

Research Article

Francesco Angelini, Massimiliano Castellani*, Gery Andrés Díaz Rubio, Simone Giannerini and Greta Goracci

Strategic play and home advantage: coaches' tactical impact in Serie A

<https://doi.org/10.1515/jqas-2025-0113>

Received August 14, 2025; accepted June 22, 2026;

published online July 29, 2026

Abstract: We analyze how coaching strategies and in-game tactics can possibly affect goal difference and home-win probabilities using hand-coded Serie A match commentary (2011/12–2013/14). We analyze home-away goal difference, home-win probability, and home points using generalized linear, logit, and proportional-odds models. By coupling such models with robust and bootstrap inference, we find stable results: aggressive opening schemes are positively associated with performance, shots and goal-kicks exhibit consistent positive effects, crosses display a persistent negative association, and referee decisions-especially red cards and penalties-strongly shape outcomes. These relationships remain robust to alternative specifications, extensive subset selection, and model-averaging procedures that account for model uncertainty. Beyond confirming home advantage, we show that coaching decisions are systematically associated with match performance conditional on pre-match strength and market expectations. Our approach sheds light on the rationale of real-time coaching and introduces a scalable, text-based framework for studying decision-making under uncertainty in competitive environments.

Keywords: sports statistics; home advantage; coaching strategies and tactics

JEL Classification: C25; L83; Z29

*Corresponding author: **Massimiliano Castellani**, Department of Statistical Sciences, “Paolo Fortunati”, University of Bologna, Piazza Teatini 10, 47921, Rimini, Italy, E-mail: m.castellani@unibo.it

Francesco Angelini and Gery Andrés Díaz Rubio, Department of Statistical Sciences, “Paolo Fortunati”, University of Bologna, Piazza Teatini 10, 47921, Rimini, Italy, E-mail: francesco.angelini7@unibo.it (F. Angelini), geryandre.diazrubio2@unibo.it (G. A. D. Rubio)

Simone Giannerini, Department of Economics and Statistics, University of Udine, via Tomadini, 30/A, 33100, Udine, Italy, E-mail: simone.giannerini@uniud.it

Greta Goracci, Faculty of Economics and Management, Free University of Bozen-Bolzano, Piazza Università, 1, 39100, Bozen-Bolzano, Italy, E-mail: greta.goracci@unibz.it

1 Introduction

The influence of coaches on sports outcome has received growing attention in the sports analytics and sports statistics literature. Existing studies highlight that coaches' tactical choices are associated with team performance, although the impact may vary across contexts and competitive environments (Bruinshoofd and ter Weel 2003; Fry et al. 2021). Recent empirical work focused on identifying observable proxies of coaching decisions, often leveraging advances in data collection and match monitoring technologies. Another well-established empirical regularity in team sports is home advantage (HA), which refers to the systematic difference in performance between home and away teams. A large body of evidence documents that playing at home is associated with a higher probability of winning and favorable match outcomes [e.g., Jamieson 2010; Kuper and Szymanski 2009; Pollard 2008; Carron et al. 2005; Courneya and Carron 1992]. The mechanisms underlying HA are heterogeneous and include psychological and contextual factors such as crowd presence, familiarity with the environment, and travel-related constraints.

Typically, HA is associated with referee decisions, crowd support, and team strategies. There is evidence of referee home bias [e.g., Holder et al. 2021; Scoppa 2008] and crowd influence on HA [e.g., Fischer and Haucap 2021; Ponzo and Scoppa 2018]. Few studies have analyzed the relationship between HA and team tactics, such as Dewenter and Namini (2013), who found that home teams often adopt a more offensive style. Recent pandemic-related studies [e.g., Fischer and Haucap 2022; Benz and Lopez 2023; Wunderlich et al. 2021] used ghost games to explore the causal impact of HA, yielding mixed results. Studies on coaching often examine tactical decisions and their associations with performance. For example, Muehlheusser et al. (2018) assessed managerial ability in the Bundesliga but did not account for match-specific tactics. Mesoudi (2020) analyzed formation decisions influenced by past strategies, while Goes et al. (2021b) used tracking data to evaluate player movements and team strategies. Recent work,

[see e.g. Bauer et al. 2023], employed machine learning to classify dynamic formations but focused only on distinct phases of play, overlooking continuous in-game adjustments. Also, Lopez-Valenciano et al. (2022) and Lorenzo-Martínez et al. (2022) explored playing styles and substitutions, emphasizing their roles in shaping performance. A related strand of the literature focuses on substitutions as means used by coaches to influence match dynamics. Empirical and survey-based evidence shows that substitutions reflect both tactical intentions and contextual constraints, and that observed substitution behavior may differ from stated coaching preferences (Wittkugel et al. 2022). Institutional changes, such as the introduction of additional substitutions, can also affect managerial behavior and match outcomes (Meyer and Klatt 2024). Relatedly, crowd presence has been shown to shape managerial decisions in professional football, highlighting the interaction between contextual factors and coaching behavior (Magee 2025). The tactical formation is widely recognized as an important factor in soccer. However, in-game changes in formation and the distinction between offensive and defensive setups have rarely been considered in previous studies on match performance (Forcher et al. 2023). Interest in tactical analysis in soccer has grown in recent years, driven by the increasing availability of player tracking data. For instance, the compact organization of defending teams has been a frequent subject of study (Forcher et al. 2024). Table 8 in Appendix C contains a synthesis of the main empirical studies on these two topics.

While both coaching decisions and HA have been widely documented in empirical research, studying them jointly requires detailed information on match dynamics and tactical choices. The availability of granular match commentary data provides a way to construct observable proxies of coaching decisions at the match level. In this work we study how observable tactical proxies and contextual factors such as home advantage are associated with match outcomes in professional football. The analysis is explicitly associational and does not aim at causal identification. In particular, we investigate the association between observable proxies of coaching decisions and match outcomes while accounting for home advantage. The contribution of this study is not merely to reaffirm the existence of HA, but to show that coaching-related variables remain informative when HA is included in the model. To this end, we build a unique, hand-collected dataset comprising 157,985 event-level observations from Italian Serie A matches spanning the 2011/12 to 2013/14 seasons, sourced from original, non-lemmatized match commentary. We analyze multiple

dimensions of performance, including home-away goal differences, the probability of a home win, and the number of points earned at home. These outcomes are modeled using three distinct approaches within the generalized linear models framework. To ensure robustness and reliability, we include model averaging and bootstrap standard errors. Moreover, we incorporate a model averaging procedure – departing from standard practice – as an additional step to enhance the robustness of our results. Our analysis shows systematic associations between observable tactical proxies, contextual factors such as home advantage, and multiple dimensions of match outcomes. Moreover, while substitutions are not explicitly modeled in this study, our formation-based offensiveness index captures a broader class of tactical adjustments that partly reflects similar strategic considerations.

The paper is structured as follows. Section 2 describes the construction of the database and its classes of variables, divided in: responses, predictors, controls and fixed-effects. Section 3 illustrates the statistical modeling framework and presents the results of the model fit, diagnostics and checking. Section 4 provides a robustness exercise to evaluate the stability of the results across alternative specifications. Section 5 concludes, including a discussion of the impact of the study and possible future investigations.

2 Data and variables description

We collected non-lemmatized soccer textual commentary data from the sports section of the Virgilio website, dedicated to the Italian Serie A league.¹ All commentaries were transformed into dummy variables. For each event, we matched players' names with their respective teams. Consequently, each event is linked to a team, either home or away. Using information about the official role within the team for each season, we also infer the scheme chosen by the coach based on the players present at any given minute. The example in Table 1 illustrates how raw textual commentary

¹ <https://sport.virgilio.it/calcio/serie-a>. 15 matches had no comments, while the Cagliari-Roma match in the 2012/13 season was awarded to Roma by default, with a score of 0–3. Therefore, the effective sample consists of 1,139 matches over 1,140. Commentary information on the other 14 matches was collected from Serie A (<https://www.legaseriea.it/it>) and ESPN (<https://www.espn.com/soccer/league>). For the seasons between 2011/12 and 2013/14, we obtained 157,985 events (or observations). Additional sources were used to integrate missing information that influenced the results, such as penalizations following the *Calciopoli* scandal in 2011 and 2012 or to resolve ambiguous information like homonymy of players. The datasets were cross-checked for possible errors in the commentary and corrected.

Table 1: Illustrative example of the data construction process. Each match is assigned a unique identifier based on the combination of the competing teams and the match date. Event attribution is performed by matching player names appearing in the live-ticker commentary with the corresponding team roster for that match. In the example, Rossi and Ferreira Pinto are identified as Atalanta players; since Atalanta is the home team, the cross and the shot events are attributed to the home side and recorded as home-team actions.

Step	Illustration
Original live-ticker commentary	Italian: 48': <i>Cross dalla destra di Rossi. Tiro di sinistro di Ferreira Pinto. Il portiere interviene. – Atalanta vs Cesena, 21/12/2011</i> English translation (for exposition only): 48': <i>Rossi crosses from the right. Ferreira Pinto shoots with the left foot. Goalkeeper saves. – Atalanta vs Cesena, 21/12/2011</i>
Event coding	<i>match:</i> Atalanta–Cesena <i>date:</i> December 21, 2011 <i>minute:</i> 48 <i>team:</i> Home <i>cross:</i> 1 <i>shot:</i> 1
Minute-level aggregation	<i>match ID:</i> 153 <i>minute:</i> 48 $H_{\text{CROSS}} = 1$ $H_{\text{SHOT}} = 1$
Derived match-level variables	$w_{153,1}$: home-away difference in crosses $w_{153,6}$: home-away difference in shots $\widetilde{w}_{153,1}, \widetilde{w}_{153,6}$: weighted counterparts

is converted into event-level observations and subsequently aggregated into the variables used in the empirical analysis.

To create a balanced dataset, we aggregate multiple events occurring within a given minute, ensuring that every minute of a match contains at least one event. Conversely, some minutes may feature multiple events. This process involved distinguishing between variables that can be aggregated (e.g., the dummy variable for fouls, which can be summed if multiple fouls are called by the referee within the same minute) and those that cannot (e.g., a variable representing a team's formation, which is assumed to remain constant unless explicitly changed by the coach). In the underlying event data, matches exhibit variable numbers of observations beyond the 45th minute of the first half and beyond the 90th minute of the second half, reflecting referee-determined stoppage time. To ensure temporal comparability across matches and to avoid introducing artificial variation driven solely by differences in recorded match length, we allocate the events of the extra-time to minute 45, for the first half, and to minute 90, for the second half. This approach standardizes the time dimension while preserving the ordering of events within each half. Since the amount of added time is itself a referee decision and may be correlated with match circumstances or home-away status, we explicitly include the referee-assigned ending time as a control variable in the empirical models. This allows us to account for systematic differences in stoppage time across matches and mitigates the risk that the aggregation of

added-time events mechanically reflects implicit officiating bias rather than tactical or match-related events. Finally, to ensure a balanced panel structure, we added the missing minutes for each match, resulting in a dataset comprising 1,139 matches, each standardized to 90 min.

To account for the variable number of events (intensity) during a match, we add a weighted version of the explanatory variables to the models. The weight is determined as follows:

$$\omega_{it} \equiv \frac{n_{it}}{n_i} + 1,$$

where n_{it} denotes the number of events at minute t of match $i = 1, \dots, N$, and n_i is the total number of events of match i . For time points $t \in \{45, 90\}$, the weight incorporates the number of events occurring during extra time. The product of the number of events and the weights is normalized to 91. From now on, for a generic variable X , its corresponding weighted version is denoted as $\tilde{X}$.

Our sample contains information on $N = 1,139$ matches. We denote by m_{ij} , the match i , $i = 1, \dots, N$, of the j -th championship season, $j \in \{2011, 2012, 2013\}$. The dataset contains the following variables (all the vectors are row-vectors, unless otherwise indicated):

- Response variables (see Section 2.1)
 - $y_i^{(1)}$: Goal difference between home and away teams (integer valued);
 - $y_i^{(2)}$: Home team win. Binary variable equal to 1 if the home team wins the match, 0 otherwise;

- $y_i^{(3)}$: Points gained in the match by the home team. Takes the values 0 (loss), 1 (tie), 3 (win).
- Predictors (see Sections 2.2.1-2.2.3)
 - $\mathbf{s}_i = (s_{i,1}, s_{i,2}, s_{i,3}, s_{i,1}^*, s_{i,2}^*, s_{i,3}^*)$: coach scheme at the beginning(no asterisk)/end(asterisk) of the match;
 - $\mathbf{w}_i = (w_{i,1}, \dots, w_{i,8})$: team actions;
 - $\mathbf{z}_i = (z_{i,1}, \dots, z_{i,5})$: referees' actions.
- Control variables $\mathbf{c}_i = (c_{i,1}, c_{i,2}, c_{i,3}, c_{i,4}, c_{i,5}, c_{i,6})$ (see Section 2.2.4)
 - $c_{i,1}$: home stadium filling index;
 - $c_{i,2}$: extra time at the end of match;
 - $c_{i,3}$: difference between the home and away teams ranking points before the match.
 - $c_{i,4}$: normalized version of $c_{i,3}$ to lie in $[-1, 1]$.
 - $c_{i,5}$: team market value normalized by the number of players.
 - $c_{i,6}$: average home win betting odds (BbAvH).
- Additional fixed effects $\mathbf{k}_i = (\mathbf{k}_i^s, \mathbf{k}_i^d, \mathbf{k}_i^e)$ (see Section 2.2.5)
 - $\mathbf{k}_i^s = (k_i^{(11)}, k_i^{(12)}, k_i^{(13)})$: season effect;
 - $\mathbf{k}_i^d = (k_{i,1}^d, k_{i,2}^d)$: league day effect;
 - $\mathbf{k}_i^e = (k_{i,1}^e, k_{i,2}^e, k_{i,3}^e)$: extreme event effect.
 - $\mathbf{k}_i^T$: team effect.

When appropriate the variables are expressed as differences between the home and the away team. Only in some cases, we have included in the models measurements for the home (H) or the away (A) team, rather than its difference, and denoted these with the corresponding superscript. For instance, w_1 is the difference in the number of crosses between home and away team and it results $w_1 = w_1^H - w_1^A$. Also, weighted versions of the variables are denoted with an added tilde. Table SM-1 in Supplementary Material (SM) Section 2 contains the variables codebook. For descriptive statistics, see also Tables SM-7 and SM-8 in SM Section 2.

2.1 Response/dependent variables

We select three variables to capture the key aspects of match results and identify three different modeling strategies to account appropriately for their nature.

The first dependent variable $y_i^{(1)}$ is the goal difference between home and away teams (Heuer and Rubner 2009) for match i . In our sample, this variable ranges from -7 to 6 , with mean and median close to 0 indicating symmetry, as also confirmed by skewness tests. The standard deviation (1.67) and coefficient of variation (4.35) suggest high variability relative to the mean. Although the

distribution exhibits a slight leptokurtosis, it is close to a normal distribution. This suggests that the goal difference variable behaves approximately as a continuous variable with a symmetric distribution. Hence, we use a ordinary least squares (OLS) regression for this first analysis.

The second dependent variable $y_i^{(2)}$ is dichotomous and takes the value 1 when the home team wins match i , and 0 otherwise (either tie or loss).² In the seasons spanning from 2011/12 to 2013/14, out of a total of $1,139$ matches, the home team was victorious in 531 instances, accounting for 46.58% of the matches. Out of 609 non-winning matches (53.42%), the home team either faced a defeat (311 matches, 27.28%) or settled for a draw (298 matches, 26.14%). These descriptive statistics indicate a moderate home advantage, as home teams win a marginally larger share of matches than away teams. Nevertheless, while HA plays a role, it is not a decisive factor in every match and our generalized linear models approach allows us to uncover how this is associated with coaches' strategies, team playing home, and in-game factors.

The third dependent variable $y_i^{(3)}$ represents the match outcome for the home team and is treated as an ordered categorical variable with three levels: loss (0), draw (1), and win (3). This response variable is ordinal and has consistent cardinal differences between the values. This makes it suitable for an ordered logit model, which assumes that the dependent variable reflects an underlying latent continuous variable.³

Overall, the three specifications assess different aspects of the match dynamics. On the one hand, logit and ordered logit models are appropriate for categorical outcomes (e.g., win, draw, or loss). On the other hand, multiple regression on the goal difference $y_i^{(1)}$ allows us to analyze the intensity of the match outcome as it accounts for the pseudo-continuous nature of the goal difference.

2.2 Explanatory, control, and additional variables

In this section, we present the explanatory variables associated with coaches, teams and referees decision-making, as well as some control and some additional fixed-effect variables. For the complete codebook see Table SM-1 in SM Section 2.

² In literature such dependent variable has been already employed to study HA (Jamieson 2010), while, over the last 30 years other variables have been considered [Picazo-Tadeo et al. 2017, p. 448].

³ Note that the ordered logit model assumes that the relationship between the independent variables and the dependent variable is the same across the different thresholds (e.g., between loss and tie, and between tie and win), i.e. *parallel- or proportional-odds assumption*.

2.2.1 Coach actions/schemes

The analysis uses an offensiveness index derived from team formations as a proxy for pre-match and in-match tactical scheme.⁴ The measure summarizes observable differences in role allocation across formations, rather than fully capturing underlying tactical behavior during the match. As such, it relies on some simplifying assumptions. First, this index does not capture within-role tactical variation, such as fullbacks pushing high, defensive forwards, or fluid positional play, nor does it allow the offensive contribution of a given role to vary across teams, coaches, or seasons. Formations with identical role distributions are therefore treated as observationally equivalent. In addition, formation changes are identified from commentary-based event streams and may not perfectly coincide with actual tactical adjustments on the field. The offensiveness index should therefore be interpreted as a coarse, formation-based proxy for tactical scheme rather than a complete representation of team behavior. Furthermore, the offensiveness index applies a common mapping from formations to tactical intensity, implicitly assuming that the tactical meaning of a given role distribution is comparable across teams and coaches. In practice, this mapping is likely heterogeneous: teams with identical formations may implement them differently depending on coaching style, player characteristics, and organizational constraints (e.g., attacking fullbacks versus conservative wide defenders). As a result, the index may measure offensive orientation with varying precision across teams.⁵

Formally, for each match i , we construct three versions of the offensiveness differential, denoted by $(s_{i,1}, s_{i,2}, s_{i,3})$, and include only one version at a time in each model. The measures are defined as the difference between the home and away teams' offensiveness indices. The indices are based on the nominal roles implied by the reported formation (defender, midfielder, forward), as detailed in Appendix A. This approach treats formations as observable summaries of coaches' role assignments, while abstracting from within-role variation in positioning or behavior.

⁴ Additional player indices can be explored to construct an offensiveness index for coaches' strategies, as exemplified in Cefis (2023) and Cefis and Carpita (2024).

⁵ The estimated coefficients should therefore be interpreted as averages over heterogeneous tactical implementations and may be attenuated where the proxy is less informative. While team-level controls partially address persistent differences, fully capturing this heterogeneity would require richer positional or tracking data beyond the scope of the current analysis.

Descriptive statistics are reported in Tables SM-2 and SM-3 in SM Section 3.

As a simple validation exercise, we examine whether higher values of the offensiveness index are associated with attacking outcomes. Teams with higher offensiveness indices exhibit higher average shot counts and goal totals, providing evidence that the index is positively correlated with observable attacking activity. While this does not validate the index as a causal measure of tactical intent, it supports its use as a proxy for offensive orientation in the empirical analysis.⁶

2.2.2 Team actions

The variables $\mathbf{w}_i = (w_{i,h})$, with $h = 1, \dots, 8$ measure the activities of the teams through 8 different actions. They are computed as:

$$w_{i,h} \equiv \sum_{t=1}^{90} w_{i,h,t}, \quad h = 1, \dots, 8.$$

Specifically, the actions are: total number of crosses ($h = 1$), corners ($h = 2$), crosses from corners ($h = 3$), crosses not from corners ($h = 4$), corners without crosses ($h = 5$), shots ($h = 6$), goal kicks ($h = 7$) and offsides ($h = 8$). Note that $w_{i,3}, w_{i,4}, w_{i,5}$ are derived from the other variables. Table SM-4 in SM Section 2, presents some descriptive statistics for these variables. They reveal insights into the tactical execution and strategic decisions made by teams in soccer matches. The differences between home and away team actions provide a quantitative measure of the HA in offensive output. The mean differences across most actions are positive, indicating that home teams, on average, undertake more offensive actions than away teams. This HA is most pronounced in crosses and shots, essential elements of attacking play. The distribution of these differences again shows variability in match strategies, with skewness and kurtosis values suggesting a range of outcomes from balanced to highly aggressive home performances. The differences in goal kicks and offsides between home and away performances are minimal, indicating that teams maintain consistent counter-offensive and defensive tactics regardless of playing at home or away. A slightly negative value for home-away difference in goal kicks indicates that, on average, more home offensive actions tend to generate away goal kicks rather than the opposite.

⁶ These results are available from the authors upon request.

2.2.3 Referee actions

The variables denoted as $\mathbf{z}_i = (z_{i,h})$, with $h = 1, \dots, 5$ represent referee decisions for match i . Specifically, they measure the number of yellow cards ($h = 1$), red cards ($h = 2$), free kicks ($h = 3$), penalty kicks ($h = 4$), and total fouls ($h = 5$) leading to free and penalty kicks. These are calculated analogously to $\mathbf{w}_i$. Table SM-5 in SM Section 2 reports some descriptive statistics of the referee actions. The distribution of yellow and red cards shows a nuanced picture of match discipline. The statistics for fouls leading to free kicks and penalties offer further insight into the tactical and disciplinary dimensions. The differences in yellow and red card distributions between home and away teams are relatively small, indicating that, while there may be slight biases or strategic differences in play style, overall disciplinary actions are evenly distributed across venues. The statistics for fouls leading to free kicks and penalties provide further information on the tactical and disciplinary dimensions. Even though there is a slight indication of HA influencing disciplinary measures, the descriptive statistics suggest that the teams maintain consistent levels of discipline irrespective of venue.

2.2.4 Control variables

The control variables are denoted as $\mathbf{c}_i = (c_{i,h})$, with $h = 1, \dots, 6$. In particular, the home stadium filling index, ($c_{i,1} \in [0, 1]$) is a continuous control variable computed as the proportion of spectators present at the home stadium relative to its capacity.⁷ This captures the influence of home fans compared to away fans.⁸ The variable $c_{i,2} \in [0, 15]$ is the extra time at the end of the match, measured in added minutes.⁹ This variable, determined by referees to compensate for lost time during the match, introduces an element of randomness to its duration. Furthermore, to account for the team effect, we introduce the control variable $c_{i,3}$ and its normalized version $c_{i,4}$. $c_{i,3}$ is the difference of points

between the home and away teams before the match. $c_{i,4} \in [-1, 1]$ is the ratio of ranking points over the maximum attainable ranking points before the match. We also include the control variable $c_{i,5}$ as a proxy of teams' economic strength by measuring, for each season, the average market value normalized by the number of players.¹⁰ Finally, we include average home win betting odds as an additional control variable $c_{i,6}$. These odds measure how betting markets assess the probability that the home team will win a match. Bookmakers quote decimal odds for each possible outcome (home win, draw, away win). For a given match, multiple bookmakers typically provide odds for a home win, and the average home-win odds are computed as the mean across bookmakers. This variable summarizes pre-match expectations formed using a wide range of publicly and privately available information, including team strength, recent performance, injuries, and other contextual factors.¹¹

The home stadium filling index, $c_{i,1}$, presents a median of 0.56 and mean of 0.59. Its standard deviation is 0.19 and its skewness is 0.15, suggesting a slightly right-skewed but nearly mesokurtic distribution. The variable extra time $c_{i,2}$ has a mean of 5.71 min, with a small standard deviation and relatively small coefficient of variation. Its distribution is slightly right-skewed and leptokurtic, suggesting some heavier tail activity. Its weighted version $\tilde{c}_{i,2}$ ranges from 0.00 to 15.51, with a mean of 5.91 and a standard deviation of 1.79, indicating moderate dispersion around the mean, slight skewness and moderate kurtosis. In addition, $c_{i,4}$ has a mean of -0.01 and a standard deviation of 0.27, reflecting some disparity in team strengths. Its distribution exhibits moderate skewness and heavier tails compared to the Gaussian distribution. Descriptive statistics for home-away differences in team market value, in both levels and logarithms, show distributions centered around zero, with medians equal to zero and means near zero in logs, indicating broad balance on average between home and away teams. At the same time, the wide range of values reflects substantial heterogeneity in relative team strength across matches. By contrast, betting odds display markedly different distributional properties. The average home-win odds are right-skewed and highly leptokurtic. This reflects the bounded nature of odds on the left and the presence of rare matches with very low perceived home-win

⁷ Attendance figures and stadium capacity data were collected from www.stadiapostcards.com.

⁸ It is assumed that, on average, home spectators occupy at least half of the available seats as shown by the percentage of spectators who have season seats for home matches.

⁹ For the weighted extra time variable, we identified the number of events in each half's extra time ($N_{i,1} + N_{i,2} = N_i$), and then divided these counts by the total number of events in the i -th match N_i . We added 1 to each of these ratios and multiplied by the respective extra-time minutes ($c_{i,2,1}$ and $c_{i,2,2}$). Summing the two weighted components gives: $\tilde{c}_{i,2} = c_{i,2,1} \times \left(1 + \frac{N_{i,1}}{N_i}\right) + c_{i,2,2} \times \left(1 + \frac{N_{i,2}}{N_i}\right)$. This increases extra-time minutes for halves that have more events relative to the overall match.

¹⁰ The market value covariates were sourced from www.transfermarkt.it.

¹¹ Average home-win betting odds (BbAvH) are the mean of decimal odds quoted by multiple bookmakers for a home-team victory in a given match; lower values indicate a higher market-assessed probability of a home win. Betting odds were sourced from <https://www.kaggle.com/datasets/mexwell/historical-football-resultsbetting-odds-data>.

probabilities. Compared with market value differences, odds exhibit much lower relative dispersion, suggesting that betting markets compress information more tightly despite underlying heterogeneity in team strength.

2.2.5 Other fixed effects: season, league day, extreme events, team

This group includes other fixed effects denoted as $\mathbf{k}_i = (\mathbf{k}_i^s, \mathbf{k}_i^d, \mathbf{k}_i^e)$. First, $\mathbf{k}_i^s = (k_i^{(11)}, k_i^{(12)}, k_i^{(13)})$ are three dummies that capture the season effects for 2011, 2012, and 2013. Second, $\mathbf{k}_i^d$ represents the league day effect that accounts for the specific match day within the championship. Specifically, we assign a rank to each distinct date of a match, giving the same rank to identical dates and incrementing by one for the next distinct date, as to create consecutive ranks without gaps. $k_{i,1}^d$ uses the absolute ranks, whereas in $k_{i,2}^d$ the ranks are scaled from 0 to 1, corresponding to the earliest and latest ranked dates, respectively. These variables are needed because the actual match date often differs from the scheduled date for reasons such as security or adverse weather. By referencing the actual playing date instead of the originally scheduled date, the ranking and relative date calculations accurately reflect the true chronological order of matches and thus ensure consistency in determining each team's standings. Third, we use $\mathbf{k}_i^e = (k_{i,1}^e, k_{i,2}^e, k_{i,3}^e)$ to control for anomalous results, i.e. matches that ended with a difference of at least five goals between the home and away teams. The first of these dummy variables, $k_{i,1}^e$ indicates extreme goal difference matches in absolute value, whereas $k_{i,2}^e$ and $k_{i,3}^e$ account for positive and negative goal differences for the home team, respectively. The variable $k_{i,1}^e$ has a median and mean of about 0.51, near-zero skewness, and slightly lower-than-normal kurtosis. The two dummy variables $k_{i,2}^e$ and $k_{i,3}^e$, (extreme positive and extreme negative goal differences for the home team), are rare as they occur only nine times overall. The dummy variables $\mathbf{k}_i^T$ add a fixed effect for each of the 26 teams. This is done to control for persistent, unobserved heterogeneity across the 26 Serie A teams in our dataset, as each team has distinct characteristics that can systematically influence both strategic choices and match outcomes. Tables SM-6 and SM-7 in SM Section 2 present details for controls and additional fixed effects, while Figure 1 shows the scatterplot of relative team effects c_4 against goal difference $y^{(1)}$, highlighting extreme matches $k_{i,1}^e$.

3 Statistical modeling

This section is devoted to the modeling exercise. The aim is to study how coaching decisions, team actions, and referee

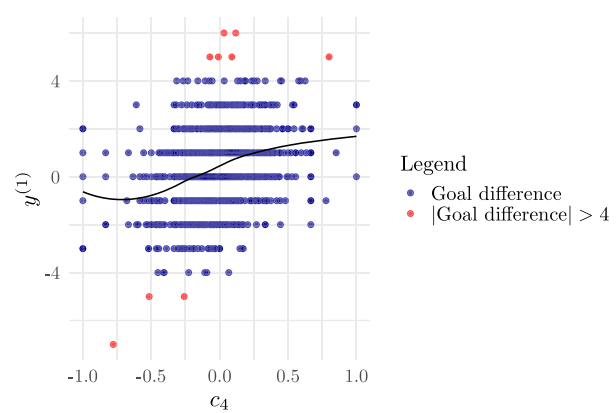


Figure 1: Scatterplot of the relative team effect, measured as the relative difference between ranking points up to the previous match (c_4), against the goal difference ($y^{(1)}$). Matches classified as extreme ($k_{i,1}^e$) are highlighted in red.

events are associated with match performance. To capture different dimensions of match outcomes, we adopt three complementary specifications. Specifically, we estimate: (i) a multiple regression model for goal difference (Model 1), (ii) a binary logit model for the probability that the home team wins (Model 2), and (iii) an ordered logit model for the ordered outcomes win, draw, and loss (Model 3). Model 1 uses goal difference as a continuous outcome and focuses on the intensity of match results. This specification is informative on how covariates relate to changes in performance beyond winning or losing, and it allows us to distinguish between narrow and large victories or defeats and to assess. Model 2 considers a binary outcome indicating whether the home team wins. Binary win/loss specifications are widely used in the sports-econometrics literature because they provide a transparent and easily interpretable framework for analyzing match outcomes and winning probabilities directly (see, e.g., Cross and Uhrig 2023; Ponzo and Scoppa 2018). In this setting, estimated effects can be interpreted in terms of marginal changes in winning chances without imposing an ordinal structure on the outcome space. For this reason, Model 2 serves as a natural benchmark specification and facilitates comparison with previous empirical studies. Model 3 extends the analysis by treating match outcomes as an ordered categorical variable (win, draw, loss), preserving more information than the binary specification while imposing an ordinal structure on the outcome space. This model allows us to examine how covariates shift probability mass across outcome categories, subject to the proportional odds assumption. The two nonlinear specifications are specified to provide complementary rather than redundant perspectives on match outcomes. In particular, comparing Models 2 and 3 allows us to assess whether incorporating draws and the ordinal outcome structure

changes the estimated relationships between covariates and match performance. Taken together, Model 2 functions both as a robustness benchmark and as a specification closely aligned with the existing sports-econometrics literature.¹²

Section 3.1 details the specifications used, whereas Section 3.2 discusses the results of model fitting. Section 3.3 refines the full specifications via AIC and BIC to retain only significant effects and performs residuals diagnostics to confirm model validity. Section 4 uses Akaike-weight averaging over top models to yield robust shrinkage estimates.

3.1 Model specifications

In the following, for a generic match i , with $i = 1, \dots, N$, and the response $y_i^{(j)}$, with $j = 1, 2, 3$, we adopt the following notation where the variable $\mathbf{X}_i$ includes all the predictors

$$\mathbf{X}_i = (\mathbf{s}_i, \mathbf{w}_i, \mathbf{z}_i, \mathbf{c}_i, \mathbf{k}_i).$$

Recall that the vectors of explanatory variables $\mathbf{s}_i$, $\mathbf{w}_i$, and $\mathbf{z}_i$ correspond to the three decisional levels, *coach*, *team*, and *referee* actions, whereas $\mathbf{c}_i$ and $\mathbf{k}_i$ are the vector of control variables and additional fixed effects, respectively. Since the three specifications do not necessarily include the whole set of predictors $\mathbf{X}_i$, we define the set of predictors for the j -th model as $\mathbf{X}_i^{(j)} = \mathbf{X}_i \mathbf{R}^{(j)}$, with $j = 1, 2, 3$ where we use the block diagonal selector matrix $\mathbf{R}^{(j)}$. We denote the corresponding parameter (column) vector as $\boldsymbol{\theta}^{(j)}$ so that the linear predictor for match i and model j results $\mathbf{X}_i^{(j)} \boldsymbol{\theta}^{(j)}$. We assume the parameter vectors to be column-vectors.

Model 1. The response variable $y_i^{(1)}$ represents home-away goal differences so that we specify a OLS regression model for it:

$$\mathbb{E} \left[y_i^{(1)} \mid \mathbf{X}_i^{(1)} \right] = \mathbf{X}_i^{(1)} \boldsymbol{\theta}^{(1)}, \quad i = 1, \dots, N, \quad (1)$$

$$y_i^{(1)} = \alpha_0 + \mathbf{s}_i^{(1)} \boldsymbol{\alpha}_s + \mathbf{w}_i^{(1)} \boldsymbol{\alpha}_w + \mathbf{z}_i^{(1)} \boldsymbol{\alpha}_z + \mathbf{c}_i^{(1)} \boldsymbol{\eta}_c + \mathbf{k}_i^{(1)} \boldsymbol{\eta}_k + \varepsilon_i, \quad (2)$$

with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The parameter vector results $\boldsymbol{\theta}^{(1)} = (\alpha_0, \boldsymbol{\alpha}_s^\top, \boldsymbol{\alpha}_w^\top, \boldsymbol{\alpha}_z^\top, \boldsymbol{\eta}_c^\top, \boldsymbol{\eta}_k^\top)^\top$.

¹² To formally evaluate this issue, we extend the classical Brant test through a bootstrap-based procedure for logistic estimators. The results provide mixed evidence, suggesting that the binary specification may serve as a parsimonious summary in some cases, while in others the ordered structure captures additional information associated with draws.

Model 2. Here $y_i^{(2)}$ is a binary variable indicating home-team-win so we model the probability of winning the i -th match for the home team:

$$\begin{aligned} \mathbb{E} \left[y_i^{(2)} \mid \mathbf{X}_i^{(2)} \right] &= \mathbb{P} \left(y_i^{(2)} = 1 \right) = g \left(\mathbf{X}_i^{(2)} \boldsymbol{\theta}^{(2)} \right); \\ g(x) &= \frac{e^x}{1 + e^x} \in [0, 1], \\ \mathbf{X}_i^{(2)} \boldsymbol{\theta}^{(2)} &= \beta_0 + \mathbf{s}_i^{(2)} \boldsymbol{\beta}_s + \mathbf{w}_i^{(2)} \boldsymbol{\beta}_w + \mathbf{z}_i^{(2)} \boldsymbol{\beta}_z \\ &\quad + \mathbf{c}_i^{(2)} \boldsymbol{\phi}_c + \mathbf{k}_i^{(2)} \boldsymbol{\phi}_k, \end{aligned} \quad (3)$$

where the function $g(\cdot)$ represents the standard logistic cumulative distribution function.

The parameter vector results $\boldsymbol{\theta}^{(2)} = (\beta_0, \boldsymbol{\beta}_s^\top, \boldsymbol{\beta}_w^\top, \boldsymbol{\beta}_z^\top, \boldsymbol{\phi}_c^\top, \boldsymbol{\phi}_k^\top)^\top$.

Model 3. We model the probability of match win, draw and loss for the home team using an ordered logit model for cross-sectional data:

$$\begin{aligned} \mathbb{P} \left(y_i^{(3)} \leq h \right) &= g \left(\pi_h - \mathbf{X}_i^{(3)} \boldsymbol{\theta}^{(3)} \right), \quad h = 0, 1; \\ g(x) &= \frac{e^x}{1 + e^x}; \\ \mathbb{P} \left(y_i^{(3)} = 3 \right) &= 1 - \mathbb{P} \left(y_i^{(3)} \leq 1 \right), \\ \mathbf{X}_i^{(3)} \boldsymbol{\theta}^{(3)} &= \mathbf{s}_i^{(3)} \boldsymbol{\gamma}_s + \mathbf{w}_i^{(3)} \boldsymbol{\gamma}_w + \mathbf{z}_i^{(3)} \boldsymbol{\gamma}_z + \mathbf{c}_i^{(3)} \boldsymbol{\psi}_c + \mathbf{k}_i^{(3)} \boldsymbol{\psi}_k. \end{aligned} \quad (4)$$

Here the response variable $y_i^{(3)} \in \{0, 1, 3\}$ (loss, draw, or win) is treated as an ordinal outcome with two cutpoints π_0 and π_1 (satisfying $\pi_0 \leq \pi_1$). As above, $g(\cdot)$ is the standard logistic cumulative distribution function. The cumulative probabilities $\mathbb{P}(y_i^{(3)} \leq h)$ for $h = 0, 1$ are modeled via logistic curves. In this specification, the *proportional-odds* assumption is retained. One can visualize the shifts in the conditional distribution of $y_i^{(3)}$ as $\mathbf{X}_i^{(3)}$ changes, where an increasing $\mathbf{X}_i^{(3)} \boldsymbol{\theta}^{(3)}$ shifts probability mass towards higher-ordered categories. The parameter vector results $\boldsymbol{\theta}^{(3)} = (\pi_0, \pi_1, \boldsymbol{\gamma}_s^\top, \boldsymbol{\gamma}_w^\top, \boldsymbol{\gamma}_z^\top, \boldsymbol{\psi}_c^\top, \boldsymbol{\psi}_k^\top)^\top$.

To assess the validity of the parallel regression (proportional-odds) assumption, we augment the classical test of Brant (1990) by incorporating a nonparametric bootstrap for the estimators (Schlegel and Steenbergen 2020). Each cutpoint π_h is treated separately, yielding a sequence of binary logistic regressions of the form $\pi_h - \mathbf{X}_i^{(3)} \boldsymbol{\theta}^{(3)}$. The resulting threshold-specific slope estimates are stacked and evaluated jointly via a Wald-type statistic to test the equality of the unconstrained coefficient vectors $\boldsymbol{\beta}$ across cutpoints, as implied by the proportional-odds restriction. Under the null hypothesis of parallel regression, both test statistics

converge in distribution to chi-squared limits, with degrees of freedom determined by the number of slope parameters and cutpoints.

Under the proportional-odds restriction, the probability of a win (i.e., the top outcome category) implied by Model 3 shares a common slope vector with the binary logit specification and, when included, features an intercept equal to the negative of the first cutpoint, $-\pi_1$. Let $\hat{\Delta} = \hat{\beta} - \hat{\gamma}$ denote the vector of slope differences between Model 2 and the implied win equation from Model 3. We test the null hypothesis $H_0: \hat{\Delta} = \mathbf{0}$. When Model 2 includes an intercept, we additionally test the threshold-intercept restriction $\beta_0 + \pi_1 = 0$, which is necessary for full equivalence between the binary and ordered specifications, given that Model 3 necessarily includes cutpoints. This procedure provides a formal test of whether exploiting the additional ordinal information in Model 3 induces significant deviations in slope estimates relative to the binary specification. Failure to reject compatibility implies that Model 2 delivers a parsimonious representation of the win component embedded in the ordered model, whereas rejection isolates the role of draws in shaping covariate effects.

3.2 Results: main models

We begin by estimating Model 1 via a Gaussian generalized linear model (GLM), equivalent to ordinary least squares under the identity link. Maximum likelihood estimation yields the coefficient vector together with heteroskedasticity consistent (HC) variance estimates. Similarly, Model 2 is also estimated by maximum likelihood with robust standard errors. For Model 3, we implement an ordinal logit specification that jointly estimates the regression coefficients and threshold terms, by maximizing the likelihood under a logistic link. Standard errors are derived via nonparametric bootstrap resampling to accommodate potential model misspecification. For each model we start by fitting six main specifications, corresponding to the 6 different choices for the scheme variable, i.e. ($s_1, s_2, s_3, \tilde{s}_1, \tilde{s}_2, \tilde{s}_3$). Table 2 reports the results for the scheme $\tilde{s}_2$ for Model 1, 2, 3, since it consistently resulted the best fit among the 6 choices, and includes the two additional pre-match controls c_5 and c_6 . To evaluate model performance, we compute Akaike's information criterion (AIC) (Akaike 1974), Schwarz's Bayesian information criterion (BIC) (Schwarz 1978), and goodness-of-fit measures including adjusted R^2 , $\bar{R}^2$ (for pseudo-continuous outcomes), McFadden's R^2_{MF} , Nagelkerke's R^2_N , deviance, and classification metrics (for discrete outcomes).

Across all specifications, s_2 is consistently positive (p -values < 0.001) and $\tilde{s}_2^*$ is significantly negative (p -values < 0.001). For the team variables, $\tilde{w}_1$ is significantly negative

(p -values < 0.005), whereas $\tilde{w}_2$ is also negative but fails to reach significance. However, under the current specification, $\tilde{w}_2$ is statistically significant in Models 1 and 3 but not in Model 2. Both $\tilde{w}_6$ and $\tilde{w}_7$ are positive and statistically significant across Models 1-3, while $\tilde{w}_8$ is significant in Models 1 and 3, but the significance is not striking in Model 2 (p -value = 0.05). Among referee variables, $\tilde{z}_2$ remains significantly negative and $\tilde{z}_4$ significantly positive in all specifications; $\tilde{z}_1$ and $\tilde{z}_3$ are not statistically supported outside Model 1. The controls c_1 (and its powers) are insignificant, and $\tilde{c}_2$ is modestly negative. Pre-match controls enter with stable and significant signs: $c_5 > 0$ (log market value difference) and $c_6 < 0$ (betting odds), with c_6 significant at the 5 % level in Model 2. The relative team effect c_4 is not supported. Additional fixed effects k_2^e and k_3^e are significant with opposite signs, and selected season indicators are significant only in Model 1. In the ordered-logit specification, thresholds -3.25 and -1.35 imply predicted probabilities (at zero covariates) of 0.037 (loss), 0.168 (draw), and 0.796 (win); the Brant test ($p = 0.971$) does not reject proportional odds. Overall fit is comparable across models, with modest differences in classification performance between Models 2 and 3. These specifications provide stable benchmarks for assessing coaching, team, and referee effects while controlling for pre-match strength and market expectations.

Beyond the inclusion of classification metrics and information criteria, we evaluated our regression models using several diagnostic tests. For Model 1, we computed the Shapiro-Wilk (SW) test of Gaussianity of the residuals, the Breusch-Pagan (BP) test for heteroskedasticity and Ramsey's RESET (RR) test for model specification (Ramsey 1969) (i.e. omitted variables or incorrect functional forms). For Model 2, we computed the above RR test, and the Hosmer-Lemeshow (HL) test of goodness-of-fit (Hosmer et al. 2013). As for Model 3, we computed the generalized HL test, the Lipsitz et al. (1996) (LI) goodness-of-fit tests and the Brant test (Brant 1990) that verifies the null hypothesis of proportional-odds. Selected diagnostics are reported in Table 1 (bottom panels). In the main specifications, Model 1 does not reject SW or RR but rejects BP; Model 2 does not reject HL and shows weak evidence against RR; Model 3 does not reject the LI, HL, or Brant tests. Table 2 reports the full-statistic paired bootstrap cross-model equality tests for three main variants. The direct main (Specification-1) rejects equality, whereas both the alternative specification replacing k_2^e and k_3^e with k_1^e (Baseline-2) and the specification excluding all extreme-result indicators (k_1^e, k_2^e, k_3^e ; Baseline-3) fail to reject the null (Table 3).

Table 2: Fit and performance of main regression models, including the two controls c_5 and c_6 , with robust standard error. Standard errors for Model 1 and 2 use the sandwich estimator HC3, those for Model 3 use a nonparametric bootstrap ($B = 1,000$). Baseline for Model 3 is $k^{(11)} = 1$. Number of observations/matches $N = 1,139$.

Models				Model 1		Model 2		Model 3	
Description		Variable	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	
Scheme									
Initial scheme 2		s_2	0.24	(0.000)	0.47	(0.000)	0.46	(0.000)	
Final weighted scheme 2		$\tilde{s}_2^*$	−0.22	(0.000)	−0.48	(0.000)	−0.47	(0.000)	
Team									
Weighted crosses		$\tilde{w}_1$	−0.02	(0.000)	−0.04	(0.000)	−0.04	(0.000)	
Weighted corners		$\tilde{w}_2$	−0.03	(0.015)	−0.03	(0.171)	−0.05	(0.006)	
Weighted shots		$\tilde{w}_6$	0.03	(0.000)	0.05	(0.000)	0.05	(0.000)	
Weighted goal kicks		$\tilde{w}_7$	0.03	(0.000)	0.06	(0.002)	0.05	(0.002)	
Weighted offsides		$\tilde{w}_8$	0.03	(0.023)	0.06	(0.050)	0.05	(0.039)	
Referee									
Weighted yellow cards		$\tilde{z}_1$	−0.03	(0.125)	−0.07	(0.180)	−0.06	(0.162)	
Weighted red cards		$\tilde{z}_2$	−1.02	(0.000)	−2.12	(0.000)	−2.06	(0.000)	
Weighted free kicks		$\tilde{z}_3$	−0.02	(0.019)	−0.01	(0.633)	−0.01	(0.312)	
Weighted penalty kicks		$\tilde{z}_4$	0.30	(0.000)	0.46	(0.018)	0.44	(0.008)	
Controls									
Linear stadium filling		c_1	−1.01	(0.660)	−3.55	(0.550)	−2.85	(0.578)	
Quadratic stadium filling		c_1^2	2.31	(0.593)	7.39	(0.505)	5.90	(0.534)	
Cubic stadium filling		c_1^3	−1.49	(0.551)	−4.57	(0.473)	−3.68	(0.495)	
Weighted extra minutes		$\tilde{c}_2$	−0.04	(0.123)	−0.08	(0.092)	−0.03	(0.496)	
Relative team effect		c_4	0.19	(0.309)	−0.02	(0.959)	0.03	(0.937)	
Log average market value difference		c_5	0.47	(0.000)	0.81	(0.000)	0.75	(0.000)	
Average home Win betting odds		c_6	−0.19	(0.000)	−0.36	(0.016)	−0.44	(0.000)	
Additional fixed effects									
Season 2011/12		$k^{(11)}$	1.19	(0.006)	1.45	(0.198)	−	−	
Season 2012/13		$k^{(12)}$	1.11	(0.012)	1.48	(0.193)	−0.07	(0.666)	
Season 2013/14		$k^{(13)}$	1.10	(0.013)	1.52	(0.176)	−0.02	(0.924)	
Scaled ranked league day		k_2^d	0.04	(0.740)	0.09	(0.735)	0.05	(0.828)	
Extreme positive goal difference		k_2^e	3.71	(0.000)	13.80	(0.000)	12.28	(0.000)	
Extreme negative goal difference		k_3^e	−4.56	(0.001)	−13.36	(0.000)	−14.66	(0.000)	
Thresholds									
0 1		τ_0	−	−	−	−	−3.25	(0.001)	
1 3		τ_1	−	−	−	−	−1.35	(0.160)	
Model	AIC	BIC	Deviance	$\bar{R}^2$ $\bar{R}_N^2$	Accuracy	Sensitivity	Specificity	F1	Brant
1	3,665	3,791	1,593	0.51	−	−	−	−	−
2	1,074	1,195	1,026	0.51	0.79	0.80	0.78	0.80	−
3	1,761	1,887	1,711	0.52	0.68	0.67, 0.38, 0.85	0.90, 0.85, 0.73	0.69, 0.43, 0.79	0.971

Our findings are broadly consistent with the existing literature while adding new evidence on the role of tactical variables. We replicate earlier mixed results (De Paola and Scoppa 2012) by showing that more aggressive opening tactics are positively associated with match performance, conditional on end-of-match attacking levels and team fixed effects. We also recover established empirical patterns, including the negative association between crossing frequency and scoring outcomes (the “cross paradox”

Sarkar 2018) as well as red cards and goal opportunities (Vecer et al. 2009), and the strong association between penalties and match outcomes (Sarmiento et al. 2018). In addition, the results show stable associations between match outcomes and variables such as goal-kicks and offsides, which have received less attention in the literature. More generally, the estimates indicate that in-match tactical variables are systematically associated with performance outcomes, suggesting that observed tactical adjustments co-vary with

Table 3: Paired bootstrap cross-model equality test (full statistic) for the main model and two variations (1: specification; 2. with k_1^e instead of k_2^e and k_3^e ; 3. without k_1^e, k_2^e, k_3^e), with market value and betting odds. Bootstrap replications: 1,000.

Paired bootstrap test	W	df	$p\text{-value}(\chi^2)$	$p\text{-value}(\text{boot})$
Baseline-1	68.98	22	0.00	0.00
Baseline-2	21.16	21	0.45	0.45
Baseline-3	19.20	41	0.99	0.97

Sample: $N = 1,139$, WIN rate = 0.465.

match events. These results describe how tactical choices and in-game events evolve alongside outcomes over the course of a match, but they should be interpreted as conditional associations rather than evidence of causal effects or optimal decision-making under uncertainty.

3.3 Results: extended models

We extend the main models of Section 3.2 by adding team market value (c_5) and betting odds (c_6) and performing a full subset search over block-nested specifications. Table 4 reports the AIC- and BIC-ranked specifications for Models 1-3, ordered from best (rank-1) to worst (rank-4). As before, AIC favors richer specifications, whereas BIC selects more parsimonious alternatives. For Model 1, AIC excludes team fixed effects but retains key action and referee variables, while BIC includes an intercept capturing home advantage. For Models 2-3, inclusion of c_5 and c_6 preserves the qualitative structure of the preferred specifications, with only minor reshuffling among referee and extreme-score indicators.

The second coaching scheme (s_2) remains frequently selected under both criteria: the initial scheme is positive and highly significant, whereas the final scheme is negative and highly significant. Team actions display stable patterns: crosses (w_1) are negative, shots (w_6) and goal kicks (w_7) positive and significant, and offsides (w_8) selected in all specifications (weaker under Model 2 AIC).

Both pre-match controls are strongly supported. Market value (c_5) enters positively and significantly across all models, while betting odds (c_6) enter negatively and significantly. Information criteria and goodness-of-fit measures indicate stable performance, and classification metrics for Models 2-3 remain largely unchanged. Substantively, estimated effects of coaching, team actions, referee decisions, and home advantage are unaffected.

Diagnostics are broadly supportive. In Model 1 (AIC), BP rejects homoskedasticity (addressed via HC3 standard errors), while SW and RR do not reject. In Model 2, HL does

not reject correct fit, though the BIC specification rejects RR. In Model 3, HL, LI, and Brant tests do not indicate misspecification or proportional-odds violations. Table 5 further shows that extreme-result indicators (k_2^e, k_3^e) drive rejection under certain AIC intercept specifications; replacing or excluding them yields non-rejection, motivating the reporting of both logit and ordered-logit benchmarks.

Overall, Table 4 confirms that extending the models with market value and betting odds improves fit modestly while leaving the core empirical conclusions unchanged, providing robustness to richer pre-match information and to AIC/BIC model-selection variation.

4 Model averaging

Selecting from a broad pool of candidate models may induce selection bias and compromise the stability of the results. To mitigate this concern, we adopt the model-averaging framework of Lukacs et al. (2010).¹³ For each main model (Models 1-3), we retain the best-fitting AIC specifications (excluding interaction terms). The resulting candidate pools contain 579 specifications for Model 1, 263 for Model 2, and 101 for Model 3. Each candidate set is partitioned into specifications using home-away differences for coaching variables (Set 1) and those with separate home and away components (Set 2). Model averaging is performed separately within each set, treating each as an independent candidate pool for inference. The procedure is implemented separately for each main model $m \in \{1, 2, 3\}$ and, within each model, separately for Set 1 and Set 2.

Formally, let $\theta^{(m)}$ denote the parameter vector associated with main model m . For a generic parameter $\theta \in \theta^{(m)}$, with $m \in \{1, 2, 3\}$, let $\hat{\theta}_l$ denote its estimate under candidate specification l , $l = 1, \dots, L$. The model-averaged (shrinkage) estimator $\tilde{\theta}$ and its variance approximation are given by:

$$\tilde{\theta} = \sum_{l=1}^L p_l \hat{\theta}_l, \quad \widehat{\text{Var}}(\tilde{\theta}) = \sum_{l=1}^L p_l \left[\widehat{\text{Var}}(\hat{\theta}_l) + (\hat{\theta}_l - \tilde{\theta})^2 \right],$$

where p_l is the Akaike weight (interpretable as the posterior probability that model l minimizes the Kullback-Leibler

¹³ Model averaging combines estimates across a set of plausible model specifications, weighting them according to information criteria (AIC/BIC), rather than selecting a single “best” model this approach accounts for model uncertainty and allows us to assess whether the estimated associations are sensitive to specification choice. The model-averaged results do not introduce a new identification strategy, but instead provide a robustness exercise that evaluates the stability of coefficient signs, magnitudes, and statistical significance across alternative admissible models. We thank an anonymous reviewer for this suggestion.

Table 4: Rank-1 AIC- and BIC-selected models for Models 1–3 with unweighted in-game actions including block-interactions adding the controls for Market Value and BbAvH (Betbrain average home win odds). Significance levels in parentheses. Model 1 uses robust HC3 SE. Below, p -values for SW (Shapiro-Wilk), KS (Kolmogorov-Smirnov), JB (Jarque-Bera), BP (Breusch-Pagan), and RR (Ramsey RESET). Model 2 also has robust HC3 SE, with p -values for HL (Hosmer-Lemeshow). Model 3 employs nonparametric bootstrap ($B = 1,000$) for the SEs and p -values also on Brant (proportional-odds) test. “–” = excluded variable. Baseline is Juventus for Model 1 if Intercept. For Models 2–3, R^2_{MF} = McFadden’s R^2 ; for Model 3, R^2_N = Nagelkerke’s R^2 . For Model 3, standard p -values for LI = Lipsitz.

Model		Model 1		Model 2		Model 3	
Description	Variable	AIC	BIC	AIC	BIC	AIC	BIC
	(Intercept)	–	0.84 (0.000)	–	–	–	–
Coach							
Initial scheme 2	S_2	–	0.25 (0.000)	0.49 (0.000)	0.49 (0.000)	0.47 (0.000)	0.47 (0.000)
Final scheme 2	S_2^*	–	–0.23 (0.000)	–0.50 (0.000)	–0.51 (0.000)	–0.49 (0.000)	–0.49 (0.000)
Home initial scheme 2	S_2^H	0.31 (0.000)	–	–	–	–	–
Away initial scheme 2	S_2^A	–0.19 (0.000)	–	–	–	–	–
Home final scheme 2	S_2^{*H}	–0.26 (0.000)	–	–	–	–	–
Away final scheme 2	S_2^{*A}	0.20 (0.000)	–	–	–	–	–
Team							
Crosses	w_1	–0.02 (0.000)	–0.02 (0.000)	–0.05 (0.000)	–0.04 (0.000)	–0.04 (0.000)	–0.04 (0.000)
Corners	w_2	–0.03 (0.014)	–0.03 (0.016)	–0.04 (0.122)	–0.03 (0.159)	–0.06 (0.004)	–0.06 (0.005)
Shots	w_6	0.03 (0.000)	0.03 (0.000)	0.06 (0.000)	0.06 (0.000)	0.05 (0.000)	0.05 (0.000)
Goal kicks	w_7	0.03 (0.000)	0.03 (0.000)	0.07 (0.001)	0.07 (0.000)	0.05 (0.002)	0.05 (0.001)
Offsides	w_8	0.03 (0.019)	0.03 (0.016)	0.05 (0.088)	0.07 (0.025)	0.05 (0.051)	0.05 (0.039)
Referee							
Yellow cards	z_1	–0.03 (0.137)	–0.04 (0.112)	–	–0.09 (0.068)	–0.06 (0.156)	–0.06 (0.141)
Red cards	z_2	–1.04 (0.000)	–1.05 (0.000)	–	–2.27 (0.000)	–2.11 (0.000)	–2.11 (0.000)
Free kicks	z_3	0.02 (0.023)	0.02 (0.017)	–	–	0.01 (0.362)	0.01 (0.327)
Penalty kicks	z_4	–0.30 (0.000)	–0.30 (0.000)	–	–	0.43 (0.009)	–0.44 (0.006)
Fouls	z_5	–	–	–	0.01 (0.621)	–	–
Home yellow cards	z_1^H	–	–	–0.18 (0.006)	–	–	–
Away yellow cards	z_1^A	–	–	–0.03 (0.610)	–	–	–
Home red cards	z_2^H	–	–	–2.09 (0.000)	–	–	–
Away red cards	z_2^A	–	–	2.31 (0.000)	–	–	–
Home fouls	z_5^H	–	–	–0.02 (0.294)	–	–	–
Away fouls	z_5^A	–	–	–0.02 (0.231)	–	–	–
Controls							
Weighted extra minutes	$\tilde{c}_2$	–0.03 (0.127)	–	–	–	–	–
Relative team effect	c_4	–	–	0.00 (0.969)	–	0.01 (0.460)	–
Quadratic relative team effect	c_4^2	–	–	0.00 (0.041)	–	0.00 (0.050)	–
Log average market value difference	c_5	0.49 (0.000)	0.50 (0.000)	0.66 (0.000)	1.00 (0.000)	0.62 (0.000)	0.75 (0.000)
Average home Win betting odds	c_6	–0.20 (0.000)	–0.19 (0.000)	–0.51 (0.002)	–0.13 (0.000)	–0.53 (0.000)	–0.46 (0.000)
Additional fixed effects							
Extreme goal difference	k_1^e	–	–	–	0.84 (0.250)	–	0.09 (0.977)
Extreme positive goal difference	k_2^e	3.71 (0.000)	3.82 (0.000)	13.81 (0.000)	–	12.24 (0.000)	–
Extreme negative goal difference	k_3^e	–4.65 (0.001)	–4.51 (0.001)	–13.53 (0.000)	–	–14.80 (0.000)	–
Season	$k^{(1,3)}$	–	–	0.00 (0.000)	–	–	–
Thresholds							
0 1	τ_0	–	–	–	–	–2.81 (0.000)	–2.71 (0.000)
1 3	τ_1	–	–	–	–	–0.91 (0.002)	–0.82 (0.002)
Goodness-of-fit							
AIC		3,651	3,654	1,055	1,069	1,744	1,750
BIC		3,747	3,740	1,156	1,135	1,840	1,830
DEV		1,591	1,601	1,015	1,043	1,706	1,718
$\bar{R}^2$		0.51	0.49	–	–	–	–
R^2_{MF}		–	–	0.35	0.34	0.29	0.29
R^2_N		–	–	0.52	0.50	0.53	0.52

Table 4: (continued)

Model Description	Variable	Model 1		Model 2		Model 3	
		AIC	BIC	AIC	BIC	AIC	BIC
Accuracy		–	–	0.79	0.78	0.67	0.68
Sensitivity		–	–	0.80	0.80	0.67, 0.37, 0.84	0.67, 0.39, 0.84
Specificity		–	–	0.77	0.77	0.89, 0.85, 0.74	0.90, 0.86, 0.74
Precision		–	–	0.80	0.80	0.72, 0.47, 0.74	0.71, 0.49, 0.74
F1		–	–	0.80	0.80	0.70, 0.42, 0.78	0.69, 0.43, 0.79
Residual diagnostics tests (p-values)							
SW		0.49	0.25	–	–	–	–
BP		0.00	0.49	–	–	–	–
RR		0.46	0.78	0.21	0.00	–	–
HL		–	–	0.42	0.46	0.67	0.63
LI		–	–	–	–	0.62	0.91
Brant		–	–	–	–	0.86	0.94

Table 5: Paired-bootstrap cross-model equality test (full statistic) for the top-4 AIC- and BIC-ranked *intercept* specifications, including c_5 and c_6 . Panels: AIC (Intercept), AIC and BIC (both with Intercept, and without k_2^e, k_3^e). $B = 1,000$.

Paired bootstrap test	W	df	p -value(χ^2)	p -value(boot)
AIC (with intercept)				
Rank-1	79.22	20	0.00	0.00
Rank-2	82.15	21	0.00	0.00
Rank-3	25.46	19	0.15	0.18
Rank-4	25.60	20	0.18	0.20
AIC (with intercept, without k_2^e & k_3^e)				
Rank-1	25.46	19	0.15	0.18
Rank-2	25.60	20	0.18	0.20
Rank-3	29.39	20	0.08	0.12
Rank-4	25.81	20	0.17	0.20
BIC (with intercept, without k_2^e & k_3^e)				
Rank-1	14.90	15	0.46	0.43
Rank-2	15.86	14	0.32	0.31
Rank-3	15.73	14	0.33	0.32
Rank-4	14.81	15	0.47	0.43
Sample: $N = 1,139$, WIN rate = 0.465.				

divergence).¹⁴ If θ_l is not selected in a given specification, we set $\hat{\theta}_l = 0$ and $\widehat{\text{Var}}(\hat{\theta}_l) = 0$. Inference relies on a Student's t distribution with degrees of freedom approximated by the

model-averaged effective degrees of freedom across specifications in the candidate set.

Table 6 reports the results. Model-averaged estimates closely match the signs and magnitudes obtained in individual specifications, indicating stability across modeling approaches. Coaching variables (s_2 and s_2^*), team actions, and referee decisions remain systematically associated with match outcomes. The quadratic relative team-effect proxy appears only in Set 2 for Models 2-3 (weak in Model 2; stronger in Model 3). The home-away split in Set 2 preserves the asymmetry observed in Set 1 ($s_2^H > 0$, $s_2^A < 0$, $s_2^{*H} < 0$, $s_2^{*A} > 0$), all highly significant. Marginal effects evaluated at the mean using BCa bootstrap estimates (see Appendix B) indicate that higher initial home-team offensiveness increases the predicted win probability by approximately 8–15 %, conditional on controls.

Across specifications, shots (w_6) and goal kicks (w_7) remain positive and highly significant; offsides (w_8) is positive but weaker in Model 2; crosses (w_1) lack robust support. Red cards and penalties (z_2, z_4) exhibit strong and stable associations, while yellow cards show weaker effects. Pre-match controls are consistently significant ($c_5 > 0$, $c_6 < 0$), confirming the relevance of relative team strength and market expectations.

Varying the retained share of top-ranked specifications affects the significance of marginal regressors (e.g., w_1, w_8, z_1), but core effects (s_2 terms, $w_6, w_7, z_2, z_4, c_5, c_6$) remain qualitatively stable. Residual diagnostics from the selected main models do not indicate major misspecification. Overall, the model-averaged results confirm the stability of the estimated associations between tactical variables, team characteristics, and match outcomes.

¹⁴ BIC weights may also be used (Schomaker and Heumann 2020) to construct confidence intervals (e.g., Table 7 in Appendix B).

Table 6: Model averaging results for Model 1, 2, and 3 using Akaike weights for the top AIC-ranked models (Sets 1 and Set 2). Model 1: top 2.5 % (579 specifications); Model 2: top 5 % (263 specifications); Model 3: top 5 % (101 specifications). “–” = excluded variable. Coefficients with p -values > 0.15 are omitted.

Sets		Set 1						Set 2					
Models		Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
Description	Variable	Est.	(p -val)	Est.	(p -val)	Est.	(p -val)	Est.	(p -val)	Est.	(p -val)	Est.	(p -val)
Initial scheme 2	S_2	0.25	(0.000)	0.48	(0.000)	0.43	(0.002)	–	–	–	–	–	–
Final scheme 2	S_2^*	–0.24	(0.000)	–0.48	(0.000)	–0.45	(0.001)	–	–	–	–	–	–
Home initial scheme 2	S_2^H	–	–	–	–	–	–	0.30	(0.000)	0.52	(0.000)	0.55	(0.000)
Away initial scheme 2	S_2^A	–	–	–	–	–	–	–0.19	(0.000)	–0.46	(0.000)	–0.40	(0.000)
Home final scheme 2	S_2^{*H}	–	–	–	–	–	–	–0.26	(0.000)	–0.49	(0.000)	–0.56	(0.000)
Away final scheme 2	S_2^{*A}	–	–	–	–	–	–	0.20	(0.000)	0.52	(0.000)	0.43	(0.000)
Crosses	w_1	–0.02	(0.039)	–	–	–	–	–0.01	(0.136)	–	–	–0.04	(0.031)
Corners	w_2	–0.02	(0.118)	–	–	–	–	–	–	–	–	–	–
Shots	w_6	0.03	(0.000)	0.06	(0.000)	0.05	(0.002)	0.03	(0.000)	0.06	(0.000)	0.06	(0.000)
Goal kicks	w_7	0.03	(0.000)	0.06	(0.001)	0.05	(0.011)	0.03	(0.000)	0.07	(0.001)	0.04	(0.018)
Offsides	w_8	0.03	(0.021)	0.06	(0.071)	0.05	(0.069)	0.03	(0.030)	0.05	(0.075)	0.06	(0.029)
Yellow cards	z_1	–0.04	(0.111)	–	–	–	–	–0.03	(0.143)	–	–	–	–
Home yellow cards	z_1^H	–	–	–0.15	(0.098)	–	–	–	–	–0.17	(0.034)	–	–
Red cards	z_2	–1.03	(0.000)	–	–	–1.91	(0.000)	–1.03	(0.000)	–	–	–2.41	(0.000)
Home red cards	z_2^H	–	–	–1.64	(0.048)	–	–	–	–	–1.89	(0.002)	–	–
Away red cards	z_2^A	–	–	1.73	(0.054)	–	–	–	–	2.15	(0.001)	–	–
Free kicks	z_3	0.02	(0.017)	–	–	–	–	0.02	(0.021)	–	–	–	–
Penalty kicks	z_4	–0.30	(0.000)	–	–	–0.39	(0.035)	–0.30	(0.000)	–	–	–0.50	(0.006)
Extreme positive goal difference	k_2^e	3.79	(0.000)	–	–	11.01	(0.000)	3.74	(0.000)	–	–	10.51	(0.000)
Extreme negative goal difference	k_3^e	–4.53	(0.001)	–	–	–13.46	(0.000)	–4.58	(0.001)	–	–	–16.98	(0.000)
Relative team effect (quadratic)	c_4^2	–	–	–	–	–	–	–	–	0.00	(0.135)	0.00	(0.000)
Log avg. market value diff.	c_5	0.48	(0.000)	0.69	(0.000)	0.66	(0.000)	0.47	(0.000)	0.68	(0.000)	0.71	(0.000)
Average home win betting odds	c_6	–0.19	(0.000)	–0.49	(0.006)	–0.51	(0.000)	–0.19	(0.000)	–0.48	(0.004)	–0.46	(0.002)
L (# models)		194		231		95		385		32		6	

5 Conclusions

The findings consistently indicate that coaching-related tactical variables are significantly associated with match outcomes. Across model-averaging procedures and robustness checks, aggressive offensive tactics – particularly when deployed early – are positively associated with performance differences, both at home and away. Not all attacking actions contribute equally: crosses display a persistent negative association (the “cross paradox”), whereas shots and goal-kicks exhibit stable positive associations across specifications. Referee variables, especially red cards and penalties, are strongly and consistently related to outcomes, while yellow cards and free-kicks show weaker and less stable effects, pointing to heterogeneity in the informational content of match events.

Extreme-score controls and team effects highlight the importance of persistent team-level heterogeneity. Home advantage remains robust: the BIC-optimal intercept in Model 1 is positive, and Model 3 confirms its persistence

after controlling for tactical, referee, and contextual variables. Market value and betting odds enter with expected signs (positively and negatively, respectively), but provide only modest improvements in fit and do not alter the estimated tactical or behavioral effects. Thus, while market-based measures summarize ex ante expectations, match outcomes remain systematically associated with tactical deployment and a set of actions observed during play.

An additional finding concerns the persistence of home advantage across specifications. Even after controlling for coaching schemes, in-game team actions, referee decisions, relative team strength, market value, and betting odds, home advantage remains statistically significant. This suggests that the variables considered in the present study do not fully account for the mechanisms underlying home advantage. While our analysis identifies several factors associated with match outcomes, the residual home advantage may reflect additional behavioral, psychological, or contextual influences that are not directly observable in live-ticker data.

Finally, the analysis remains associative. Coach selection is potentially endogenous, as managerial appointments reflect team expectations and strategic objectives. Future work could incorporate richer positional data, instrumental-variable strategies, or panel methods to strengthen causal identification. Extending the database to include tracking or context-dependent role measures would further refine the interpretation of tactical effects.

Acknowledgement: The authors acknowledge the participants of the 14th EAI INTETAIN 2023 at IMT Lucca, for their insightful comments.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: The authors state no conflict of interest.

Research funding: None declared.

Data availability: Research data are not shared.

A Scheme variables based on the offensiveness index

In the following we sketch the construction of the scheme variables. Building on the role of each player, we assume that a team's formation is defined by the number of defenders, midfielders, and forwards among its outfield players. In doing so, we do not account for role changes that players may undergo within the same match or across the same season. Additionally, we assume that teams employing identical formations adhere to the same tactical scheme. To convert this categorical variable into a numeric one, we employ an offensiveness index that evaluates the goal-scoring potential of 10 outfield players, excluding goalkeepers. Specifically, the index weights players based on their roles, linked to the average goals scored historically by players in those positions. The weighted sum of these points is assigned according to player positions: defenders receive a weight of 1, midfielders 2, and forwards 3.¹⁵ The resulting weighted sum ranges from 10 to 30 – representing the theoretical extremes

where all players are either defenders or forwards. One advantage of this transformation is that it enables an intuitive interpretation of the index. An increase of one point reflects an elevation in a team's offensive setup and vice versa. Furthermore, this metric accounts for a team's potential to create scoring opportunities, providing a quantifiable measure of offensive capability.

The first variable, $s_{i,1}$, computes the offensiveness index using all 10 outfield players and is invariant to red card events, and incorporates substitutions. The second variable, $s_{i,2}$, adjusts dynamically for dismissals by recalculating the index based on active players only. The third variable, $s_{i,3}$, normalizes $s_{i,2}$ by its attainable maximum, defined as the case in which all outfield players are assigned attacking roles, yielding a bounded index in $[0, 1]$. Formally, for the home team,

$$s_{i,t,3}^H \equiv \frac{s_{i,t,2}^H}{\text{active players at } t} \times \frac{10}{30}, \quad (5)$$

with an analogous definition for the away team. The home-away differential $s_{i,t,3} = s_{i,t,3}^H - s_{i,t,3}^A$ is computed at each minute t . To distinguish between ex ante and ex post tactical scheme, we define an initial scheme based on the starting lineup ($t = 0$) and a final scheme based on the reported formation at the end of the match ($t = 90$). The initial scheme reflects pre-match role allocation, while the final scheme reflects the formation observed after in-match events have unfolded.

B Bootstrap confidence intervals

We extend our investigation of coaching strategies beyond the use of HC estimators by incorporating automated bootstrap standard errors for the parameters in linear, logistic, and ordered logistic regressions, as detailed in Efron and Narasimhan (2020). This approach generalizes the Bias-Corrected and Accelerated (BCa) bootstrap (Efron 1987) by employing an innovative acceleration factor that simplifies computations.¹⁶ Table 7 offers a subset of our findings, emphasizing the 95 % asymptotic and BCa confidence intervals (CIs) for the initial scheme, s_2 , across the top AIC and BIC specifications in each of the three models. All bootstrap CIs are based on $B = 2,000$ resamples. The results align with those obtained via asymptotic theory and HC standard errors (in Model 1 and 2), as well as bootstrapped standard

¹⁵ The observed distribution of goals across player roles, with forwards accounting for approximately 55–60 % of total goals, midfielders for around 25–30 %, and defenders for roughly 10–15 % (<https://www.gazzetta.it/>).

¹⁶ Providing second-order adjustments to address both the non-normality of the estimator and its inherent bias.

Table 7: Asymptotic, BCa bootstrap, and model averaging CIs for different estimators at the 95 % level for the initial scheme s_2 , based on the best AIC and BIC-ranked specifications. All bootstrap CIs are computed using $B = 2,000$ resamples. L_1/L_2 = regularized/penalized estimation. Avg. = averaging. Model Avg. for Model 3 uses $B = 1,000$ replications.

Response	Estimator (s_2)	AIC		BIC	
		CI _{2.5}	CI _{97.5}	CI _{2.5}	CI _{97.5}
$y^{(1)}$	Classical	0.194	0.303	0.194	0.303
	BCa	0.192	0.302	0.192	0.302
	BCa L_1/L_2	0.178	0.287	0.186	0.276
	Model avg.	0.190	0.302	0.144	0.306
$y^{(2)}$	Classical	0.361	0.615	0.367	0.616
	BCa	0.333	0.591	0.349	0.604
	BCa L_1/L_2	0.361	0.615	0.368	0.615
	Model avg.	0.268	0.684	0.224	0.706
$y^{(3)}$	Classical	0.368	0.574	0.365	0.570
	BCa	0.373	0.561	0.362	0.560
	BCa L_1/L_2	0.377	0.570	0.364	0.562
	Model avg.	0.161	0.691	0.048	1.063

errors (in Model 3). In particular, for the best-AIC specification of Model 2, a one-unit increment in the disparity of initial schemes between home and away teams, s_2 , is linked to a marginal effect at the mean ranging from 8.26 % to 14.63 %, and an average marginal effect varying from 5.10 % to 8.56 %, on home win – at the 95 % confidence level, conditional on other in-game actions. Figure 2 provides an additional comparison by illustrating BCa confidence intervals derived from a classical asymptotic theory, BCa bootstrap,

and BCa bootstrap using penalized Gaussian regression for Model 1, a penalized logistic regression for Model 2 (with regularization via k -fold cross-validation), and an elastic net regularization (Wurm et al. 2021) for Model 3.

The shrinkage procedure described by Efron and Narasimhan (2020) is designed to reduce the width of the confidence interval, which remains tighter than its counterpart obtained via BCa under Gaussian and logistic link functions. In the updated results, reductions in CI width are heterogeneous. The BCa intervals are largely unchanged relative to classical ones (approximately 0–2%) in Models 1–2, whereas BCa L_1/L_2 delivers modest tightening in selected cases (up to –17 % under BIC in Model 1 and about –3% to –8% in Model 3). By contrast, accounting for model multiplicity through model averaging widens confidence intervals (+3 % to +49 % in Model 1; +63 % to +94 % in Model 2; and +158 % in Model 3 under AIC), consistent with the incorporation of between-specification uncertainty. For initial schemes s_2 , all parameter estimates indicate a positive effect; similarly, for the final scheme estimates suggest a negative effect. There is no conflicting evidence among classical, BCa, regularized BCa, or model averaging methods. These results collectively suggest significant dynamic influences of schemes on home outcomes, measured by goal difference and winning probability.

C Literature review

Table 8 reports a selective review of the literature on coaching, strategy and HA effects in soccer teams' performances.

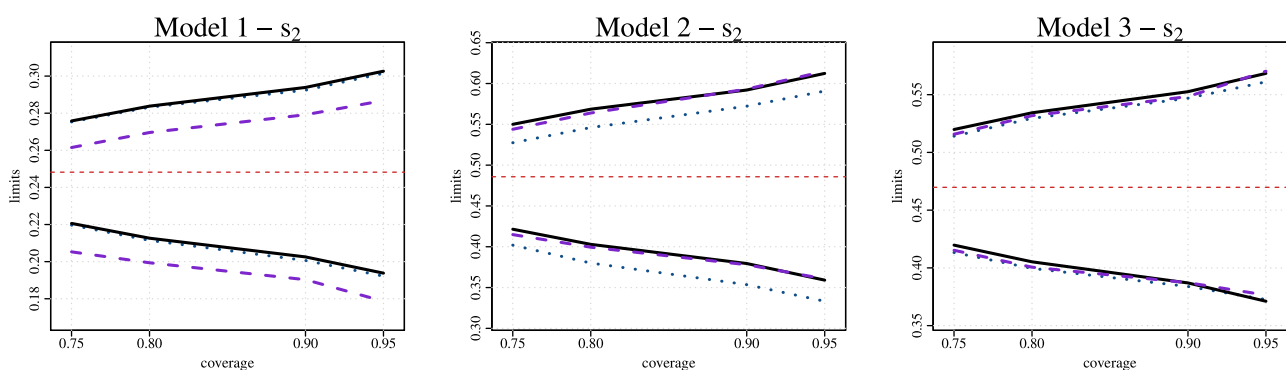


Figure 2: Comparison of CIs for best-AIC ranked Model 1, 2, and 3 for the initial scheme parameter. Red dotted line: classical post estimate.

Table 8: Illustrative contributions from the literature on coaching, strategy and HA effects in soccer teams' performances. The selected studies reported in this table do not represent an outcome of a systematic literature review.

Reference	Main focus	Method	League	Matches	Season(s)
Muehlheusser et al. (2018)	Strategies/tactics	OLS with fixed effects	Bundesliga	6,426	1993/94–2013/14
Mesoudi (2020)	Strategies/tactics	Bayesian multi-level regression	European leagues	9,127	2012/13–2016/17
Goes et al. (2021a)	Strategies/tactics	4 machine learning classifiers	Eredivisie	118	n/a
Lopez-Valenciano et al. (2022)	Strategies/tactics	Pearson's correlation and PCA	LaLiga	760	2017/18–2018/19
Lorenzo-Martínez et al. (2022)	Strategies/tactics	Linear mixed model	Bundesliga	234	2016/17
Bauer et al. (2023)	Strategies/tactics	Convolutional neural network	Bundesliga	2,142	2013/14–2019/20
Scoppa (2008)	HA/HB effect	OLS regression	Serie A	686	2003/04–2004/05
Pettersson-Lidbom and Priks (2010)	HA/HB effect	Natural experiment	Serie A-B	842	2006/07
Dewenter and Namini (2013)	HA/HB effect	Conditional FE negative Binomial regressions	Bundesliga	8,029	1982/83–2007/08
Hlasny and Kolaric (2017)	HA/HB effect	OLS and fixed-effect regressions	English football leagues	34,692	2000/01–2013/14
Picazo-Tadeo et al. (2017)	HA/HB effect	Random effects panel-data regressions	LaLiga	2,651	2002/03–2009/10
Ponzo and Scoppa (2018)	HA/HB effect	Ordered Probit	Serie A	7,398	1991/92–2012/13
Fischer and Haucap (2021)	HA/HB effect	OLS regression	Bundesliga 1,2,3	2,976	2017/18–2019/20
Holder et al. (2021)	HA/HB effect	OLS regression	Bundesliga, Serie A	2,744	2015/16–2018/19
Fischer and Haucap (2022)	HA/HB effect	Difference-in-differences	Bundesliga 1-2	6,120	2011/12–2020/21
Wunderlich et al. (2021)	HA/HB effect	Linear mixed regressions models	10 European professional leagues	37,888	2010/11–2019/20
Price et al. (2022)	HA/HB effect	Hierarchical causal model	Premier league	380	2020/21
Reade et al. (2022)	HA/HB effect	Poisson regression	UEFA and European top leagues	51,333	2011/12–2020/21
Benz and Lopez (2023)	HA/HB effect	Poisson regression	17 European leagues	25,327	2015/16–2020/21

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* AC-19: 716–723.
- Bauer, P., Anzer, G., and Shaw, L. (2023). Putting team formations in association football into context. *J. Sports Anal.* 9: 39–59.
- Benz, L. and Lopez, M. (2023). Estimating the change in soccer's home advantage during the Covid-19 pandemic using bivariate Poisson regression. *AStA Adv. Stat. Anal.* 107: 205–232.
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*: 1171–1178, <https://doi.org/10.2307/2532457>.
- Bruinshoofd, A. and ter Weel, B. (2003). Manager to go? Performance dips reconsidered with evidence from Dutch football. *Eur. J. Oper. Res.* 148: 233–246.
- Carron, A., Loughhead, T., and Bray, S. (2005). The home advantage in sport competitions: Courneya and Carron's (1992) conceptual framework a decade later. *J. Sports Sci.* 23: 395–407.
- Cefis, M. (2023). Observed heterogeneity in players' football performance analysis using pls-pm. *J. Appl. Stat.* 50: 3088–3107.
- Cefis, M. and Carpita, M. (2024). The higher-order PLS-SEM confirmatory approach for composite indicators of football performance quality. *Comput. Stat.* 39: 93–116.
- Courneya, K. and Carron, A. (1992). The home advantage in sport competitions: a literature review. *J. Sport Exerc. Psychol.* 14: 13–27.
- Cross, J. and Uhrig, R. (2023). Do fans impact sports outcomes? A covid-19 natural experiment. *J. Sports Econ.* 24: 3–27.
- De Paola, M. and Scoppa, V. (2012). The effects of managerial turnover: evidence from coach dismissals in Italian soccer teams. *J. Sports Econ.* 13: 152–168.
- Dewenter, R. and Namini, J.E. (2013). How to make soccer more attractive? Rewards for a victory, the teams' offensiveness, and the home bias. *J. Sports Econ.* 14: 65–86.
- Efron, B. (1987). Better bootstrap confidence intervals. *J. Am. Stat. Assoc.* 82: 171–185.
- Efron, B. and Narasimhan, B. (2020). The automatic construction of bootstrap confidence intervals. *J. Comp. Graph. Stat.* 29: 608–619.
- Fischer, K. and Haucap, J. (2021). Does crowd support drive the home advantage in professional football? Evidence from German ghost games during the Covid-19 pandemic. *J. Sports Econ.* 22: 982–1008.
- Fischer, K. and Haucap, J. (2022). Home advantage in professional soccer and betting market efficiency: the role of spectator crowds. *Kyklos* 75: 294–316.
- Forcher, L., Forcher, L., Altmann, S., Jekauc, D., and Kempe, M. (2024). Is a compact organization important for defensive success in elite soccer? Analysis based on player tracking data. *Int. J. Sports Sci. Coach.* 19: 757–768.
- Forcher, L., Preine, L., Forcher, L., Wäsche, H., Jekauc, D., Woll, A., Gross, T., and Altmann, S. (2023). Shedding some light on in-game formation changes in the German Bundesliga: frequency,

- contextual factors, and differences between offensive and defensive formations. *Int. J. Sports Sci. Coach.* 18: 2051–2060.
- Fry, J., Serbera, J.-P., and Wilson, R. (2021). Managing performance expectations in association football. *J. Bus. Res.* 135: 445–453.
- Goes, F., Kempe, M., Van Norel, J., and Lemmink, K. (2021a). Modelling team performance in soccer using tactical features derived from position tracking data. *IMA J. Manag. Math.* 32: 519–533.
- Goes, F.R., Brink, M.S., Elferink-Gemser, M.T., Kempe, M., and Lemmink, K.A. (2021b). The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *J. Sports Sci.* 39: 523–532.
- Heuer, A. and Rubner, O. (2009). Fitness, chance, and myths: an objective view on soccer results. *Eur. Phys. J. B* 67: 445–458.
- Hlasny, V. and Kolaric, S. (2017). Catch me if you can: referee–team relationships and disciplinary cautions in football. *J. Sports Econ.* 18: 560–591.
- Holder, U., Ehrmann, T., and König, A. (2021). Monitoring experts: insights from the introduction of video assistant referee (VAR) in elite football. *J. Bus. Econ.* 92: 285–308.
- Hosmer, D., Lemeshow, S., and Sturdivant, R. (2013). *Applied logistic regression*, Vol. 398. John Wiley & Sons.
- Jamieson, J. (2010). The home field advantage in athletics: a meta-analysis. *J. Appl. Soc. Psychol.* 40: 1819–1848.
- Kuper, S. and Szymanski, S. (2009). *Soccernomics*. Nation Books, Philadelphia; PA.
- Lipsitz, S.R., Fitzmaurice, G.M., and Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models. *J. Roy. Stat. Soc. C: Appl. Stat.* 45: 175–190.
- Lopez-Valenciano, A., Garcia-Gómez, J.A., López-Del Campo, R., Resta, R., Moreno-Perez, V., Blanco-Pita, H., Valés-Vázquez, Á., and Del Coso, J. (2022). Association between offensive and defensive playing style variables and ranking position in a national football league. *J. Sports Sci.* 40: 50–58.
- Lorenzo-Martínez, M., Rein, R., Garnica-Caparrós, M., Memmert, D., and Rey, E. (2022). The effect of substitutions on team tactical behavior in professional soccer. *Res. Q. Exerc. Sport* 93: 301–309.
- Lukacs, P.M., Burnham, K.P., and Anderson, D.R. (2010). Model selection bias and freedman's paradox. *Ann. Inst. Stat. Math.* 62: 117–125.
- Magee, C. (2025). Crowds and managerial decisions in European football leagues. *J. Sports Econ.* 26: 744–765.
- Mesoudi, A. (2020). Cultural evolution of football tactics: strategic social learning in managers' choice of formation. *Evol. Hum. Sci.* 2: e25.
- Meyer, J. and Klatt, S. (2024). Additional substitutions in elite European football. *Int. J. Sports Sci. Coach.* 19: 769–778.
- Muehlheusser, G., Schneemann, S., Sliwka, D., and Wallmeier, N. (2018). The contribution of managers to organizational success: evidence from German soccer. *J. Sports Econ.* 19: 786–819.
- Pettersson-Lidbom, P. and Priks, M. (2010). Behavior under social pressure: empty Italian stadiums and referee bias. *Econ. Lett.* 108: 212–214.
- Picazo-Tadeo, A., González-Gómez, F., and Guardiola, J. (2017). Does the crowd matter in refereeing decisions? Evidence from Spanish soccer. *Int. J. Sport Exerc. Psychol.* 15: 447–459.
- Pollard, R. (2008). Home advantage in football: a current review of an unsolved puzzle. *Open Sports Sci. J.* 1: 12–14.
- Ponzo, M. and Scoppa, V. (2018). Does the home advantage depend on crowd support? Evidence from same-stadium derbies. *J. Sports Econ.* 19: 562–582.
- Price, K., Cai, H., Shen, W., and Hu, G. (2022). How much does home field advantage matter in soccer games? A causal inference approach for English Premier League analysis, arXiv preprint arXiv:2205.07193.
- Ramsey, J. (1969). Tests for specification errors in classical linear least-squares regression analysis. *J. Roy. Stat. Soc. B: Stat. Methodol.* 31: 350–371.
- Reade, J., Schreyer, D., and Singleton, C. (2022). Eliminating supportive crowds reduces referee bias. *Econ. Inq.* 60: 1416–1436.
- Sarkar, S. (2018). Paradox of crosses in association football (soccer)—a game-theoretic explanation. *J. Quant. Anal. Sports* 14: 25–36.
- Sarmiento, H., Clemente, F.M., Araújo, D., Davids, K., McRobert, A., and Figueiredo, A. (2018). What performance analysts need to know about research trends in association football (2012–2016): a systematic review. *Sports Med.* 48: 799–836.
- Schlegel, B. and Steenbergen, M. (2020). *Brant: test for parallel regression assumption*. R package version 0.3-0.
- Schomaker, M. and Heumann, C. (2020). When and when not to use optimal model averaging. *Stat. Pap.* 61: 2221–2240.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6: 461–464.
- Scoppa, V. (2008). Are subjective evaluations biased by social factors or connections? An econometric analysis of soccer referee decisions. *Empir. Econ.* 35: 123–140.
- Vecer, J., Kopriva, F., and Ichiba, T. (2009). Estimating the effect of the red card in soccer: when to commit an offense in exchange for preventing a goal opportunity. *J. Quant. Anal. Sports* 5, <https://doi.org/10.2202/1559-0410.1146>.
- Wittkugel, J., Memmert, D., and Wunderlich, F. (2022). Substitutions in football—what coaches think and what coaches do. *J. Sports Sci.* 40: 1668–1677.
- Wunderlich, F., Weigelt, M., Rein, R., and Memmert, D. (2021). How does spectator presence affect football? Home advantage remains in European top-class football matches played without spectators during the Covid-19 pandemic. *PLoS One* 16: e0248590.
- Wurm, M.J., Rathouz, P.J., and Hanlon, B.M. (2021). Regularized ordinal regression and the ordinalnet R package. *J. Stat. Software* 99: 1–42.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/jqas-2025-0113>).