THE INFLUENCE OF NEUTRAL GROUNDS ON MATCH OUTCOMES

Authors: Liam Kneafsey and Stefan Müller
Reproduced by Mason Ricci

INTRODUCTION

Much research has been conducted to explore the effects of home or away matches in sports. This article explores the impact of neutral ground on outcomes. Gaelic football and hurling are unique in that a large portion of games are played on neutral ground. The chances of a favored hurling or Gaelic football team winning when at, away, or in a neutral field are observed while controlling for important covariates.
This article suggests that the neutral grounds are a more fair venue for sporting matches especially in important elimination games.

HYPOTHESIS

Is the probability of the favored team winning less likely at a neutral venue compared to a home game?

DATA

Statistics for matches and players are available on the GAA Rankings and GAA website pages. All variables for the models are derived from these datasets. Further steps are made to subset, pivot, and manipulate data to create dummy variables for wins and losses (see code).

```
# Import data
data_raw_football <- rio::import("/Users/riccimason99
  /Downloads/ gaa_results_football.xlsx") %>%
  mutate(sport = "Gaelic Football")
data_raw_hurling <- rio::import("/Users/riccimason99
  /Downloads/gaa_results_hurling.xlsx") %>%
  mutate(sport = "Hurling")
# Combinde the data
data_raw <- bind_rows(data_raw_football, data_raw_hurling)
```

ANALYSIS

Regression and Variables

Binary logistic regression is used to test the hypothesis where the outcome variable is win or loss. The treatment variable is the location of the match (home, away, or neutral). Other variables that are controlled for include the importance of the match (league game or championship game), year, sport, and "elo difference," which is the strength of the favored team compared to the opponent. Since "elo difference" is highly skewed, the study uses the logged difference between the ranking of the favored vs. underdog team. However, the results do not change in either case. Interactions are considered between the importance of the match, location, and sport.

```
# Model for both sports
glmer_total <- glmer(result_dummy ~ competition_dummy_factor
   * match_place * sport + elo_diff_adjusted_log + year +
  (1 | team_sport), family = binomial,
  data = data_long_one_team_dummy)

# Model for football sports
glmer_football <- glmer(result_dummy ~
  elo_diff_adjusted_log +
  competition_dummy_factor * match_place +
  year + (1 | team),
  family = binomial(link="logit"),
  data = filter(data_long_one_team_dummy,
  sport == "Gaelic Football"))

# Model for hurling sports
glmer_hurling <- glmer(result_dummy ~
  elo_diff_adjusted_log +
  competition_dummy_factor * match_place +
  year + (1 | team),
  family = binomial(link = "logit"),
  data = filter(data_long_one_team_dummy,
  sport == "Hurling"))
```

Coarsened Exact Matching

Coarsened Exact Matching is a method employed to compare cases (games) based on similar values for key covariates, aside from the variable of interest. This method allows for the comparison between two groups of games. The groups would possess similar elo differences between teams, and the games would be of the same type (league or championship); however, they would differ in the treatment variable which is the location of the match.

This code adjusts the data frame to remove away matches only comparing neutral and home. Then it uses the "cem()" function to perform coarsened exact matching.

```
data_matching_not_away <- data_matching %>%
    filter(match_place != "Away") %>%
    select(-match_place) %>%
    as.data.frame()

q1.grp <- list(c("Hurling"), c("Gaelic Football"))

mat_home_neutral_weight <- cem(treatment = "match_place_num",
                               data = data_matching_not_away,
                               grouping = list(sport = q1.grp),
                               drop = "result_dummy_num")
```

Random Intercepts

Random Intercepts are used in the model to account for unknown aspects of sub-categories of the sample. For example, are teams from different counties affected differently by home-away and neutral games due to unaccounted-for variables such as team cohesion, coaching style, etc.

RESULTS

The study first tests whether the previous scholarly work which identifies that home-field advantage as a relevant factor applies to hurling and football. The effect of home field advantage on in both sports is between 0.55 and 0.6. These values are highly statistically significant with coefficients for neutral ground at p less than 0.001 in all models.

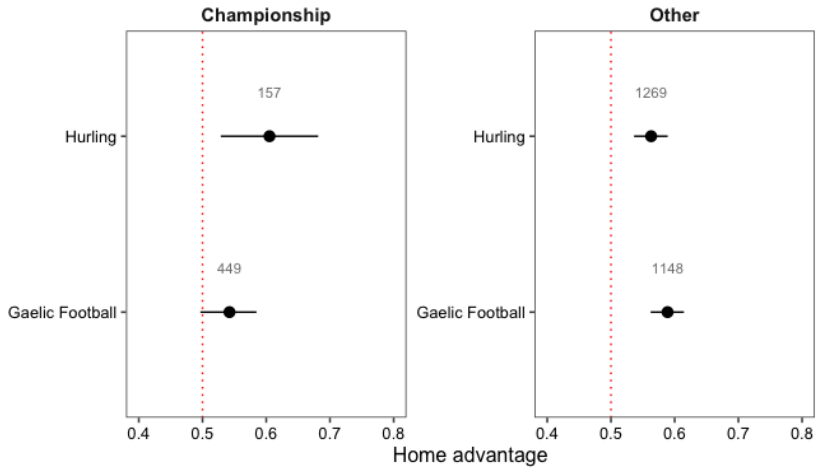

Figure 1. As can be seen, the favored team is more likely to win at home than when away. 0.5 indicates that the home and away team are equally likely to win while 1 indicates that the home team always wins. The number above the circle is the number of analyzed matches for each category

Next, the study observes the probability of the favored team (team with the highest elo ranking) winning the game away and neutral fields. The results of the findings remain the same when including draws, as is displayed in the supplementary material. At home the probability of the favored team winning is 0.8 (football) and 0.9 (hurling). For away matches the probability of the favored team winning is much lower, .5(football) and .55(hurling). Notably the probability of the favored team winning in neutral ground is between 0.6 and 0.7 for both sports. This value is significantly lower

than the probability of the favored team winning at home displaying the neutral grounds are more fair.

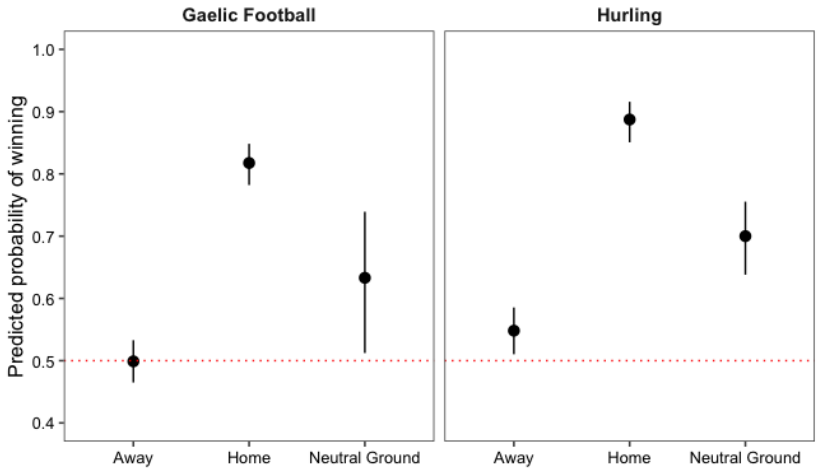

Figure 2. This image illustrates the probability of the favored team winning, depending on the type of ground, based on 95 percent confidence intervals. Independent variables are maintained at their respective mean (for continuous variables) or mode (for ordinal variables). Random intercepts are utilized.

Finally, coarsened exact matching is used to group observations that have similar values for covariates; observations that cannot be matched are removed. The "sample average treatment effects on the treated" (SATT) is the effect that the treatment variable (home/away/neutral) has on the outcome (win or loss). Each variable is assigned weights to control for imbalance between groups, as there are many more home and away games than games on neutral fields. The SATT represents the change in log odds when the venue type shifts from away/home to neutral. A positive value signifies that playing at neutral venues raises the chance of winning, while a negative value suggests a decrease in the likelihood of victory.

It is anticipated that there will be an increase in the logged odds of winning when transitioning from Away to Neutral grounds, whereas a shift from Home to Neutral is expected to result in a decrease in the logged odds. The model and figure below confirms that the direction of the treatment variable is as predicted and is highly significant.

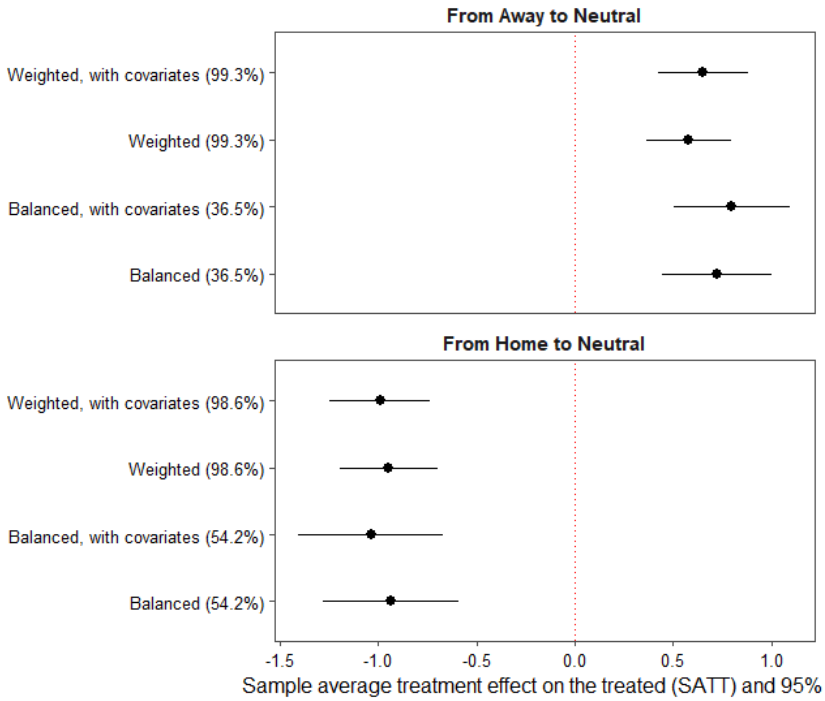

Figure 3. Average treatment effect based on corsened exact matching.

My Twist

In my twist to the paper I test the effect of home turf when teams playing in northern Ireland are playing against teams in the republic of Ireland. Due to teams from the north playing in a different country with a potentially different fan base I wanted to see if the effect was different. First I remove all games in which northern teams do not play. After that I remove games were northern teams play each other. Then I run the exact same model.

In the modified version of my model, both the direction and the magnitude of the covariates remain largely unchanged. However, there is a noticeable reduction in the statistical significance of certain values. Specifically, the coefficients representing the favored team's likelihood of winning in "away" scenarios compared to "home" or "neutral" games have reduced statistical significance for both sports examined. However, this reduction could potentially stem from a smaller data set. Overall, the adjustments introduced in this model do not lead to any substantial changes in the findings.

```
# list of teams in northern Ireland
values_to_keep <- c("Antrim", "Armagh", "Derry",
    "Down", "Fermanagh", "Tyrone")

# subset teams that are in the north
north_df <- data_long_one_team_dummy %>%
    filter(Team_1 %in% values_to_keep | Team_2
```

```
    %in% values_to_keep)

# remove games where two northern  teams play
north_df <- north_df %>%
    filter(!(Team_1 %in% values_to_keep & Team_2
    %in% values_to_keep))

# run model on football and hurling teams in north
twist <- glmer(result_dummy ~  competition_dummy_factor
    * match_place * sport +
    elo_diff_adjusted_log +year + (1 | team_sport),
    family = binomial,
    data = north_df)

# run model on football teams in north
twist_football <- glmer(result_dummy ~
    elo_diff_adjusted_log +
    competition_dummy_factor * match_place +
    year + (1 | team),
    family = binomial(link="logit"),
    data = filter(north_df,
    sport == "Gaelic Football"))


# run model on football teams in north
twist_hurling <- glmer(result_dummy ~
    elo_diff_adjusted_log +
    competition_dummy_factor * match_place +
    year + (1 | team),
    family = binomial(link="logit"),
    data = filter(north_df,
    sport == "Hurling"))
```

	(1) Combined	(2) Football	(3) Hurling
(Intercept)	-2.62 (0.27)***	-2.49 (0.36)***	-3.08 (0.39)***
Other Competition (ref: Championship)	-0.41 (0.16)**	-0.42 (0.16)**	0.13 (0.23)
Home (ref: Away)	1.28 (0.25)***	1.26 (0.26)***	2.39 (0.54)***
Neutral Ground (ref: Away)	0.67 (0.23)**	0.65 (0.23)**	1.01 (0.31)**
Hurling (ref: Gaelic Football)	-0.22 (0.25)		
Elo Difference (log)	0.56 (0.04)***	0.49 (0.05)***	0.66 (0.06)***
2010	-0.16 (0.16)	-0.00 (0.23)	-0.37 (0.24)
2011	0.30 (0.17)	0.46 (0.24)	0.10 (0.24)
2012	0.21 (0.17)	0.40 (0.23)	0.01 (0.25)
2013	0.10 (0.17)	0.27 (0.24)	-0.14 (0.25)
2014	0.23 (0.17)	0.61 (0.24)*	-0.22 (0.25)
2015	0.14 (0.17)	0.34 (0.24)	-0.11 (0.24)
2016	0.22 (0.17)	0.37 (0.24)	0.03 (0.25)
2017	0.09 (0.17)	0.38 (0.24)	-0.23 (0.24)
2018	0.35 (0.17)*	0.90 (0.26)***	-0.20 (0.24)
Other Competition * Home	0.30 (0.30)	0.28 (0.30)	-0.58 (0.56)
Other Competition * Neutral Ground	-0.16 (0.41)	-0.15 (0.41)	-0.38 (0.37)
Hurling * Other Competition	0.56 (0.27)*		
Hurling * Home	1.04 (0.58)		
Hurling * Neutral Ground	0.30 (0.38)		
Other competition * Home * Hurling	-0.92 (0.62)		
Other competition * Neutral Ground * Hurling	-0.27 (0.55)		
Log likelihood	-1906.74	-983.07	-915.03
N	3302	1687	1615
N (Team/Sport)	69		
N (Team)		32	37

***p < 0.001; **p < 0.01; *p < 0.05

Statistical models

Figure 4. Table for the paper.

	(1) Combined	(2) Football	(3) Hurling
(Intercept)	-2.92 (0.47)***	-2.92 (0.63)***	-3.64 (0.86)***
Other Competition (ref: Championship)	-0.09 (0.27)	-0.08 (0.27)	0.77 (0.65)
Home (ref: Away)	1.18 (0.41)**	1.13 (0.42)**	2.84 (1.28)*
Neutral Ground (ref: Away)	0.99 (0.39)*	0.96 (0.40)*	16.81 (105.62)
Hurling (ref: Gaelic Football)	-0.34 (0.64)		
Elo Difference (log)	0.51 (0.07)***	0.41 (0.09)***	0.71 (0.11)***
2010	0.11 (0.28)	0.56 (0.39)	-0.49 (0.43)
2011	0.51 (0.29)	1.13 (0.42)**	-0.22 (0.43)
2012	0.54 (0.30)	1.27 (0.41)**	-0.34 (0.45)
2013	0.48 (0.29)	1.07 (0.41)**	-0.28 (0.45)
2014	0.19 (0.29)	0.43 (0.41)	-0.11 (0.46)
2015	0.52 (0.29)	1.33 (0.42)**	-0.41 (0.43)
2016	0.53 (0.30)	1.37 (0.44)**	-0.38 (0.43)
2017	0.25 (0.29)	0.95 (0.43)*	-0.55 (0.43)
2018	0.71 (0.31)*	1.51 (0.43)***	-0.27 (0.46)
Other Competition * Home	0.43 (0.49)	0.40 (0.49)	-0.99 (1.30)
Other Competition * Neutral Ground	-0.33 (0.68)	-0.40 (0.68)	-16.60 (105.62)
Hurling * Other Competition	0.80 (0.66)		
Hurling * Home	1.52 (1.30)		
Hurling * Neutral Ground	15.04 (1029.20)		
Other competition * Home * Hurling	-1.46 (1.36)		
Other competition * Neutral Ground * Hurling	-15.57 (1029.20)		
Log likelihood	-637.44	-331.75	-294.32
N	1086	556	530
N (Team/Sport)	64		
N (Team)		30	34

***p < 0.001; **p < 0.01; *p < 0.05

Statistical models

Figure 5. Table for twisted model.

Bibliography

Stefan Müller Liam Kneafsey. (2021) Evidence for the irrelevance of irrelevant events. Political Science Research and Methods 11:2, pages 311-327.