Department of Economics- FEA/USP

Second Leg Home Advantage: A study on the Brazilian soccer cup

Pedro Acioly Teixeira

Márcio Issao Nakane

WORKING PAPER SERIES Nº 2025-04

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2025-04

Second Leg Home Advantage: A study on the Brazilian soccer cup

Pedro Acioly Teixeira (acioly_pedro@usp.br)

Márcio Issao Nakane (minakane@usp.br)

Abstract:

This study examines the existence of Second Leg Home Advantage (SLHA) in soccer knockout matches, focusing on the Brazilian Cup. Using an instrumentalvariables approach, we construct a team quality control variable based on betting odds. Our findings indicate a significant SLHA effect, with home teams in the second leg winning between 53.9% and 55.8% of matches, after controlling for team strength. We further explore the conditional impact of first-leg results, showing that a draw in the first leg increases a second-leg home team's probability of advancing to over 64%. Contrary to common belief, the away goals rule does not significantly affect outcomes. These results contribute to the ongoing debate on SLHA and suggest that playing the second leg at home provides a meaningful competitive advantage. The findings have implications for tournament design and team strategy in knockout competitions.

Keywords: home advantage, second leg home advantage, soccer, Brazilian soccer cup

JEL Codes: Z2, Z21

Second Leg Home Advantage: A study on the Brazilian soccer cup*

Pedro Acioly Teixeira† Márcio Issao Nakane‡

February 18, 2025

[1]

Abstract

This study examines the existence of Second Leg Home Advantage (SLHA) in soccer knockout matches, focusing on the Brazilian Cup. Using an instrumental variables approach, we construct a team quality control variable based on betting odds. Our findings indicate a significant SLHA effect, with home teams in the second leg winning between 53.9% and 55.8% of matches, after controlling for team strength. We further explore the conditional impact of first-leg results, showing that a draw in the first leg increases a second-leg home team's probability of advancing to over 64%. Contrary to common belief, the away goals rule does not significantly affect outcomes. These results contribute to the ongoing debate on SLHA and suggest that playing the second leg at home provides a meaningful competitive advantage. The findings have implications for tournament design and team strategy in knockout competitions.

Keywords: home advantage, second leg home advantage, soccer, Brazilian soccer cup

JEL Codes: Z2, Z21

*Pedro Teixeira acknowledges financial support by CAPES and by FIPE. The authors thank Mauro Rodrigues Jr., Bruno de Paula Rocha, Márcio Alves Diniz, and Ricardo Avelino for useful comments and suggestions. Remaining errors are our own responsibility.

†Brazilian Agency for Industrial Development, Brazil

‡Department of Economics, University of Sao Paulo, Brazil

1 Introduction

Soccer stands unequivocally as the preeminent global sport, captivating enthusiasts worldwide. Beyond its universal appeal, the game is widely studied in fields such as health sciences, sociology, history, and economics. Evidently, growing interest in soccer within the social sciences stems from its substantial social impact and expanding financial significance.

In light of this, there are several factors that render soccer a subject of considerable relevance for economists. For instance, soccer strategies closely mirror economic decision making, as both involve rational agents seeking to maximize their payoffs. In economic theory, firms and consumers select strategies to maximize profit and utility, respectively, taking into account their intrinsic constraints. In a soccer match, embedded within the context of competition, teams adopt strategies with the aim of achieving the best possible outcome, considering their level of strength and other relevant constraints.

In addition to this, the soccer context allows for the study of various dimensions involving the interaction between different agents in competitive environments, under pressure and in the presence of numerous factors that can influence the expected payoff. For example, two opposing teams can achieve completely different results when playing against each other, depending on which team has the home-field advantage. In other words, a change in a single variable (the venue of the match) is a determining factor in the game. Similar dynamics occur in economic relationships among firms, government, and individuals: changes in interest rates, intrinsic shifts in individual preferences, among other factors, completely alter interactions between agents in economic environments.

In this sense, understanding the phenomena that unfold in the soccer environment contributes to elucidating discussions in economic science. The focus of this article is to explore the phenomenon popularly known as Second-Leg Home Advantage

(SLHA), a hypothesis in which the team hosting the second leg of a knockout match enjoys an advantage over the opponent playing the first leg at home.

The possible existence of SLHA is a direct result of the teams' behavior at each moment of the match, revealing a dilemma in the strategic choices of the teams. Consequently, this article aims to examine and detail the aforementioned theme, using the editions of the Brazilian soccer cup as a basis. Our empirical results indicate a positive effect of the home-field advantage in the second leg of a knockout match, with a range varying between 53.9% and 55.8%, taking into account the difference in the strength of the teams.

In addition to this introductory section, this article is divided as it follows: section 2 contains the literature review . Section 3 describes the methodology and the dataset construction, as well as the development of the control variables, including the use of instrumental variables. This methodological contribution has advantages over the adoption of control variables commonly used in the literature, which are developed with a certain degree of arbitrariness, as it will be discussed later. Section 4 introduces the first descriptive statistics. Section 5 presents the econometric estimations, while section 6 approaches the models' extensions. Finally, section 7 presents the main conclusions.

2 Literature Review

The existence of Home Field Advantage (HA) is recognized in the literature on numerous sports, including football (Dobson & Goddard, 2011). Concisely, this phenomenon can be defined as the relative advantage enjoyed by a team when playing a match in its own ground compared to playing the same match in the opponent's field.

Often, according to Dobson and Goddard (2011), the determinants of HA can

be divided into four different categories: familiarity, distance, supporter effects and rules which may favor the home team. Familiarity refers to the natural knowledge that the home team has about its own stadium - or lack of knowledge on the part of the visitors. Distance, in turn, can interfere with the away team's preparation and fatigue levels for the match, given the need to travel. As for crowd effects, these can occur through three different channels, via encouragement of the home team, intimidation of the visitors and influence on the referee (referee bias). Finally, rules such as the qualified away goal in knockout matches can also interfere with HA.

Moreover, other possible effects have also been explored as an explanation for the existence of Home Advantage. Specific tactics chosen by the teams, leading the away teams to play more defensively in general, could imply a psychological advantage for the home team (Pollard, 1986). One piece of evidence pointed out by Pollard and Pollard (2005) is that Home Advantage is considerably higher in knockout duels than in domestic league matches, considering the scenario of European competitions. This could be attributed to the fact that an away defeat by a small goal difference is generally seen as reversible in the second-leg, which would lead away teams to adopt even more defensive tactics during the first-leg.

Psychological factors are also discussed as possibles determinants of HA. Dobson and Goddard (2011) argue that since the beginning of English football - back in the 19th century - there has been a notable Home Advantage. Teams' beliefs about familiarity and other factors could reinforce this advantage, regardless of whether these beliefs are actually accurate or rational. Finally, Neave and Wolfson (2003) argue for territoriality as an additional reason for HA. According to the authors, territoriality is defined as the sense of protecting one's territory from invaders. Interestingly, the players' testosterone levels, measured before matches, were significantly higher before home matches compared to away matches.

While the causes of Home Advantage are well-documented, the existence of

Second-Leg Home Advantage (SLHA) remains debated. SLHA can be defined as the advantage of playing the second-leg of a knockout match as the home team compared to playing the same leg as the away team. In other words, SLHA implies a greater HA effect in the second-leg of a tournament, all else being equal.

Page and Page (2007) observed the existence of SLHA when they analysed the European Cups over 51 years, concluding that the Second-Leg Home Team (SLHT) has on average more than 50% chance of advancing to the next round. Lidor et al. (2010) also examined European competitions and obtained similar conclusions. However, Eugster et al. (2011) and Amez et al. (2020) point out that there is no SLHA and that the difference can be explained by the clubs' performance in the Champions League group stage. This conclusion is corroborated by Abad et al. (2017) using data from the South-American Libertadores Cup (Copa Libertadores) matches.

Hence, there is no consensus on the subject. It is therefore necessary to investigate the methodological differences that led to the differences in the results found.

Page and Page (2007) began investigating this phenomenon using data from the three most important European competitions between 1955 and 2006. The tournaments examined were the Champions League (Champions Cup until 1998), the UEFA Cup (Inter-Cities Fairs Cup), and the Cup Winners' Cup. It is worth noting that, in general, the match draws for the knockout stages were not randomized over the period studied, with better teams hosting the second-leg in their ground. Therefore, the authors decided to control for the difference in ability between the teams using UEFA's team ability index - which UEFA itself uses for grouping teams.

The results estimated by the authors indicate the presence of SLHA in all three competitions analyzed. Furthermore, they found that although this advantage exists, it has been decreasing over time.

Additionally, their results indicate that part of this advantage actually comes

from extra time, where the SLHT have an extra 30 minutes of HA. This advantage is also present in penalty shoot-outs. However, it must be pointed out that these two factors are not sufficient to explain the whole phenomenon. Even removing those matches decided in extra time and penalty shoot-outs, the SLHT still has around a 54% chance of progressing, all else being equal.

Lidor et al. (2010) explore the same question, relating SLHA to the number of goals scored in each match. That is, whether this effect is linked to the number of goals scored by the SLHT. The variance analysis of the number of goals scored in each of the two games revealed that the average number of goals scored by the home team was higher in the second game. Therefore, the authors reach a similar conclusion to Page and Page (2007). However, no control variables were directly used for controlling for the differences in ability between the teams.

In general, other authors explore the data in a similar way to the model proposed by Page and Page (2007), with a few modifications, but which lead to different results. For example, Eugster et al. (2011) add other controls for the difference in ability between teams. Firstly, the authors consider the performance of the teams throughout the tournaments group stage, along with the UEFA coefficients. By adding a control for the clubs' performance in the group stage of the competition, a more up-to-date measure of the teams' performance is obtained. Using these controls, the authors conclude that there is no advantage in playing the second-leg of a knockout match at home, and propose that the best-placed team in the group stage should then be free to choose the order of where to play the first and second legs.

Waquil et al. (2020) adopt a similar analysis for the Brazilian Cup, but construct a proxy for the ability of the teams based on their performance in the Brazilian Championship and in the cup itself. The estimated econometric model, as the authors themselves argue, is an extension of Page and Page (2007). The conclusion of the article points to the existence of SLHA in the tournament, however, the authors select

a different sample of matches from the present study, in which they also considered matches where the elimination rule due to two or more away goals in the first match was applied. This rule will be detailed later.

In contrast to most studies, Geenens and Cuddihy (2018) opted for a non-parametric approach to test for the existence of SLHA. Also using information from the Champions League and Europa League, the authors indicate that there is indeed an advantage at playing the second-leg at home. Notwithstanding, the most important contribution of this article is its methodology.

The authors argue that, although widespread, the use of logistic regression models may not be appropriate on certain occasions. In addition, studies such as Page and Page (2007) and Eugster et al. (2011) assume that the specified models fit the data without first observing how the data is distributed. Therefore, Geenens and Cuddihy (2018) decided to adopt a non-parametric approach, using the NW estimator (Nadaraya-Watson) and different confidence intervals (Wilson and Agresti-Coull).

In any case, it is noteworthy that a large part of the debate about the SLHA presence falls within a controversial context: the choice of the measure of ability between the teams. To some degree, the control variables chosen will always be arbitrary, since the difference in level between the teams is not directly observable and there is no consensus on how to define it.

Further, the possible causes of SLHA are still unclear. What factors would lead a team to enjoy a greater Home Advantage effect in the second-leg compared to the first leg? Some possible explanations are superficially presented in the literature, among which are the different strategies used by each team in each leg of a knockout match, the away goal rule and the possibility of having an extra time in the second-leg.

Regarding the first explanation above, it is argued that in the first-leg the teams generally opt for the strategy of playing in the "best possible way", in order to

maximize their pay-offs (Lidor et al., 2010). On the other hand, in the second-leg of the match, the teams adopt the strategy of seeking the "minimum result necessary" to advance to the next stage, and therefore select a defined goal target.

In turn, the qualified away goal rule can also imply a greater advantage for playing the last match at home. If the SLHT manages to score goals during the first-leg, in the second-leg this SLHT will be benefitted by the advantage of losing (winning) by the same goal difference as they won (lost) previously in order to advance to the next stage, as long as they don't concede a greater number of goals at home.

In addition to these factors, the existence of extra time in the second-leg can also affect the final match. If, at the end of the 180 minutes, the match ends in a draw (even in terms of the tie-break criteria), the SLHT will still have an extra 30 minutes playing on its own field to decide the match, i.e. benefiting from more time under the effect of HA. Even if the match is ultimately decided on penalty kicks, it is possible that the SLHT will also enjoy an additional advantage.

As a result, this article sought to investigate the Brazilian Cup in order to contribute to the debate on SLHA. The idea is that the Brazilian Cup is possibly the closest to the ideal tournament for assessing the presence, effects and causes of SLHA.

Firstly, the tournament's format since its inception in 1989 has been partially homogeneous. Although there have been a number of changes, both in the definition of who qualifies for the tournament and in the way the first stages of the competition are played, the two-legged tie system has always been adopted, at least from the round of 16 onwards. Furthermore, extra time has never been used as a tie-breaker, which excludes the artificial advantage of this criterion. Finally, between 1989 and 2017, the qualified away goal criterion was adopted, but has ceased since the 2018 edition. In addition, the selection of which team will play at home in the second-leg has been random for several editions.

Hence, the Brazilian Cup can be considered a competition that allows the ob-

servation of the SLHA phenomenon in an empirical scenario close to a controlled experiment. It becomes relatively straightforward and simple to test for the existence of the SLHA effect. Consequently, it is possible to test which measures would be accurate for team's ability. Finally, such a scenario also permits to examine the causes, or at least the singularities, that implies the presence or absence of SLHA.

Therefore, using the data from the tournament editions, the following points will be addressed: i) The existence of SLHA; ii) The accuracy of the selected control variables for ability. ii) The accuracy of the selected control variables for ability.

3 Empirical Strategy

3.1 Data

The data was collected from the following sources: Bola na Área (*bolanaarea.com*), *oddsportal.com*,*betexplorer.com*, CBF (*cbf.com.br*) and RSSSF Brasil (*rsssf-brasil.com*) websites and, finally, from Sofascore (*sofascore.com.*).

The observations corresponding to the results of the matches from the 1989 to 2008 editions of the Copa do Brasil were extracted from the first website mentioned above, while the data corresponding to the years 2009 to 2020 were obtained from *oddsportal*, a website that aggregates the odds of the matches. The information for the 2021 and 2022 editions was obtained from *betexplorer*, which also provides aggregate odds from bookmakers. These odds represent the average of the values set by the bookmakers, and were used in the presented regressions as measures of relative ability between the teams, which will be further detailed in the empirical strategy section.

From 2012 onwards, information from the CBF website was also used to compose the database. The RSSSF Brasil website was used in particular to consult the regulations of each edition and any missing observation.

Finally, the teams standings in the Brazilian championship were extracted from Sofascore, since 2008 for Serie A (first division) and Serie B (second division), and since 2013 for Serie C (third division). The 2009 to 2012 editions of the third division were excluded during the dataset's construction, as it is understood that the rules at the time - dividing into 4 groups of 5 teams - would not permit constructing an accurate measure for control variables. In addition, from Bola na Área's website, the teams' standings in the national championship were collected, regarding the years 1988 to 2007, for the top two divisions of the tournament. Bola na Área uses its own criteria to calculate the final standings of each club after the end of the editions prior to the implementation of the round-system in 2003. This criteria takes into account both the performance in the first stages and the knockout phases matches from the Brazilian championship, prior to the adoption of this current system. The following section will detail the use of this information for the empirical part of the analysis.

3.2 Brazilian Cup Rules

Historically, different rules have been used in the competition, with different criteria both for qualifying for the tournament and for the tournament's itself. The greatest advantage of using the Brazilian Cup as an object of study, however, is that the random draw mechanism was adopted a long time ago for those matches in which two games were mandatory. Thus, the best and worst teams had their home and away fixtures decided randomly, which makes the following analysis more straightforward.

It should also be noted that the Copa do Brasil has never adopted extra time as a tie-breaker. As discussed in the literature review, the adoption of extra time creates an (artificial) bias for the SLHA study: the SLHT will be under the effect of HA for an extra 30 minutes after a draw in regulation time. In addition, the FLHT may enjoy an extra advantage if they score during extra time in the second-leg. Since the exact magnitude of these effects is unknown, it becomes imprecise how to calculate

the real effect of SLHA.

Another relevant historical event was the implementation of the qualified away goal rule, applied in all editions between 1989 and 2017. In 2018, this system was completely dismissed, though it's important to note that since 2015, the final matches no longer followed this criterion. In addition, two other matches were also played without the qualified away goal rule: the 2014 match between Atlético Mineiro and Cruzeiro, and the 2006 match between Flamengo and Vasco. Both first and second legs were played in the same stadium (Mineirão and Maracanã, respectively) and it was understood that the field should be considered neutral. Therefore, CBF decided beforehand not to apply this criterion to these duels.

Additionally, regarding historical changes in the tournament, from 1989 to 1994, only the state champions qualified for the Brazilian Cup. Although some states lacked clubs representing them at that time, more states gradually joined the competition. In addition to the state champions, other teams also qualified based on different criteria, such as the CBF ranking. All stages were contested in two mandatory matches: a first leg and a return leg.

Moreover, the number of participants in the competition has significantly increased over the years due to changes in the classification criteria and invitations to additional teams. While the details of these regulations are interesting, they fall outside the scope of this text. Consequently, more stages were introduced to the tournament to accommodate the larger number of teams.

From 1995 until the 2016 edition, an additional tie-breaker criterion was in place during the first two stages of the tournament. If the away team won by two or more goals in the first leg (three or more in the 1995 edition), it would automatically qualify for the next round without the need for a second leg. Moreover, during this period, the draw was not random: the teams with the highest CBF rankings played the first match away from home. In fact, this was the only criterion used to determine

the home and away fixtures in the Copa do Brasil during this time.

From 2017 onwards, the first two rounds were contested as single matches. In the first round, the away team (the one with the highest CBF ranking) advanced even in the event of a draw, while in the second round, any ties were decided by penalty shoot-outs.

3.3 Dataset construction

Let "match" be defined as the set of games between team i and team j in the same stage of the Brazilian Cup during a given year. In other words, a match can consist of two games (a first leg and a second leg) or a single leg, as previously explained.

To create the final data set, a series of filters were applied to the matches. First, any matches that faced issues in their dispute or were affected by off-field decisions were excluded. It is important to note that if the problem occurred in the second-leg, both games were disregarded. In total, 18 observations were removed from the sample.

As described in the previous section, to build the database, all observations from the first two rounds between 1995 and 2022 were removed from the sample, as these matches were not necessarily played as two legs.

With the exclusions mentioned above, the final dataset comprises 1,492 games, or 746 matches played over two legs, from 1989 to 2022. The available information includes: i) match score; ii) home team; iii) away team; iv) tournament stage; v) indication of whether the match was the first or second leg of the knockout stage; vi) penalty shoot-out winner; vii) from 2009 onwards, the average odds from bookmakers for each team's win, draw, and loss; and viii) the teams' standings in the national championship (available from 1989 for Serie A and Serie B, and from 2013 for Serie C).

3.4 Construction of control variables

3.4.1 Team standings in the Brazilian Championship

As previously discussed, the standings in the Brazilian championship were used as a control for the ability of each team. For the years between 2006 and 2022, the variable was constructed as follows: first, the Serie A ranking followed the standard order from first to twentieth place, while for Serie B, 20 positions were added to each rank. In other words, the leader of the second division would be in 21st place, and the last team would be in 40th place, for instance.

Regarding Serie C, in the editions without observations, the teams were considered as being in 41st place in the standings. Starting in 2013, 40 positions were added to each rank. Since the tournament is organized into 2 groups of 10 teams, for each position between 41 and 50, there are two teams. Teams competing in Serie D or those not participating in any division were considered as last place.

For the tournaments between 1989 and 2005, the construction of the variable was approached differently due to the peculiarities of the national tournament. Since the regulations of the Brazilian championship frequently varied in terms of format, number of participants, and the number of rounds played by each team, certain criteria were established to rank the teams.

First, all teams not participating in either of the two higher divisions were considered as last place. In other words, teams without a division or those competing in Serie C were classified as having the same ability. This choice was made due to the fact that in several years the third division was not contested, and moreover, in many editions several teams played a very limited number of matches, which makes a ranking based on overall standings less informative. For example, in the 1995 edition of the third division, teams ranked from 41st to 107th played six or fewer matches.

For the teams in Serie A, as well as in the editions from 2008 onwards, the

positions were considered based on the natural ranking in the tournament. For Serie B, the number of available positions in Serie A was added to each team's placement in the second division. It is important to note that the number of participants in the top division varied significantly over time, reaching a maximum of 32 participants in the 1993 edition. This observation also applies to Serie B, where, for example, the 1989 edition had 96 participants.

In this way, with the intention of making the proposed ranking more uniform, it was determined that the worst possible position for a team would be 51st. This decision stems from the same reasoning previously discussed regarding the exclusion of Serie C standings in the construction of the variable for the aforementioned years. The limited number of matches in many editions with several teams (especially for the lowest-ranked teams) does not provide much explanatory value for the variable. For example, in the 1989 edition of the second division, only 3 points separate the team in 54th place (Sobradinho Esporte Clube) from the team in 90th place (ACEC Baraúnas). Therefore, including the exact position of these clubs would likely introduce more bias into the linear relationship of this variable than treating both teams as having the same level of ability.

Next, for each match, the difference in standings between the SLHT and the FLHT was calculated, resulting in the following variable:

$$\Delta S = S_{SLHT} - S_{FLHT} \tag{1}$$

Where ΔS is the control variable for league position used in the regressions, S_{SLHT} is the SLHT's standing and S_{FLHT} is the FLHT's position in the league table.

The same variable was also defined lagged by one period:

$$\Delta S_{-1} = S_{SLHT-1} - S_{FLHT-1} \tag{2}$$

Where the subscript -1 indicates the lag to the previous edition. It can be shown that $-50 \leq \Delta S \leq 50$, where a lower value indicates a stronger SLHT in relation to the FLHT. Specifically, a negative value of the variable signifies that the SLHT possesses a higher technical level than the FLHT, while a positive value indicates the opposite. Consequently, the higher the values of ΔS and ΔS_{-1}, the lower the probability of the SLHT winning a match.

3.4.2 *Odds*

Another measure of the teams' relative strength was developed using the information on match odds. First, the odds for win, draw, and loss were converted into probabilities by inverting the odds and normalizing the sum of these values so that they totaled 1.

However, this approach presents a problem: the probabilities for the first game are conditioned by whether the team is playing at home or not. Thus, using this variable would overestimate the strength of the away team in the second match. The opposite happens if we use the probabilities from the second match, with the added complication that the probabilities are also conditioned by the SLHA effect and, certainly, by the result of the first match. This issue arises in cases where bookmakers and bettors believe in the existence of SLHA.

Combinations of probabilities from both the first and second legs can resolve part of this problem. It is reasonable to assume that averaging the odds mitigates the HA effects for each team in the match, but it is still needed to address the endogeneity caused by considering SLHA in the odds. Amez et al. (2020) do not account for this aspect in their measure of relative strength, which may reduce the explanatory power of the variable of interest.

Considering the points mentioned above, we propose an instrumental variables approach. Below, the creation of the relative strength variable and the instrument

variable are detailed.

Let $\Delta F_i = P_{SLHTi} - P_{FLHTi}$, where P_{SLHTi} indicates the probability of the SLHT winning in match i and P_{FLHTi} indicates the probability of the FLHT winning in that match. Additionally, i indicates whether the encounter refers to the first or second leg of the match ($i = \{FL, SL\}$). Also, let $P\mu_i$ represent the probability of a draw.

i) First, let the strength variable be defined as follows:

$$\Delta F = \Delta F_{FL} + \Delta F_{SL} \tag{3}$$

Where $-2 \leq \Delta F \leq 2$. As mentioned above, it is assumed that the HA effects for both teams are equal, on average. Therefore, over two matches, both effects would cancel out due to the large number of observations in the sample. Also, it should be noted that this strength measure is still endogenous.

ii) Consider the following linear regression:

$$\Delta F_{SL} = \beta_0 + \beta_1 \Delta F_{FL} + \beta_2 P\mu_{FL} + \epsilon \tag{4}$$

Observe that the error term ϵ in equation (8) will include, among all other factors, the component related to SLHA.

iii) Notice that the predicted values from the equation above will be exogenous to the SLHA effect:

$$\hat{\Gamma_j} = \hat{\beta}_0 + \hat{\beta}_1 \Delta F_{FLj} + \hat{\beta}_2 P\mu_{FL} \tag{5}$$

Where Γ is the instrument defined for ΔF and j is the index representing the matchup.

It is noted that Γ satisfies the two requirements of a good instrument (Wooldridge, 2010): i) it meets the exclusion restriction, i.e., it affects the dependent variable only

through ΔF, and ii) it is strongly correlated with the endogenous variable. The first-stage estimation can be found in Table 10 in the appendix. Finally, note that the higher ΔF, the higher the home team's technical level compared to the away team.

Arguably, using odds as a control has clear advantages. Bookmakers estimate match probabilities using up-to-date information and take into account several relevant variables, including recent team performance and quality of available players. This approach captures more relevant factors than rankings based on league standings. Furthermore, for these same reasons, odds are less subject to the arbitrariness problem of creating team rankings, especially when using the average odds from different bookmakers, as is the case in this study.

In addition, odds are frequently used in the sports literature, including in football. For example, Xu (2011) explores the efficiencies of using odds as predictors of match outcomes in the Premier League. The article's results show that the betting probabilities for the 2006-2007 season of the tournament are effective forecasts of football match results.

3.5 Model

In general, the studies that address SLHA estimate the existence of the effect using binary response models —*logit, probit*, or linear probability models. In this article, the results are reported from *probit* models, although the results from logistic regressions were very similar.

Let p be the probability of the home team winning the match. Consider the following probit:

$$p = \Phi(\alpha + \beta \Delta Strenght) \tag{6}$$

Where $\Phi(.)$ represents the cumulative density function of the standard normal distribution. For two teams with the same level of strength ($\Delta Strenght = 0$), SLHA exists if $p > 0.5$, indicating that the home team in the second leg has more than a 50% chance of advancing, even though it has the same level of ability as its opponent. This is equivalent to obtaining $\alpha > 0$ in the estimated model. Consequently, $\alpha \leq 0$ indicates the absence of SLHA. $\Delta Strenght$ refers to the control variables discussed in Subsection 3.4.

4 Descriptive Statistics

4.1 Evidence of *Second Leg Home Advantage*

The descriptive statistics shown below indicate the existence of SLHA in the Brazilian Cup. Over the entire history of the tournament, the home teams in the second leg advanced in 56.4% of the opportunities. Considering only the matches won by goal difference, home teams defeated their opponents 57.1% of the time. Both results are significant at the 0.1% level, holding up to both the one-tailed and two-tailed t-tests for mean difference, against the null hypothesis of 50%. Table 1 summarizes these statistics:

Table 1: Match Winners by Different Criteria

	SLHT	FLHT	Total	SLHT %	FLHT %	Variable
1	421	325	746	56.43%	43.57%	Total
2	350	263	613	57.1%	42.9%	Goal Difference
3	34	35	69	49.28%	50.72%	Away Goals
4	37	28	65	56.92%	43.08%	Penalties

Note: SLHT shows the number of matches won by the team playing the second match at home. FLHT shows the number of matches won by the team playing the first match at home. Prepared by the authors.

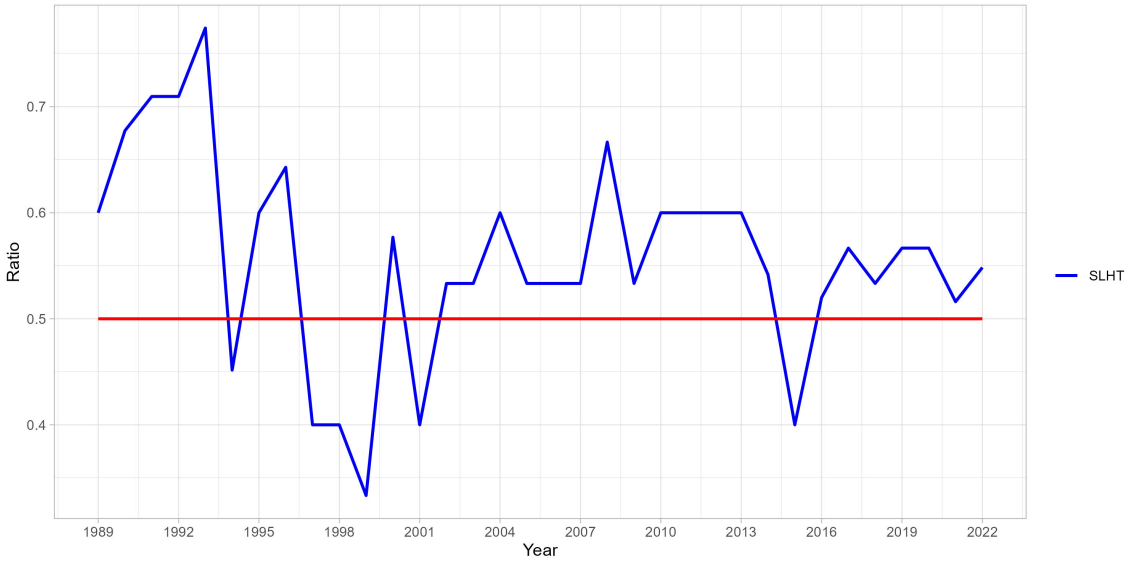
It can be observed that, considering the matches decided by the away goals rule,

apparently this rule did not benefit either group. On the other hand, there is an advantage for the home teams in the second leg during penalty shootouts. However, it is possible that this difference is explained by the imbalance in the early years of the sample, as previously noted. Thus, it is plausible that teams with a higher technical level have more frequently taken penalties on their home ground. Nevertheless, such differences are not statistically significant, perhaps due to the small number of observations.

Furthermore, Table 11 in the Appendix reports the same results as above but considers the subsample starting from 2009, as the data during this period are free from potential biases from non-random pairings.

Additionally, the year-by-year presence of this effect in the tournament was evaluated. Figure 1 summarizes how often home teams defeated their opponents in each edition. SLHA appears to be a persistent effect over time, though with varying magnitudes. Out of the 34 years in our sample, in 7 of them, the away teams from the second match advanced more often than the home teams: 1994, 1997, 1998, 1999, 2001, 2007, and 2015.

Figure 1: Win ratio in each edition - SLHT



Prepared by the authors.

A more detailed qualitative analysis of the matches in the years when SLHA was not observed in the Copa do Brasil revealed that, in the 1990s editions, some upsets occurred, and some of the random pairings placed the stronger teams to play the second leg away from home. For instance, in 1994, Ceará eliminated Palmeiras and Internacional as the away team in the second match. The Ceará team was playing in the second division at the time but reached the final of the Copa do Brasil that year. These two matches, if won by the strongest teams at the time, would have confirmed the SLHA effect in that edition on their own.

In the triennium from 1997 to 1999, and in 2001, there were also some upsets, such as Vasco's elimination in 1997, the team that would go on to become Brazilian champions that year, by Atlético Paranaense. However, it is especially noteworthy that the sample size between 1995 and 2012 (with the exception of the year 2000) is smaller, restricted to 15 matchups in our database for each of those years due to the regulations in place at the time. As mentioned earlier, during that period, the first two stages were played under the rule that allowed the away team to eliminate the home team if they won the first leg by two or more goals, which led to the exclusion of those matches from the dataset.

As a result, the sample is more susceptible to significant variations in results caused by random pairings, where stronger teams were drawn to play the second leg away from home, or in the case of upsets. Nonetheless, the apparent absence of SLHA in those editions remains notable. This issue will be revisited later, taking into account the imbalances in the sample.

Furthermore, the effect can be broken down by tournament stage. Since the stages have been officially named in different ways throughout the tournament editions, a nomenclature was created based on the distance of each stage from the final. In other words, the following stages were considered: final, semifinals, quarterfinals, round of 16, and round of 32. Once again, results were reported for both the full sample and

for the sample of the editions starting from 2009.

Tables 2 and 12 (in the Appendix) report these results. It is noteworthy that the presence of SLHA does not occur during the finals, contrasting with what is observed in the other tournament stages.

Table 2: Match Winners by Tournament Stage

Stage	SLHT	FLTH	Total	SLHT %	FLHT %
Round of 32	140	99	239	58.58%	41.42%
Round of 16	149	120	269	55.39%	44.61%
Quarterfinals	74	62	136	54.41%	45.59%
Semifinals	43	25	68	63.24%	36.76%
Final	15	19	34	44.12%	55.88%

Note: SLHT shows the number of matches won by the team playing the second match at home. FLHT shows the number of matches won by the team playing the first match at home. Prepared by the authors.

Additionally, before delving into the estimation results, it is important to qualify the matches described in the dataset to be used in the regressions. Table 3 characterizes the matches according to the division in which both teams were located. It is expected that, due to random seeding, there is a similar number of matchups in each group, considering the division of the home and away teams.

The group "No Division" includes teams from Série D and those without a division from 2013 to 2022. Between 1989 and 2012, this group also includes teams from Série C, as previously discussed. Additionally, the balance of the observations can be checked by using the average of the home team's ranking difference compared to the opponents' average (previously defined as ΔS).

The sample mean of this variable is -2.1. This result indicates a slight imbalance in the allocation of teams, with home teams in the second match occupying a slightly better ranking. For the subsample starting in 2009, the variable has a mean of 0.4, indicating a negligible difference in the seeding process.

Table 13 in the Appendix replicates the results of the previous table, considering

Table 3: Matchups: Division of Clubs in the Sample

	SLHT X FLHT	Total	Sample %
1	Série A X Série A	337	45.17%
2	Série A X Série B	111	14.88%
3	Série B X Série A	76	10.19%
4	Série A X No Division	72	9.65%
5	No Division X Série A	42	5.63%
6	Série B X Série B	27	3.62%
7	No Division X No Division	16	2.14%
8	Série C X Série A	16	2.14%
9	Série B X No Division	14	1.88%
10	No Division X Série B	13	1.74%
11	Série A X Série C	7	0.94%
12	Série B X Série C	7	0.94%
13	Série C X Série C	3	0.40%
14	Série C X Série B	3	0.40%
15	No Division X Série C	1	0.13%
16	Série C X No Division	1	0.13%

Note: SLHT indicates the team playing the second match at home. FLHT indicates the team playing the first match at home. Prepared by the authors.

only the subsample starting from 2009.

5 Results

In this section, the results of the estimations are divided into two groups. The first tables report the results obtained using the complete observations of the dataset. Later, the values found for observations from 2009 onwards are presented, intending to compare the standing control variables (ΔS and ΔS_{-1}) with the variable derived from betting odds (ΔF).

Table 4: Probit Estimations

		Win	
Model:	(1)	(2)	(3)
	Probit	Probit	Probit
Variables			
Constant	0.0986*	0.1303***	0.1109**
	(0.0490)	(0.0488)	(0.0495)
ΔS_{-1}	-0.0260***		-0.0126***
	(0.0025)		(0.0043)
ΔS		-0.0268***	-0.0163***
		(0.0025)	(0.0044)
Statistics			
Observations	746	746	746
Corr^2	0.15018	0.16161	0.16961
Pseudo R^2	0.11903	0.12452	0.13282
BIC	912.93	907.33	905.46

IID Standard errors in parentheses. Significance Codes: ***: 0.01, **: 0.05, *: 0.1 *Note:* The models use observations from 1989 onwards. Columns (1) to (3) represent different specifications, including the various controls presented. Column (1) presents the lagged specification of the control variable. Column (3) includes both variables. Prepared by the authors.

Table 4 presents the selected results of the regressions. As observed, in all specifications, the constant is positive and significant at least at the 10% level, indicating the existence of SLHA. Furthermore, it is noted that the control variables have the expected sign: the higher the SLHT's technical level compared to the FLHT (lower ΔS), the greater the probability of a match win by the SLHT. In terms of interpretation, the focus is on the probability of a home win in a match, considering equal

team strengths ($\Delta S = 0$ and $\Delta S_{-1} = 0$).

Therefore, to calculate the estimated effect of SLHA, it is necessary to obtain $\hat{p_0} = \Phi(\hat{\alpha})$, where $\hat{\alpha}$ is the estimated value of the constant in the models above. Table 5 compares these values with the average win percentage of the home team found in the sample—56.6% of SLHTs advanced between 1989 and 2022.

Table 5: Predicted SLHA and Difference from the Sample SLHA

Model	$\hat{p_0}$	Difference
(1) Probit	0.5393	-0.0264
(2) Probit	0.5518	-0.0138
(3) Probit	0.5442	-0.0215
Observed	0.5657	

Note: $\hat{p_0}$ is the predicted SLHA from the respective Probit model. Prepared by the authors.

Interestingly, all three models underestimate the SLHA effect compared to the sample data. This is expected, as team pairings are slightly unbalanced, as discussed earlier, and the control variables help correct for this issue.

The estimates presented in Table 6 below, in turn, report the results with control variables derived from the odds - ΔF - including estimates with the instrumental variable, with the Γ instrument previously defined. The time window includes information from 2009 onwards. These regressions also support the hypothesis of the existence of SLHA.

In all models, the constant is positive and significant at least at the 10% level. Additionally, the signs of the control variables have the expected direction: the higher the technical level of the home team compared to the visiting team (higher ΔF and lower ΔS), the higher the probability of winning the match. Notably, even in the endogenous specification of the first model, the null hypothesis of no effect is rejected, which, in other words, indicates that bookmakers (and thus bettors) underestimate

the magnitude of SLHA. Moreover, there is clear superiority in the odds-derived model in terms of model fit, with the highest Pseudo R^2 above 0.2, indicating good predictive power (McFadden, 1974).

Table 6: Estimates - Models with *odds*

	Win					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	Probit (IV)	Probit	Probit	Probit	Probit (IV)
Variables						
Constant	0.1322*	0.1267*	0.1393*	0.1460**	0.1450*	0.1254*
	(0.0761)	(0.0760)	(0.0732)	(0.0738)	(0.0740)	(0.0761)
ΔF	1.296***	1.322***				1.3678***
	(0.1423)	(0.1426)				(0.2384)
ΔS_{-1}			-0.0311***		-0.0104	0.0015
			(0.0040)		(0.0078)	(0.0070)
ΔS				-0.0323***	-0.0235***	
				(0.0040)	(0.0078)	
Statistics						
Observations	340	340	341	341	341	340
Corr^2	0.27496		0.19132	0.21860	0.22173	
Pseudo R^2	0.22374		0.15331	0.16967	0.17337	
BIC	375.48		409.53	401.84	405.93	

IID Standard errors in parentheses. Significance Codes: ***: 0.01, **: 0.05, *: 0.1 *Note:* The models use observations from a dataset of 341 matchups played between 2009 and 2022. Columns (1) to (6) represent different model specifications, including the presented controls. Column (1) presents the endogenous model specification, as previously discussed. Column (2) specifies the same control but instruments the variable ΔF. The match between Grêmio and Bahia in 2012 did not have odds available for the first-leg. Prepared by the authors.

Table 7 shows the estimated effects, $\hat{p}_0$, found for the specifications shown in

Table 6. For the years in question, the Second-Leg Home Teams advanced in 54.6% of the cases in the sample.

Table 7: Difference between predicted SLHA and sample SLHA - Models with *odds*

Model	$\hat{p_0}$	Difference
(1) Probit	0.5526	0.0071
(2) Probit (IV)	0.5504	0.0050
(3) Probit	0.5553	0.0099
(4) Probit	0.5580	0.0126
(5) Probit	0.5576	0.0122
(6) Probit (IV)	0.5498	0.0044
Observed	0.5455	

Note: $\hat{p_0}$ is the predicted SLHA from the respective Probit model. Prepared by the authors.

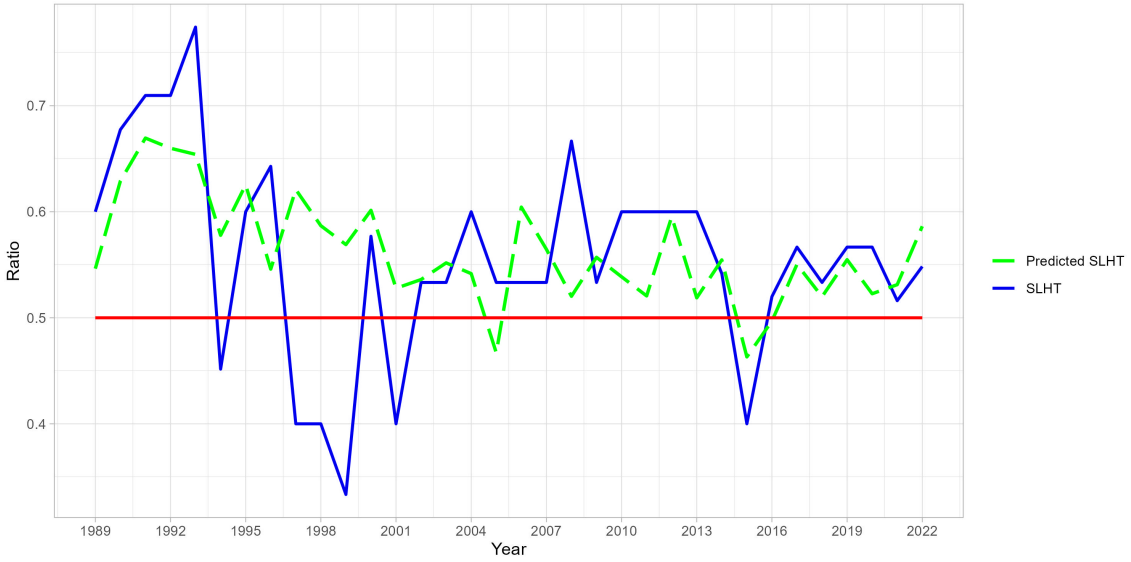
From the values reported above, it is noteworthy that the models using the ΔF control present the estimated coefficient $\hat{p_0}$ numerically closer to the observed sample.

Returning to the results presented in Table 4, one of the important questions to be addressed in this study is the evolution of the SLHA effect over the years. A quick analysis of Figure 1 shows, in addition to the large variance in its magnitude, an apparent decrease over the years in the impact of playing the second leg of a knockout match at home. Given that the observations are slightly unbalanced, especially in the early editions of the Copa do Brasil, one way to approach this issue is through the predicted values from the regressions. Figure 2 compares the annual results from Figure 1 with the average predicted by model (3) in Table 4. The choice of this model specification was due to its greater predictive capacity in terms of Pseudo R^2.

Figure 2 shows that, when controlling for the difference in team strength, the existence of SLHA is not observed in only two editions of the tournament. In other words, this effect remains relatively constant over time in the Copa do Brasil, with

just two occasions where its presence cannot be confirmed.

Figure 2: Win ratio in each edition and predicted values - SLHT



Prepared by the author.

Therefore, the empirical evidence presented here reveals that clubs playing the second leg at home enjoy an additional advantage, with estimated chances of victory ranging from 53.9% to 55.8%, holding other variables constant. Furthermore, the calculated effect has historically been present throughout the tournament's editions.

6 Extension: A Good Result in the First-leg

This section aims to explore what happens in the second part of the match, conditioned on the result of the first leg. The idea is that, given the goal difference between the two teams in the first match of the dispute, one can observe the probability of victory. The basic model estimated, therefore, is defined as follows.

$$p_{|FL} = \Phi(\alpha + \beta_1 \Delta Strenght + \beta_2 \Delta G_{FL}) \tag{7}$$

Where $p_{|FL}$ is the probability of victory in the match conditioned on the result of

the first leg, and $\Delta G_{FL} = GS_{FL} - GF_{FL}$, where GS_{FL} and GF_{FL} are the number of goals scored in the first leg by the SLHT and the FLHT, respectively. It is immediate to see that if $\Delta G_{FL} < 0$, the visitor has won the first match.

Table 8 shows the results. All the control variables kept their expected signs. For the new control variable included in the specification, GF_{FL}, the greater the score difference in the first match, the higher is the SLHA, as expected.

Table 8: Home Advantage Conditional

	Victory					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	Probit (IV)	Probit	Probit	Probit	Probit (IV)
Variables						
Constant	0.3639***	0.3687***	0.3783***	0.3961***	0.3811***	0.3709***
	(0.0895)	(0.0892)	(0.0599)	(0.0595)	(0.0602)	(0.0894)
ΔG_{FL}	0.4966***	0.5024***	0.5895***	0.5892***	0.5834***	0.5034***
	(0.0705)	(0.0717)	(0.0498)	(0.0500)	(0.0501)	(0.0718)
ΔF	0.9477***	0.9293***				0.8172***
	(0.1579)	(0.1605)				(0.2664)
ΔS_{-1}			-0.0175***		-0.0094*	-0.0041
			(0.0029)		(0.0048)	(0.5823)
ΔS				-0.0176***	-0.0100***	
				(0.0029)	(0.0048)	
Statistics						
Observations	340	340	746	746	746	340
Corr^2	0.43861		0.36731	0.37015	0.37430	
Pseudo R^2	0.35990		0.31432	0.31484	0.31861	
BIC	317.49		720.73	716.37	719.17	

IID Standard errors in parentheses. Significance Codes: ***: 0.01, **: 0.05, *: 0.1 *Note:* Sample is from 2009 onwards when ΔF is included. Columns (1) to (6) represent various specifications of the presented model, including the different controls introduced. Column (1) presents the endogenous specification of the model, as discussed earlier. Column (2) specifies the same control, but instrumentalizes the variable ΔF. Prepared by the authors.

Consider $\hat{p}_{|FL=x}$ representing the conditional probability of victory for the home team in the match, given $x = \Delta G_{FL}$ before the return leg. Analogous to the interpretation of the coefficients from Tables 4 and 6, one can obtain the home team's chance of victory conditional on the goal difference in the first match through the estimated coefficients above. Considering the strength difference equal to 0, in the case of a 0-0 draw (or any other tie, $\Delta G_{FL} = 0$), we have $\hat{p}_{|FL=0} = \Phi(\hat{\alpha})$. In the case of a home team victory by one goal difference in the first leg, the probability is given by: $\hat{p}_{|FL=1} = \Phi(\hat{\alpha} + \hat{\beta}_2)$. Table 9 calculates this information using the specification of model (2) from Table 8.

Table 9: Probability of winning conditional on the first leg result

ΔG_{FL}	$\hat{p}_{\|FL=x}$	Effect(%)	$\hat{q}_{\|FL=x}$
-5	0.0160		0.9840
-4	0.0504	3.44%	0.9496
-3	0.1275	7.71%	0.8725
-2	0.2624	13.49%	0.7376
-1	0.4468	18.44%	0.5532
0	0.6438	19.70%	0.3562
1	0.8082	16.44%	0.1918
2	0.9152	10.70%	0.0848
3	0.9697	5.45%	0.0303
4	0.9913	2.16%	0.0086
5	0.9980	0.67%	0.0020

Note: Prepared by the authors.

The column *Effect(%)* measures the impact of scoring one additional goal— all else being equal — on the calculated probability of win for the SLHT. This measure is expressed in percentage terms. Meanwhile, $\hat{q}_{|FL=x} = 1 - \hat{p}_{|FL=x}$ is the probability

of victory for the FLHT.[2].

Still considering teams with the same technical level, it is notable that, for the SLHT, a draw in the first leg increases the chances of advancing to the next stage to 64.4%. However, a defeat by a one-goal difference in the first leg already reverses this advantage. In addition, the unconditional specification of model (2) presented in Table 6 shows the presence of SLHA with a magnitude of 55.0%. In other words, the result of the conditional model also indicates that, on average, the visitors are defeated in the first leg by a difference of 0.4817 goals, all else being equal [3].

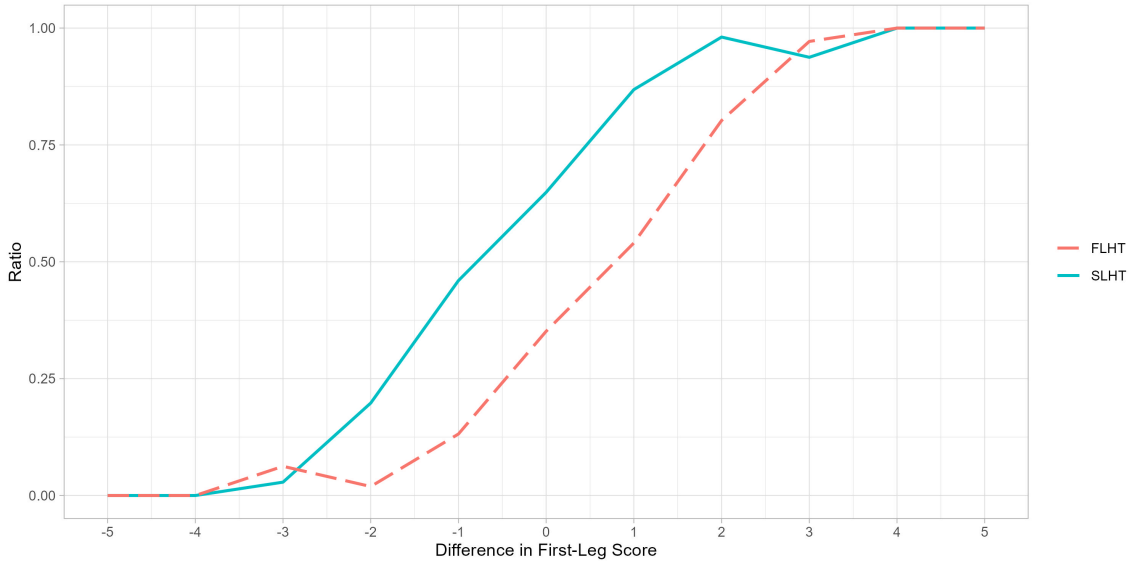
Alternatively, based on the results above, it is possible to calculate the goal difference needed in the first leg so that no SLHA exists. If the home teams were defeated, on average, by 0.7339 goals, the probability of the visiting teams advancing would be 50%. Any SLHT result higher than this value in the first match reveals the effect of Second-leg Home Advantage.

To illustrate what has been discussed in this section, one can graphically represent the frequency of home and away victories after the first-leg score in the sample. Figure 3 reports this information.

[2]It could be argued that, in addition to the score difference variable in the first match played (ΔG_{FL}), the number of away goals scored in that same match could influence the chance of a win, in the cup editions where away goals criteria were adopted. When the control for those editions when the away goals rule was adopted as a tiebreaker was included in the specification, such variable did not return statistical significance. Furthermore, the coefficients of the other variables remained very close in magnitude to the values estimated by equation (7).

[3]From $0.5504 = \Phi(0.3687 + \hat{\beta}_2 \overline{\Delta G_{FL}})$, one obtains $\overline{\Delta G_{FL}} = -0.4817$.

Figure 3: Sample proportion of wins conditional on the goal difference in the first leg



Presented by the author.

In summary, this section presented an empirical approach to estimate the probability of winning a match conditional on the result of the first game. For instance, a first-leg draw favors the second-leg home team, increasing its win probability to over 64%, assuming equal team strength. Additionally, the away goals rule does not seem to have a significant impact on these probabilities, contrary to the common belief that "away goals count double," which is often argued in popular discourse.

7 Conclusions

This study examines the advantage of playing the second leg at home, using data from the Brazilian Cup. Our study found evidence for the existence of such an effect, with a range varying between 53.9% and 55.8%, taking into account the difference in the strength of the teams. The values predicted by the models also indicate that SLHA has been observed since the beginning of the sample period, except for only two years. Models that use *odds* to calculate the relative strength show a better fit to

the data. An instrumental variable approach was used to control for the endogeneity of second-leg odds. The IV estimations are the closest estimates to the SLHA value observed in the sample.

In addition, we also investigated what happens in the second match, conditional on the result observed in the first match. A relevant result is that, considering opponents with the same technical level, a draw away from home in the first game implies a chance of victory greater than 64% (in the most conservative model) for the home team in the return match.

Furthermore, the difference in goals between the clubs in the first game is an important determinant of the probability of winning the match, but the number of goals scored away from home is not as relevant in any of the estimated models. This a notable result, given that it is commonly assumed that a 0-0 draw away from home in the first game is a much better scenario for the visitor in the second game than a draw with goals.

References

Abad, C. C. C., Terra, B. P., de Toledo Ananias, M. D., da Silva, R. A., & Diniz, M. A. (2017). Having the second leg at home: Advantage in the UEFA champions league knockout phase? Motriz. Revista de Educação Física, 23(3), 1–8.

Amez, S., Baert, S., Neyt, B., & Vandemaele, M. (2020). No evidence for second leg home advantage in recent seasons of European soccer cups. Applied Economics Letters, 27(2), 156–160. https://EconPapers.repec.org/RePEc:taf:apeclt:v:27:y:2020:i:2:p:156-160

Dobson, S., & Goddard, J. (2011). The economics of football. Cambridge University Press.

Eugster, M. J., Gertheiss, J., & Kaiser, S. (2011). Having the second leg at home - advantage in the UEFA champions league knockout phase? Journal of Quantitative Analysis in Sports, 7(1), 1–11. https://EconPapers.repec.org/RePEc:bpj:jqsprt:v:7:y:2011:i:1:n:6

Geenens, G., & Cuddihy, T. (2018). Non-parametric evidence of second-leg home advantage in European football. Journal of the Royal Statistical Society Series A, 181(4), 1009–1031. https://EconPapers.repec.org/RePEc:bla:jorssa:v:181:y:2018:i:4:p:1009-1031

Lidor, R., Bar-Eli, M., Arnon, M., & Bar-Eli, A. A. (2010). On the advantage of playing the second game at home in the knockout stages of European soccer cup competitions.". International Journal of Sport and Exercise Psychology, 8(3), 312–325.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. Frontiers in econometrics, 105–142.

Neave, N., & Wolfson, S. (2003). Testosterone, territoriality, and the "home advantage". Physiol Behav, 78, 269–275.

Page, L., & Page, K. (2007). The second leg home advantage: Evidence from European football cup competitions. Journal of Sports Sciences, 14, 1547–1556.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. Journal of Sports Sciences, 4, 237–248.

Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876-2003). Journal of Sports Sciences, 4, 337–350.

Waquil, A. P., Horta, E., & Moraes, J. C. (2020). Home advantage and away goals rule: An analysis from Brazil cup. Journal of Sports Analytics, 6(1), 13–24.

Wooldridge, J. (2010). Econometric analysis of cross section and panel data. MIT Press.

Xu, J. S. (2011). Online sports gambling: A look into the efficiency of bookmakers' odds as forecasts ir
[Doctoral dissertation, University of California].

A Additional Tables

Table 10: First stage with instrument

Model:	ΔF (1)
Variables	
Constant	-0.4258***
	(0.0092)
Γ	2.132***
	(0.0259)
Statistics	
Observations	340
R^2	0.95236
R^2 Adjusted	0.95222

IID Standard errors in parentheses. Significance Codes: ***: 0.01, **: 0.05, *: 0.1 *Note:* The models use observations from a dataset of 341 matchups played between 2009 and 2022. The match between Grêmio and Bahia in 2012 did not have odds available for the first-leg. Prepared by the authors.

Table 11: Match Winners by Criteria - From 2009 Onwards

	SLHT	FLHT	Total	SLHT %	FLHT %	Variable
1	185	156	341	54.25%	45.75%	Total
2	155	126	281	55.16%	44.84%	Goal Difference
3	9	13	22	40.91%	59.09%	Away Goals
4	21	18	39	53.85%	46.15%	Penalties

Note: SLHT shows the number of matches won by the team playing the second match at home. FLHT shows the number of matches won by the team playing the first match at home. Prepared by the authors.

Table 12: Match Winners by Tournament Stage - From 2009 Onwards

Stage	SLHT	FLTH	Total	SLHT %	FLHT %
Round of 32	66	66	132	50.00%	50.00%
Round of 16	58	53	111	52.25%	47.75%
Quarterfinals	35	21	56	62.50%	37.50%
Semifinals	20	8	28	71.43%	28.57%
Final	6	8	14	42.86%	57.14%

Note: SLHT shows the number of matches won by the team playing the second match at home. FLHT shows the number of matches won by the team playing the first match at home. Prepared by the authors.

Table 13: Matchups: Division of Clubs in the Sample - From 2009 Onwards

	SLHT X FLHT	Total	Sample %
1	Série A X Série A	155	45.45%
2	Série A X Série B	45	13.20%
3	Série B X Série A	36	10.56%
4	Série A X No Division	23	6.74%
5	No Division X Série A	21	6.16%
6	Série C X Série A	16	4.69%
7	Série B X Série B	9	2.64%
8	No Division X Série B	7	2.05%
9	Série A X Série C	7	2.05%
10	Série B X Série C	7	2.05%
11	Série B X No Division	5	1.47%
12	Série C X Série C	3	0.88%
13	Série C X Série B	3	0.88%
14	No Division X No Division	2	0.59%
15	No Division X Série C	1	0.29%
16	Série C X No Division	1	0.29%

Note: SLHT indicates the team playing the second match at home. FLHT indicates the team playing the first match at home. Prepared by the authors.