The Split: Analysing Contest Design in the Scottish Premier League

Jessica K. Hargreaves* and Johan M. Rewilak**

*Department of Mathematics, University of York, York, YO10 5DD, UK: jessica.hargreaves@york.ac.uk
** Department of Sport and Entertainment Management, University of South Carolina, South Carolina, USA.

Abstract

This paper examines whether the policy to split the Scottish Premier League (SPL) into two after 33 games for post-season play generated negative externalities. Using a regression discontinuity (RD) design, it tests whether the policy reduced attendance for teams finishing in the lower half of the standings. The analysis uses data from 23 seasons (2000/01 to 2023/24, excluding pandemic-impacted seasons) in which the league has operated under this structure. The results show that teams just below The Split experience lower attendances compared to those just above, driven by the lost opportunity to play against the "top" teams such as Celtic and Rangers. This implies the new structure harmed a subset of clubs. Furthermore, this work highlights how large market teams subsidise smaller teams in sports leagues.

1 Introduction

The Scottish Football League (SFL) is a professional football competition similar to other open European football leagues, featuring multiple divisions with promotion and relegation. The Scottish Premier League (SPL) is the top division and, in the 2000/01 season, the SPL expanded from 10 to 12 teams [6]. The addition of these teams had the potential to create fixture congestion as, traditionally, SPL teams played each other four times in a round-robin format, twice at home and twice away, with 36 fixtures played in total. With the new format, teams would have to complete 44 games.

To avoid fixture congestion and to make the total number of fixtures comparable to other top leagues in Europe, the SPL altered its tournament design. It "split" the season into two: the "Regular Season" and the "Play-offs". During the Regular Season, teams play one another three times for a total of 33 matches. Then the league is "split" in two, creating two mini-leagues. Teams who finish in the **top six** places after 33 games play one another (one more time) in the "**Championship Play-off**" and the **bottom six** teams play one another in the "**Relegation Play-off**". This leads to a total of 38 games being played by each team, the same number of matches as other European leagues [8].

Figure 1 shows home attendances as a proportion of stadium capacity for all 12 teams in the 2023/24 SPL season. Celtic, Rangers and, to a lesser extent, Hearts, display stable attendance, nearly selling out all matches. In contrast, the other teams show fluctuating attendances. To investigate this further, Figure 2 shows home attendances for two SPL teams in 2023/24 (Dundee and Aberdeen). Attendance fluctuations are driven by the opposition, with higher attendances when Celtic/ Rangers (or a historic rival) visit. In particular, matched attendances (Pre- and Post-Split) are very similar.

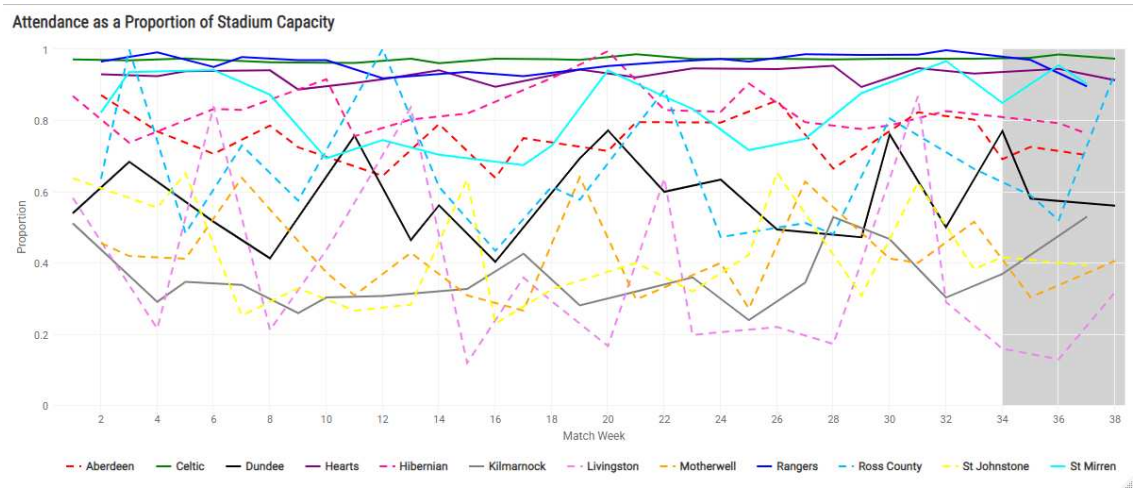

Figure 1: Attendances as a proportion of stadium capacity for all teams in the 2023/24 SPL Season. Solid lines indicate teams in the Championship Play-off. Dashed lines indicate teams in the Relegation Play-off.

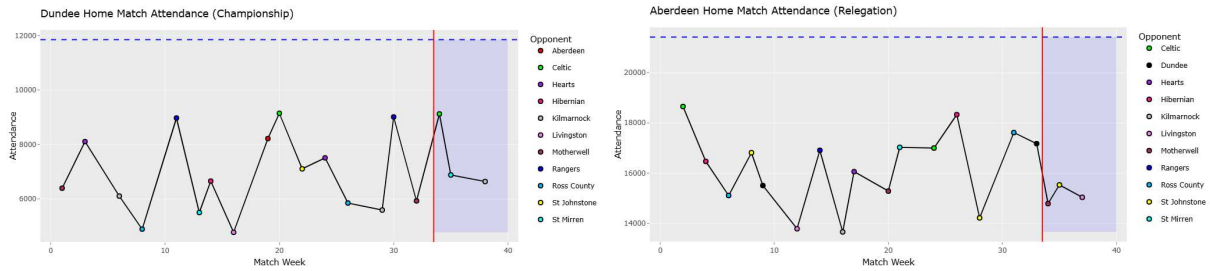

Figure 2: Home attendances for two example SPL Teams in the 2023/24 Season. The reported stadium capacity is shown as a horizontal dashed blue line. Blue background denotes play-off matches.

In this paper, we investigate whether The Split generated any negative externalities. Using a regression discontinuity (RD) design approach, we empirically test whether the "split" impacted teams above the cut-off more than those below in terms of (home) attendance and whether any home attendance differences may be explained via "superstar" effects.

2 Background and Motivation

In sporting contest design, to maintain sporting integrity, we require that participants utilise costly effort to achieve success [14]. Similarly, tournament designers often face multiple objectives when designing an optimal sporting contest and face difficult trade-offs that can create wrong incentives [4]. By using appropriate mechanisms – often financial – contest designers try to ensure that the competition is incentive compatible and teams take the correct actions [9].

The theory of superstars is well established [10, 11]. The superstar effect in sport has been extensively studied, with individual players or teams driving fan attendance [3, 13]. Since 1984-85, only Rangers and

Celtic have won the SPL title, making them dominant forces in the league. These “Old Firm" clubs drive fan interest, suggesting the league functions as a duopoly with a competitive fringe. As the Old Firm always play in the Championship segment of the SPL split, teams finishing in the top half will play an additional game against them. This could boost attendance and profits for those teams.

To investigate the impact of The Split, we adopt a Regression Discontinuity (RD) Design. The use of RD design in sports is increasing due to its ability to provide causal interpretations by estimating the local average treatment effect. It has been applied across a wide range of sports, including professional football, covering topics from contest design to on-field performance [9]. [7] and [8] use an RD design approach to investigate the impact of league design in the SPL on spectator attendance and club revenues.

3 Data and Methods

3.1 Data Description

We obtain data from the SPL and World Football websites[1]. The data spans from 2000/01 to 2023/24. Following [8], we exclude the 2019/20-2021/22 seasons, which were affected by restrictions due to the Covid-19 pandemic. Following other RD design studies in sport (e.g. [1], [9], [12]), we use **annual home attendance** data (i.e. one observation per club per season).

3.2 Methods

To comprehensively investigate the impact of The Split, we use Regression Discontinuity (RD) Design. For a detailed survey of the RD approach, we direct the reader to [5].

Following [7], the dependent variable is the natural logarithm of home match day attendance. We also construct three variations of this variable (see Section 4) to further develop the work by [7]. These include various changes in attendance Pre- and Post-Split, to capture like-for-like factors that vary solely due to the SPL split. These variables are normally distributed and may provide a better measure to examine the attendance effect of finishing in the Championship Play-off versus the Relegation Play-off.

The remaining data is manually constructed. The treatment variable is a dummy equal to one for teams finishing seventh to twelfth in a season, with teams finishing first to sixth coded as zero. In addition, the running variable is the position a team finishes Pre-Split, centred around zero, similar to other regression discontinuity studies [1, 9, 12].

An RD design should not require additional control variables but, in practice, the most relevant confounders are included [5]. This study includes team and season dummies. Club fixed effects are included as some teams have larger supporter bases than others and time fixed effects are used to capture season-wide shocks affecting all teams, such as Gretna’s liquidation and Rangers’ reformation as a new club.

Equation (1) outlines the RD design in its linear form. Subscript (i) represents individual teams and (t) indexes time. Club fixed effects are represented as (α_i) and period fixed effects (τ_t). The conditioning variables are shown in X.

$$Y_{i,t} = \alpha_i + \beta_1(Rank-7) + \beta_2(Rank-7) * Treat_{i,t} + \beta_3 X_{i,t} + \tau_t + \varepsilon_{i,t} \quad (1)$$

The RD design has three requirements [5]. In this context, they are as follows:

[1] urls: https://spfl.co.uk/league/premiership and https://www.worldfootball.net/

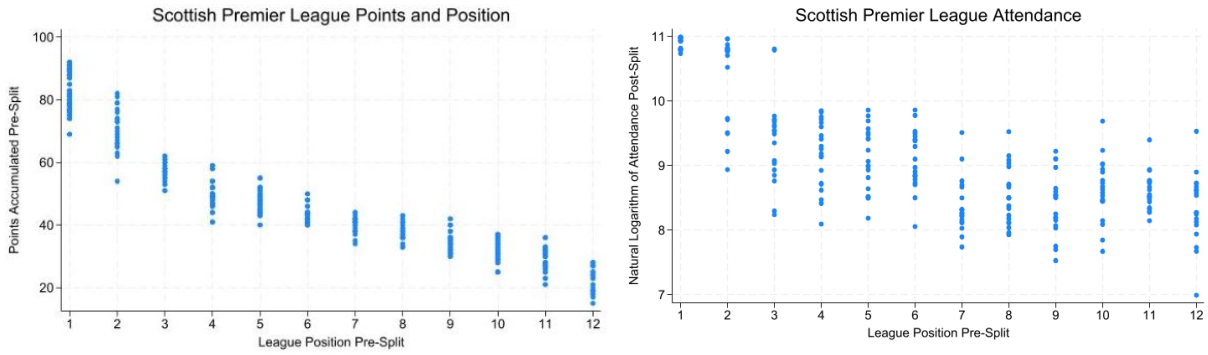

Figure 3: Points accumulated Pre-Split (**Left**) and Post-Split attendance (**Right**).

1. There is a threshold, with teams randomly assigned above and below this cut-off.
2. Teams on either side of the threshold are similar in characteristics, forming a good treatment and comparator group.
3. There is a significant jump in the dependent variable at the threshold.

After 33 games in the SPL, teams finishing seventh (sixth) or below (above) are placed in the lower (upper) half of The Split with 100% probability. Therefore, as teams finishing seventh and below are always treated, we observe a sharp RD design. Teams also have incomplete control over their allocation above or below the threshold, as they cannot influence other match results, referee errors or other random factors that may affect their position. Therefore, final allocations around the cut-off are random, satisfying condition 1.

As a team's final position is based on points from three rounds of round-robin matches, it is anticipated that teams are similar in terms of on-field performance around the cut-off. Indeed, the average points difference between teams finishing sixth and seventh is three- the number of points awarded for a win. Figure 3 shows the points accumulated Pre-Split for all teams in the league. Figure 3 shows a smooth curve for points accumulated Pre-Split, with no jump near the threshold, supporting the second assumption. However, Figure 3 also reveals a large jump between the top two and third place, indicating that Celtic and Rangers should be omitted from the analysis.

To examine Condition 3, a scatter plot of the data is presented in Figure 3, providing graphical evidence of a discontinuity between sixth and seventh place and thus supporting the econometric design. Furthermore, Figure 4 shows an RD plot, with a linear polynomial (Equation (1)), using three teams on either side of the threshold. The plot shows a clear "jump" in (log) attendance at the threshold.

4 Results

Firstly, we investigate the impact of The Split on home attendance using a local linear estimator and subsequently calculate the change in attendance following [2]. The results show that teams finishing below the cut-off face a statistically significant ($p < 0.01$) drop in attendance of approximately 30%, relative to those who finish above the threshold. This supports the findings of [7] who report a 24% drop in attendance for SPL teams in the Relegation Play-off compared to those in the Championship Play-off.

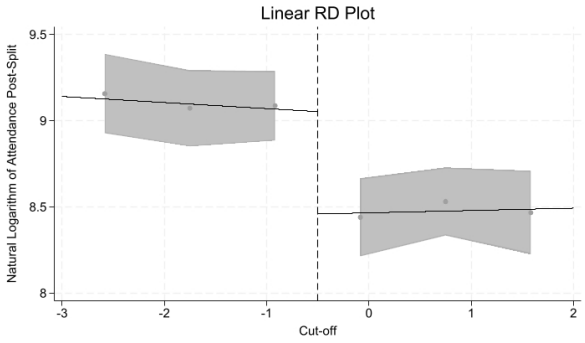

Figure 4: Regression discontinuity plot (linear regression) where the cut-off (zero) represents seventh place.

Y-Transform	Pre-Split - Post-Split	Last 5 Pre-Split - Post-Split	Matched By Pre/Post Split Fixture
Treatment	-1074.50*** (349.37)	-1154.90*** (439.05)	381.03 (257.09)
Club Fixed Effects	Yes	Yes	Yes
Time Dummies	Yes	Yes	Yes
Observations	126	126	126

Table 1: **Sensitivity Analysis.** Each column represents a separate regression. The dependent variable is the change in attendance Pre- and Post-Split (calculated in three ways). Standard errors are reported in parentheses and *** denotes statistical significance at the 1% level.

Furthermore, we conduct robustness tests using the change in attendance before and after The Split as the dependent variable. For all three tests, we focus on teams three places either side of the cut-off (positions 4-9), but alter how we calculate Pre-Split attendance. In the first test, we use all the data. However, in the second test we only use the last five home fixtures before The Split (subtracting the average Post-Split attendance from the average of the last five Pre-Split home games). This accounts for potential fan disengagement before The Split (if a team is destined to finish mid-table). Finally, we match any home fixtures against the teams played after The Split with the same respective fixtures before The Split (subtracting the average attendance). This overcomes issues of teams playing smaller sides before The Split and larger ones afterwards, as well as game-specific characteristics like derby matches that might influence the findings.

The results in Table 1 show that only in the final column, where fixtures Pre- and Post-Split are matched, is the treatment variable not statistically significant. Figure 2 illustrates this, highlighting that matches against the same opponent have similar attendance throughout the season. For all clubs, fixtures against Celtic/ Rangers (at any point in the season) attract the highest attendance. This suggests that a reason why attendance is lower for teams in the Relegation Play-off is that they miss out on these lucrative Old Firm fixtures.

5 Conclusion

In 2000, the SPL expanded from 10 to 12 teams. To avoid fixture congestion, it introduced a policy splitting the league into two halves after 33 matches, with teams in each half playing each other once more for a final

five matches. This study uses a regression discontinuity (RD) design and finds that “The Split" generated several externalities. In Section 4, we found that teams just finishing in the Relegation Play-off faced a 30% attendance drop compared to teams just qualifying for the Championship Play-off. However, when fixtures were matched Pre- and Post-Split, this negative effect disappeared. This suggests that the opposition, particularly the two superstar clubs Rangers and Celtic, drive these attendance differences.

References

[1] İ. Güner and M. Hamidi Sahneh. Dancing with the stars: Does playing in elite tournaments affect performance? *Oxford Bulletin of Economics and Statistics*, 85(1):1–34, 2023.

[2] R. Halvorsen and R. Palmquist. The interpretation of dummy variables in semilogarithmic equations. *American economic review*, 70(3), 1980.

[3] J. A. Hausman and G. K. Leonard. Superstars in the national basketball association: Economic value and policy. *Journal of Labor Economics*, 15(4):586–624, 1997.

[4] G. Kendall and L. J. Lenten. When sports rules go awry. *European Journal of Operational Research*, 257(2):377–394, 2017.

[5] D. S. Lee and T. Lemieux. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355, 2010.

[6] L. J. Lenten. Unbalanced schedules and the estimation of competitive balance in the scottish premier league. *Scottish Journal of Political Economy*, 55(4):488–508, 2008.

[7] B. Reilly and R. Witt. The effect of league design on spectator attendance: A regression discontinuity design approach. *Journal of Sports Economics*, 22(5):514–545, 2021.

[8] B. Reilly and R. Witt. The effect of league design on club revenues in the scottish premier league. *Eastern Economic Journal*, 50(1):1–28, 2024.

[9] J. Rewilak. Dancing with the stars revisited: does dropping out of the champions league, into the europa league, impact domestic performance? *Managing Sport and Leisure*, pages 1–5, 2022.

[10] S. Rosen. The economics of superstars. *The American economic review*, 71(5):845–858, 1981.

[11] S. Rosen and A. Sanderson. Labour markets in professional sports. *The economic journal*, 111(469):47–68, 2001.

[12] J. D. Speer. The consequences of promotion and relegation in european soccer leagues: A regression discontinuity approach. *Sports Economics Review*, 1:100003, 2023.

[13] H. Sung and B. M. Mills. Estimation of game-level attendance in major league soccer: Outcome uncertainty and absolute quality considerations. *Sport Management Review*, 21(5):519–532, 2018.

[14] S. Szymanski. The economic design of sporting contests. *Journal of economic literature*, 41(4):1137–1187, 2003.