A Systematic Approach to Footy Tipping

Grant Holtes

August 26, 2024

1 Introduction

Office AFL tipping competitions are commonplace in Victoria, and my office is no exception. For the uninitiated, the Victorian Gambling and Casino Control Commission (VGCCC) provides a helpful definition: A "tipping" competition as a competition "that allow entrants to pick the winning team/s in each week of a sporting season to earn points if successful", where "all of the money that has been collected is returned as prizes". "The most common method is to award prizes to the three entrants who have selected the most winners over the season"(VGCCC, 2024).

As I have little[1] knowledge of AFL teams and their relative abilities, I thought I could probably have more success by building a model to predict the outcome of each game, and perhaps learn something along the way.

2 Methodology

The approach used is heavily inspired by ELO scores, which are updating "goodness" scores assigned to each team in a tournament, from which a win probability can be calculated. Rather than use the original ELO specification, we estimate the scores directly using linear and non-linear regression models, which allow for other variables to be included to compliment the scores and boost predictive power.

2.1 Background on ELO Scores

An ELO score is a metric that measures how "good" a team (or individual) is, with the probability of a win by team A in a match given they are playing team B is given by the formula below. The use of this approach was heavily inspired by the work of Nate Silver, 2022.

$$P(A_{win}|B) = \frac{1}{10^{-\frac{S_A - S_B}{n}} + 1} \tag{1}$$

where:

- S_A is the ELO score for team A, and similar for team B
- n is a scale adjuster give the size of the ELO scores assigned, which is usually 400.

[1] None whatsoever

The ELO scores are calculated through an iterative update process:

1. Every team starts with the same given number of points, eg 1300.
2. After a game, the winner takes points from the loser, with the number of points being determined by the difference in their ELO scores and the win margin (in games with a margin). A high score player would only take a few points off a weaker player, but if a low score player wins in an upset, they would take more points. The max number of points traded is given as K, a parameter.[2]

This approach has the advantage of not requiring two teams to have played each other to be able to estimate the outcome of the game. In the case that the two teams have played each other, that outcome is fused with the outcome of all other matches they have both played to further enrich the prediction.

However, the ELO score has a few parameters which cause issues: The value of K used determines how quickly the scores reflect new information, and the value of n used greatly influences forecast win probability. Both are subject to tuning and debate.

2.2 Maximum Likelihood Estimation

A similar outcome to an ELO score method can be achieved by explicitly predicting win probability using a logistic regression, with the aim of avoiding issues around selecting an optimal value for n and K.

The specification of a logistic regression is structurally very similar to the ELO model, but in place of the score difference we have a linear combination of coefficients and inputs.

$$P(Y|X) = \frac{1}{e^{-(B_0 + B_1 X_1 + .. + B_n X_n)} + 1} \tag{2}$$

The aim is to transform the linear combination of coefficients and inputs to be the difference in some optimal scores of the team's playing each other. This can be achieved by letting X be a vector where every element is zero, except the home team's element is 1 and the away team's is -1.

$$X = [0, ..., 1, ..., -1, ..., 0] \tag{3}$$

The linear combination now collapses to $B_0 + B_h - B_a$ where h and a denote the indices of the home and away teams, leaving the model as:

$$P(Y|X) = \frac{1}{e^{-(B_0 + B_h - B_a)} + 1} \tag{4}$$

When Y is set as a binary "home team win" indicator, the estimated model coefficients themselves are the optimal score values given the historical data, while the constant term B_0 can be interpreted as an estimate of the home ground advantage.

[2] The exact details on the update mechanism is more complex, but there are many chess websites that explain it well. See "Wikipedia ELO rating system", 2024

2.3 Win margin prediction

In some cases it is preferable to have an estimate of the win margin rather than an explicit win probability. The modelling setup developed in the previous section can be easily migrated from producing a probability to a margin by substituting the logistic regression for a linear regression and setting Y as the historical win margin.

2.4 Incorporating other variables

The neat aspect to these model specifications is that incorporating other explanatory variables is straightforward - simply add the variables to the regression. The obvious addition is to use the home team win probability implied by sports betting odds.

$$P(A|B) = \frac{1}{O_A(\frac{1}{O_A} + \frac{1}{O_B})} \tag{5}$$

where O_A is the dollar payout being offered on team A. The $\frac{1}{O_A} + \frac{1}{O_B}$ term corrects for the bookmakers margin.

This probability can also be used as a sensible baseline model.

2.5 Exponential weighting of observations

Logically, more recent games should provide better information about a team's ability, and so should be weighted more highly in the estimation of the model. This was tested by exponentially weighting the observations, however, it was seen to decrease performance in backtests, so was removed.

3 Results

3.1 Backtests

The model specifications described in the previous sections were tested using a backtest methodology, the results of which are given in Table 1 for the win margin prediction models and Table 2 for the win probability prediction models. The models all have an accuracy of approximately $\frac{2}{3}$.

Model	Median absolute error	Win accuracy
score only	22.73	64.46
score and odds	21.95	67.47
odds only	21.93	67.6

Table 1: Results for the margin predicting regression model specification. Win accuracy is the percentage of game outcomes predicted correctly

Model	Log loss	Win accuracy
score only	0.6337	64.73
score and odds	0.6026	67.74
odds only	0.6026	67.74

Table 2: Results for the win probability predicting logistic model specification. Win accuracy is the percentage of game outcomes predicted correctly

The preferred model is the "score and odds" margin prediction model, as this model does not simply follow the odds and so provides an opportunity to rank better than competition mates that simply follow the betting market, effectively introducing some small "tracking error" against the betting market.

To further critique the performance of this model we analysed its performance over over the last 8 seasons, as in Figure 1. This suggests that the performance of the model is stable over time.

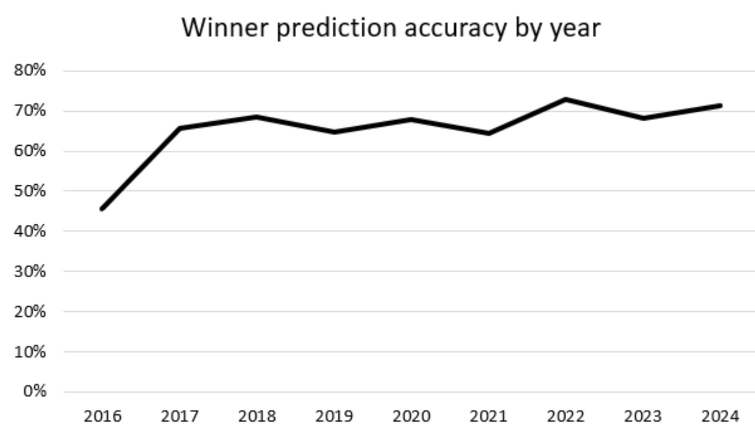

Figure 1: Average backtest accuracy of the "score and odds" margin prediction model by season

We also evaluated its average accuracy in each week of these seasons, as in Figure 2. This seems to show that model performance is most volatile in beginning and end weeks, but this is likely just a product of the smaller sample sizes in these groups of weeks.

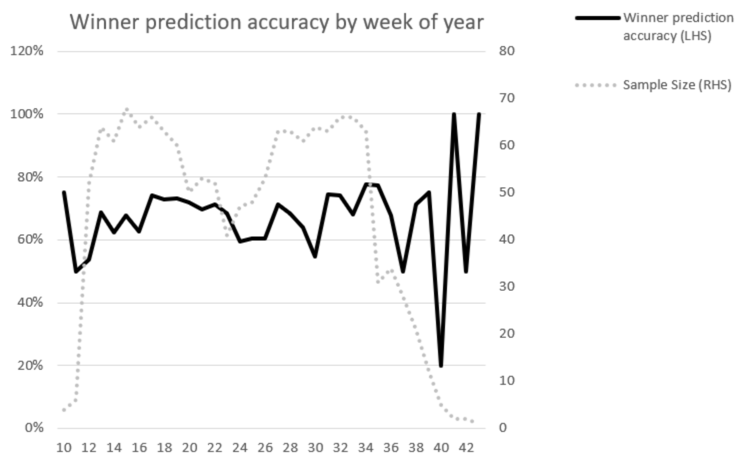

Figure 2: Average backtest accuracy of the "score and odds" margin prediction model by week of season

3.2 Implemented Model

The "score and odds" margin prediction model was used to produce forecasts for the 2024 AFL season, which were used in my office footy tipping competition. Unlike the backtests which predict one game ahead, the actual implementation predicts the entire week's games at once. However, the performance is seemingly unchanged, with a realised accuracy of 72.2% compared to a backtest accuracy of 71.3% over the 2024 season [3]. This has included a number of predictions that go against the betting odds, and have resulted in the model placing 3^{rd} in the office pool of 57 (top 6%) and 21,035 of 877,916 (top 2.4%) in the footy tipping system used (ESPN).

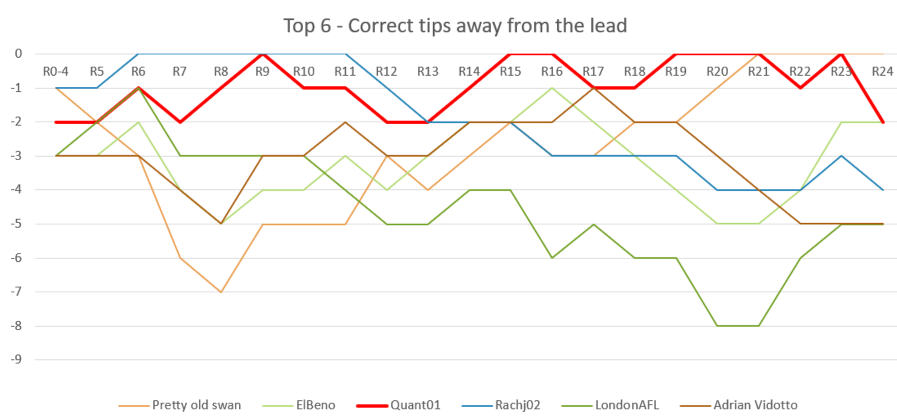

Figure 3: Number of correct tips from the lead

Frustratingly 2^{nd} place was lost due to two bad predictions in the final round, with the model

[3]Backtest was conducted mid season

holding first or second place for many weeks in the later stage of the season (see Figure 5) and was consistently two or fewer tips from the lead (see Figure 3).

Figure 4: Competition rank by round for the model

Another interesting observation from the season was how quickly the field thinned out into a small group of leaders, with that volume holding quite steady from round 13 onwards.

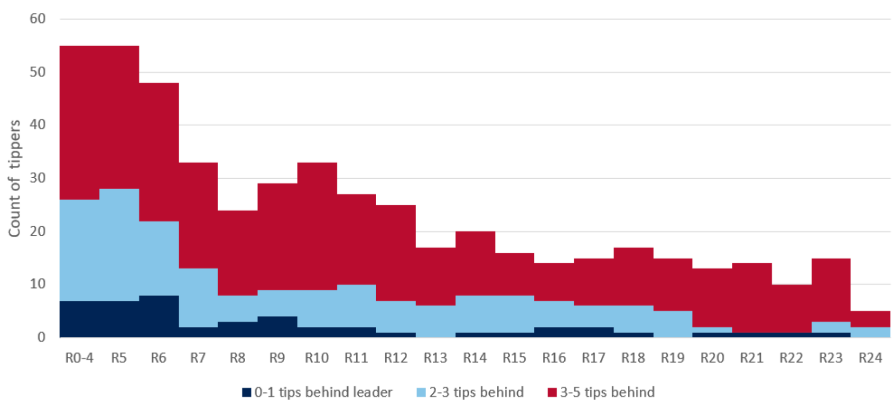

Figure 5: Field depth as measured by number of tippers varying numbers of tips from the lead

4 Appendix: Other ideas

There are a number of other approaches which I would like to explore if time allows in the future.

4.1 Embedding approaches

Both the ELO score and regression approaches aim to condense all useful information about the team into a single value, their "goodness" score. It would be interesting to see if providing the model with more values to store data in provides superior results. The natural conclusion of this would be to develop a model that represents each team as a n dimensional vector or embedding, a common technique in machine learning.

4.2 Player level data

Developing player scores as well as team level scores should allow the model to account for player trades between teams, allowing for longer lookback periods in the training data to be valid.

4.3 Information ratio optimisation

The current models optimise for raw predictive power, which is a fair approach if we believe the other players in the competition are operating in this manner as well. However, if we assume that the other players are aiming to follow the betting market predictions, it may be beneficial to optimise for matching or beating the betting market only, so the model only deviates from the "market" when it has a strong reason to do so. The inclusion of betting odds in the implemented model was intended to promote this behavior, but explicitly optimising for this outcome would be a good next step if the other competition participants are seen to be closely following the betting odds.

References

Nate Silver, N. P., Jay Boice. (2022). 538: How our nfl predictions work [accessed at https://fivethirtyeight.com/methodolo our-nfl-predictions-work/].

VGCCC. (2024). Sports tipping [accessed at https://www.vgccc.vic.gov.au/gambling/community-and-charitable-gaming/permitee-resources/footy-tipping].

Wikipedia elo rating system [accessed at https://en.wikipedia.org/wiki/Elo$_{r}ating_{s}ystem/$]. (2024).