Dynamic programming in one-day cricket-optimal scoring rates. Journal of the Operational Research Society, 39(4), 331-337.
- da Silva Curiel, R. S. (2018). World football Elo ratings. eloratings.net.
- Elo, A. E. (1978). The rating of chessplayers, past and present. Arco Pub.
- Glickman, M. E. (1995). The Glicko system. Boston University.
- Goddard, J. and Asimakopoulos, I. (2004). Modelling football match results and the efficiency of fixed-odds betting. Journal of Forecasting, 23:51 – 66.
- Herbrich, R., Minka, T., & Graepel, T. (2007). TrueSkill™: a Bayesian skill rating system. In Advances in neural information processing systems (pp. 569-576).
- Hucaljuk, J. and Rakipovic, A. (2011). Predicting football scores using machine learning techniques. In Proceedings of the 34th International Convention MIPRO, pages 1623–1627.

- Hvattum, L. M. and Arntzen, H. (2009). Using ELO ratings for match result prediction in association football. International Journal of Forecasting, 26(3):460 – 470. Sports Forecasting.
- Ingram, M. (2019). Gaussian Process Priors for Dynamic Paired Comparison Modelling. arXiv preprint arXiv:1902.07378.
- Király, F. J. and Qian, Z. (2017). Modelling competitive sports: Bradley-Terry-Élő models for supervised and on-line learning of paired competition outcomes. CoRR. arxiv.org/abs/1701.08055.
- Lasek, J., Szlávik, Z., and Bhulai, S. (2013). The predictive power of ranking systems in association football. International Journal of Applied Pattern Recognition, 1(1):27–46.
- Levin, A. (2017). Ranking the Skills of Golfers on the PGA TOUR using Gradient Boosting Machines and Network Analysis. MIT Sloan Sports Analytics Conference.
- McHale, I., & Morton, A. (2011). A Bradley-Terry type model for forecasting tennis match results. International Journal of Forecasting, 27(2), 619-630.
- Morris, B. and Bialik, C. (2015). Serena Williams and the difference between all-time great and greatest of all time. www.fivethirtyeight.com.
- Odachowski, K. and Grekow, J. (2013). Using bookmaker odds to predict the result of football matches. In Knowledge Engineering, Machine Learning and Lattice Computing with Applications, pages 196–205, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Opisthokonta (2016). Tuning the Elo ratings: Initial ratings and inter-league matches. opisthokonta.net/?p=1412.
- Patel. A. K., Bracewell. P.J., & Bracewell, M.G. (2018). Estimating expected total in the first innings of T20 cricket using gradient boosted learning. Paper presented at The Proceedings of the 14th Australian Conference on Mathematics and Computers in Sports. Sunshine Coast, Queensland, Australia: ANZIAM MathSport. ISBN: 978-0-646-95741-8.
- Scarf, P., & Shi, X. (2005). Modelling match outcomes and decision support for setting a final innings target in test cricket. IMA Journal of Management Mathematics, 16(2), 161-178.
- Silver, N. (2006). Lies, damned lies: We are Elo? www.baseballprospectus.com.
- Silver, N. (2014). Introducing NFL Elo ratings. fivethirtyeight.com/features/introducing-nfl-elo-ratings.
- Silver, N. and Fischer-Baum, R. (2015). How we calculate NBA Elo ratings. www.fivethirtyeight.com.
- Singh, T., Singla, V., & Bhatia, P. (2015, October). Score and winning prediction in cricket through data mining. In 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI) (pp. 60-66). IEEE.
- Stefani, R. (2011). The methodology of officially recognized international sports rating systems. Journal of Quantitative Analysis in Sports, 7(4).
- United, O. (2018). Weekly tennis ELO rankings. tenniseloranking.blogspot.com.
- Moore. W. E., Rooney. S.J., & Bracewell, P.J. (2018). An Elo rating system for rugby union. Paper presented at The Proceedings of the 14th Australian Conference on Mathematics and Computers in Sports. Sunshine Coast, Queensland, Australia: ANZIAM MathSport. ISBN: 978-0-646-95741-8.

# ANZIAM Mathsport Australia & NZ - May 26-28 in Wellington. Hosted by Victoria University of Wellington