# An analysis of factors impacting team strengths in the Australian Football League using time-variant Bradley-Terry models

Carlos Rafael González Soffner, Manuele Leonelli

[a]School of Science and Technology, IE University, Madrid, Spain

**Abstract**

Australian Rules Football is a field invasion game where two teams compete to score the most points. While complex machine learning models can predict match outcomes post-game, their lack of interpretability limits understanding of the factors influencing team performance. Using data from the male Australian Football League (2015–2023), we estimate team strengths and their determinants by fitting flexible Bradley-Terry models. We identify teams significantly stronger or weaker than average, with stronger teams placing higher in the previous season's ladder and leading in Forward 50 activity, goal shots, and scoring. Playing at home consistently creates an advantage, regardless of strength. When used for forward prediction, models incorporating team-specific, time-variant features correctly anticipate up to 71.5% of outcomes. Our approach thus enables interpretable strength estimation and competitive forecasting, supporting data-driven strategies and training.

*Keywords:*
Australian football, Bradley-Terry models, Feature selection, Predictive modeling, Sports analytics

## 1. Introduction

Australian Rules Football (AF) is a field invasion game played by two teams of 18 players over four twenty-minute quarters. Teams score by kicking the ball (the footy) between goalposts: a goal (6 points) is scored through the central posts, and a behind (1 point) if the ball passes between a central and side post, touches a post, or is touched by an opponent. The team with the highest score wins; draws occur if scores are equal.

Players advance the ball by running and bouncing it or by kicking and handballing (punching to a teammate). Opponents contest possession through tackling, intercepting disposals, or taking a mark (catching a kicked ball, earning a free kick). The oval-shaped field is divided into three zones: Defensive 50 (nearest a team's goal), Forward 50 (nearest the opponent's goal), and Midfield.

The male Australian Football League (AFL) is the premier AF competition, comprising 18 teams competing weekly in a 23–24 round Home-and-Away season. Teams earn 4 points for a win and 2 for a draw. The top 8 qualify for a Finals Series culminating in the Grand Final.

A growing body of research uses AFL match data, readily available online, to model match outcomes, player performance, and network effects. Game winners and score margins have been predicted using action counts, e.g., handballs, disposals (Robertson et al., 2016; Young et al., 2019a), network measures (Braham and Small, 2018; Fransen et al., 2022; Sargent et al., 2022; Young et al., 2020), and game features like home advantage (Robertson and Joyce).

Most outcome models are post-game, predicting winners or scores using in-game performance. While often accurate (Robertson et al., 2016; Young et al., 2019a, 2020), such models offer limited insight into pre-game factors. Pre-game models, though less accurate (Fahey-Gilmour et al., 2019; Manderson et al., 2018; Sargent et al., 2022), confirm that recent team performance can predict future outcomes.

Complex machine learning models have performed well (Fahey-Gilmour et al., 2019; Manderson et al., 2018; Young et al., 2019a, 2020), but their lack of interpretability makes them less suitable for informing coaching and training decisions. This creates a trade-off between predictive power and model transparency.

We investigate flexible Bradley-Terry (BT) models (Agresti, 2012) for estimating team strength and predicting AFL match outcomes using only pre-game features. Designed for pairwise comparisons, BT models express a team's probability of winning as a function of relative strength. We extend the basic formulation to include time-varying and team-specific features, following Tsokos et al. (2019), and implement our models using the `BradleyTerry2` R package (Firth and Turner, 2012).

Although widely used in other sports (e.g., football, cricket, tennis), BT models have not yet been applied to Australian football. This paper introduces the first BT-based predictive framework for AFL, demonstrating that these models can offer interpretable, data-efficient alternatives to black-box machine learning tools. We go beyond previous work by incorporating cumulative and recent team performance metrics, testing multiple training horizons, and simulating realistic round-by-round forecasting. Our results show that BT models can provide accurate predictions while revealing structural patterns in evolving team strength, offering both practical utility for analysts and theoretical insights into match dynamics.

## 2. Related work

### 2.1. Game features

The variety of movement and scoring actions in AF, combined with the complex fixture structure of the AFL, has led to a broad pool of features used to model game outcomes. Most studies focus on fixture-related variables (e.g., home advantage) or team-level aggregates of individual actions (e.g., total kicks, goals), though network-based team features have received growing attention.

Following Fahey-Gilmour et al. (2019), features can be divided into 'strength' variables, which reflect relative team quality, and 'match difficulty' variables, which affect outcomes independently of team strength (e.g., location, travel).

#### 2.1.1. Technical strength

Technical features refer to the individual and aggregated actions of players, as well as team outcomes. The most prominent are Performance Indicators (PIs): individual player actions aggregated to the team level. Modelling outcomes based on absolute PI counts (e.g., total kicks) is discouraged, as this ignores the opponent's performance (Robertson et al., 2016). Instead, Young et al. (2019a) showed that relative PIs were significantly more explanatory of game outcomes and score margins.

Higher relative counts of kicks and goal conversions during a game appear indicative of winning. For instance, in the 2013 season, 81% of teams with kick differentials $\geq -1$ won, and 90% of teams with higher additional kicks and a 4.2% better goal conversion also won (Robertson et al., 2016). Teams with more activity in the Forward 50, the zone closest to the opponent's goalposts, tend to score more. Influential variables include differentials in inside 50s, marks (general and in the Forward 50), and contested possessions. These findings were reinforced by Young et al. (2019a), who also highlighted time in possession and metres gained as especially important.

Most PI-based studies are descriptive, using post-game features to explain results. While useful for retrospective analysis, few investigate the predictive power of cumulative PIs before the game. This is a notable gap, as the evolution of indicators over a season or recent matches is a simple yet powerful signal of relative strength.

Beyond PIs, some studies consider physical or demographic features. Older and heavier teams, i.e., those with higher average player age and weight, have been found slightly more likely to win, possibly due to greater experience (Lazarus et al., 2018).

Team-level indicators such as recent match form also play a role. Wins over the previous four games ("momentum") are significantly associated with better outcomes (Robertson and Joyce). Similarly, current and previous season ladder positions correlate with winning probability (Fahey-Gilmour et al., 2019; Robertson and Joyce).

Finally, player-based team ratings (e.g., Elo scores), and the availability of top-performing players are significant predictors (Fahey-Gilmour et al., 2019). However, apart from Elo, there is no agreed-upon standard for ranking players pre-game in a way usable by all teams.

*2.1.2. Tactical strength*

Recently, outcome prediction based on aggregated PIs has been complemented by tactical analyses, which model AF as a social network of player interactions. These typically take the form of passing networks, where edges represent passes between players.

One of the earliest examples is Sargent and Bedford (2013), who found that teams with higher connectedness had better score margins and could even estimate the impact of missing players. Braham and Small (2018) extended this approach and confirmed that higher values in network metrics such as betweenness and out-degree were associated with stronger teams, particularly when averaged over one or four previous games.

However, Young et al. (2019b) found no significant differences in betweenness between winners and losers, and only weak correlations for other measures. This suggests that while network metrics may reflect strategy, they are sensitive to seasonal changes, coaching decisions, or rule adaptations. As with PIs, absolute values may miss important contextual information.

Efforts to combine tactical and technical features remain limited and inconclusive. Challenges include the lack of freely available network data and the computational cost of constructing networks for every match. Moreover, diverse design choices (e.g., what constitutes a pass or a node) lead to metrics that capture different facets of play and may not be comparable across studies.

Indeed, Young et al. (2020) found that tactical features were slightly more important than technical ones, but did not improve classification accuracy. In contrast, Fransen et al. (2022) showed that a one-unit increase in connectedness raised the probability of winning by 5.3%, with tactical metrics explaining 14–27% of outcome variance. This suggests that while promising, tactical features require further methodological refinement.

*2.1.3. Match difficulty*

Match difficulty features are pre-game factors that influence the outcome independently of team strength. These include home/away designation, travel, venue familiarity, and rest periods.

Home advantage is well-documented. AFL teams playing away, especially interstate, are less likely to win (Robertson et al., 2016; Lazarus et al., 2018). This holds even when teams share venues or play at neutral grounds. Broader sports literature links this effect to crowd support and emotional arousal (Leitner et al., 2023).

Travel may also impair recovery. Australian teams often cross time zones or have reduced rest. Sargent et al. (2022) showed that interstate travel reduces player sleep, though the link to in-game performance remains underexplored.

### *2.2. Eras in AFL*

Like many sports, AF has evolved through changes in rules, tactics, and athlete profiles. The modern game is faster and features taller, heavier players (Woods et al., 2017). Scoring has also declined in recent years (Lane et al., 2020).

A growing body of research suggests that meaningful structural changes can be observed in medium-term blocks or "eras". For example, Woods et al. (2017) identified three eras based on PI trends: a possession-focused era (2005–09), a defensive era (2010–13), and a turnover-focused era (2014–onward). Lane et al. (2020) found a similar shift from offensive to increasingly defensive strategies around 2007–11, reflected in higher tackle and disposal counts and lower scoring.

These findings challenge the assumption that historical match data can be pooled across years without adjustment. If team strategies shift over time, models trained across multiple seasons may conflate distinct underlying dynamics. Young et al. (2019a) raised similar concerns and noted the lack of consensus on how to segment the data into meaningful eras. While no standard methodology exists, the literature supports the use of season-specific models or time-weighted approaches that give more influence to recent data.

### *2.3. Machine learning models*

Non-linear models often perform best when modelling AFL outcomes, particularly when feature selection is applied (Fahey-Gilmour et al., 2019). These methods are common in team invasion sports, where the high dimensionality of potential predictors makes selection essential. However, in pre-game prediction settings, black-box models can be problematic, as their limited interpretability reduces their practical value for coaches and analysts.

Most studies focus on classification models to predict wins and losses, excluding draws due to their rarity. Some attention has also been given to models predicting score margins. Surprisingly, few studies model the number of goals and behinds directly, whether jointly or separately.

Decision trees tend to perform well, based on the idea that thresholds in differential PIs can identify likely winners. Young et al. (2020) achieved 89% accuracy using technical and tactical features in a decision tree. Similarly, random forests yielded accuracies between 85% and 89% across different eras (Young et al., 2019a).

Generalised linear models (GLMs) often outperform more complex methods. Young et al. (2020) reported 93.2% accuracy when modelling binary outcomes and an RMSE of 6.9 when predicting score margins. GLMs have also been used for pre-game prediction: with elastic net regularisation, Fahey-Gilmour et al. (2019) achieved 73.3% accuracy for the 2018 season and 71.6% via cross-validation over 2013–2017.

Finally, Manderson et al. (2018) used a Bayesian hierarchical model based on the Skellam distribution to predict goals and behinds separately, then aggregate scores to identify winners. The model correctly predicted 132 games (about two-thirds), slightly underperforming relative to betting markets.

A full summary of the literature reviewed in this section, including features, models, and data sources, is provided in Appendix A.

## 3. Methodology

### *3.1. The data*

Data were collected using the `fitzRoy` R package, which provides access to official AFL data from sources including AFL Tables, the AFL website, FootyWire, and Squiggle (Nguyen et al., 2021). As the most comprehensive non-commercial tool for AFL data, `fitzRoy` has been used in prior research (Fahey-Gilmour et al., 2019). Match results and ladder rankings for the 2015–2023 seasons were retrieved via `fitzRoy`, while final ladder positions accounting for Finals Series outcomes were added manually. Performance indicator (PI) data for each game were also retrieved using `fitzRoy`. Network data were excluded, as they are not publicly available.

In total, data from 1826 games were used. Most seasons had 207 games, except 2015 (206, one cancelled), 2020 (162, shortened due to Covid), and 2023 (216, one round added). Following Fahey-Gilmour et al. (2019) and Fransen et al. (2022), draws were excluded due to their rarity (14 games total).

Appendix B provides a glossary of all features used to predict outcomes, and Appendix C summarizes the data pipeline and processing steps used to construct the modelling datasets. The code to replicate the experiments is available at the following link: https://github.com/charlieceratops/AFL_BradleyTerry.

### *3.2. The Bradley-Terry model*

The Bradley-Terry (BT) model is a generalised linear model (GLM) for pairwise comparisons, such as AFL matches where two teams compete for a win. Following Tsokos et al. (2019), outcomes can be modelled as:

$$Y_{ijr} = \begin{cases} 1, & \text{if team } i \text{ beats team } j \text{ at round } r, \\ 0, & \text{if team } j \text{ beats team } i \text{ at round } r, \end{cases}$$

for $i, j = 1, \ldots, 18$, $i \neq j$, and $r = 1, \ldots, N$, where $N$ is the total number of rounds. The probability that team $i$ defeats team $j$ in round $r$ is given by:

$$P(Y_{ijr} = 1) = \frac{\pi_{ir}}{\pi_{ir} + \pi_{jr}},$$

where $\pi_{ir} = \exp(\lambda_{ir})$ and $\lambda_{ir}$ is team $i$'s latent strength in round $r$ (Tsokos et al., 2019). This formulation models outcomes as proportional to the relative strength of the two teams, providing interpretable insight into performance.

*3.2.1. Standard Bradley-Terry*

The standard BT model assumes team strengths are fixed over time (Tsokos et al., 2019). It excludes match- or team-specific predictors and estimates strengths solely from win–loss outcomes. Assuming independence of games, it models a binomial target (Team $i$, Team $j$) based on head-to-head results (Firth and Turner, 2012). The log-odds of team $i$ beating team $j$ in round $r$ is:

$$\text{logit}(P(Y_{ijr} = 1)) = \lambda_i - \lambda_j,$$

where the $r$ subscript is dropped because strengths are assumed constant across rounds.

*3.2.2. Bradley–Terry model with a contest-specific (home advantage) effect*

This extension introduces a bias term (or "order effect") to account for advantages unrelated to team strength, such as playing at home:

$$\text{logit}(P(Y_{ijr} = 1)) = \lambda_i - \lambda_j + \delta Z,$$

where $Z = 1$ if team $i$ has the advantage, and $-1$ otherwise (Agresti, 2012; David, 1988; Firth and Turner, 2012; Yan, 2025). This formulation corresponds to the standard Bradley–Terry model with a home-field (or order) effect and can be viewed as a special case within more general Bradley–Terry models with covariates.

In the AFL, the main contest-level feature is the Home/Away designation. Other contextual effects (e.g., travel, rest) vary by team or over time and are excluded. The model estimates $\delta$ jointly with team strengths $\lambda_i$, which are assumed to be independent of this contest-specific bias. Parameter estimation is carried out by maximum likelihood under the logistic regression representation of the model, as implemented in the `BradleyTerry2` R package (Firth and Turner, 2012).

*3.2.3. Team-specific, time-variant Bradley-Terry*

This more flexible formulation allows team strength to depend on team-specific, time-varying pre-game features (e.g., cumulative wins, recent form, or performance indicators). Match outcomes are modelled as

$$\operatorname{logit}(P(Y_{ijr} = 1)) = \sum_{k=1}^{p} \beta_k \big(X_{ikr} - X_{jkr}\big) + \sum_{\ell=1}^{q} \gamma_\ell Z_{ijr,\ell},$$

where $X_{ikr}$ denotes the value of the $k$-th team-specific predictor for team $i$ observed prior to round $r$, and $Z_{ijr,\ell}$ denotes contest-level difficulty variables (e.g., home/away status or interstate travel). The coefficients $\beta_1, \ldots, \beta_p$ and $\gamma_1, \ldots, \gamma_q$ are estimated jointly by maximum likelihood.

In this specification, no static team-specific intercepts are included. Instead, team strength varies across rounds through predictors whose values evolve over the season, such as cumulative or rolling performance indicators and form measures. Randomness enters the model through the Bernoulli likelihood for match outcomes, as in standard generalised linear models.

Although the model does not impose a latent dynamic evolution equation on team strength (e.g., autoregressive or state-space dynamics), it captures time variation by allowing the covariates entering the linear predictor to change from contest to contest depending on the stage of the season. In this sense, team strength is time-variant through observed information rather than through an unobserved stochastic process. By contrast, the model of Section 3.2.2 is time-invariant, as team strengths do not depend on evolving predictors.

From a theoretical perspective, Bradley–Terry models and their extensions have been studied under sparse and high-dimensional paired-comparison regimes, including results on existence, inference, and likelihood-based testing as the number of teams grows (Yan et al., 2012; Yan, 2016; Yan et al., 2025). In this work, we focus on the applied use of these models for AFL data.

*3.3. Construction of features*

The ladder position after the previous game is added to the results and transformed into differentials, as this improves predictive performance over absolute rankings (Robertson and Joyce). Following Fahey-Gilmour et al. (2019); Robertson and Joyce, differentials in the final ladder position from the previous season are also included.

Team form features capture short- and long-term performance, including whether a team won their last game, the number of wins over the last four games, cumulative season wins, and streaks of consecutive wins or losses. Only the differential in cumulative wins is used.

Points in favour, points against, and their ratio (percentage) are transformed into absolute differentials. Performance Indicators (PIs) are aggregated by game and team, computed as cumulative totals over the season and previous four games, and then transformed into differentials.

Finally, binary indicators are included for match difficulty: whether a team plays at home, at their home ground, or interstate.

### *3.4. Description of the experiments*

#### *3.4.1. Experiment 1: Standard Bradley-Terry*

We first apply the standard BT model using the `BradleyTerry2` R package. Models are trained separately on each season, as well as on aggregated datasets combining 2, 3, or 4 consecutive seasons, and on the full 2015–2023 dataset. This allows us to explore the consistency and evolution of team strengths across time.

The outcome variable `WIN` equals 1 if the team won the game and 0 otherwise (draws are excluded, as noted earlier). Model fit is assessed using the Akaike Information Criterion (AIC).

To aid interpretation, the reference team is selected as the one with approximately average performance, typically, a team with a balanced win-loss record within each training window.

Prediction accuracy is computed in two ways: on the training set and on the test set (i.e., the season immediately following the training period). A team is predicted to win if its estimated win probability exceeds 0.5. Figure D.3 summarises the training and testing strategy.

#### *3.4.2. Experiment 2: Contest-specific effects*

This experiment follows the same training and testing procedure as Experiment 1 but includes the binary variable `AT_HOME` as a contest-level covariate. This variable is modelled as an order effect, estimated jointly with team strengths.

#### *3.4.3. Experiment 3: Team-specific, time-variant features*

This experiment introduces team-specific, time-varying predictors into the BT model. Models are trained using one or two seasons of data and tested on the subsequent season. The outcome variable and evaluation metric (classification accuracy) remain the same as in the previous experiments.

Predictors are grouped into three categories:

- **Match difficulty** (e.g., home/away status),

- **Performance indicators (PIs)** (e.g., cumulative disposals, inside 50s),
- **Form indicators** (e.g., recent wins, ladder position).

Each model is trained on two variants of these features: cumulative values over the entire season, and cumulatives over the last four games. To reduce dimensionality and improve interpretability, we apply feature selection as recommended by Fahey-Gilmour et al. (2019). Each feature is tested individually on the training set. Those significant at the 5% level are added to a candidate set. A final model is then built using backward elimination: features with the highest non-significant p-values are iteratively removed until only significant predictors remain. We record both the final selected features and those that were individually significant but excluded from the final model. Figure D.4 illustrates this process.

*3.4.4. Experiment 4: Round-by-round predictions*

The fourth experiment aims to mimic real-world forecasting by updating models progressively throughout the season. Rather than predicting the entire test season in advance, we generate predictions one round at a time, allowing model retraining after each round based on newly available data. This is closer to a dynamic modelling approach that accommodates time-varying effects.

We test four strategies for round-by-round forecasting:

- **Addition**: A model is trained on the full previous season and used to predict Round 1 of the next season. After each round, the new data are added to the training set, and the model is retrained before predicting the next round.
- **Substitution**: This is similar to Addition, but in each retraining step, the corresponding round from the previous season is dropped. This gives more weight to the current season while keeping the training set size constant.
- **Incremental**: The model from the previous season is used to predict the first three rounds of the test season, during which many differentials are zero and cannot be used for training. Starting from Round 4, a new model is trained using only the completed rounds of the test season. If no significant features are identified, predictions default to the previous model. Otherwise, the newly fitted model is used. Figures D.5–D.7 illustrate these three strategies.
- **Majority Voting**: This ensemble strategy combines predictions from the three strategies above, as well as from the models in Experiments 2 and 3. For each game, the final prediction corresponds to the majority vote (i.e., if at least three of the models predict a win, that outcome is selected).

Table 1: Estimated strength coefficients with the standard Bradley-Terry model, fitted to single seasons (2015-2023) and all available data. Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2015-23 |
| Adelaide | 0.14 | **1.55** | 1.15 | 0.04 | -0.29 | -1.58 | -0.77 | -1.07 | 0.00 | 0.00 |
| Brisbane Lions | **-2.34** | **-2.08** | -1.27 | **-1.74** | 0.72 | **1.80** | 0.75 | 0.78 | **1.31** | -0.09 |
| Carlton | **-2.26** | -0.94 | -1.00 | **-3.03** | -0.99 | -0.28 | -0.56 | 0.04 | 0.61 | **-0.63** |
| Collingwood | -0.79 | -0.21 | -0.26 | 0.66 | 0.81 | 0.43 | -0.95 | 0.82 | **1.75** | 0.25 |
| Essendon | **-1.57** | **-2.12** | 0.06 | 0.10 | 0.09 | -0.40 | -0.06 | -0.95 | -0.13 | -0.35 |
| Fremantle | 1.03 | -1.55 | -0.53 | -0.91 | -0.43 | -0.28 | -0.13 | 0.76 | -0.36 | -0.16 |
| Gold Coast | **-2.20** | -1.35 | -1.17 | **-2.27** | **-2.02** | -0.74 | -0.76 | -0.58 | -0.48 | **-0.95** |
| Geelong | 0.00 | **1.86** | 0.90 | 0.20 | 0.84 | 1.22 | 1.14 | **1.68** | -0.10 | **0.66** |
| Greater Western Sydney | -0.67 | **1.58** | 0.89 | 0.40 | 0.56 | 0.00 | 0.39 | -1.20 | 0.45 | 0.24 |
| Hawthorn | 0.92 | **1.63** | -0.07 | 0.41 | 0.08 | -0.87 | -0.65 | -0.96 | -0.97 | 0.00 |
| Melbourne | -1.36 | -0.01 | 0.17 | 0.42 | -1.32 | 0.28 | **2.04** | 0.83 | 0.73 | 0.20 |
| North Melbourne | 0.20 | 0.75 | -1.09 | -0.25 | -0.16 | -1.58 | **-1.48** | **-2.83** | **-2.25** | **-0.66** |
| Port Adelaide | 0.02 | 0.00 | 0.46 | 0.00 | 0.00 | **1.80** | **1.36** | -0.36 | 0.96 | **0.39** |
| Richmond | 0.31 | -0.49 | 1.09 | 1.33 | 1.25 | **1.67** | -0.13 | 0.07 | -0.21 | **0.49** |
| Saint Kilda | **-1.60** | 0.44 | 0.07 | **-1.86** | -0.68 | 0.61 | 0.00 | 0.00 | 0.15 | -0.22 |
| Sydney | 0.50 | **1.82** | 0.60 | 0.44 | -0.70 | -0.87 | 0.75 | 0.96 | 0.14 | 0.36 |
| Western Bulldogs | 0.05 | **1.76** | 0.00 | -0.93 | 0.12 | 0.43 | 1.21 | 0.05 | 0.09 | 0.26 |
| West Coast | 1.05 | 1.38 | 0.29 | 1.15 | 0.70 | 0.96 | -0.17 | **-2.79** | **-2.22** | 0.14 |

All strategies are implemented using both feature encodings (season cumulative and last-4-game cumulative), with feature selection applied as in Experiment 3. Classification accuracy is used to evaluate performance, consistent with Experiments 1–3.

## 4. Results

### *4.1. Team Strengths Based on Game Outcomes (Experiment 1)*

The estimated team strength coefficients from the standard BT model trained on individual seasons are summarised in Table 1. Models are fitted by maximum likelihood using the `BradleyTerry2` R package. For each season considered in Experiment 1, after excluding drawn matches, the strong-connectivity condition of Yan (2016) is satisfied, guaranteeing the existence of the maximum likelihood estimator. Tables E.12–E.14 report the corresponding estimates for windows of two, three, and four seasons.

Not all coefficients are statistically significant in each season. For instance, no team was significantly stronger or weaker than the reference team in 2017. The direction of the coefficients also varies across seasons: in 2015, all significant strengths were negative, while in 2020 they were positive. Models trained on the full 2015–2023 window identify teams with consistently strong or weak historical performance (e.g., Geelong and Gold Coast), but estimate near-zero strengths for teams with changing trajectories over time (e.g., Brisbane Lions, West Coast).

Table 2: AIC and classification accuracy for the standard Bradley-Terry model, fitted to windows of one season (2015-2023).

| Window: 1 season | | | | |
|---|---|---|---|---|
| Train Season | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015 | 2016 | 237.07 | 71.20% | 61.20% |
| 2016 | 2017 | 224.21 | 78.38% | 57.46% |
| 2017 | 2018 | 269.16 | 65.38% | 63.04% |
| 2018 | 2019 | 233.99 | 75.54% | 60.54% |
| 2019 | 2020 | 269.21 | 69.73% | 63.83% |
| 2020 | 2021 | 196.88 | 74.83% | 60.22% |
| 2021 | 2022 | 254.23 | 70.33% | 58.15% |
| 2022 | 2023 | 237.35 | 75.00% | 61.58% |
| 2023 | | 261.10 | 72.90% | |

Using longer training windows modifies which teams are identified as strong or weak. For example, Richmond appears significantly stronger from 2017 to 2020 in the two-season window model, though it was only significant in 2020 under single-season training. Carlton, weak in both 2015 and 2018 in the season-specific model, is revealed as consistently weak from 2015 to 2019 in the multi-season models. Three- and four-season windows reinforce previous patterns (e.g., Richmond, Geelong), and better capture evolving performance in teams like Brisbane Lions.

Table 2 shows AIC and classification accuracy for models trained on single seasons. The AIC values are broadly consistent within this setup, except for 2020, where a substantial decrease is observed—likely due to the shortened season. AIC increases with longer training windows (Table E.15), as expected when fitting more generalized models.

Classification accuracy on the training seasons ranges from 65.38% (2017) to 78.38% (2016), and typically declines as the training window expands. Test accuracy is consistently lower, averaging around 60% for one-season models and 60.98% when using all available data. Notably, certain seasons (e.g., 2018 at 65.76%, 2023 at 65.26%) achieve their highest accuracy with a three-season training window.

### *4.2. Effect of home team designation (Experiment 2)*

The AT_HOME effect was estimated alongside team strength coefficients in the contest-specific BT model. Results are shown in Table 3 and Tables E.17–E.19. While not significant in every season, AT_HOME is positive and significant when using all available data, indicating a consistent advantage for home teams under the BT formulation. Strength estimates remain broadly unchanged from Experiment 1.

The AT_HOME effect becomes consistently significant when training on windows of three or more seasons, suggesting that its influence is more robustly detected with larger datasets.

Table 3: Estimated AT_HOME Coefficient with the Bradley-Terry expansion for contest-specific effects, fitted to single seasons (2015-2023) and all available data. Significant coefficients at the 5% level are in bold.

| **Season** | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2015-23 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Feature (AT_HOME)** | 0.15 | **0.63** | **0.43** | 0.25 | **0.35** | 0.35 | 0.08 | **0.62** | **0.37** | **0.29** |

Table 4: AIC and classification accuracy for the Bradley-Terry expansion for contest-specific effects, fitted to windows of one season (2015-2023).

| Window: 1 season | | | | |
|---|---|---|---|---|
| Train Season | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015 | 2016 | 238.33 | 72.28% | 63.93% |
| 2016 | 2017 | 214.80 | 75.68% | 62.43% |
| 2017 | 2018 | 263.75 | 65.93% | 62.50% |
| 2018 | 2019 | 233.93 | 77.17% | 59.46% |
| 2019 | 2020 | 266.42 | 68.65% | 63.12% |
| 2020 | 2021 | 195.64 | 76.22% | 60.77% |
| 2021 | 2022 | 255.98 | 71.98% | 58.70% |
| 2022 | 2023 | 227.32 | 78.80% | 67.37% |
| 2023 | | 257.97 | 72.43% | |

Table 4 shows the AIC and classification accuracy for models trained on single seasons. AIC improves slightly compared to Experiment 1, indicating better model fit. Training accuracy is also marginally higher when using all data (61.92%; see Table E.20).

Test accuracy improves across most seasons. The highest improvement occurs in 2023, where the test accuracy reaches 67.37%, representing a meaningful boost over the standard BT model.

### *4.3. Effect of team-specific, time-variant features (Experiment 3)*

The results from the team-specific, time-variant BT models are summarised in Table 5 and Tables E.21–E.23, including both the final selected features and those found significant at the individual level but excluded from the models.

**Match difficulty features.** These features are significant in most seasons, with at least one included in all models except those trained on 2018 and 2021. In the model trained on all available data, INTERSTATE has a strong negative coefficient and AT_HOME a strong positive one, confirming home advantage and interstate disadvantage. HOMEGROUND, although not included in the all-data model, is individually significant in many cases and included in four two-season window models. Coefficients are generally stable across both feature encodings. Interactions between `AT_HOME` and other difficulty variables were tested but found significant only in 2015, and thus excluded from subsequent models.

**Form features.** Most form-related variables behave similarly under both encodings. LADDER_POSITION_DIFF is individually significant with a positive sign across many seasons but is never selected in the final model. In contrast, LADDERLY_POSITION_DIFF (from the previous season) appears in more final models, especially under last-4-game encodings, but only up to the 2020–21 window. PERCENTAGE_DIFF is often individually significant but rarely selected. POINTSFOR_DIFF and POINTSAGAINST_DIFF are frequently included in one- and two-season models and in both encoding types, also appearing in the all-data model with positive coefficients.

Some variables vary more across seasons. LG_WON (win in last game) is included only in the 2023 model but is individually significant in the full dataset. WINS_CUMULATIVE_DIFF is individually significant in several seasons but occasionally appears in final models with a negative coefficient.

**Performance indicators – last 4 games.** Most PIs are significant at the individual level across training windows and seasons. For last-4-game cumulative encodings:

- INSIDE50_L4_CSUM_DIFF appears in some one-season models with positive coefficients and is individually significant in most two-season windows.
- GOALS_SHOTS_L4_CSUM_DIFF is significant from 2018–2020 and included in the 2019 model, with confirmation from two-season windows.
- GETS_GROUNDBALL50_L4_CSUM_DIFF becomes significant from 2020 onward and is included in several models with a positive coefficient.
- INTERCEPTS_L4_CSUM_DIFF is consistently significant (but not selected) between 2021–2023.

Other relevant indicators include:

- METRES_GAINED_L4_CSUM_DIFF, significant in 2016–2019;
- POSSESSIONS_CONTESTED_L4_CSUM_DIFF, significant from 2018–2022;
- TACKLES_INSIDE50_L4_CSUM_DIFF, significant between 2020–2022;
- SCORE_LAUNCHES_L4_CSUM_DIFF, consistently significant from 2018, and included in models for 2022–23.

**Performance indicators – season cumulative.** Under full-season encodings, several patterns remain consistent:

- GOALS_SHOTS_CSUM_DIFF is significant and positively associated with winning in nearly every season from 2018 to 2022, though interestingly included with a negative coefficient in the 2017 model.
- Indicators like INSIDE50_CSUM_DIFF and MARKS_INSIDE50_CSUM_DIFF, reflecting activity in the Forward 50, are often selected with positive coefficients.
- REBOUND_INSIDE50S_CSUM_DIFF is significant with a negative sign in several seasons.
- SCORE_LAUNCHES_CSUM_DIFF is frequently significant and positively associated with winning.

Other season-level features showing consistent individual significance in two-season windows include:

- CONTEST_DEFENSIVE_LOSS_CSUM_DIFF (negative, 2018–21),
- GETS_GROUNDBALL50_CSUM_DIFF (positive, 2016–22),
- METRES_GAINED_CSUM_DIFF and TACKLES_CSUM_DIFF, both positively associated with outcomes in earlier windows (up to 2018–19).

**Model performance.** Classification accuracies are summarised in Table 6 and Table E.24. Compared to previous experiments, model fit improves, especially when trained on all available data,achieving 68.54% accuracy for both encodings.

Training accuracy for individual seasons is often lower than in Experiment 2, although both encodings perform similarly within each window. For test accuracy, the last-4-game encoding performs better until 2019, after which the full-season encoding tends to outperform it, with a maximum of 69.38% in 2020. In two-season windows, both encodings produce similar test accuracies, with some outperforming single-season models.

### *4.4. Round-by-Round Prediction Accuracy (Experiment 4)*

Table 7 compares the classification accuracy of the round-by-round prediction strategies described in Section 3.4.4 with the models from previous experiments. Under the last-4-games encoding, the Majority Voting strategy achieves the highest accuracy in 2016 (69.57%), 2017, 2021, and 2022. The Addition strategy performs best in 2018 and 2019. The highest overall accuracy (70.05%) is achieved by the

Table 5: Estimated coefficients for the time-variant Bradley-Terry fitted to windows of one season and with PIs encoded as cumulatives over the previous four games (2015-2023). Significant coefficients are in bold and underlined. Significant coefficients at the individual level only are in italics.

| | Season | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **FEATURE** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** | **2015-23** |
| MATCH DIFFICULTY | | | | | | | | | | |
| AT_HOME | | **0.609** | **0.404** | | *0.303* | | | *0.521* | **0.527** | **0.187** |
| HOMEGROUND | *0.381* | *0.438* | | | **0.615** | | | **0.716** | *0.409* | *0.363* |
| INTERSTATE | **-0.545** | *-0.443* | *-0.469* | | *-0.530* | **-0.705** | | *-0.515* | *-0.437* | **-0.268** |
| FORM | | | | | | | | | | |
| CONSECUTIVE_LOSSES | | | | | | | **0.416** | | *-0.164* | *-0.139* |
| CONSECUTIVE_WINS | | | | | | | | | | *0.106* |
| L4G_WINS | | | | | | | *-0.280* | | *0.230* | *0.299* |
| LADDER_POSITION_DIFF | *0.039* | *0.061* | *0.038* | *0.030* | | *0.029* | | *0.050* | | *0.044* |
| LADDERLY_POSITION_DIFF | **0.047** | **0.058** | **0.025** | **0.045** | **0.031** | **0.040** | *0.036* | | *0.045* | **0.026** |
| LG_WON | | | | | | | | | **0.557** | *0.278* |
| PERCENTAGE_DIFF | | | *0.009* | *0.007* | | | **0.011** | *0.012* | | *0.009* |
| POINTSAGAINST_DIFF | | | **0.002** | | | | *0.002* | **0.002** | **0.001** | **0.001** |
| POINTSFOR_DIFF | **0.001** | | | **0.002** | *0.003* | **0.003** | | **0.002** | *0.001* | *0.001* |
| PIs | | | | | | | | | | |
| BOUNCES_L4_CSUM_DIFF | | | | | | | | | | *0.004* |
| CLANGERS_L4_CSUM_DIFF | | | | | | | | | | |
| CLEARANCES_CENTRE_L4_CSUM_DIFF | | | | | | | | | | *0.009* |
| CLEARANCES_L4_CSUM_DIFF | | | | | | | | | | *0.006* |
| CLEARANCES_STOPPAGE_L4_CSUM_DIFF | | | | | | | *0.012* | | | *0.005* |
| CONTEST_DEFENSIVE_LOSS_L4_CSUM_DIFF | | | | | *-0.018* | | | | | |
| CONTEST_DEFENSIVE_LOSS_RATE_L4_CSUM_DIFF | | | | | **-0.023** | | | | | *-0.008* |
| CONTEST_OFFENSIVE_WIN_L4_CSUM_DIFF | | | | | | | | | | *0.013* |
| CONTEST_OFFENSIVE_WIN_RATE_L4_CSUM_DIFF | | | | | | | | **0.025** | | |
| DISPOSALS_EFFECTIVE_L4_CSUM_DIFF | | | | | | | | | | |
| DISPOSALS_EFFICIENCY_L4_CSUM_DIFF | | | | | | | | | | |
| DISPOSALS_L4_CSUM_DIFF | | | | | | | | | | *0.001* |
| FREES_AGAINST_L4_CSUM_DIFF | | | | | | | | | | |
| GETS_GROUNDBALL_L4_CSUM_DIFF | | | | | | | *0.006* | | | *0.002* |
| GETS_GROUNDBALL50_L4_CSUM_DIFF | | | | | | | *0.018* | *0.014* | | *0.012* |
| GOALS_ACCURACY_L4_CSUM_DIFF | | | | | | | | | | |
| GOALS_SHOTS_L4_CSUM_DIFF | | | | *0.020* | **0.016** | *0.015* | | | | *0.013* |
| HANDBALLS_L4_CSUM_DIFF | | | | | | | | | | |
| HITOUTS_ADVANTAGE_L4_CSUM_DIFF | | | | | | | | | **-0.013** | |
| HITOUTS_ADVANTAGE_RATE_L4_CSUM_DIFF | | | | | | | | | | *0.005* |
| HITOUTS_WIN_RATE_L4_CSUM_DIFF | **0.005** | | | | | | | | *-0.006* | *0.002* |
| INSIDE50_L4_CSUM_DIFF | | **0.013** | | *0.014* | | | | | **0.009** | **0.004** |
| INTERCEPTS_L4_CSUM_DIFF | | | | | | | | | | *0.005* |
| KICK2HANDBALL_L4_CSUM_DIFF | | | | | | | | | | *0.400* |
| KICKS_EFFECTIVE_L4_CSUM_DIFF | | | | | | | | | | *0.001* |
| KICKS_EFFICIENCY_L4_CSUM_DIFF | | | | | | | | | | |
| KICKS_L4_CSUM_DIFF | | | | *0.004* | | | | | | *0.002* |
| MARKS_CONTESTED_L4_CSUM_DIFF | | | | *0.017* | | | | | | *0.010* |
| MARKS_INSIDE50_L4_CSUM_DIFF | | | | | | | | | | *0.010* |
| MARKS_INTERCEPT_L4_CSUM_DIFF | | | | | | | | | | *0.009* |
| MARKS_L4_CSUM_DIFF | | | | | | | | | | |
| MARKS_ONLEAD_L4_CSUM_DIFF | | | | | | | | | | |
| METRES_GAINED_L4_CSUM_DIFF | | | | *0.000* | | | | | | *0.000* |
| ONE_PERCENTERS_L4_CSUM_DIFF | | | | | | | | | | *0.002* |
| POSSESSIONS_CONTESTED_L4_CSUM_DIFF | | | | | | | | **0.008** | | *0.004* |
| POSSESSIONS_CONTESTED_RATE_L4_CSUM_DIFF | | | | | | | | | | *0.019* |
| POSSESSIONS_L4_CSUM_DIFF | | | | | | | | | | *0.001* |
| POSSESSIONS_UNCONTESTED_L4_CSUM_DIFF | | | | | | | | | | |
| PRESSURE_DEFENSEHALF_L4_CSUM_DIFF | | *-0.002* | | *-0.003* | | | | | | *-0.002* |
| PRESSURE_L4_CSUM_DIFF | | | | | | | | | | |
| REBOUND_INSIDE50S_L4_CSUM_DIFF | *-0.009* | | | | | | | | | *-0.006* |
| SCORE_LAUNCHES_L4_CSUM_DIFF | | | | *0.016* | *0.012* | | | | *0.012* | *0.014* |
| SPOILS_L4_CSUM_DIFF | | | | | | | | | | |
| TACKLES_INSIDE50_L4_CSUM_DIFF | | | | | | | | | | *0.009* |
| TACKLES_L4_CSUM_DIFF | | | | | | | | | | |
| TURNOVERS_L4_CSUM_DIFF | | | | | | | | | | |

Table 6: Classification accuracy for the Bradley-Terry team-specific, time-variant expansion, fitted to windows of one season (2015-2023) with both encodings of PIs.

| Window: 1 season | | | | | |
|---|---|---|---|---|---|
| | | Last 4 games cumulative | | Season cumulative | |
| Train Season | Test Season | Train Accuracy | Test Accuracy | Train Accuracy | Test Accuracy |
| 2015 | 2016 | 71.08% | 67.15% | 67.16% | 62.80% |
| 2016 | 2017 | 71.50% | 60.29% | 74.40% | 59.31% |
| 2017 | 2018 | 62.75% | 66.02% | 70.10% | 46.60% |
| 2018 | 2019 | 71.36% | 61.84% | 65.53% | 59.00% |
| 2019 | 2020 | 70.05% | 65.62% | 66.18% | 69.38% |
| 2020 | 2021 | 68.75% | 62.25% | 73.12% | 65.20% |
| 2021 | 2022 | 68.63% | 61.17% | 66.67% | 58.25% |
| 2022 | 2023 | 72.82% | 64.02% | 72.33% | 64.95% |
| 2023 | | 70.56 | | 69.16% | |

Incremental strategy in 2015, though this method performs considerably worse in most other seasons. The Substitution and Contest-Specific models do not outperform any of the others in any season.

When using season cumulative encoding, Majority Voting yields lower accuracy overall, though it still achieves the best results in 2019 and 2022. Full-season models (from Experiment 3) produce the highest accuracy in four seasons, outperforming the other strategies in 2020 and 2021. The Incremental approach again achieves the best score for 2015 (71.5%) but performs worst in 2021 (50.49%).

Table E.25 summarises the number of correctly predicted games in the Finals Series. With a few exceptions (notably 2019 and 2022), models generally struggle to predict these games as accurately as those in the Home-and-Away season, likely due to the more balanced strength of qualifying teams.

Finally, team-level accuracy (Table E.26) shows that models tend to predict outcomes more reliably for the strongest and weakest teams. For example, predictions for North Melbourne (2021–23), Brisbane Lions (2015–18), and Richmond (2018–19) are notably more accurate, likely due to their consistent performance profiles during those periods. However, predictions for mid-ranked teams are sometimes equally or more accurate, suggesting that the model's performance by team strength is inconclusive at this stage.

## 5. Discussion

### *5.1. Summary and discussion of findings*

The results from the standard BT model confirm that game outcomes alone are insufficient to fully explain team strength. However, the model successfully identifies teams significantly stronger or weaker than average. For instance, it captures North

Table 7: Classification accuracy for the Bradley-Terry team-specific, time-variant (TS-TV) expansion, fitted to windows of one season (2015-2023) with both encodings of PIs. The highest accuracies for each season are underlined and in bold.

| | | | Last 4 games cumulative | | | | |
|---|---|---|---|---|---|---|---|
| Train Season | Test Season | Contest-Specific | TS-TV | Addition | Substitution | Incremental | Majority Voting |
| 2015 | 2016 | 63.93% | 67.15% | 66.18% | 67.15% | **70.05%** | 69.57% |
| 2016 | 2017 | 62.43% | 60.29% | 61.76% | 59.31% | 51.96% | **63.76%** |
| 2017 | 2018 | 62.50% | 66.02% | **67.48%** | 61.65% | 64.56% | 64.56% |
| 2018 | 2019 | 59.46% | 61.84% | **66.18%** | 62.80% | 58.94% | 65.70% |
| 2019 | 2020 | 63.12% | **65.62%** | 65.00% | 60.00% | 55.63% | 61.88% |
| 2020 | 2021 | 60.77% | 62.25% | 61.27% | 60.78% | 57.85% | **63.73%** |
| 2021 | 2022 | 58.70% | 61.17% | 61.65% | 60.68% | 60.68% | **65.05%** |
| 2022 | 2023 | 67.37% | 64.02% | **67.76%** | 64.49% | 53.27% | 64.49% |
| | | | Season cumulative | | | | |
| Train Season | Test Season | Contest-Specific | TS-TV | Addition | Substitution | Incremental | Majority Voting |
| 2015 | 2016 | 63.93% | 62.80% | 67.15% | 64.25% | **71.50%** | 65.22% |
| 2016 | 2017 | **62.43%** | 59.31% | 57.35% | 61.27% | 50.49% | 60.29% |
| 2017 | 2018 | 62.50% | 46.60% | 65.53% | 58.74% | **66.50%** | 64.08% |
| 2018 | 2019 | 59.46% | 59.90% | **64.25%** | 62.32% | 57.49% | **64.25%** |
| 2019 | 2020 | 63.12% | **69.38%** | 68.12% | 61.88% | 60.63% | 66.25% |
| 2020 | 2021 | 60.77% | **65.20%** | 60.29% | 55.39% | 50.49% | 60.78% |
| 2021 | 2022 | 58.70% | 58.25% | 66.99% | 65.06% | 63.11% | **67.96%** |
| 2022 | 2023 | **67.37%** | 64.95% | 62.62% | 61.21% | 60.75% | 62.62% |

Melbourne's poor performance between 2021–23, the Brisbane Lions' evolution from weak to strong, and the best and worst teams in the 2023 season. Multi-season windows (e.g., three or four seasons) help uncover medium-term trends, such as Richmond's dominance between 2017–2020. However, wider windows may smooth over performance changes, as seen with Brisbane or West Coast. This suggests team strengths evolve gradually, typically within a 2–3 year window, and that draft-based advantages for weaker teams do not manifest immediately.

As expected, predictive accuracy declines as training windows widen, likely because the standard model considers only outcomes and cannot account for performance shifts between seasons.

The AT_HOME effect is consistently positive and significant, aligning with findings in Lazarus et al. (2018) and Robertson and Joyce. Its inclusion improves model fit and accuracy with minimal distortion to team strength estimates, confirming that home advantage is a meaningful, independent effect.

Other difficulty features, such as INTERSTATE and HOMEGROUND, also show consistent influence across encodings, suggesting they operate independently of performance features. Playing interstate reduces win probability, while playing at a team's home ground increases it.

Form-related predictors show mixed behaviour. While current ladder position differentials are individually significant, they are rarely selected in final models. In contrast, previous season rankings are more predictive, especially in models trained on shorter windows. Differentials in points for and against, reflecting offensive and defensive capabilities, are both positively associated with winning.

The performance indicators (PIs) demonstrate strong predictive value. Across both encodings and training windows, cumulative differentials between teams are significant for many key actions. Indicators related to Forward 50 activity (e.g., entries, marks, tackles), goal shots, and scoring chains are especially influential. These reflect a team's ability to penetrate the opponent's defensive zone and convert opportunities into points. Other relevant features include metres gained (signalling territorial dominance) and rebounds, which are often higher for weaker teams under pressure.

### *5.2. Changing coefficients and eras*

Some performance indicators exhibit inconsistent effects across seasons. For instance, WINS_CUMULATIVE_DIFF is occasionally associated with lower win probabilities, counterintuitive at first glance. This may stem from early-season dynamics, when differentials are low, or scheduling asymmetries (e.g., a losing team with more cumulative wins due to easier fixtures). Finals Series matches may also contribute to coefficient shifts, as they feature strong teams facing each other in atypical matchups.

These findings also support the presence of evolving "eras" in AFL. Between 2018–2020, scoring-related advantages (e.g., goal shots, score launches) were more predictive, suggesting an era favouring offensive output. From 2020 onwards, performance in the Forward 50, particularly in contested possessions, gained relevance, indicating a strategic shift toward pressure and territory control. These results reinforce the literature on medium-term tactical trends in AFL.

### *5.3. Predictive ability of the Bradley-Terry expansions*

Incorporating time-variant, team-specific features significantly improves model fit and predictive accuracy. Compared to the standard and contest-specific models, these expansions better capture the dynamic nature of team strength.

Across seasons, prediction accuracy reaches up to 71.5%. Round-by-round strategies—particularly those retraining the model during the season—often yield higher test accuracy than static models. However, in some cases, full-season models still outperform dynamic ones. Majority voting strategies offer consistent accuracy above 60%, peaking at 67.96%, and demonstrate the benefit of ensemble forecasting in this context.

Overall, the results confirm that interpretable BT models, augmented with pregame features, can deliver meaningful and competitive predictive performance in AFL forecasting.

### *5.4. Limitations and future lines of research*

This study focuses exclusively on the men's AFL competition from 2015 to 2023. While this period captures recent tactical shifts and includes a rich dataset, the findings may not generalize to earlier seasons, the AFL Women's league, or regional competitions. Future work could assess whether key predictive features and model performance vary across leagues or genders.

Prediction accuracy is lower in the Finals Series, likely due to more balanced matchups between high-performing teams. While separating these games from the main season may improve clarity, the small sample size limits the feasibility of modelling them independently.

Our results suggest that BT models perform particularly well when teams are clearly stronger or weaker than their opponents. However, the models' ability to predict outcomes for evenly matched teams remains inconclusive. This could reflect limitations in the current encoding of features or genuine volatility in those contests. Investigating predictive variance by team strength or introducing smoothed or weighted inputs may improve performance.

There is also potential to refine the modelling framework itself. BT expansions could be compared with post-game prediction models, tested with more flexible feature selection (e.g., LASSO regularization or mixed-effect formulations), or extended with proprietary data sources such as player tracking or network metrics. An additional extension would be the use of dynamic Bradley–Terry models, in which team strengths evolve according to an explicit temporal process (e.g. Du et al., 2023). Such approaches could borrow strength across seasons and smooth year-to-year variation in estimated abilities, as proposed in related work on dynamic networks.

Finally, adapting BT models to predict score margins or component scoring (goals and behinds) could offer a richer understanding of performance, especially in a high-scoring sport like AFL.

## 6. Conclusions

This study shows that Bradley-Terry models, especially those incorporating time-variant and team-specific features, can offer interpretable, data-efficient tools to predict AFL match outcomes in advance. Unlike complex machine learning models, BT models provide transparent estimates of team strength and the effect of pre-game conditions, making them well suited for coaching and performance analysis.

Our results demonstrate that strength evolves gradually across seasons, rather than shifting dramatically from year to year. This has implications for list management and long-term strategic planning. Historical ladder positions, rather than recent wins alone, are often more predictive of future success.

Pre-game difficulty factors, such as playing away, interstate, or at unfamiliar grounds, show consistent, independent effects. Their impact remains even after accounting for relative strength, highlighting their importance in fixture analysis and match preparation. Performance indicators, particularly those capturing Forward 50 activity and scoring intent, are robust predictors of success and could inform targeted training or scouting strategies.

Overall, this work suggests that interpretable models like BT are not only statistically sound but practically valuable. They can help clubs understand their evolving strengths, anticipate upcoming challenges, and design data-informed approaches to improve competitive performance.


## Funding Statement

The authors report that there is no funding to declare.


## Conflict of Interest Statement

The authors report there are no competing interests to declare.

## Data Availability Statement

The data is available in R. The code to download and analyze it is available at https://github.com/charlieceratops/AFL_BradleyTerry.

## References


Agresti, A., 2012. Categorical Data Analysis. 3rd ed., Wiley, New York.

Braham, C., Small, M., 2018. Complex networks untangle competitive advantage in Australian football. Chaos: An Interdisciplinary Journal of Nonlinear Science 28.

David, H., 1988. The Method of Paired Comparisons. 2nd ed., Oxford University Press, Oxford.

Du, Y., Qu, L., Yan, T., Zhang, Y., 2023. Time-varying $\beta$-model for dynamic directed networks. Scandinavian Journal of Statistics 50, 1687–1715.

Fahey-Gilmour, J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., 2019. Multifactorial analysis of factors influencing elite Australian football match outcomes: A machine learning approach. International Journal of Computer Science in Sport 18, 100–124.

Firth, D., Turner, H., 2012. Bradley-Terry models in R: The BradleyTerry2 package. Journal of Statistical Software 48.

Fransen, J., Tribolet, R., Sheehan, W.B., McBride, I., Novak, A.R., Watsford, M.L., 2022. Cooperative passing network features are associated with successful match outcomes in the Australian Football League. International Journal of Sports Science & Coaching 17, 1101–1108.

Lane, J.C., van der Ploeg, G., Greenham, G., Norton, K., 2020. Characterisation of offensive and defensive game play trends in the Australian Football League (1999–2019). International Journal of Performance Analysis in Sport 20, 557–568.

Lazarus, B.H., Hopkins, W.G., Stewart, A.M., Aughey, R.J., 2018. Factors affecting match outcome in elite Australian football: A 14-year analysis. International Journal of Sports Physiology & Performance 13.

Leitner, M.C., Daumann, F., Follert, F., Richlan, F., 2023. The cauldron has cooled down: A systematic literature review on home advantage in football during the COVID-19 pandemic from a socio-economic and psychological perspective. Management Review Quarterly 73, 605–633.

Manderson, A., Murray, K., Turlach, B., 2018. Dynamic Bayesian forecasting of AFL match results using the Skellam distribution. Australian & New Zealand Journal of Statistics 60, 174–187.

Nguyen, R., Day, J., Warton, D., Lane, O., 2021. fitzRoy-An R package to encourage reproducible sports analysis. The R Journal 12, 82–106.

Robertson, S., Back, N., Bartlett, J.D., 2016. Explaining match outcome in elite Australian rules football using team performance indicators. Journal of Sports Sciences 34, 637–644.

Robertson, S., Joyce, D., . Evaluating strategic periodisation in team sport. Journal of Sports Sciences 36, 279–285.

Sargent, C., Rogalski, B., Montero, A., Roach, G.D., 2022. The sleep behaviors of elite Australian rules footballers before and after games during an entire season. International Journal of Sports Physiology and Performance 17, 932–942.

Sargent, J., Bedford, A., 2013. Evaluating Australian football league player contributions using interactive network simulation. Journal of Sports Science & Medicine 12, 116.

Tsokos, A., Narayanan, S., Kosmidis, I., Baio, G., Cucuringu, M., Whitaker, G., Király, F., 2019. Modeling outcomes of soccer matches. Machine Learning 108, 77–95.

Woods, C.T., Robertson, S., Collier, N.F., 2017. Evolution of game-play in the Australian Football League from 2001 to 2015. Journal of Sports Sciences 35, 1879–1887.

Yan, T., 2016. Ranking in the generalized Bradley–Terry models when the strong connection condition fails. Communications in Statistics-Theory and Methods 45, 340–353.

Yan, T., 2025. Inference in a generalized Bradley-Terry model for paired comparisons with covariates and a growing number of subjects. arXiv preprint arXiv:2507.22472 .

Yan, T., Li, Y., Xu, J., Yang, Y., Zhu, J., 2025. Likelihood ratio tests in random graph models with increasing dimensions. Journal of the American Statistical Association , 1–14.

Yan, T., Yang, Y., Xu, J., 2012. Sparse paired comparisons in the Bradley-Terry model. Statistica Sinica , 1305–1318.

Young, C., Luo, W., Gastin, P., Tran, J., Dwyer, D., 2019a. Modelling match outcome in Australian football: Improved accuracy with large databases. International Journal of Computer Science in Sport 18, 80–92.

Young, C.M., Luo, W., Gastin, P., Lai, J., Dwyer, D.B., 2019b. Understanding effective tactics in Australian football using network analysis. International Journal of Performance Analysis in Sport 19, 331–341.

Young, C.M., Luo, W., Gastin, P.B., Dwyer, D.B., 2020. Understanding the relative contribution of technical and tactical performance to match outcome in Australian football. Journal of Sports Sciences 38, 676–681.

# Appendix A. Summary of works on machine learning approaches in AF

| Work | Models | Features | Data | Comments |
|---|---|---|---|---|
| Braham and Small (2018) | Linear Regression | Differentials in mean betweenness last 4 games, outdegree | 2014 AFL Season (207 games). From Champion Data | Games of AF can be modelled as passing networks, with players as nodes and weighted edges as passes |
| Fahey-Gilmour et al. (2019) | GLM with elastic net regularization | Differentials ladder position last season, player-based and team-based rating. Pre-game features | 2013-18 AFL Season (1,241 games). From fitzRoy package. Player data from Champion Data | Classify features as team-related and game-related. Non-linear models and approaches including feature selection perform better |
| Fransen et al. (2022) | Binomial generalized linear mixed effects regression | Linear combinations of network variables representing connectedness, in-degree variability, out-degree variability | Network data from Champion Data (1,629 observations) | Combine network measures of passing networks as connectedness, in-degree variability and out-degree variability. One-unit increases in connectedness increase the probability of winning by 5.3% |
| Lane et al. (2020) | Calculate z-scores of offensive and defensive variables | Offensive (inside 50s, metres gained, etc.) and defensive features (tackles, rebound 50s, clangers, etc.) | Performance Indicator Data from AFL seasons 1999-2019 | Identify eras in the AFL with significant differences in PIs. They identify an era of an offensive game (2007-08) moving to a defensive game (until 2011) with more tackles and disposals but lower scores |
| Lazarus et al. (2018) | Logistic mixed-effect regression | Playing Home/Away, travelling, age and weight differentials, days between games. | Data for AFL seasons 2000-13 (5,109 games). From AFL Tables | Older and heavier teams are found to have significantly yet slightly higher chances of winning. Playing as the away team significantly yet slightly decreases the chances of winning |
| Manderson et al. (2018) | Bayesian hierarchical model with Skellam distribution | Score, goals and behinds for each team. | Home-and-Away data from AFL seasons 2013-15. From Footywire | Pre-game prediction of the number of goals and behinds separately. It performed slightly worse in the last games of the season |
| Robertson et al. (2016) | Logistic regression | Differentials in Performance Indicators (kicks, goal conversions, inside 50s, etc.) | Performance Indicator data from AFL 2013-14 Season (396 games). Unspecified freely available sources | 81% of teams with a kick differential above -1 won their games. 90% of teams with higher additional kicks and goal conversion won their games. Differentials in the inside 50s, marks, marks inside 50s and contested possessions were the most important predictors. |
| Robertson and Joyce | Logistic regression | Number of wins over the past 4 games, and differentials in ladder rankings from the current and previous seasons | AFL difficulty factors from the 2014 pre-season (198 games). From Champion Data and AFL Stats | Teams playing away show decreased chances of winning, especially if they play interstate |
| Sargent and Bedford (2013) | Linear regression | Team ratings based on eigenvector centrality | Games from AFL 2011 season, Geelong Football Club | Higher team connectedness increases the score margin. The score margin can be predicted with pre-game connectedness values, to the point that the individual impact of players can be estimated |
| Sargent et al. (2022) | Linear mixed effects model | Sleep time, sleep efficiency, time between sleeping and waking up | Sleep data from 37 male AFL players | Confirm that players sleep less when playing interstate. Mention the lack of studies relating poorer sleep to game outcomes |
| Woods et al. (2017) | Non-metric multidimensional scaling | Handballs, disposals, uncontested possessions, clangers, marks | Performance Indicators from AFL 2001-15 seasons. From Champion Data | Identify eras in the AFL: a 'possession' era (2005-09 seasons), a 'defensive' era (2010-13) and a 'repossession' (2014-onwards) |
| Young et al. (2019b) | Parametric tests for network measures | Edge count, edge density, average path length, degree centrality, eigenvector centrality, betweenness centrality, transitivity | Games from AFL 2009-16 seasons (1,516 games). From Champion Data | Create passing networks based on player interactions. No significant differences are found in the betweenness centrality. Other network measures were significant but only slightly correlated to the score margin. |
| Young et al. (2019a) | Random Forest | Differentials in the time in possession and meters gained, along with relative disposals, Inside 50s, Marks Inside 50, along the regular Marks and Contested Possessions | Performance Indicator data from AFL 2001-16 seasons. From Champion Data | Raise concerns about the lack of consensus on the identification of eras in the AFL |
| Young et al. (2020) | Decision trees and GLMs | Differentials in metres gained, time in possession, kicks | Performance Indicator data (52 PIs) from AFL 2009-16 seasons (1,516 games). From Champion Data. 14 Tactical Indicators form event data representing passes | Tactical features are more important but they do not increase accuracy |

# Appendix B. Glossary of the used features

Table B.8: Glossary of the used features - Part I.

| Type | Variable | Description |
|---|---|---|
| TARGET | WIN | Indicates whether the team won (1) or lost (0) |
| TARGET | AWAY_WIN | Indicates whether the Away team won (1) or lost (0) |
| TARGET | HOME_WIN | Indicates whether the Home team won (1) or lost (0) |
| MATCH DIFFICULTY | AT_HOME | Indicates whether the team was designated as the Home Team (1) or the Away Team (0) |
| MATCH DIFFICULTY | HOMEGROUND | Indicates whether the team was played at their Home Ground (1) or not (0) |
| MATCH DIFFICULTY | INTERSTATE | Indicates whether the team was played outside their State (1) or not (0) |
| FORM | CONSECUTIVE_LOSSES | Number of consecutive losses |
| FORM | CONSECUTIVE_WINS | Number of consecutive wins |
| FORM | L4G_WINS | Number of wins over the previous 4 games |
| FORM | LADDER_POSITION_DIFF | Differential in ladder position compared to the other team at the time of the game |
| FORM | LADDERLY_POSITION_DIFF | Differential in ladder position of last year compared to the other team |
| FORM | LG_WON | Indicates whether the team won their previous game (1) or not (0) |
| FORM | PERCENTAGE_DIFF | Differential in Points For to Points Against ratio compared to the other team |
| FORM | POINTSAGAINST_DIFF | Differential in cumulative points scored against the team compared to the other team |
| FORM | POINTSFOR_DIFF | Differential in cumulative points scored in favour of the team compared to the other team |
| FORM | WINS_CUMULATIVE_DIFF | Differential in cumulative wins compared to the other team |
| PI | BOUNCES_L4_CSUM_DIFF | Differential in the cumulative sum of bounces over the previous 4 games, compared to the other team |
| PI | CLANGERS_L4_CSUM_DIFF | Differential in the cumulative sum of clangers (mistakes) over the previous 4 games, compared to the other team |
| PI | CLEARANCES_CENTRE_L4_CSUM_DIFF | Differential in the cumulative sum of centre clearances (clearing the centre area) over the previous 4 games, compared to the other team |
| PI | CLEARANCES_L4_CSUM_DIFF | Differential in the cumulative sum of clearances (clearing the centre or stoppage area) over the previous 4 games, compared to the other team |
| PI | CLEARANCES_STOPPAGE_L4_CSUM_DIFF | Differential in the cumulative sum of stoppage clearances (clearing the stoppage area) over the previous 4 games, compared to the other team |
| PI | CONTEST_DEFENSIVE_LOSS_L4_CSUM_DIFF | Differential in the cumulative losses of one-to-one contests (moments where two opposing players can get the ball) over the previous 4 games, compared to the other team |
| PI | CONTEST_DEFENSIVE_LOSS_RATE_L4_CSUM_DIFF | Differential in the rate of losses of one-to-one contests (moments where two opposing players can get the ball) over the previous 4 games, compared to the other team |
| PI | CONTEST_OFFENSIVE_WIN_L4_CSUM_DIFF | Differential in the cumulative wins of one-to-one contests (moments where two opposing players can get the ball) over the previous 4 games, compared to the other team |
| PI | CONTEST_OFFENSIVE_WIN_RATE_L4_CSUM_DIFF | Differential in the rate of wins of one-to-one contests (moments where two opposing players can get the ball) over the previous 4 games, compared to the other team |
| PI | DISPOSALS_EFFECTIVE_L4_CSUM_DIFF | Differential in the cumulative sum of successful disposals (kicks and handballs) over the previous 4 games, compared to the other team |
| PI | DISPOSALS_EFFICIENCY_L4_CSUM_DIFF | Differential in the rate of successful disposals (kicks and handballs) over the previous 4 games, compared to the other team |
| PI | DISPOSALS_L4_CSUM_DIFF | Differential in the cumulative sum of disposals (kicks and handballs) over the previous 4 games, compared to the other team |
| PI | FREES_AGAINST_L4_CSUM_DIFF | Differential in the cumulative sum of free kicks against the team over the previous 4 games, compared to the other team |
| PI | GETS_GROUNDBALL_L4_CSUM_DIFF | Differential in the cumulative sum of contested possessions won in the ground over the previous 4 games, compared to the other team |
| PI | GETS_GROUNDBALL50_L4_CSUM_DIFF | Differential in the cumulative sum of contested possessions won in the ground of the Forward 50 area over the previous 4 games, compared to the other team |
| PI | GOALS_ACCURACY_L4_CSUM_DIFF | Differential in the goals-to-goal shots ratio over the previous 4 games, compared to the other team |

Table B.9: Glossary of the used features - Part II.

| **Type** | **Variable** | **Description** |
|---|---|---|
| PI | GOALS_SHOTS_L4_CSUM_DIFF | Differential in the cumulative sum of goals over the previous 4 games, compared to the other team |
| PI | HANDBALLS_L4_CSUM_DIFF | Differential in the cumulative sum of handballs over the previous 4 games, compared to the other team |
| PI | HITOUTS_ADVANTAGE_L4_CSUM_DIFF | Differential in the cumulative sum of hitouts to advantage (knocking the ball to a teammate in a contest after a stoppage) over the previous 4 games, compared to the other team |
| PI | HITOUTS_ADVANTAGE_RATE_L4_CSUM_DIFF | Differential in the ratio of hitouts to advantage to hitouts over the previous 4 games, compared to the other team |
| PI | HITOUTS_WIN_RATE_L4_CSUM_DIFF | Differential in the ratio of contests with a hitout win to contests over the previous 4 games, compared to the other team |
| PI | INSIDE50_L4_CSUM_DIFF | Differential in the cumulative sum of entries into the Forward 50 area over the previous 4 games, compared to the other team |
| PI | INTERCEPTS_L4_CSUM_DIFF | Differential in the cumulative sum of intercepts (taking the ball from the other team) over the previous 4 games, compared to the other team |
| PI | KICK2HANDBALL_L4_CSUM_DIFF | Differential in the ratio of kicks to handballs over the previous 4 games, compared to the other team |
| PI | KICKS_EFFECTIVE_L4_CSUM_DIFF | Differential in the cumulative sum of successful kicks over the previous 4 games, compared to the other team |
| PI | KICKS_EFFICIENCY_L4_CSUM_DIFF | Differential in the ratio of effective kicks to all kicks over the previous 4 games, compared to the other team |
| PI | KICKS_L4_CSUM_DIFF | Differential in the cumulative sum of kicks over the previous 4 games, compared to the other team |
| PI | MARKS_CONTESTED_L4_CSUM_DIFF | Differential in the cumulative sum of marks taken under pressure over the previous 4 games, compared to the other team |
| PI | MARKS_INSIDE50_L4_CSUM_DIFF | Differential in the cumulative sum of marks taken in the Forward 50 area over the previous 4 games, compared to the other team |
| PI | MARKS_INTERCEPT_L4_CSUM_DIFF | Differential in the cumulative sum of marks taken from the opponent over the previous 4 games, compared to the other team |
| PI | MARKS_L4_CSUM_DIFF | Differential in the cumulative sum of marks over the previous 4 games, compared to the other team |
| PI | MARKS_ONLEAD_L4_CSUM_DIFF | Differential in the cumulative sum of marks taken under no pressure over the previous 4 games, compared to the other team |
| PI | METRES_GAINED_L4_CSUM_DIFF | Differential in the cumulative sum of metres gained over the previous 4 games, compared to the other team |
| PI | ONE_PERCENTERS_L4_CSUM_DIFF | Differential in the cumulative sum of one-percenters (defensive but rare actions) over the previous 4 games, compared to the other team |
| PI | POSSESSIONS_CONTESTED_L4_CSUM_DIFF | Differential in the cumulative sum of contested possessions (possessions gained under dispute) over the previous 4 games, compared to the other team |
| PI | POSSESSIONS_CONTESTED_RATE_L4_CSUM_DIFF | Differential in the ratio of possessions won in a dispute to all disputes over the previous 4 games, compared to the other team |
| PI | POSSESSIONS_L4_CSUM_DIFF | Differential in the cumulative sum of possessions over the previous 4 games, compared to the other team |
| PI | POSSESSIONS_UNCONTESTED_L4_CSUM_DIFF | Differential in the cumulative sum of uncontested possessions (possessions gained under no pressure) over the previous 4 games, compared to the other team |
| PI | PRESSURE_DEFENSEHALF_L4_CSUM_DIFF | Differential in the cumulative sum of pressure acts in the Defense Half of the ground over the previous 4 games, compared to the other team |
| PI | PRESSURE_L4_CSUM_DIFF | Differential in the cumulative sum of pressure acts over the previous 4 games, compared to the other team |
| PI | REBOUND_INSIDE50S_L4_CSUM_DIFF | Differential in the cumulative sum of rebound 50s (moving the ball from the Defensive 50 to the midfield) over the previous 4 games, compared to the other team |
| PI | SCORE_LAUNCHES_L4_CSUM_DIFF | Differential in the cumulative sum of score launches (scoring chains after successful clearances, intercepts, free kicks, hitouts) over the previous 4 games, compared to the other team |
| PI | SPOILS_L4_CSUM_DIFF | Differential in the cumulative sum of spoils (preventing opponents from taking a mark) over the previous 4 games, compared to the other team |
| PI | TACKLES_INSIDE50_L4_CSUM_DIFF | Differential in the cumulative sum of tackles in the Forward 50 area over the previous 4 games, compared to the other team |
| PI | TACKLES_L4_CSUM_DIFF | Differential in the cumulative sum of tackles over the previous 4 games, compared to the other team |
| PI | TURNOVERS_L4_CSUM_DIFF | Differential in the cumulative sum of turnovers (losses of possession) over the previous 4 games, compared to the other team |

Table B.10: Glossary of the used features - Part III.

| **Type** | **Variable** | **Description** |
|---|---|---|
| PI | BOUNCES_CSUM_DIFF | Differential in the cumulative sum of bounces over the season, compared to the other team. |
| PI | CLANGERS_CSUM_DIFF | Differential in the cumulative sum of clangers (mistakes) over the season, compared to the other team |
| PI | CLEARANCES_CENTRE_CSUM_DIFF | Differential in the cumulative sum of centre clearances (clearing the centre area) over the season, compared to the other team |
| PI | CLEARANCES_CSUM_DIFF | Differential in the cumulative sum of clearances (clearing the centre or stoppage area) over the season, compared to the other team |
| PI | CLEARANCES_STOPPAGE_CSUM_DIFF | Differential in the cumulative sum of stoppage clearances (clearing the stoppage area) over the season, compared to the other team |
| PI | CONTEST_DEFENSIVE_LOSS_CSUM_DIFF | Differential in the cumulative losses of one-to-one contests (moments where two opposing players can get the ball) over the season, compared to the other team |
| PI | CONTEST_DEFENSIVE_LOSS_RATE_CSUM_DIFF | Differential in the rate of losses of one-to-one contests (moments where two opposing players can get the ball) over the season, compared to the other team |
| PI | CONTEST_OFFENSIVE_WIN_CSUM_DIFF | Differential in the cumulative wins of one-to-one contests (moments where two opposing players can get the ball) over the season, compared to the other team |
| PI | CONTEST_OFFENSIVE_WIN_RATE_CSUM_DIFF | Differential in the rate of wins of one-to-one contests (moments where two opposing players can get the ball) over the season, compared to the other team |
| PI | DISPOSALS_CSUM_DIFF | Differential in the cumulative sum of successful disposals (kicks and handballs) over the season, compared to the other team |
| PI | DISPOSALS_EFFECTIVE_CSUM_DIFF | Differential in the rate of successful disposals (kicks and handballs) over the season, compared to the other team |
| PI | DISPOSALS_EFFICIENCY_CSUM_DIFF | Differential in the cumulative sum of disposals (kicks and handballs) over the season, compared to the other team |
| PI | FREES_AGAINST_CSUM_DIFF | Differential in the cumulative sum of free kicks against the team over the season, compared to the other team |
| PI | GETS_GROUNDBALL_CSUM_DIFF | Differential in the cumulative sum of contested possessions won in the ground over the season, compared to the other team |
| PI | GETS_GROUNDBALL50_CSUM_DIFF | Differential in the cumulative sum of contested possessions won in the ground of the Forward 50 area over the season, compared to the other team |
| PI | GOALS_ACCURACY_CSUM_DIFF | Differential in the goals-to-goal shots ratio over the season, compared to the other team |
| PI | GOALS_SHOTS_CSUM_DIFF | Differential in the cumulative sum of goals over the season, compared to the other team |
| PI | HANDBALLS_CSUM_DIFF | Differential in the cumulative sum of handballs over the season, compared to the other team |
| PI | HITOUTS_ADVANTAGE_CSUM_DIFF | Differential in the cumulative sum of hitouts to advantage (knocking the ball to a teammate in a contest after a stoppage) over the season, compared to the other team |
| PI | HITOUTS_ADVANTAGE_RATE_CSUM_DIFF | Differential in the ratio of hitouts to advantage to hitouts over the season, compared to the other team |
| PI | HITOUTS_WIN_RATE_CSUM_DIFF | Differential in the ratio of contests with a hitout win to contests over the season, compared to the other team |
| PI | INSIDE50_CSUM_DIFF | Differential in the cumulative sum of entries into the Forward 50 area over the season, compared to the other team |
| PI | INTERCEPTS_CSUM_DIFF | Differential in the cumulative sum of intercepts (taking the ball from the other team) over the season, compared to the other team |
| PI | KICK2HANDBALL_CSUM_DIFF | Differential in the ratio of kicks to handballs over the season, compared to the other team |
| PI | KICKS_CSUM_DIFF | Differential in the cumulative sum of kicks over the season, compared to the other team |
| PI | KICKS_EFFECTIVE_CSUM_DIFF | Differential in the cumulative sum of successful kicks over the season, compared to the other team |
| PI | KICKS_EFFICIENCY_CSUM_DIFF | Differential in the ratio of effective kicks to all kicks over the season, compared to the other team |
| PI | MARKS_CONTESTED_CSUM_DIFF | Differential in the cumulative sum of marks taken under pressure over the season, compared to the other team |
| PI | MARKS_CSUM_DIFF | Differential in the cumulative sum of marks over the season, compared to the other team |

Table B.11: Glossary of the used features - Part IV.

| Type | Variable | Description |
|---|---|---|
| PI | MARKS_INSIDE50_CSUM_DIFF | Differential in the cumulative sum of marks taken in the Forward 50 area over the season, compared to the other team |
| PI | MARKS_INTERCEPT_CSUM_DIFF | Differential in the cumulative sum of marks taken from the opponent over the season, compared to the other team |
| PI | MARKS_ONLEAD_CSUM_DIFF | Differential in the cumulative sum of marks taken under no pressure over the season, compared to the other team |
| PI | METRES_GAINED_CSUM_DIFF | Differential in the cumulative sum of metres gained over the season, compared to the other team |
| PI | ONE_PERCENTERS_CSUM_DIFF | Differential in the cumulative sum of one-percenters (defensive but rare actions) over the season, compared to the other team |
| PI | POSSESSIONS_CONTESTED_CSUM_DIFF | Differential in the cumulative sum of contested possessions (possessions gained under dispute) over the season, compared to the other team |
| PI | POSSESSIONS_CONTESTED_RATE_CSUM_DIFF | Differential in the ratio of possessions won in a dispute to all disputes over the season, compared to the other team |
| PI | POSSESSIONS_CSUM_DIFF | Differential in the cumulative sum of possessions over the season, compared to the other team |
| PI | POSSESSIONS_UNCONTESTED_CSUM_DIFF | Differential in the cumulative sum of uncontested possessions (possessions gained under no pressure) over the season, compared to the other team |
| PI | PRESSURE_CSUM_DIFF | Differential in the cumulative sum of pressure acts over the season, compared to the other team |
| PI | PRESSURE_DEFENSEHALF_CSUM_DIFF | Differential in the cumulative sum of pressure acts in the Defense Half of the ground over the season, compared to the other team |
| PI | REBOUND_INSIDE50S_CSUM_DIFF | Differential in the cumulative sum of rebound 50s (moving the ball from the Defensive 50 to the midfield) over the season, compared to the other team |
| PI | SCORE_LAUNCHES_CSUM_DIFF | Differential in the cumulative sum of score launches (scoring chains after successful clearances, intercepts, free kicks, hitouts) over the season, compared to the other team |
| PI | SPOILS_CSUM_DIFF | Differential in the cumulative sum of spoils (preventing opponents from taking a mark) over the previous 4 games, compared to the other team |
| PI | TACKLES_CSUM_DIFF | Differential in the cumulative sum of tackles over the season, compared to the other team |
| PI | TACKLES_INSIDE50_CSUM_DIFF | Differential in the cumulative sum of tackles in the Forward 50 area over the season, compared to the other team |
| PI | TURNOVERS_CSUM_DIFF | Differential in the cumulative sum of turnovers (losses of possession) over the season, compared to the other team |

## Appendix C. Data preparation

Following Firth and Turner (2012), game outcomes were turned into binomial frequencies, where each entry pairs the Home and the Away Team retrieving the absolute frequencies of wins for each team against the other. For instance, for the AFL data used in this thesis, there is an entry pairing the Brisbane Lions as the Home Team against Collingwood as the Away Team, and the number of wins each team achieved against the other with this pairing. Likewise, there will be another entry with Collingwood as the Home Team and the Brisbane Lions as the Away Team. Since the AFL pairs 18 teams through less than 30 rounds for our limited number of seasons, the binomial frequencies are expected to be low for all pairs. FIGURE B.2.1. summarises the data processing for this expansion. The data is also expanded to include an AT_HOME effect, which will be 1 for the Home Team and 0 for the Away Team, keeping the data design as binomial counts. This process is reflected in C.1.

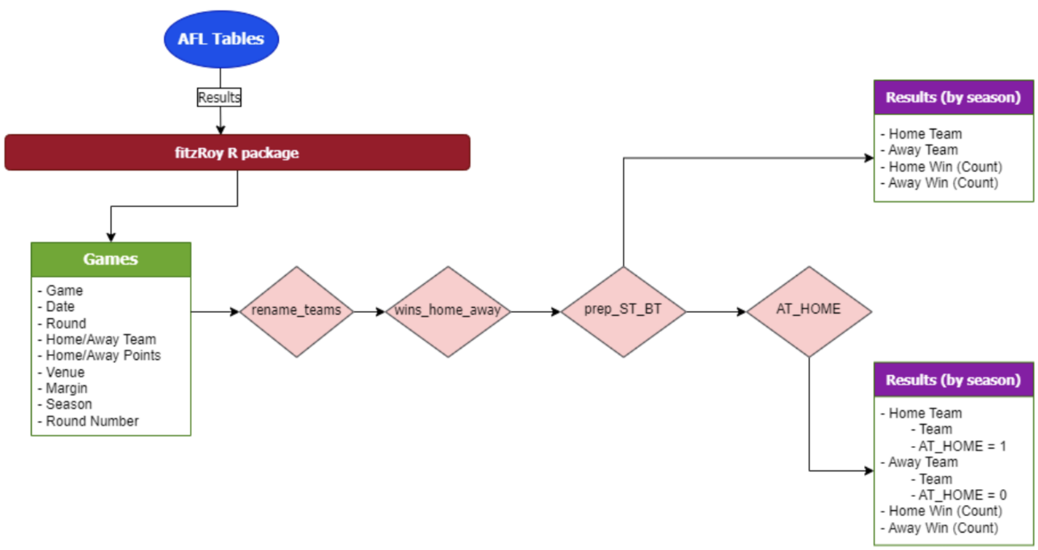


Figure C.1: Data processing for standard and contest-specific Bradley-Terry models.

For models fitted using time-variant features, the data has to be structured in the form of three data frames, since contests cannot be grouped by the Home and Away pairs: one with a set of teams playing Home, their features and results; another with the set of Away teams; and a third dataset with time-invariant features which, for this expansion, are none. All datasets were complete and contained no missing values.

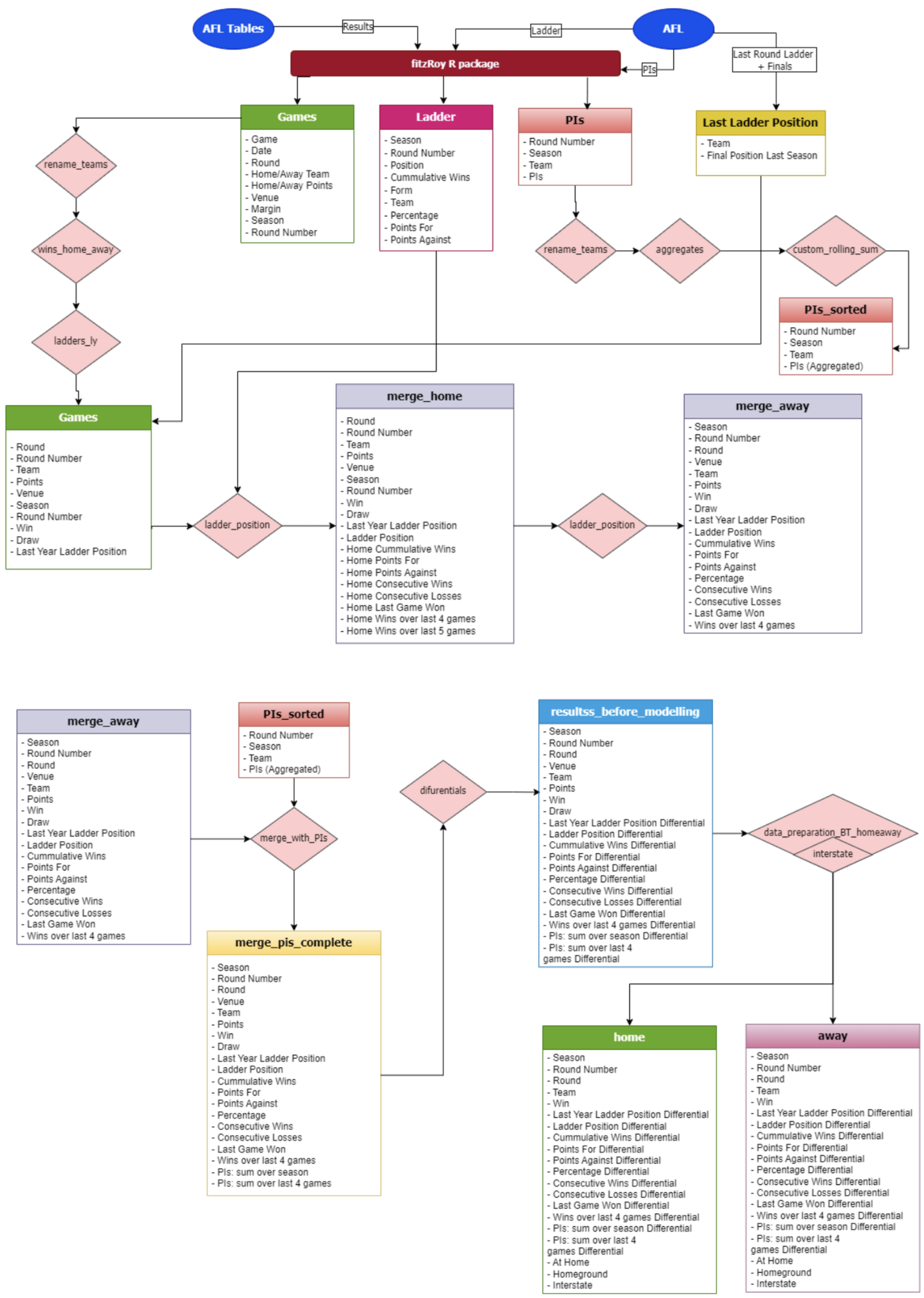

Figure C.2: Data processing for the team-specific, time-variant Bradley-Terry model. Data from games, merge_away and merge_PIs_complete contains entries from both Home and Away teams. Data from PIs, PIs_sorted, ladder and last ladder position contains entries by round and team.

# Appendix D. Description of Experiments

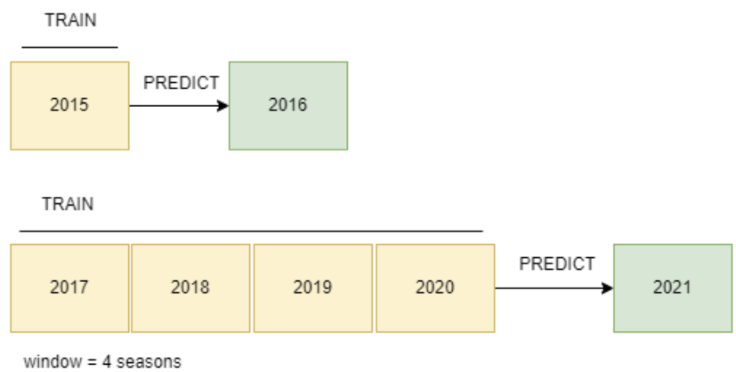


Figure D.3: Examples of training and testing of models in Experiment 1. In the first example, a model is trained on 2015 results and predicts all the games in 2016. In the example below, a model is trained on a window of 4 seasons, from 2017 to 2020, to predict all the games in 2021.

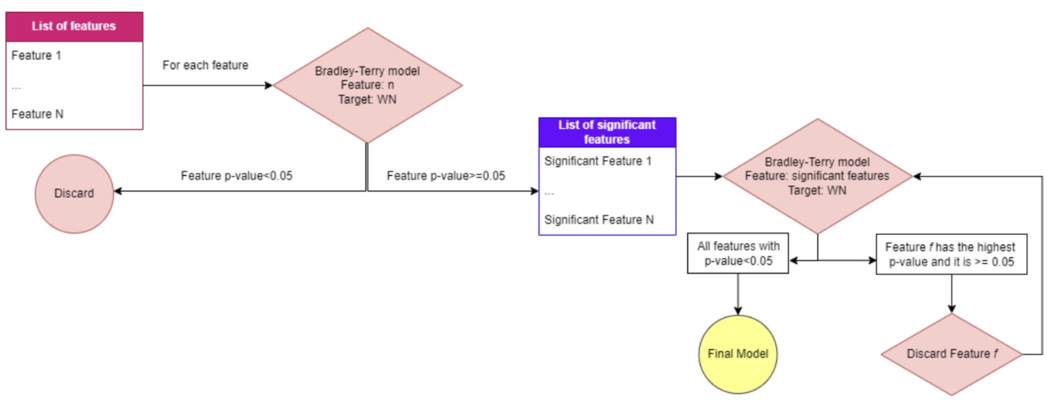


Figure D.4: Process for training models in Experiment 3.

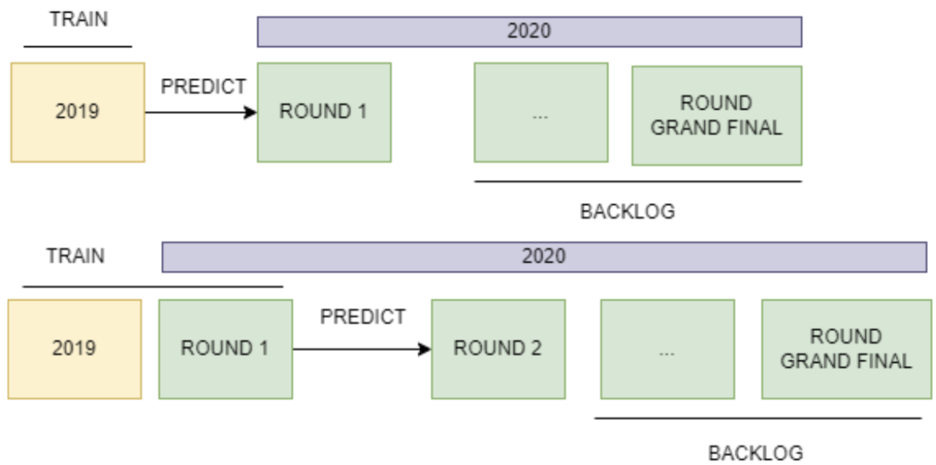


Figure D.5: Example of the Addition strategy. The model trained on 2019 games is used to predict Round 1 of the 2020 season. Then, the model is retrained with that season and used to predict Round 2. The process continues until the Grand Final.

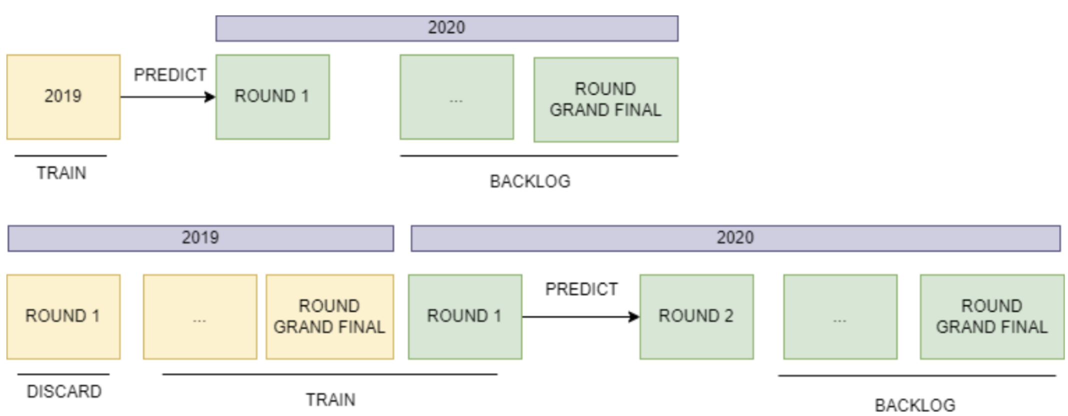


Figure D.6: Example of the Substitution strategy. The model trained on 2019 games is used to predict Round 1 of the 2020 season. Then, the first round of 2019 is discarded, and the model is retrained with the remaining rounds and the first round of 2020 and used to predict Round 2. The process continues until the Grand Final.

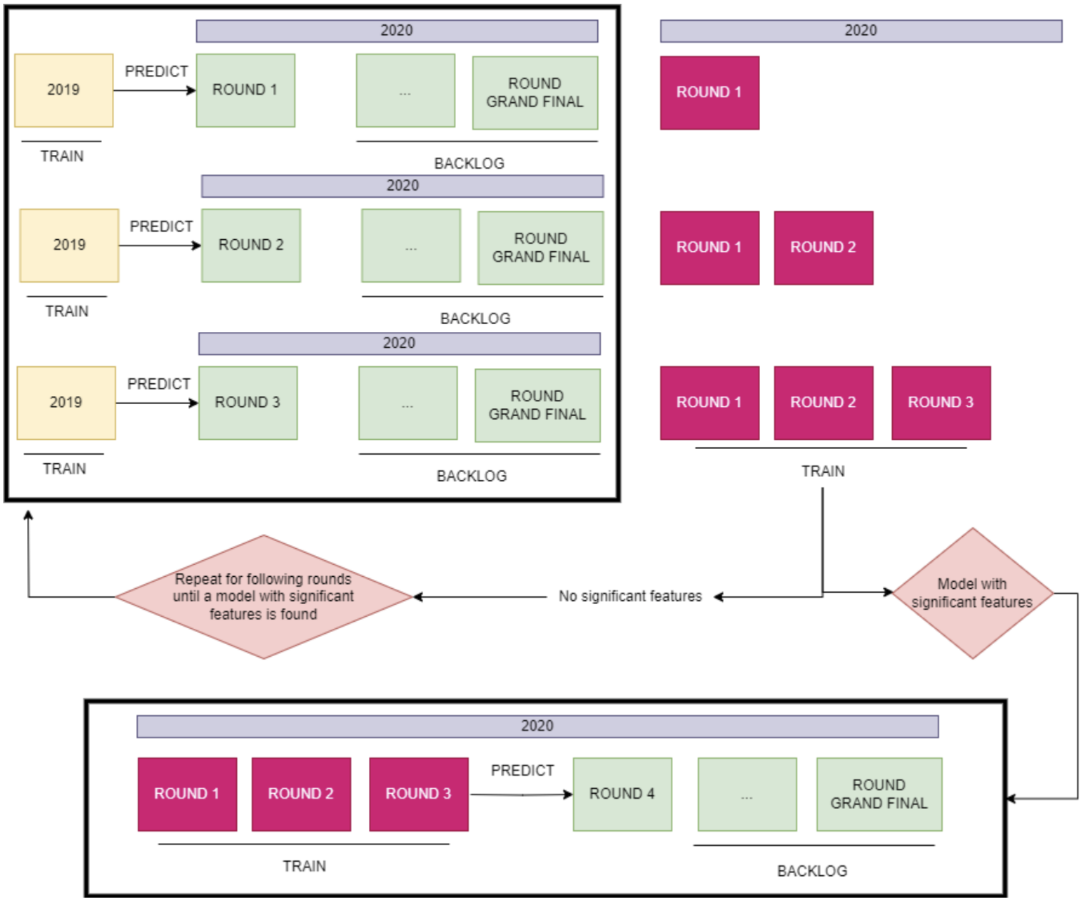


Figure D.7: Example of the Incremental strategy. The model trained on 2019 games is used to predict the first three rounds of the 2020 season. If the model fitted to those rounds has no significant features, the process is repeated for the following round. Otherwise, the newly trained model is used to predict the subsequent rounds.

## Appendix E. Result Tables

Table E.12: Estimated Strength Coefficients with the Standard Bradley-Terry Model, Fitted to Windows of Two Seasons (2015-2023). Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2015-16 | 2016-17 | 2017-18 | 2018-19 | 2019-20 | 2020-21 | 2021-22 | 2022-23 |
| Adelaide | 0.83 | **0.96** | 0.42 | -0.03 | -0.58 | **-1.13** | -0.83 | -0.50 |
| Brisbane Lions | **-1.97** | **-1.75** | **-1.63** | -0.20 | **1.20** | **1.00** | 0.81 | **0.94** |
| Carlton | **-1.40** | **-1.11** | **-1.92** | **-1.58** | -0.50 | -0.49 | -0.21 | 0.30 |
| Collingwood | -0.40 | -0.45 | 0.02 | 0.84 | 0.72 | -0.36 | -0.02 | **1.14** |
| Essendon | **-1.61** | **-0.93** | -0.13 | 0.22 | 0.00 | -0.28 | -0.42 | -0.53 |
| Fremantle | 0.00 | **-1.11** | -0.84 | -0.50 | -0.21 | -0.26 | 0.34 | 0.21 |
| Gold Coast | **-1.50** | **-1.29** | **-1.75** | **-1.87** | **-1.24** | -0.81 | -0.57 | -0.53 |
| Geelong | 0.84 | **0.97** | 0.35 | 0.59 | **1.00** | **0.96** | **1.30** | 0.67 |
| Greater Western Sydney | 0.49 | 0.85 | 0.44 | 0.55 | 0.43 | 0.10 | -0.27 | -0.30 |
| Hawthorn | **1.17** | 0.44 | 0.00 | 0.32 | -0.20 | -0.81 | -0.70 | **-0.96** |
| Melbourne | -0.57 | -0.12 | 0.13 | -0.21 | -0.44 | **1.06** | **1.33** | 0.72 |
| North Melbourne | 0.37 | -0.48 | -0.81 | -0.06 | -0.55 | **-1.57** | **-1.91** | **-2.49** |
| Port Adelaide | 0.05 | 0.06 | 0.08 | 0.10 | 0.79 | **1.34** | 0.54 | 0.29 |
| Richmond | -0.02 | 0.17 | **0.94** | **1.34** | **1.42** | 0.55 | 0.06 | -0.04 |
| Saint Kilda | -0.47 | 0.00 | -0.87 | **-0.99** | 0.03 | 0.18 | 0.00 | 0.00 |
| Sydney | **1.04** | 0.84 | 0.29 | 0.00 | -0.60 | -0.00 | 0.85 | 0.50 |
| Western Bulldogs | 0.88 | 0.54 | -0.58 | -0.28 | 0.33 | 0.72 | 0.59 | 0.04 |
| West Coast | **1.17** | 0.46 | 0.45 | 1.00 | 0.87 | 0.21 | **-1.05** | **-2.45** |

Table E.13: Estimated Strength Coefficients with the Standard Bradley-Terry Model, Fitted to Windows of Three Seasons (2015-2023). Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2015-17 | 2016-18 | 2017-19 | 2018-20 | 2019-21 | 2020-22 | 2021-23 |
| Adelaide | **0.94** | 0.59 | 0.18 | -0.40 | -0.65 | **-1.10** | -0.56 |
| Brisbane Lions | **-1.58** | **-1.73** | **-0.70** | 0.22 | **0.93** | **0.84** | **0.87** |
| Carlton | **-1.13** | **-1.55** | **-1.51** | **-1.15** | -0.52 | -0.37 | 0.07 |
| Collingwood | -0.24 | -0.14 | 0.29 | 0.64 | 0.15 | -0.07 | 0.50 |
| Essendon | **-0.82** | -0.66 | -0.04 | 0.00 | -0.06 | -0.55 | -0.32 |
| Fremantle | -0.10 | **-1.06** | -0.66 | -0.46 | -0.21 | 0.02 | 0.14 |
| Gold Coast | **-1.20** | **-1.53** | **-1.72** | **-1.49** | **-1.05** | -0.73 | -0.55 |
| Geelong | **0.89** | 0.61 | 0.48 | 0.62 | **0.89** | **1.03** | **0.79** |
| Greater Western Sydney | 0.68 | 0.61 | 0.45 | 0.33 | 0.31 | -0.37 | -0.06 |
| Hawthorn | **0.82** | 0.35 | 0.00 | -0.02 | -0.38 | **-0.87** | **-0.77** |
| Melbourne | -0.20 | 0.04 | -0.26 | -0.13 | 0.30 | **0.85** | **1.04** |
| North Melbourne | 0.00 | -0.43 | -0.57 | -0.42 | **-0.86** | **-1.93** | **-1.97** |
| Port Adelaide | 0.28 | 0.00 | 0.05 | 0.44 | **0.87** | 0.66 | 0.63 |
| Richmond | 0.42 | 0.43 | **1.02** | **1.24** | **0.83** | 0.34 | -0.01 |
| Saint Kilda | -0.16 | -0.53 | **-0.76** | -0.54 | 0.00 | 0.00 | 0.00 |
| Sydney | **0.94** | 0.61 | 0.00 | -0.27 | -0.13 | 0.24 | 0.58 |
| Western Bulldogs | 0.66 | 0.04 | -0.35 | -0.14 | 0.52 | 0.36 | 0.40 |
| West Coast | **0.86** | 0.58 | 0.53 | **0.88** | 0.45 | -0.60 | **-1.36** |

Table E.14: Estimated Strength Coefficients with the Standard Bradley-Terry Model, Fitted to Windows of Four Seasons (2015-2023). Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | |
|---|---|---|---|---|---|---|
| | 2015-18 | 2016-19 | 2017-20 | 2018-21 | 2019-22 | 2020-23 |
| Adelaide | **0.73** | 0.38 | 0.00 | -0.43 | **-0.67** | **-0.75** |
| Brisbane Lions | **-1.54** | **-0.93** | -0.13 | 0.34 | **0.88** | **0.92** |
| Carlton | **-1.38** | **-1.31** | **-1.10** | **-0.93** | -0.33 | -0.06 |
| Collingwood | 0.02 | 0.12 | 0.40 | 0.27 | 0.27 | 0.40 |
| Essendon | -0.55 | -0.42 | 0.00 | 0.00 | -0.24 | -0.38 |
| Fremantle | -0.24 | **-0.85** | -0.47 | -0.34 | 0.04 | 0.00 |
| Gold Coast | **-1.34** | **-1.54** | **-1.35** | **-1.22** | **-0.82** | -0.62 |
| Geelong | **0.72** | **0.65** | **0.64** | **0.70** | **1.04** | **0.77** |
| Greater Western Sydney | **0.63** | 0.56 | 0.44 | 0.31 | 0.00 | -0.12 |
| Hawthorn | **0.73** | 0.26 | -0.05 | -0.12 | -0.47 | **-0.83** |
| Melbourne | 0.03 | -0.23 | -0.07 | 0.34 | 0.42 | **0.81** |
| North Melbourne | 0.00 | -0.37 | -0.59 | -0.60 | **-1.18** | **-1.90** |
| Port Adelaide | 0.26 | 0.00 | 0.44 | **0.63** | 0.58 | **0.74** |
| Richmond | **0.64** | **0.62** | **1.14** | **0.87** | **0.69** | 0.25 |
| Saint Kilda | -0.44 | -0.52 | -0.38 | -0.36 | 0.00 | 0.05 |
| Sydney | **0.82** | 0.31 | -0.06 | 0.00 | 0.16 | 0.25 |
| Western Bulldogs | 0.34 | 0.04 | -0.10 | 0.19 | 0.39 | 0.33 |
| West Coast | **0.92** | **0.61** | **0.68** | **0.62** | -0.10 | **-0.87** |

Table E.15: AIC and classification accuracy for the standard Bradley-Terry model, fitted to windows of two to four seasons (2015-2023) and all available data.

| Window: 2 seasons | | | | |
|---|---|---|---|---|
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-16 | 2017 | 400.55 | 72.05% | 58.79% |
| 2016-17 | 2018 | 418.26 | 67.30% | 62.70% |
| 2017-18 | 2019 | 424.11 | 70.96% | 60.54% |
| 2018-19 | 2020 | 415.88 | 69.29% | 56.64% |
| 2019-20 | 2021 | 397.24 | 66.16% | 54.14% |
| 2020-21 | 2022 | 387.31 | 68.52% | 60.22% |
| 2021-22 | 2023 | 413.76 | 69.13% | 63.68% |
| 2022-23 | | 422.02 | 70.86% | |
| Window: 3 seasons | | | | |
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-17 | 2018 | 540.34 | 67.71% | 65.76% |
| 2016-18 | 2019 | 535.43 | 67.82% | 58.38% |
| 2017-19 | 2020 | 562.96 | 68.55% | 56.64% |
| 2018-20 | 2021 | 522.77 | 65.43% | 50.83% |
| 2019-21 | 2022 | 535.23 | 65.16% | 58.15% |
| 2020-22 | 2023 | 521.41 | 67.26% | 65.26% |
| 2021-23 | | 548.09 | 69.06% | |
| Window: 4 seasons | | | | |
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-18 | 2019 | 632.71 | 66.94% | 57.30% |
| 2016-19 | 2020 | 653.18 | 65.99% | 53.19% |
| 2017-20 | 2021 | 645.30 | 65.80% | 57.46% |
| 2018-21 | 2022 | 634.16 | 65.17% | 59.67% |
| 2019-22 | 2023 | 641.68 | 64.44% | 63.30% |
| 2020-23 | | 634.52 | 68.10% | |
| Window: All Available Data | | | | |
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-23 | | 978.62 | 60.98% | |

Table E.16: Estimated strength and AT_HOME coefficients with the Contest-Specific Bradley-Terry model, fitted to single seasons (2015-2023) and all available data. Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2015-23 |
| Adelaide | 0.16 | **1.61** | 1.06 | 0.05 | -0.30 | -1.65 | -0.77 | -1.12 | 0.00 | 0.00 |
| Brisbane Lions | **-2.35** | **-2.22** | -1.32 | **-1.72** | 0.70 | **1.73** | 0.74 | 0.90 | **1.35** | -0.09 |
| Carlton | **-2.29** | -0.98 | -1.01 | **-3.01** | -1.02 | -0.31 | -0.57 | 0.06 | 0.63 | **-0.63** |
| Collingwood | -0.80 | -0.28 | -0.27 | 0.68 | 0.84 | 0.51 | -0.95 | 0.99 | **1.73** | 0.26 |
| Essendon | **-1.58** | **-2.28** | 0.07 | 0.08 | 0.08 | -0.45 | -0.06 | -1.05 | -0.15 | -0.35 |
| Fremantle | 1.03 | **-1.71** | -0.58 | -0.93 | -0.45 | -0.39 | -0.13 | 0.79 | -0.37 | -0.17 |
| Gold Coast | **-2.21** | -1.50 | -1.20 | **-2.29** | **-2.08** | -0.86 | -0.75 | -0.57 | -0.46 | **-0.96** |
| Geelong | 0.00 | **1.79** | 0.84 | 0.24 | 0.85 | 1.25 | 1.14 | **1.67** | -0.11 | **0.67** |
| Greater Western Sydney | -0.67 | **1.60** | 0.94 | 0.44 | 0.60 | 0.00 | 0.39 | -1.30 | 0.49 | 0.26 |
| Hawthorn | 0.94 | **1.73** | -0.06 | 0.41 | 0.08 | -0.90 | -0.66 | -0.96 | -0.95 | 0.01 |
| Melbourne | -1.36 | -0.03 | 0.18 | 0.43 | -1.36 | 0.25 | **2.02** | 0.80 | 0.76 | 0.20 |
| North Melbourne | 0.21 | 0.76 | -1.09 | -0.23 | -0.18 | -1.64 | **-1.47** | **-3.10** | **-2.26** | **-0.66** |
| Port Adelaide | 0.01 | 0.00 | 0.48 | 0.00 | 0.00 | **1.75** | **1.34** | -0.40 | 0.98 | 0.39 |
| Richmond | 0.30 | -0.58 | 1.13 | 1.35 | 1.23 | **1.71** | -0.13 | 0.12 | -0.21 | **0.50** |
| Saint Kilda | **-1.59** | 0.47 | 0.09 | **-1.87** | -0.70 | 0.59 | 0.00 | 0.00 | 0.16 | -0.21 |
| Sydney | 0.50 | **1.81** | 0.64 | 0.45 | -0.72 | -0.93 | 0.75 | 1.03 | 0.20 | 0.37 |
| Western Bulldogs | 0.05 | **1.97** | 0.00 | -0.94 | 0.16 | 0.45 | 1.22 | 0.04 | 0.13 | 0.29 |
| West Coast | 1.03 | 1.48 | 0.34 | 1.13 | 0.71 | 0.93 | -0.17 | **-2.85** | **-2.23** | 0.14 |
| **Feature (AT_HOME)** | 0.15 | **0.63** | **0.43** | 0.25 | **0.35** | 0.35 | 0.08 | **0.62** | **0.37** | **0.29** |

Table E.17: Estimated strength and AT_HOME coefficients with the Contest-Specific Bradley-Terry model, fitted to windows of two seasons (2015-2023). Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2015-16 | 2016-17 | 2017-18 | 2018-19 | 2019-20 | 2020-21 | 2021-22 | 2022-23 |
| Adelaide | 0.88 | 0.91 | 0.39 | -0.03 | -0.60 | -1.13 | -0.84 | -0.52 |
| Brisbane Lions | **-1.99** | **-1.86** | **-1.67** | -0.22 | **1.18** | **0.99** | 0.82 | **1.00** |
| Carlton | **-1.44** | **-1.15** | **-1.93** | **-1.61** | -0.52 | -0.48 | -0.21 | 0.32 |
| Collingwood | -0.41 | -0.49 | 0.04 | 0.85 | 0.78 | -0.34 | -0.02 | **1.18** |
| Essendon | **-1.65** | **-1.00** | -0.14 | 0.21 | 0.00 | -0.27 | -0.43 | -0.56 |
| Fremantle | 0.00 | **-1.19** | -0.88 | -0.53 | -0.25 | -0.27 | 0.34 | 0.22 |
| Gold Coast | **-1.54** | **-1.38** | **-1.78** | **-1.91** | **-1.31** | -0.82 | -0.58 | -0.52 |
| Geelong | 0.84 | **0.91** | 0.35 | 0.61 | **1.02** | **0.98** | **1.31** | 0.67 |
| Greater Western Sydney | 0.51 | 0.87 | 0.47 | 0.59 | 0.46 | 0.12 | -0.28 | -0.31 |
| Hawthorn | **1.25** | 0.44 | 0.00 | 0.33 | -0.18 | -0.80 | -0.72 | **-0.96** |
| Melbourne | -0.58 | -0.14 | 0.14 | -0.22 | -0.46 | **1.05** | **1.29** | 0.73 |
| North Melbourne | 0.40 | -0.52 | -0.82 | -0.07 | -0.56 | **-1.55** | **-1.92** | **-2.59** |
| Port Adelaide | 0.05 | 0.01 | 0.08 | 0.10 | 0.78 | **1.32** | 0.53 | 0.31 |
| Richmond | -0.01 | 0.17 | **0.95** | **1.33** | **1.44** | 0.57 | 0.07 | -0.01 |
| Saint Kilda | -0.47 | 0.00 | -0.88 | **-1.02** | 0.03 | 0.19 | 0.00 | 0.00 |
| Sydney | **1.06** | 0.86 | 0.30 | 0.00 | -0.62 | 0.00 | 0.86 | 0.56 |
| Western Bulldogs | **0.94** | 0.58 | -0.60 | -0.28 | 0.37 | 0.75 | 0.61 | 0.05 |
| West Coast | **1.18** | 0.48 | 0.45 | **1.00** | 0.88 | 0.21 | **-1.7** | **-2.51** |
| **Feature (AT_HOME)** | **0.35** | **0.47** | **0.35** | **0.29** | **0.34** | 0.18 | **0.29** | **0.47** |

Table E.18: Estimated strength and AT_HOME coefficients with the Contest-Specific Bradley-Terry model, fitted to windows of three seasons (2015-2023). Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2015-17 | 2016-18 | 2017-19 | 2018-20 | 2019-21 | 2020-22 | 2021-23 |
| Adelaide | **0.94** | 0.58 | 0.16 | -0.40 | -0.66 | **-1.12** | -0.58 |
| Brisbane Lions | **-1.63** | **-1.78** | **-0.73** | 0.21 | **0.91** | **0.83** | **0.88** |
| Carlton | **-1.16** | **-1.56** | **-1.54** | **-1.17** | -0.53 | -0.37 | 0.07 |
| Collingwood | -0.26 | -0.12 | 0.31 | 0.67 | 0.16 | -0.06 | 0.49 |
| Essendon | **-0.84** | -0.67 | -0.04 | 0.00 | -0.06 | -0.57 | -0.33 |
| Fremantle | -0.12 | **-1.09** | -0.69 | -0.48 | -0.22 | 0.00 | 0.13 |
| Gold Coast | **-1.25** | **-1.58** | **-1.76** | **-1.53** | **-1.08** | -0.76 | -0.55 |
| Geelong | **0.87** | 0.62 | 0.48 | 0.64 | **0.91** | **1.04** | **0.79** |
| Greater Western Sydney | 0.70 | 0.65 | 0.48 | 0.36 | 0.32 | -0.37 | -0.05 |
| Hawthorn | **0.85** | 0.38 | 0.00 | 0.00 | -0.38 | **-0.88** | **-0.79** |
| Melbourne | -0.22 | 0.07 | -0.27 | -0.14 | 0.29 | **0.81** | **1.02** |
| North Melbourne | 0.00 | -0.43 | -0.59 | -0.43 | **-0.86** | **-1.95** | **-1.99** |
| Port Adelaide | 0.26 | 0.00 | 0.05 | 0.43 | **0.85** | 0.63 | 0.62 |
| Richmond | 0.43 | 0.46 | **1.02** | **1.25** | **0.84** | 0.35 | 0.00 |
| Saint Kilda | -0.16 | -0.52 | **-0.78** | -0.54 | 0.00 | 0.00 | 0.00 |
| Sydney | **0.95** | 0.64 | 0.00 | -0.28 | -0.14 | 0.24 | 0.60 |
| Western Bulldogs | 0.69 | 0.09 | -0.36 | -0.12 | 0.54 | 0.38 | 0.42 |
| West Coast | **0.87** | 0.61 | 0.54 | **0.88** | 0.45 | -0.63 | **-1.39** |
| **Feature (AT_HOME)** | **0.35** | **0.39** | **0.34** | **0.30** | **0.24** | **0.30** | **0.32** |

Table E.19: Estimated strength and AT_HOME coefficients with the Contest-Specific Bradley-Terry model, fitted to windows of four seasons (2015-2023). Coefficients found to be significant at the 5% level are in bold and underlined. The coefficient for the reference team is 0.00.

| **Team** | **Season** | | | | | |
|---|---|---|---|---|---|---|
| | 2015-18 | 2016-19 | 2017-20 | 2018-21 | 2019-22 | 2020-23 |
| Adelaide | **<u>0.73</u>** | 0.36 | -0.02 | -0.43 | **<u>-0.69</u>** | **<u>-0.75</u>** |
| Brisbane Lions | **<u>-1.58</u>** | **<u>-0.95</u>** | -0.15 | 0.33 | **<u>0.86</u>** | **<u>0.95</u>** |
| Carlton | **<u>-1.41</u>** | **<u>-1.33</u>** | **<u>-1.13</u>** | **<u>-0.94</u>** | -0.35 | -0.04 |
| Collingwood | 0.02 | 0.14 | 0.43 | 0.29 | 0.28 | 0.43 |
| Essendon | -0.57 | -0.43 | 0.00 | 0.01 | -0.25 | -0.37 |
| Fremantle | -0.26 | **<u>-0.87</u>** | -0.50 | -0.35 | 0.02 | 0.00 |
| Gold Coast | **<u>-1.39</u>** | **<u>-1.59</u>** | **<u>-1.40</u>** | **<u>-1.24</u>** | **<u>-0.86</u>** | -0.62 |
| Geelong | **<u>0.71</u>** | **<u>0.66</u>** | **<u>0.65</u>** | **<u>0.72</u>** | **<u>1.04</u>** | **<u>0.80</u>** |
| Greater Western Sydney | **<u>0.64</u>** | 0.60 | 0.47 | 0.33 | 0.00 | -0.09 |
| Hawthorn | **<u>0.75</u>** | 0.28 | -0.05 | -0.12 | -0.49 | **<u>-0.82</u>** |
| Melbourne | 0.02 | -0.23 | -0.07 | 0.34 | 0.40 | **<u>0.81</u>** |
| North Melbourne | 0.00 | -0.37 | -0.60 | -0.60 | **<u>-1.20</u>** | **<u>-1.90</u>** |
| Port Adelaide | 0.24 | 0.00 | 0.43 | 0.62 | 0.56 | **<u>0.75</u>** |
| Richmond | **<u>0.65</u>** | **<u>0.63</u>** | **<u>1.16</u>** | **<u>0.88</u>** | **<u>0.69</u>** | 0.28 |
| Saint Kilda | -0.45 | -0.53 | -0.38 | -0.36 | -0.01 | 0.08 |
| Sydney | **<u>0.82</u>** | 0.32 | -0.06 | 0.00 | 0.15 | 0.29 |
| Western Bulldogs | 0.36 | 0.07 | -0.09 | 0.21 | 0.41 | 0.38 |
| West Coast | **<u>0.91</u>** | **<u>0.63</u>** | **<u>0.68</u>** | **<u>0.62</u>** | -0.13 | **<u>-0.88</u>** |
| **Feature (AT_HOME)** | **<u>0.32</u>** | **<u>0.36</u>** | **<u>0.33</u>** | **<u>0.24</u>** | **<u>0.30</u>** | **<u>0.31</u>** |

Table E.20: AIC and classification accuracy for the Contest-Specific Bradley-Terry model, fitted to windows of two to four seasons (2015-2023) and all available data.

| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
|---|---|---|---|---|
| Window: 2 seasons | | | | |
| 2015-16 | 2017 | 393.84 | 72.33% | 59.89% |
| 2016-17 | 2018 | 403.27 | 68.66% | 64.32% |
| 2017-18 | 2019 | 416.55 | 69.86% | 60.54% |
| 2018-19 | 2020 | 411.12 | 69.57% | 60.84% |
| 2019-20 | 2021 | 390.83 | 67.68% | 56.35% |
| 2020-21 | 2022 | 386.91 | 68.52% | 61.88% |
| 2021-22 | 2023 | 408.92 | 68.85% | 65.26% |
| 2022-23 | | 407.56 | 74.33% | |
| Window: 3 seasons | | | | |
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-17 | 2018 | 527.27 | 68.26% | 65.76% |
| 2016-18 | 2019 | 519.56 | 68.73% | 62.16% |
| 2017-19 | 2020 | 550.78 | 68.18% | 55.24% |
| 2018-20 | 2021 | 514.49 | 65.62% | 55.80% |
| 2019-21 | 2022 | 530.17 | 66.14% | 61.96% |
| 2020-22 | 2023 | 513.40 | 67.26% | 67.37% |
| 2021-23 | | 537.55 | 70.50% | |
| Window: 4 seasons | | | | |
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-18 | 2019 | 618.02 | 68.59% | 61.92% |
| 2016-19 | 2020 | 632.80 | 67.58% | 56.74% |
| 2017-20 | 2021 | 629.43 | 65.51% | 56.35% |
| 2018-21 | 2022 | 626.85 | 65.61% | 61.33% |
| 2019-22 | 2023 | 628.95 | 65.17% | 64.36% |
| 2020-23 | | 620.77 | 65.95% | |
| Window: All Available Data | | | | |
| Train Seasons | Test Season | Train AIC | Train Accuracy | Test Accuracy |
| 2015-23 | | 945.57 | 61.92% | |

Table E.21: Estimated coefficients for the time-variant Bradley-Terry fitted to windows of one season and all available data with PIs encoded as cumulatives over the season (2015-2023). Significant coefficients are in bold and underlined. Significant coefficients at the individual level only are in italics.

| | Season | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **FEATURE** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **2021** | **2022** | **2023** | **2015-23** |
| MATCH DIFFICULTY | | | | | | | | | | |
| AT_HOME | | **0.599** | *0.392* | | *0.303* | | | *0.521* | **0.551** | **0.182** |
| HOMEGROUND | *0.381* | *0.316* | | | **0.653** | | | **0.746** | *0.408* | *0.363* |
| INTERSTATE | **-0.459** | *-0.443* | **-0.506** | | *-0.530* | **-0.645** | | *-0.515* | *-0.437* | **-0.268** |
| FORM | | | | | | | | | | |
| CONSECUTIVE_LOSSES | | | | | | | **0.344** | | *-0.164* | *-0.139* |
| CONSECUTIVE_WINS | | | | | | | | | | *0.106* |
| LADDER_POSITION_DIFF | *0.039* | *0.061* | *0.038* | *0.030* | | *0.029* | | *0.050* | | *0.044* |
| LADDERLY_POSITION_DIFF | **0.056** | **0.043** | *0.034* | **0.077** | *0.037* | **0.043** | *0.036* | | **0.027** | **0.025** |
| LG_WON | | | | | | | | | **0.787** | *0.278* |
| PERCENTAGE_DIFF | | | *0.009* | *0.007* | | | **0.010** | *0.012* | | *0.009* |
| POINTSAGAINST_DIFF | | | | | | | *0.002* | *0.002* | *0.001* | **0.001** |
| POINTSFOR_DIFF | *0.001* | | **0.003** | **0.003** | *0.003* | *0.003* | | **-0.005** | *0.001* | **0.001** |
| WINS_CUMULATIVE_DIFF | | | **-0.268** | **-0.148** | | *0.130* | | *0.133* | | *0.100* |
| PIs | | | | | | | | | | |
| BOUNCES_CSUM_DIFF | | | | | | | | | | *0.002* |
| CLANGERS_CSUM_DIFF | | | | | | | | | | |
| CLEARANCES_CENTRE_CSUM_DIFF | *0.012* | | | | | | | | | *0.006* |
| CLEARANCES_CSUM_DIFF | | *0.006* | | | *-0.004* | | | | | *0.003* |
| CLEARANCES_STOPPAGE_CSUM_DIFF | | | | | *-0.005* | | | | | *0.002* |
| CONTEST_DEFENSIVE_LOSS_CSUM_DIFF | | | | | **0.022** | | | | | *0.009* |
| CONTEST_DEFENSIVE_LOSS_RATE_CSUM_DIFF | | | | | *-0.031* | | | | | *-0.014* |
| CONTEST_OFFENSIVE_WIN_CSUM_DIFF | | | *0.013* | | | | | | | *0.006* |
| CONTEST_OFFENSIVE_WIN_RATE_CSUM_DIFF | | | | | | | | | | |
| DISPOSALS_CSUM_DIFF | | | | | | | | | | **0.001** |
| DISPOSALS_EFFECTIVE_CSUM_DIFF | | | | | | | | | | *0.000* |
| DISPOSALS_EFFICIENCY_CSUM_DIFF | | | | | | | | | | |
| FREES_AGAINST_CSUM_DIFF | | | **-0.011** | | | | | | | |
| GETS_GROUNDBALL_CSUM_DIFF | | | | | *-0.002* | | **0.003** | | | **-0.001** |
| GETS_GROUNDBALL50_CSUM_DIFF | | | *0.006* | | | | *0.008* | | | **0.004** |
| GOALS_ACCURACY_CSUM_DIFF | | | | | | | | | | *0.009* |
| GOALS_SHOTS_CSUM_DIFF | | | **-0.029** | *0.010* | **0.011** | **0.027** | | *0.009* | | *0.008* |
| HANDBALLS_CSUM_DIFF | | | | | | | | | | |
| HITOUTS_ADVANTAGE_CSUM_DIFF | | | | | | | | | | *0.002* |
| HITOUTS_ADVANTAGE_RATE_CSUM_DIFF | | | | | | | | | | *0.010* |
| HITOUTS_WIN_RATE_CSUM_DIFF | | | **-0.003** | | | | | | | |
| INSIDE50_CSUM_DIFF | | **0.005** | *0.003* | *0.004* | | **0.014** | | *0.006* | **0.005** | *0.005* |
| INTERCEPTS_CSUM_DIFF | | | | | | | | | | *0.002* |
| KICK2HANDBALL_CSUM_DIFF | *-1.535* | | | | | | | | | **0.371** |
| KICKS_CSUM_DIFF | | | | | | | | | | **-0.001** |
| KICKS_EFFECTIVE_CSUM_DIFF | | | | | | *0.002* | | | | *0.001* |
| KICKS_EFFICIENCY_CSUM_DIFF | | | | | *0.060* | | | | | |
| MARKS_CONTESTED_CSUM_DIFF | | *0.011* | | *0.013* | | | | | | *0.008* |
| MARKS_CSUM_DIFF | | | | | | *0.003* | | | | *0.001* |
| MARKS_INSIDE50_CSUM_DIFF | | | *0.008* | | | *0.021* | | *0.014* | | *0.008* |
| MARKS_INTERCEPT_CSUM_DIFF | | | | *0.009* | | | | | | *0.005* |
| MARKS_ONLEAD_CSUM_DIFF | | | | | | | | | | *0.003* |
| METRES_GAINED_L4_CSUM_DIFF | | | | *0.000* | | | | | | *0.000* |
| ONE_PERCENTERS_CSUM_DIFF | | | | | | | | | | *0.001* |
| POSSESSIONS_CONTESTED_CSUM_DIFF | | | | | **-0.002** | | *0.003* | | | *0.002* |
| POSSESSIONS_CONTESTED_RATE_CSUM_DIFF | | | | | | | | | | |
| POSSESSIONS_CSUM_DIFF | | | | | | | | | | *0.000* |
| POSSESSIONS_UNCONTESTED_CSUM_DIFF | | | | | | | | | | *0.000* |
| PRESSURE_CSUM_DIFF | | | | | | | | | | |
| PRESSURE_DEFENSEHALF_CSUM_DIFF | | **-0.002** | | *-0.002* | | | | | | *-0.001* |
| REBOUND_INSIDE50S_CSUM_DIFF | **-0.008** | | | | | *-0.008* | | *-0.010* | | *-0.004* |
| SCORE_LAUNCHES_CSUM_DIFF | | | **0.039** | *0.009* | *0.009* | **-0.037** | | **0.030** | *0.005* | *0.008* |
| SPOILS_CSUM_DIFF | | | | | | | | | | |
| TACKLES_CSUM_DIFF | | | | | | | *0.003* | | | |
| TACKLES_INSIDE50_L4_CSUM_DIFF | | | | | | | *0.009* | | | *0.006* |
| TURNOVERS_L4_CSUM_DIFF | | | | | | **0.009** | | | | |

Table E.22: Estimated coefficients for the time-variant Bradley-Terry fitted to windows of two seasons with PIs encoded as cumulatives over the previous four games (2015-2023). Significant coefficients are in bold and underlined. Significant coefficients at the individual level only are in italics.

| | Season | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **FEATURE** | **2015-16** | **2016-17** | **2017-18** | **2018-19** | **2019-20** | **2020-21** | **2021-22** | **2022-23** |
| | | | | MATCH DIFFICULTY | | | | |
| AT_HOME | *0.005* | **0.465** | **0.324** | *0.268* | *0.312* | | *0.279* | *0.431* |
| HOMEGROUND | *0.003* | *0.316* | *0.257* | **0.464** | **0.610** | | **0.406** | **0.563** |
| INTERSTATE | **-0.497** | *-0.463* | *-0.348* | *-0.359* | *-0.583* | **-0.351** | *-0.358* | *-0.488* |
| | | | | FORM | | | | |
| CONSECUTIVE_LOSSES | | *-0.092* | | | *-0.124* | | | |
| CONSECUTIVE_WINS | | | | | | | | *0.111* |
| L4G_WINS | | *0.161* | | | | | | |
| LADDER_POSITION_DIFF | *0.038* | *0.035* | *0.023* | *0.019* | | | *0.032* | *0.022* |
| LADDERLY_POSITION_DIFF | **0.039** | | **0.030** | **0.033** | **0.029** | | | |
| LG_WON | | | | | | | | |
| PERCENTAGE_DIFF | *0.005* | **0.005** | *0.007* | *0.005* | | | *0.009* | *0.005* |
| POINTSAGAINST_DIFF | **0.001** | **0.001** | **0.001** | *0.001* | *0.001* | **0.001** | **0.002** | **0.001** |
| POINTSFOR_DIFF | **0.001** | *0.001* | **0.001** | **0.002** | **0.002** | | *0.001* | *0.001* |
| | | | | PIs | | | | |
| BOUNCES_L4_CSUM_DIFF | | | | | | | | |
| CLANGERS_L4_CSUM_DIFF | | | | | | | | |
| CLEARANCES_CENTRE_L4_CSUM_DIFF | *0.012* | *0.012* | | | | | | |
| CLEARANCES_L4_CSUM_DIFF | | | | | | *0.010* | | |
| CLEARANCES_STOPPAGE_L4_CSUM_DIFF | | | | | | *0.011* | | |
| CONTEST_DEFENSIVE_LOSS_L4_CSUM_DIFF | | | | | **0.024** | | | |
| CONTEST_DEFENSIVE_LOSS_RATE_L4_CSUM_DIFF | | | | *-0.014* | *-0.014* | | | |
| CONTEST_OFFENSIVE_WIN_L4_CSUM_DIFF | | | | | | | | |
| CONTEST_OFFENSIVE_WIN_RATE_L4_CSUM_DIFF | | | | | | | | |
| DISPOSALS_EFFECTIVE_L4_CSUM_DIFF | | | | | | | | |
| DISPOSALS_EFFICIENCY_L4_CSUM_DIFF | | | | | | | *-0.043* | |
| DISPOSALS_L4_CSUM_DIFF | | | | | | | | |
| FREES_AGAINST_L4_CSUM_DIFF | | | | | | | | |
| GETS_GROUNDBALL_L4_CSUM_DIFF | | | | | | *0.005* | *0.005* | |
| GETS_GROUNDBALL50_L4_CSUM_DIFF | | | | | | **0.014** | **0.014** | *0.009* |
| GOALS_ACCURACY_L4_CSUM_DIFF | | | | | | | | |
| GOALS_SHOTS_L4_CSUM_DIFF | | | *0.008* | **0.009** | **0.011** | | | *0.008* |
| HANDBALLS_L4_CSUM_DIFF | | | | | | | | |
| HITOUTS_ADVANTAGE_L4_CSUM_DIFF | | | *0.008* | *0.009* | | | | |
| HITOUTS_ADVANTAGE_RATE_L4_CSUM_DIFF | | | | | | | | |
| HITOUTS_WIN_RATE_L4_CSUM_DIFF | *0.004* | | *0.003* | | | | | |
| INSIDE50_L4_CSUM_DIFF | *0.007* | *0.006* | *0.006* | *0.008* | | *0.006* | *0.005* | *0.007* |
| INTERCEPTS_L4_CSUM_DIFF | | | | *0.004* | | | *0.005* | *0.004* |
| KICK2HANDBALL_L4_CSUM_DIFF | | | | | | | | |
| KICKS_EFFECTIVE_L4_CSUM_DIFF | | | | | | | | |
| KICKS_EFFICIENCY_L4_CSUM_DIFF | | | | | | | | |
| KICKS_L4_CSUM_DIFF | | | | | | | | |
| MARKS_CONTESTED_L4_CSUM_DIFF | | | *0.011* | | | | | |
| MARKS_INSIDE50_L4_CSUM_DIFF | | | | | | | | |
| MARKS_INTERCEPT_L4_CSUM_DIFF | | | **0.009** | | | | | |
| MARKS_L4_CSUM_DIFF | | | | | | | | |
| MARKS_ONLEAD_L4_CSUM_DIFF | | | | | | | | |
| METRES_GAINED_L4_CSUM_DIFF | | *0.000* | *0.000* | *0.000* | | | | |
| ONE_PERCENTERS_L4_CSUM_DIFF | | | | | | | | |
| POSSESSIONS_CONTESTED_L4_CSUM_DIFF | | | | *0.002* | | *0.005* | *0.001* | |
| POSSESSIONS_CONTESTED_RATE_L4_CSUM_DIFF | | | | | | | *0.030* | |
| POSSESSIONS_L4_CSUM_DIFF | | | | | | | | |
| POSSESSIONS_UNCONTESTED_L4_CSUM_DIFF | | | | | | | | |
| PRESSURE_DEFENSEHALF_L4_CSUM_DIFF | **-0.002** | | | | | | | |
| PRESSURE_L4_CSUM_DIFF | | | | | | | | |
| REBOUND_INSIDE50S_L4_CSUM_DIFF | *-0.009* | | | | | | | |
| SCORE_LAUNCHES_L4_CSUM_DIFF | | | *0.008* | *0.147* | *0.012* | | *0.007* | **0.012** |
| SPOILS_L4_CSUM_DIFF | | | | | | | | |
| TACKLES_INSIDE50_L4_CSUM_DIFF | | | | | *0.009* | | *0.010* | |
| TACKLES_L4_CSUM_DIFF | | | | | | | | |
| TURNOVERS_L4_CSUM_DIFF | | | | | | | | |

Table E.23: Estimated coefficients for the time-variant Bradley-Terry fitted to windows of two seasons with PIs encoded as cumulatives over the season (2015-2023). Significant coefficients are in bold and underlined. Significant coefficients at the individual level only are in italics.

| | Season | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **FEATURE** | **2015-16** | **2016-17** | **2017-18** | **2018-19** | **2019-20** | **2020-21** | **2021-22** | **2022-23** |
| | | | | MATCH DIFFICULTY | | | | |
| AT_HOME | *0.316* | **0.475** | **0.324** | *0.268* | *0.312* | | *0.279* | *0.007* |
| HOMEGROUND | **0.458** | *0.316* | *0.257* | **0.469** | **0.688** | | **0.438** | **0.616** |
| INTERSTATE | *-0.457* | *-0.463* | *-0.348* | *-0.359* | *-0.583* | **-0.337** | *-0.358* | *-0.488* |
| | | | | FORM | | | | |
| CONSECUTIVE_LOSSES | | *-0.092* | | | *-0.124* | | | |
| CONSECUTIVE_WINS | | | | | | | | **0.133** |
| LADDER_POSITION_DIFF | *0.038* | **0.029** | *0.023* | *0.019* | | | *0.032* | *0.022* |
| LADDERLY_POSITION_DIFF | **0.047** | | **0.030** | **0.038** | **0.026** | | | |
| LG_WON | | | | | | | | |
| PERCENTAGE_DIFF | *0.005* | *0.007* | *0.007* | *0.005* | | | **0.006** | *0.005* |
| POINTSAGAINST_DIFF | *0.001* | *0.002* | **0.001** | *0.001* | *0.001* | *0.001* | *0.002* | **0.002** |
| POINTSFOR_DIFF | *0.001* | *0.001* | **0.001** | *0.002* | *0.002* | | *0.001* | *0.001* |
| WINS_CUMULATIVE_DIFF | *0.058* | *0.067* | *0.052* | *0.060* | **-0.161** | | *0.070* | **-0.104** |
| | | | | PIs | | | | |
| BOUNCES_CSUM_DIFF | | | | | | | | |
| CLANGERS_CSUM_DIFF | | | | | | | **-0.004** | |
| CLEARANCES_CENTRE_CSUM_DIFF | *0.009* | *0.007* | | | | | | |
| CLEARANCES_CSUM_DIFF | *0.004* | | | | **-0.005** | | | |
| CLEARANCES_STOPPAGE_CSUM_DIFF | | | | | *-0.006* | | | |
| CONTEST_DEFENSIVE_LOSS_CSUM_DIFF | | | *0.013* | | **0.031** | | | |
| CONTEST_DEFENSIVE_LOSS_RATE_CSUM_DIFF | | | | *-0.018* | *-0.021* | *-0.022* | | |
| CONTEST_OFFENSIVE_WIN_CSUM_DIFF | | | | | | | | |
| CONTEST_OFFENSIVE_WIN_RATE_CSUM_DIFF | | | | | | | | |
| DISPOSALS_CSUM_DIFF | *0.000* | | | | | *0.001* | | |
| DISPOSALS_EFFECTIVE_CSUM_DIFF | | | | | *0.001* | | | |
| DISPOSALS_EFFICIENCY_CSUM_DIFF | | | | | | | | |
| FREES_AGAINST_CSUM_DIFF | | | | | | | | |
| GETS_GROUNDBALL_CSUM_DIFF | | | | | | *0.002* | | |
| GETS_GROUNDBALL50_CSUM_DIFF | | *0.005* | *0.004* | *0.004* | | *0.007* | *0.005* | |
| GOALS_ACCURACY_CSUM_DIFF | | | | | | *-0.021* | | |
| GOALS_SHOTS_CSUM_DIFF | *0.004* | *0.004* | *0.005* | **0.012** | **0.024** | *0.005* | *0.006* | *0.006* |
| HANDBALLS_CSUM_DIFF | | | | | | | | |
| HITOUTS_ADVANTAGE_CSUM_DIFF | | | | | | | | |
| HITOUTS_ADVANTAGE_RATE_CSUM_DIFF | | | | | | | | |
| HITOUTS_WIN_RATE_CSUM_DIFF | | | | | | | | |
| INSIDE50_CSUM_DIFF | **0.004** | *0.003* | *0.003* | *0.003* | | **0.005** | *0.005* | *0.004* |
| INTERCEPTS_CSUM_DIFF | | *0.002* | *0.002* | | | | *0.003* | |
| KICK2HANDBALL_CSUM_DIFF | **-0.941** | | | | | | | |
| KICKS_CSUM_DIFF | | | | | | | | |
| KICKS_EFFECTIVE_CSUM_DIFF | | | | *0.001* | *0.001* | | | |
| KICKS_EFFICIENCY_CSUM_DIFF | | | | | | | | |
| MARKS_CONTESTED_CSUM_DIFF | | *0.005* | *0.006* | | | | | |
| MARKS_CSUM_DIFF | | | | | *0.001* | | | |
| MARKS_INSIDE50_CSUM_DIFF | | | *0.005* | *-0.008* | **-0.015** | | | *0.008* |
| MARKS_INTERCEPT_CSUM_DIFF | | | | | | | | |
| MARKS_ONLEAD_CSUM_DIFF | | | | | | | | |
| METRES_GAINED_L4_CSUM_DIFF | *0.000* | *0.000* | *0.000* | *0.000* | | | | |
| ONE_PERCENTERS_CSUM_DIFF | | | | | | | **0.002** | |
| POSSESSIONS_CONTESTED_CSUM_DIFF | | *0.001* | | | | *0.002* | *0.002* | |
| POSSESSIONS_CONTESTED_RATE_CSUM_DIFF | *0.001* | | | | | | | |
| POSSESSIONS_CSUM_DIFF | *0.000* | | | | | *0.001* | | |
| POSSESSIONS_UNCONTESTED_CSUM_DIFF | | | | | *0.001* | | | |
| PRESSURE_CSUM_DIFF | | | | | | | | |
| PRESSURE_DEFENSEHALF_CSUM_DIFF | **-0.001** | **-0.001** | *-0.001* | | | | | |
| REBOUND_INSIDE50S_CSUM_DIFF | *-0.004* | | *-0.004* | *-0.003* | *-0.005* | | *-0.007* | |
| SCORE_LAUNCHES_CSUM_DIFF | | *0.004* | *0.006* | *0.009* | *0.009* | *0.004* | *0.007* | **0.009** |
| SPOILS_CSUM_DIFF | | | | | | | | |
| TACKLES_CSUM_DIFF | *-0.002* | | | | | *0.003* | **0.002** | |
| TACKLES_INSIDE50_L4_CSUM_DIFF | | *0.005* | | | | *0.007* | *0.006* | |
| TURNOVERS_L4_CSUM_DIFF | | | | | **0.003** | | | |

Table E.24: Classification accuracy for the Bradley-Terry team-specific, time-variant expansion, fitted to windows of two seasons and all available data (2015-2023) with both encodings of PIs.

| Window: 2 seasons | | | | | |
|---|---|---|---|---|---|
| | | Last 4 games cumulative | | Season cumulative | |
| Train Season | Test Season | Train Accuracy | Test Accuracy | Train Accuracy | Test Accuracy |
| 2015-16 | 2017 | 69.83% | 61.27% | 70.32% | 60.29% |
| 2016-17 | 2018 | 67.64% | 65.05% | 67.64% | 65.05% |
| 2017-18 | 2019 | 66.59% | 63.29% | 65.61% | 63.77% |
| 2018-19 | 2020 | 68.28% | 68.12% | 68.28% | 66.88% |
| 2019-20 | 2021 | 68.39% | 58.82% | 70.84% | 54.90% |
| 2020-21 | 2022 | 65.93% | 67.48% | 64.84% | 67.96% |
| 2021-22 | 2023 | 68.29% | 63.55% | 68.54% | 64.02% |
| 2022-23 | | 70.00% | | 69.05% | |
| All available data | | | | | |
| | | Last 4 games cumulative | | Season cumulative | |
| Train Season | Test Season | Train Accuracy | Test Accuracy | Train Accuracy | Test Accuracy |
| 2015-23 | | 68.54% | | 68.54% | |

Table E.25: Number of correctly predicted games in the Finals Series for the Bradley-Terry team-specific, time-variant (TS-TV) expansion, fitted to windows of one season (2015-2023) with both encodings of PIs. The Finals Series has 9 games in total.

| | | | Last 4 games cumulative | | | | |
|---|---|---|---|---|---|---|---|
| Train Season | Test Season | Contest-Specific | TS-TV | Addition | Substitution | Incremental | Majority Voting |
| 2015 | 2016 | 4 | 3 | 3 | 3 | 3 | 3 |
| 2016 | 2017 | 6 | 5 | 5 | 5 | 5 | 5 |
| 2017 | 2018 | 4 | 4 | 6 | 5 | 6 | 5 |
| 2018 | 2019 | 7 | 6 | 6 | 6 | 6 | 6 |
| 2019 | 2020 | 4 | 5 | 5 | 3 | 5 | 5 |
| 2020 | 2021 | 5 | 3 | 4 | 5 | 5 | 5 |
| 2021 | 2022 | 5 | 4 | 6 | 6 | 6 | 6 |
| 2022 | 2023 | 5 | 5 | 7 | 6 | 3 | 5 |
| | | | Season cumulative | | | | |
| Train Season | Test Season | Contest-Specific | TS-TV | Addition | Substitution | Incremental | Majority Voting |
| 2015 | 2016 | 4 | 3 | 3 | 3 | 2 | 1 |
| 2016 | 2017 | 6 | 5 | 4 | 6 | 5 | 5 |
| 2017 | 2018 | 4 | 3 | 5 | 6 | 6 | 6 |
| 2018 | 2019 | 7 | 7 | 3 | 4 | 5 | 5 |
| 2019 | 2020 | 4 | 5 | 5 | 5 | 6 | 4 |
| 2020 | 2021 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2021 | 2022 | 5 | 3 | 7 | 7 | 6 | 6 |
| 2022 | 2023 | 5 | 5 | 4 | 5 | 4 | 4 |

Table E.26: Average Accuracy from Experiment 4 Predictions, by Home Team

| **Team** | **Season** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
| Adelaide | 79.17% | 65.48% | 65.15% | 65.15% | 50.00% | 60.61% | 72.73% | 68.06% |
| Brisbane Lions | 81.82% | 77.27% | 78.79% | 55.13% | 87.88% | 72.22% | 65.28% | 89.29% |
| Carlton | 68.18% | 57.58% | 66.67% | 69.70% | 43.75% | 63.64% | 39.39% | 48.61% |
| Collingwood | 40.91% | 50.00% | 76.39% | 66.67% | 80.95% | 30.30% | 47.22% | 65.56% |
| Essendon | 71.21% | 39.39% | 51.52% | 45.45% | 64.58% | 57.58% | 65.15% | 72.22% |
| Fremantle | 59.09% | 77.27% | 51.52% | 46.97% | 51.67% | 72.73% | 65.28% | 34.72% |
| Gold Coast | 78.79% | 40.91% | 66.67% | 84.85% | 61.11% | 72.73% | 59.09% | 65.15% |
| Geelong | 60.26% | 58.97% | 69.70% | 67.95% | 63.33% | 79.17% | 63.10% | 63.89% |
| Greater Western Sydney | 43.06% | 87.88% | 66.67% | 48.61% | 64.58% | 54.55% | 69.70% | 66.67% |
| Hawthorn | 81.94% | 50.00% | 58.33% | 62.12% | 59.52% | 51.52% | 62.12% | 46.97% |
| Melbourne | 60.61% | 62.12% | 51.39% | 40.91% | 64.81% | 41.03% | 47.44% | 62.50% |
| North Melbourne | 77.27% | 33.33% | 60.61% | 69.70% | 62.50% | 85.00% | 72.73% | 80.30% |
| Port Adelaide | 63.64% | 63.89% | 59.09% | 65.15% | 75.76% | 66.67% | 59.09% | 78.21% |
| Richmond | 60.61% | 68.06% | 92.31% | 82.05% | 71.67% | 51.67% | 65.00% | 53.03% |
| Saint Kilda | 69.70% | 56.06% | 63.33% | 69.70% | 27.78% | 63.64% | 63.64% | 52.78% |
| Sydney | 70.24% | 68.06% | 48.61% | 62.12% | 39.58% | 58.33% | 63.89% | 51.67% |
| Western Bulldogs | 59.09% | 63.64% | 71.21% | 42.42% | 47.92% | 62.50% | 51.52% | 60.61% |
| West Coast | 81.94% | 59.09% | 60.71% | 73.61% | 85.00% | 54.55% | 80.30% | 72.73% |