

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E ATUÁRIA
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

***Home Advantage* no Futebol Brasileiro: Uma Análise Multidimensional da Última Década**

Marcus Vinícius Carone Jovita
Orientador: Prof. Dr. Márcio Issao Nakane

São Paulo

2024

Prof. Dr. Carlos Gilberto Carlotti Junior
Reitor da Universidade de São Paulo

Profa. Dra. Maria Dolores Montoya Diaz
Diretora da Faculdade de Economia, Administração, Contabilidade e Atuária

Prof. Dr. Claudio Ribeiro de Lucinda
Chefe do Departamento de Economia

Prof. Dr. Mauro Rodrigues Junior
Coordenador do Programa de Pós-Graduação em Economia

MARCUS VINÍCIUS CARONE JOVITA

***Home Advantage* no Futebol Brasileiro: Uma Análise
Multidimensional da Última Década**

Dissertação apresentada no Departamento de Economia da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo como requisito da obtenção do título de Mestre em Ciências.

Área de Concentração: Teoria Econômica

Orientador: Prof. Dr. Márcio Issao Nakane

Versão Original

São Paulo

2024

Jovita, Marcus Vinícius Carone

Home Advantage no Futebol Brasileiro: Uma Análise Multidimensional da Última Década/ Marcus Vinícius Carone Jovita. – São Paulo, 2024.

63p.

Dissertação (Mestrado) – Universidade de São Paulo, 2024.

Orientador: Prof. Dr. Márcio Issao Nakane

1. Futebol. 2. Economia dos Esportes. 3. *Home Advantage*. 4. Econometria. I. Universidade de São Paulo. Faculdade de Economia, Administração, Contabilidade e Atuária. II. Título.

Agradecimentos

Agradeço, primeiramente, aos meus pais, Manoel e Melissa, que foram imprescindíveis ao longo de toda a minha formação pessoal e profissional e representam um exemplo vivo de dedicação e amor incondicionais.

Ao professor Márcio Nakane, meu orientador, quero expressar a minha mais sincera gratidão por sua orientação exemplar, paciência e sabedoria. Em momentos turbulentos e de incerteza, o senhor não apenas se mostrou compreensivo, mas confiou em mim e me guiou por esta jornada com maestria. Faço menção também aos professores Ricardo Avelino e Mauro Rodrigues, pela ajuda no exame de qualificação e na avaliação de progresso, e à banca examinadora, pela atenção e tempo despendidos em minha avaliação.

À minha irmã, Mayanna, este agradecimento vem do fundo do meu coração, por sempre acreditar em mim e nunca deixar eu pensar o contrário. Tenho enorme admiração por você. Obrigado por ser minha parceira nesta viagem incrível que é a vida.

Aos meus irmãos mais novos, João Pedro e Mariah, cujo amor, apoio e inspiração foram fundamentais em minha jornada. A alegria de vocês é contagiante e foi indispensável ao longo deste processo.

À querida Alice, que me apoiou em cada momento desta etapa da minha vida e a tornou mais leve, com seu carinho, conforto e companhia. Fico muito feliz de compartilhar esta vitória com você.

Aos demais familiares, que também desempenharam papéis cruciais nesta jornada, minha mais sincera gratidão. Cada um de vocês, de maneira única, contribuiu para minha caminhada, seja por meio de palavras de incentivo, gestos de carinho ou, simplesmente, pela presença e apoio constantes.

Por último, quero agradecer à Universidade de São Paulo (USP), à Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação Instituto de Pesquisas Econômicas (FIPE), por viabilizarem esta pesquisa.

Resumo

Esta dissertação investiga a influência de alguns fatores, como ocupação dos estádios, distância de viagem dos times visitantes e viés de arbitragem, sobre o fenômeno de *Home Advantage* no futebol brasileiro. Este estudo também aproveita os últimos eventos que influenciaram o mundo do futebol e examina o impacto da pandemia de Covid-19 e da implementação do VAR (Árbitro de Vídeo) sobre essa vantagem. A abordagem empírica revela que a presença de público e a distância de viagem têm efeitos distintos sobre o fenômeno na primeira e segunda divisões do Campeonato Brasileiro e que a pandemia reduziu a vantagem da equipe mandante, possivelmente devido à ausência de torcedores nos estádios. Por outro lado, verifica-se que a introdução do VAR teve um impacto limitado nas decisões de arbitragem de forma geral, com destaque para cartões amarelos, mas seu efeito parece ser praticamente inexistente para decisões que podem influenciar mais diretamente o curso de uma partida, como cartões vermelhos e pênaltis.

Palavras-chave: Economia do Esporte, Econometria, Home Advantage, Covid-19, Árbitro de Vídeo

Classificação JEL: C50; Z20; Z29

Abstract

This dissertation investigates the influence of several factors, such as stadium occupancy, travel distance of visiting teams, and referee bias, on the phenomenon of Home Advantage in Brazilian soccer. This study also takes advantage of the latest events that have influenced the world of soccer and examines the impact of the Covid-19 pandemic and the implementation of VAR (Video Assistant Referee) on this advantage. The empirical approach reveals that the presence of spectators and the travel distance have distinct effects on the phenomenon in the first and second divisions of the Brazilian Championship, and that the pandemic reduced the home team advantage, possibly due to the absence of fans in the stadiums. On the other hand, it is found that the introduction of VAR had a limited impact on refereeing decisions in general, with an emphasis on yellow cards, but its effect seems to be virtually nonexistent for decisions that can more directly influence the course of a match, such as red cards and penalties.

Keywords: Sports Economics, Econometrics, Football, Soccer, Home Advantage, Covid-19, Video Assistant Referee

JEL Classification: C50; Z20; Z29

Sumário

1	INTRODUÇÃO	17
2	REVISÃO DE LITERATURA	19
2.1	Evidências de <i>Home Advantage</i>	19
2.2	<i>Home Advantage</i> em Outros Esportes	20
2.3	<i>Home Advantage</i> no Futebol	20
2.3.1	Público	22
2.3.1.1	Covid-19	22
2.3.2	Arbitragem	24
2.3.2.1	Árbitro de Vídeo	24
2.3.3	Distância Percorrida pelo Visitante	25
2.3.4	<i>Home Advantage</i> no Brasil	26
3	DADOS E ESTATÍSTICAS DESCRITIVAS	28
3.1	Dados	28
3.2	Estatísticas Descritivas	29
4	METODOLOGIA	37
4.1	Método Ajustado de Pollard	37
4.2	Modelo Logístico Binário	37
4.3	Modelos por Mínimos Quadrados Ordinários	39
5	RESULTADOS E DISCUSSÃO	41
5.1	Método Ajustado de Pollard	41
5.2	Resultados de Desempenho da Equipe Mandante	42
5.3	Resultados de Arbitragem	49
5.3.1	Grupo 1 - Cartões:	49
5.3.2	Grupo 2 - Impedimentos e Pênaltis:	52
6	CONCLUSÃO	55
	REFERÊNCIAS	57
	APÊNDICES	61
	APÊNDICE A – TABELAS ADICIONAIS	62

Lista de ilustrações

Figura 1 – Método Ajustado de Pollard (2012/2022) - Gols	41
Figura 2 – Método Ajustado de Pollard (2012/2022) - Pontos	42
Figura 3 – Probabilidades Previstas - Diferencial de Colocação	44
Figura 4 – Histogramas por Grupo de Dados	45

Lista de tabelas

Tabela 1 – Estatísticas Mandante x Visitante - Dados Totais	30
Tabela 2 – Outras Médias	31
Tabela 3 – Distribuição Geográfica de Equipes e Partidas	31
Tabela 4 – Proporção de Resultados Time Mandante	32
Tabela 5 – Estatísticas Mandante x Visitante - Séries A e B	32
Tabela 6 – Diferencial Mandante x Visitante - Covid-19	33
Tabela 7 – Diferencial Mandante x Visitante - VAR	35
Tabela 8 – Método Ajustado de Pollard	41
Tabela 9 – Resultados Logit	43
Tabela 10 – Efeitos Marginais Médios e Probabilidade Preditada da Constante	47
Tabela 11 – Resultados MQO - Diferencial de Gols	48
Tabela 12 – Resultados MQO - Cartões	51
Tabela 13 – Resultados MQO - Impedimentos e Pênaltis	53
Tabela 14 – Resultados Logit - Público Absoluto	62
Tabela 15 – Resultados MQO - Diferencial de Gols - Público Absoluto	63

1 Introdução

O conceito de *Home Field Advantage*, ou vantagem de jogar em casa (VC), tem sido um tópico de investigação recorrente no domínio da psicologia esportiva e análise de desempenho no futebol. Este fenômeno, caracterizado pela maior probabilidade de vitória para a equipe que joga em seu próprio estádio, tem atraído interesse tanto por suas implicações práticas quanto teóricas. No contexto do futebol brasileiro, um esporte que desperta paixões intensas e movimenta multidões, entender a extensão e as causas da vantagem de jogar em casa torna-se ainda mais relevante. Por este motivo, esta pesquisa visa entender o efeito da VC no Campeonato Brasileiro de futebol, com foco em identificar e analisar os principais fatores que contribuem para este fenômeno, bem como explorar o comportamento dessas variáveis em diferentes períodos e divisões do campeonato.

O estudo se propõe a investigar como o nível de ocupação nos estádios, a distância percorrida pelos times visitantes e o viés de arbitragem influenciam a vantagem de jogar em casa no futebol brasileiro. A presença de torcedores pode criar uma atmosfera que apoia a equipe mandante e intimida a equipe visitante, potencialmente afetando o desempenho dos jogadores e as decisões dos árbitros. Além disso, viagens longas para a equipe visitante podem levar a um desempenho inferior devido à fadiga e outros fatores relacionados ao deslocamento. O estudo também explora como o viés de arbitragem, potencialmente exacerbado pela pressão da torcida, pode favorecer a equipe da casa.

O comportamento do fenômeno em períodos específicos afetados por eventos externos significativos, como a pandemia de Covid-19 e a implementação do árbitro de vídeo (VAR) no Campeonato Brasileiro também é analisado. A ausência de público nos estádios durante a pandemia proporciona uma oportunidade única para avaliar o impacto da torcida, enquanto a introdução do VAR oferece um contexto para analisar mudanças no viés de arbitragem e como isso pode influenciar a vantagem de jogar em casa, considerando a precisão das decisões dos árbitros e a percepção de todos os envolvidos na partida.

O trabalho ganha relevância diante da escassez de literatura específica ao contexto brasileiro a respeito do fenômeno da VC e pelo fato de abranger uma quantidade relevante de dados analisados. Representa, assim, uma fonte potencial de percepções que podem ser aplicadas ao aprimoramento de estratégias de equipes e ao desenvolvimento de políticas esportivas, enriquecendo os debates no âmbito do futebol nacional.

A presença de público e a distância de viagem apresentaram efeitos distintos sobre o fenômeno da vantagem de jogar em casa na primeira e segunda divisões do Campeonato Brasileiro. A pandemia do Covid-19, em ambos os casos, reduziu a vantagem da equipe mandante, possivelmente devido à ausência de torcedores nos estádios. A introdução do

VAR teve um impacto limitado sobre as decisões de arbitragem de forma geral, com destaque para cartões amarelos, mas seu efeito parece ser praticamente inexistente para decisões que podem influenciar mais diretamente o curso de uma partida, como cartões vermelhos e pênaltis.

A dissertação está estruturada da seguinte forma: esta seção introduz o tema e estabelece a justificativa para o estudo; a seção 2 revisa a literatura existente sobre a vantagem de jogar em casa, destacando diferentes esportes, campeonatos e abordagens; a seção 3 apresenta os dados coletados e estatísticas descritivas; a seção 4 descreve a metodologia utilizada na pesquisa; a seção 5 apresenta os resultados da análise e os discute; e, por fim, a seção 6 sintetiza os resultados e conclui.

2 Revisão de Literatura

2.1 Evidências de *Home Advantage*

O conceito de *Home Field Advantage*, ou vantagem de jogar em casa, é uma ideia amplamente reconhecida e debatida no mundo dos esportes, referindo-se à tendência observada de equipes jogarem melhor e obterem mais vitórias em seu próprio ambiente familiar, em comparação a jogos disputados no campo do adversário. Ela representa um fenômeno complexo, que se manifesta em diversas modalidades esportivas ao redor do mundo e tem sido um foco extenso de pesquisa nos últimos anos à medida que os analistas buscam compreender melhor as suas causas e efeitos. Ao explorar este conceito, pesquisadores procuram entender como fatores ambientais e contextuais podem afetar o desempenho atlético e, conseqüentemente, a natureza competitiva desses esportes. Algumas das modalidades nas quais este fenômeno foi documentado, entre outras, incluem o basquete, esportes de combate, o handebol e, principalmente, o futebol.

Um dos primeiros trabalhos a investigar o tema foi o de Schwartz e Barsky (1977), que investigaram a existência de *home advantage* com base em porcentagem de vitórias e, exclusivamente para o beisebol, na quantidade de pontos marcados dentro de uma partida. Os resultados foram favoráveis à existência de HA em todas as outras modalidades esportivas investigadas (basquete, futebol e hóquei de gelo). Pollard (1986) também explora o fenômeno em uma coleção de esportes com base na porcentagem de vitórias dos times mandantes em relação ao total de partidas e identifica indícios de vantagem do time da casa nas principais ligas estadunidenses (beisebol, futebol americano, hóquei de gelo, basquete e futebol) e inglesas (críquete e futebol) entre 1982 e 1984.

Diante de um panorama geral, Pollard, Prieto e Gómez (2017) conduziram um estudo que analisou as variações de VC ao longo de diversos esportes, países, e entre competições femininas e masculinas. Os dados foram coletados de 15 esportes em 65 países ao longo de cinco anos (2011-2015). O ritmo do esporte e as dimensões da arena de jogo mostraram influência sobre o fenômeno: esportes mais rápidos e com arenas menores, como basquete e handebol, mostraram uma maior vantagem do time mandante. O estudo ainda identificou uma vantagem consistentemente menor para mulheres no esporte, o que pode ser atribuído a uma menor presença de torcida e diferenças nos níveis de testosterona no corpo. O efeito também não se mostrou consistente a depender do país e do esporte escolhido.

2.2 *Home Advantage* em Outros Esportes

A influência da VC no basquete é extensivamente pesquisada, com estudos que determinam vários fatores causadores deste fenômeno. McHill e Chinoy (2020) mostraram que competições em um cenário em que os times não possuem a necessidade de viajar reduzem os efeitos típicos de fadiga e vantagem do campo mandante na porcentagem de vitórias, precisão de arremessos e rebotes. Há estudos, também, sobre o impacto do barulho da torcida, que é um fator externo ao jogo, sobre as decisões dos árbitros de basquete ao analisar vídeos de potenciais faltas cometidas (SORS et al., 2019). Além disso, Tabassum et al. (2021) investigaram o impacto psicológico do local de jogo sobre o fenômeno de VC ao evidenciarem que o local de jogo afeta os níveis de cortisol e de índices de ansiedade, o que influencia o desempenho dos jogadores no campo adversário durante a partida.

Evidências de VC em esportes de combate podem ser observadas a partir do trabalho de Franchini e Takito (2016). Este analisou o desempenho de países anfitriões em diferentes esportes de combate (boxe, esgrima, judô, taekwondo e luta livre) durante os Jogos Olímpicos de 1996 a 2012 e identificou um significativo efeito de vantagem para os países anfitriões quanto à conquista de medalhas de ouro e prata.

Algumas outras evidências de VC também foram vistas no handebol. Pic (2018) investigou a influência da localização das partidas de handebol no desempenho das equipes ao utilizar uma metodologia observacional, que revelou a existência de padrões comportamentais distintos de cada equipe a depender da localização das partidas (em casa ou fora). Ainda, identificou que as equipes mandantes apresentaram maior probabilidade de vitória se comparadas às equipes visitantes. Outro estudo neste tema é o de Aguilar, García e Romero (2012), que investigou a presença de VC em jogos de handebol a partir de uma amostra de 19.000 partidas da liga espanhola e que revelou uma vantagem significativa para o time mandante: 61% de vitórias em média.

2.3 *Home Advantage* no Futebol

Como afirmado anteriormente, o efeito da vantagem de se jogar em casa (VC) é objeto de estudo importante e recorrente no campo de diversos esportes coletivos, e o futebol tem destaque (INAN, 2018; LEITE, 2017; POLLARD, 1986; POLLARD; POLLARD, 2005). Pela complexidade deste efeito, é difícil determinar um único fator que seja responsável pela sua propagação nos mais diversos países e ligas de futebol ao redor do mundo. Por este motivo, pesquisadores interessados no tema procuram propor e estudar diversos fatores que, conjuntamente, possam explicar este efeito tão curioso: apoio da torcida, efeitos de viagem, familiaridade com as condições locais de jogo, territorialidade, altitude, viés do árbitro, técnica, tática e gestão, bem como a condição fisiológica e psicológica dos jogadores (BROCHERIE et al., 2015).

Um dos primeiros trabalhos a investigar de forma mais profunda a presença do efeito de VC no futebol é o de Pollard (1986). O autor se baseia em resultados das partidas da liga inglesa ocorridas desde o ano de sua fundação até 1984 e mostra que a vantagem em casa é consistente ao longo dos anos. Ainda, explora a diferença deste fenômeno em diferentes divisões da liga inglesa e entre competições mata-mata e de pontos corridos. Anos depois, o mesmo autor revisita o trabalho, mas cobre diferentes continentes e contextos culturais como, por exemplo, países da América do Sul e outros países da Europa. Os dados revelam variações na VC para diferentes regiões. Na Europa, países dos Bálcãs exibem maiores vantagens se comparados aos países do Norte da Europa. Na América do Sul, países andinos mostraram maior destaque. O estudo mostra, a partir de seus modelos, que o tamanho do país, a natureza da liga (profissional ou amadora), e particularidades regionais são fatores que influenciam o fenômeno (POLLARD, 2006).

García et al. (2013) também apoia a literatura ao investigar a vantagem de jogar em casa nas ligas de futebol dos países membros da *Union of European Football Associations* UEFA durante a primeira década do século XXI, calculada a partir dos pontos obtidos dentro e fora de casa. De 2000 a 2010, a vantagem em casa média foi de 55,6%, com uma pequena redução ao longo das temporadas, o que sugere que os times profissionais estão, progressivamente, superando a VC, segundo os autores. Assim como a literatura já mostrava, os autores encontraram diferenças de níveis de vantagem a depender do país escolhido ou o grupo da UEFA no qual os países se situavam. Os grupos de países com rankings mais altos (A e B) mostraram vantagens maiores que grupos de rankings mais baixos (D e E).

É possível encontrar, ao longo da extensa literatura de VC, alguns métodos menos convencionais também. A título de exemplo, Anderson et al. (2012) verificou, a partir de um questionário aplicado, as perspectivas de jogadores, árbitros e torcedores a respeito da VC. Deste estudo, o ambiente proporcionado por se jogar em casa foi o fator escolhido, por todos os grupos questionados, que mais justifica uma performance superior do time da casa. Outra forma de abordar o assunto e que está presente na literatura é a partir de uma análise do ponto de vista estratégico e tático dos clubes para cada partida jogada (CARMICHAEL; THOMAS, 2005; DEWENTER; NAMINI, 2011). O primeiro identifica, a partir de dados da liga alemã de futebol das temporadas de 1982 a 2007, que um aumento na recompensa por uma vitória na partida (alteração da regra de 2 pontos concedidos ao time vitorioso por 3 pontos) induz o time da casa a jogar menos ofensivamente pelo fato de existir um viés a seu favor simplesmente por estar jogando em casa. O segundo, a partir de dados individuais de cada partida do futebol inglês da temporada 1997-1998, conclui que um estilo de jogo mais ofensivo parece ser mais efetivo quando se joga em casa e um estilo mais defensivo parece ser mais efetivo quando se joga como visitante.

Como é um fenômeno complexo, os estudos a respeito da vantagem do time

mandante começaram a se ramificar a fim de investigar mais profundamente cada um dos fatores que, potencialmente, o explicam. Como hipotetizado por Pollard (1986), algumas causas possíveis para o fenômeno seriam apoio da torcida local, fadiga de viagem, viés de arbitragem, familiaridade com condições locais, táticas de jogo e, por fim, fatores psicológicos. As três primeiras possuem presença significativa na literatura a respeito do tema da VC.

2.3.1 Público

Quando se trata do apoio de sua torcida, imagina-se que os jogadores do time da casa possam se sentir mais motivados e apresentem uma melhor performance dentro de campo (LEITE, 2017; POLLARD, 1986). A atmosfera criada pela torcida nas arquibancadas não só pode proporcionar esta melhora de rendimento dos jogadores, como também pode influenciar indiretamente algumas decisões tomadas pelo árbitro a favor do time da casa: marcação de pênaltis e faltas e aplicação de cartões (POLLARD; ARMATAS, 2017; PONZO; SCOPPA, 2016). A partir de dados de cinco das principais ligas de futebol do mundo (alemã, francesa, espanhola, inglesa e italiana), Inan (2020) identifica que o suporte do público e densidade do público nos estádios são variáveis importantes e que contribuem para a VC. A fim de dar prioridade ao efeito da torcida do time da casa sobre a VC, estudos também foram feitos em jogos de clubes rivais que acontecem no mesmo estádio, cenário em que não existe o fator familiaridade com o estádio e fadiga decorrente de viagens, mas grande parte dos ingressos são destinados aos torcedores do time mandante. Todas as estimativas, realizadas a partir de dados das temporadas de 1991 a 2012 da liga italiana de futebol, mostraram que o time da casa, ainda assim, se aproveita de uma considerável vantagem, o que indica o papel relevante da torcida neste resultado (PONZO; SCOPPA, 2016).

2.3.1.1 Covid-19

Apesar de a literatura proporcionar diversos estudos de longa data a respeito da VC, os recentes e relevantes acontecimentos não só no mundo, como no futebol, ofereceram uma boa oportunidade para novos estudos do impacto de fatores sobre a VC. Um destes eventos foi a pandemia do Covid-19, que proporcionou a paralisação de diversos campeonatos pelo mundo. A ausência do público em partidas de futebol por um período tão longo, ao ponto de abranger o curso integral de um campeonato, é um evento valioso e sem precedentes na história do futebol, que pode auxiliar nos estudos que visam identificar os impactos da torcida sobre a VC.

Como representou um evento raro, estudiosos aproveitaram para elaborar novos estudos do impacto da ausência de público sobre a VC e os resultados obtidos foram mistos. Bryson et al. (2021), em um estudo amplo, ao analisar dados de 23 ligas profissionais de

17 países da temporada 19/20, verificou que a ausência do público não causou impacto significativo no resultado das partidas. Resultados similares foram obtidos em outros estudos, que englobaram uma amostra menor de ligas e países (ALMEIDA; LEITE, 2021; MATOS et al., 2021; WUNDERLICH et al., 2021). Fischer e Haucap (2020), em contrapartida, verificaram que a redução de VC ocorrida na principal liga alemã (Bundesliga) foi mais bem explicada pela menor taxa de ocupação nos estádios e que os times mais acostumados com a presença densa de sua torcida sofreram mais com este efeito. Em contrapartida, o mesmo estudo, que incluiu as ligas de segunda e terceira divisão alemãs, não identificou a ocupação dos estádios como um fator relevante para os resultados dos jogos nessas divisões inferiores.

Apesar destes resultados, a ausência do público não se mostrou como um fator totalmente irrelevante na dinâmica das partidas. Notou-se que variáveis atreladas à arbitragem sofreram impactos significativos a partir do momento em que os clubes não puderam contar com a presença de sua torcida devido às restrições impostas pela pandemia. Os árbitros, supostamente por não sofrerem pressão da torcida do time da casa, reduziram a diferença no número de penalidades aplicadas ao time da casa em relação ao número de penalidades aplicadas ao time visitante, o que levou a uma distribuição mais igualitária e a uma redução de VC (ENDRICH; GESCHE, 2020; TILP; THALLER, 2020). Resultados semelhantes, mas que não identificaram redução de VC mesmo com a mudança do comportamento da arbitragem também foram obtidos (BRYSON et al., 2021; WUNDERLICH et al., 2021). Dentre essas penalidades investigadas, podemos incluir cartões amarelos e vermelhos, pênaltis, faltas marcadas e impedimentos.

Além do viés de arbitragem, a performance dos times, o domínio sobre a partida e a qualidade dos times também foram fatores que serviram como objetos de estudo a fim de investigar se a ausência do público causada pela pandemia impactou o comportamento dos clubes dentro de campo e reduziu a VC. Correia-Oliveira e Andrade-Souza (2021), ao analisarem as temporadas de 2016-17 a 2019-20, verificam uma forte correlação entre VC e a qualidade dos times para as primeira e segunda divisões das ligas alemã, espanhola, inglesa e italiana e suas descobertas mostram que a VC foi reduzida no período da Covid-19. McCarrick et al. (2021), Wunderlich et al. (2021), a partir de dados de 10 e 15 ligas europeias, respectivamente, identificam uma considerável redução no domínio dos times da casa a partir da ausência de suas torcidas em seus estádios. Contudo, apenas o segundo foi acompanhado também de uma redução na VC.

Inclusive, Leitner et al. (2022), a partir de uma extensa revisão de literatura de 26 estudos sobre a VC no período da Covid-19, identifica que, dentre os artigos analisados, o viés da arbitragem e a motivação/emoções da torcida foram as duas atribuições mais mencionadas para mudanças na VC. Cada uma apresentou uma frequência de citações de 54% (14 estudos).

2.3.2 Arbitragem

A arbitragem e o público costumam aparecer juntas na literatura de VC pelo fato de que, hipoteticamente, as decisões de arbitragem seriam influenciadas pela presença da torcida nos estádios. Estádios mais cheios e com incentivos mais acalorados dos torcedores poderiam fazer com que o árbitro temesse marcar infrações contra o time mandante ou fosse incentivado a marcar infrações contra o time visitante. Apesar de a pandemia ter fornecido uma ótima oportunidade para investigar este comportamento, outros estudos também foram realizados nesta seara em outros momentos.

Dohmen e Sauermann (2015), a partir de uma extensa revisão de literatura a respeito do viés de arbitragem, oferecem uma visão abrangente do fenômeno em esportes profissionais (futebol, beisebol, hóquei no gelo, futebol americano e basquete). Eles destacam os fatores que contribuem para que este viés exista (pressão da torcida; incentivos materiais; características individuais, como idade e experiência; e preferências por discriminação, que toma como base fatores como nacionalidade, etnia ou gênero). Mais especificamente no futebol, os autores identificam essa tendência a favorecer a equipe mandante a partir do tempo adicional concedido ao jogo, que se mostrou sistematicamente maior em jogos onde a equipe da casa está em desvantagem em várias ligas de futebol. Além disso, o viés mais favorável ao time mandante também foi identificado nas decisões de arbitragem relacionadas a gols, penalidades e cartões amarelos e vermelhos.

Em um estudo de Goumas (2012), voltado exclusivamente para o futebol, o viés de arbitragem foi investigado em jogos da *Champions League* e da *Europa League* ocorridos nas temporadas de 2009/2010 e 2010/2011. Os resultados revelaram que as equipes visitantes recebem significativamente mais cartões amarelos que as equipes mandantes, com um viés maior na *Champions League*. Em seus modelos de regressão Poisson, o autor mostrou que árbitros puniram as equipes visitantes com cartões amarelos 25% a mais nesta competição e 10% a mais na *Europa League*. Além disso, também mostrou que a densidade do público presente nos estádios mostrou um impacto substancial sobre o comportamento da arbitragem, mesmo quando controlada para outros fatores do jogo.

2.3.2.1 Árbitro de Vídeo

A introdução de novas tecnologias nas partidas em anos recentes, como o Árbitro de Vídeo (VAR), também pode influenciar a dinâmica dos resultados e o comportamento dos árbitros dentro de campo, fatores que podem estar intimamente ligados com o efeito da VC. De acordo com o protocolo da tecnologia, consultado no endereço eletrônico do *International Football Association Board* (IFAB), o VAR é utilizado para revisar quatro situações de jogo: aplicações diretas de cartões vermelhos, decisões de pênalti, validações de gols e em casos de falha na identificação, pelo árbitro, de um jogador infrator (HAN et al., 2020).

Pela influência que a introdução do VAR pode ter sobre a dinâmica das partidas e sobre a VC, este evento também serviu como objeto de estudo para diversos pesquisadores da área. As metodologias adotadas para abordar o tema são semelhantes, em que se realiza uma comparação entre temporadas anteriores e temporadas posteriores à adoção do VAR a partir de testes não paramétricos. Em vez de temporadas, Kubayi, Larkin e Toriola (2021) trabalham as edições de 2014 (sem VAR) e 2018 (com VAR) da Copa do Mundo. As variáveis mais comuns encontradas nos estudos são os números de faltas, gols, impedimentos, pênaltis, cartões amarelos e vermelhos e o tempo de jogo total, assim como no primeiro e no segundo tempo. Adicionalmente, adequa-se cada uma dessas variáveis a um tipo de distribuição a fim de estimar, parametricamente, um Modelo Linear Generalizado (GLM) (LAGO-PEÑAS; REY; KALÉN, 2019; LAGO-PEÑAS; SÁNCHEZ-GÓMEZ; POLLARD, 2020; HAN et al., 2020) Os resultados são mistos, mas parece haver um consenso entre os trabalhos quanto à redução do número de impedimentos e ao aumento dos minutos jogados no primeiro tempo e segundo tempo e do tempo total de jogo.

Spitz et al. (2020) ainda traz uma abordagem diferente, em que coletou dados de partidas de 13 diferentes associações de futebol e incluiu, como benchmark, a opinião de membros do comitê nacional de arbitragem de cada federação sobre as checagens do VAR ocorridas nas 2195 partidas em análise. Segundo o modelo estimado, o nível de precisão das decisões aumentou de 92,1% para 98,3% após a introdução da tecnologia.

2.3.3 Distância Percorrida pelo Visitante

A distância também é um fator presente na literatura como um dos motivadores do fenômeno de VC. Em ligas com grandes distâncias geográficas a serem percorridas, que é o caso do Brasil, o cansaço físico e mental resultante de longas viagens pode afetar o desempenho físico e mental dos jogadores, o que oferece uma vantagem adicional para a equipe da casa.

Gilbert, Dixon e Loughead (2020) investigam como os fatores de viagem afetam o desempenho das equipes visitantes na *Major League Soccer* (MLS) a partir dos dados das temporadas de 2015 a 2017. O estudo mostra o impacto negativo de voos com conexão e viagens na direção leste sobre os resultados dos jogos para as equipes visitantes. Voos comerciais com conexão, que representam 4,7% das viagens contabilizadas na base de dados dos autores, aumentam significativamente a probabilidade de a equipe visitante perder. Além disso, viajar para o leste também afeta negativamente os times visitantes devido ao fenômeno do jetlag, que representa um distúrbio temporário do sono que ocorre quando uma pessoa viaja para uma zona horária diferente.

Damme e Baert (2019) investigaram, a partir de 2012 jogos da *UEFA Champions League* e *UEFA Europa League*, como várias dimensões de distância entre os times mandante e visitante afetam o fenômeno de VC: distâncias geográficas, climáticas, culturais e

disparidades de prosperidade econômica. Em relação à distância geográfica percorrida pelo time visitante, os autores não encontraram uma relação significativa com uma vantagem ou desvantagem de se jogar em casa. Em contrapartida, os times mandantes de seus jogos em locais com níveis de altitude maiores mostraram uma probabilidade maior de vencer os seus jogos.

Goumas (2014) também explora a influência das viagens sobre o desempenho das equipes mandantes nos jogos das temporadas 2005/06 a 2011/12 da Liga Australiana de Futebol. Em seu trabalho, a vantagem de jogar em casa, medida tanto em termos de pontos quanto de gols, está significativamente associada à distância percorrida e ao número de fusos horários cruzados pelos times visitantes.

Esses resultados destacam a influência significativa da logística de viagem e dos desafios físicos associados a longas distâncias no desempenho das equipes visitantes, como adaptação ao fuso horário, condições climáticas adversas e fadiga. Entender e gerenciar esses fatores pode ser crucial para o sucesso das equipes visitantes.

2.3.4 *Home Advantage* no Brasil

A literatura sobre o efeito de VC também já investigou o caso brasileiro em alguns trabalhos. Almeida, Oliveira e Silva (2011) identificaram, a partir do método de aproveitamento percentual de pontos proposto por Pollard em 1986, que a Série B do Campeonato Brasileiro apresentou maior VC em relação à Série A nas temporadas de 2003 a 2009, o que pode ser causado, segundo os autores, pela diferença no tamanho dos estádios e condições no campo de jogo, além dos transtornos advindos das distâncias percorridas, fatores que apresentam condições mais precárias na segunda divisão brasileira. No período da pandemia do Covid-19, Ribeiro et al. (2022) também analisam a VC nas séries A e B do Campeonato Brasileiro (2018 a 2020), principalmente a partir de testes de média entre as diferentes temporadas e divisões, e identificam uma redução para o primeiro caso e nenhuma alteração para o segundo caso. Uma explicação plausível pode ser o menor nível de competitividade e o menor comparecimento do público aos jogos da liga inferior.

A diferença de qualidade dos clubes e a região do país a que pertencem também se mostraram como fatores de possível explicação da VC (SILVA; MEDEIROS; SILVA, 2011; POLLARD; SILVA; MEDEIROS, 2008). O primeiro trabalho, que utilizou como base as partidas jogadas na Série A entre os anos 1998 e 2007, classificou os clubes em três categorias de qualidade (alta, média e baixa) a partir da pontuação obtida em cada temporada e verificou que o efeito da vantagem de jogar em casa se mostrou menos pronunciado no último grupo se comparado aos outros dois. O segundo grupo de autores, que trabalhou com os jogos das temporadas da Série A de 2003 a 2007, realizou uma comparação individual de VC entre os times da amostra a partir do percentual de pontos conquistados por cada time quando atuaram como mandantes e realizou uma análise de

regressão logística ordinal para quantificar o efeito da distância viajada pelo time visitante no resultado do jogo. Os resultados mostraram que a vantagem em casa para as equipes das regiões Norte, Nordeste e Sul foi maior do que para os times da região central. Além disso, a distância viajada pelo time visitante teve um efeito pequeno, mas significativo no resultado do jogo, que correspondeu a 0,12 gol a mais para o time mandante para cada 1000 km viajados pela equipe visitante.

Em contrapartida, a presença do público demonstrou exercer impacto nulo ou pouco significativo sobre o comportamento dos jogadores e sobre a VC em outros estudos (BRAGA; GUILLEN, 2012; BELCHIOR, 2020). No primeiro trabalho, os autores utilizam uma abordagem de variável instrumental para contornar um possível problema de causalidade reversa entre o desempenho do time mandante e a presença de público nos estádios. Esta variável foi gerada a partir de uma campanha de propaganda realizada na Série A do Brasileirão e que tornou os ingressos 50% mais baratos em algumas partidas a fim de fomentar uma maior presença do público. O segundo estudo também utilizou uma abordagem de variável instrumental, que representava o horário alternativo de jogos designado pela Confederação Brasileira de Futebol (CBF) em 2015, que consistia em jogos realizados às 11h da manhã dos domingos. Essa medida aumentou significativamente a presença do público nesses jogos e, como foi designada de forma aleatória pela CBF, isso permitiu o uso de uma variável instrumental que isolasse o efeito exógeno da presença de público sobre a VC. Apesar de o aumento observado da presença do público ter sido em torno de 46% para estas partidas, a análise não encontrou evidências de que o aumento tenha alterado a VC.

É notável que os estudos sobre a VC no futebol brasileiro não são numerosos, principalmente quanto a análises de dados individuais das partidas referentes à conduta da arbitragem e da performance dos times. Os acontecimentos recentes no futebol nacional, como campeonatos inteiros sendo jogados com portões fechados devido à pandemia, e a introdução do VAR, com o objetivo de minimizar os erros de arbitragem, são eventos que motivam e proporcionam oportunidades valiosas de estudo da VC no Brasil.

3 Dados e Estatísticas Descritivas

3.1 Dados

Os dados utilizados neste estudo foram reunidos a partir de diversas fontes e formas de coleta. Ela se propôs a cobrir as temporadas de 2012 a 2022 das Séries A e B do Campeonato Brasileiro de Futebol (Brasileirão). A base de dados possui 8335 observações ao todo, em que cada observação representa uma partida. Importante mencionar que, por ausência de alguns tipos de dados, principalmente para partidas mais antigas, nem todas as observações da base possuem valores para todas as variáveis exploradas.

Dados gerais referentes à partida, como número da rodada, time mandante, time visitante, gols do time mandante (primeiro e segundo tempo), gols do time visitante (primeiro e segundo tempo), estádio, público presente, público máximo, árbitro e colocações do time mandante e visitante no momento da partida foram coletados a partir do uso de um *software* de *webscrapping* denominado *WebHarvy*, em que é possível configurar um robô para acessar as páginas *web* e capturar os dados presentes em posições padronizadas da página. O endereço eletrônico utilizado como referência para a captura dos dados expostos acima foi o *Transfermarkt* (<https://www.transfermarkt.com.br>), referência de dados em vários trabalhos da literatura a respeito do tema. Adicionalmente, a partir dos dados de público presente na partida e de lotação máxima do estádio, obteve-se, em porcentagem, a ocupação do estádio para a partida em questão, que representa uma medida relativa (não absoluta) de presença de público.

Em relação às estatísticas de cada partida, foram coletados posses de bola do time mandante e visitante (primeiro e segundo tempo), chutes do time mandante e chutes do time visitante (primeiro e segundo tempo), escanteios concedidos ao time mandante ou ao time visitante (primeiro e segundo tempo), impedimentos marcados contra o time mandante ou o time visitante (primeiro e segundo tempo), faltas cometidas pelo time mandante ou pelo time visitante, além dos cartões amarelos e vermelhos aplicados ao time mandante ou ao time visitante (primeiro e segundo tempo). Essas estatísticas individuais de cada partida foram coletadas manualmente a partir do endereço eletrônico *Sofascore* (<https://www.sofascore.com>), que possui uma ampla base de dados dos Campeonatos Brasileiros Séries A e B.

O número de pênaltis marcados dentro de cada partida foi obtido a partir do uso de um pacote do *software* estatístico *R Studio* denominado *worldfootballR*, que coletou dados do endereço eletrônico do *FBRef* (<https://fbref.com/pt>). A base de dados originada a partir deste método apontava o tipo de evento (um pênalti, por exemplo) que havia

ocorrido para cada partida do campeonato de determinado ano e, portanto, ainda precisou ser tratada a fim de descobrir a quantidade de vezes que esse tipo de evento ocorreu. Este tratamento foi feito a partir do *R Studio*.

Por fim, coletou-se os dados de longitude e latitude dos times mandante e visitante manualmente via *Google Maps* a fim de se descobrir a distância de deslocamento que a equipe visitante precisou realizar para enfrentar a equipe mandante. A premissa utilizada para a escolha da localização atribuída a cada equipe foi a localização do estádio em que ela mais jogou em jogos que atuou como time mandante. A partir de um pacote de georreferência do *R Studio*, denominado *geosphere*, calculou-se a distância a partir da fórmula de *haversine*, que fornece a distância entre dois pontos de uma esfera.

3.2 Estatísticas Descritivas

Com o intuito de identificar evidências de VC e, a exemplo de como acontece em alguns trabalhos da literatura, as estatísticas extraídas das partidas foram consolidadas em uma tabela, que apresenta a média de cada uma delas relativas ao time mandante e ao time visitante, além da diferença entre os dois tipos de médias. Por fim, a fim de verificar se esta diferença é estatisticamente significativa na amostra coletada, realizou-se um teste de diferença entre médias (teste de t).

Como é possível observar na tabela 1, as diferenças entre as médias das variáveis que indicam o domínio ou imposição sobre a partida, como gols, pontos conquistados, posse de bola, chutes, escanteios e impedimentos, se mostraram positivas e estatisticamente significantes, o que mostra uma tendência maior de o time mandante dominar as ações ofensivas em relação ao time visitante. A tabela também mostra que essa vantagem existe tanto no primeiro tempo quanto no segundo tempo, com aparente redução na segunda etapa, em que é possível notar uma diferença menor entre as variáveis, com exceção do número de gols marcados.

Em relação às variáveis ligadas à arbitragem, como a aplicação de cartões (amarelos e vermelhos) e a marcação de pênaltis, os dados também mostram vantagens para o time que joga em casa: são penalizados, em média, com menos cartões amarelos e menos cartões vermelhos, independentemente da etapa do jogo, e recebem, em média, um número maior de pênaltis por partida em relação ao seu adversário. Este diferencial de cartões existe mesmo com a ausência de uma diferença significativa entre o número de faltas cometidas por uma ou outra equipe.

Apesar de os dados de todas as partidas, conjuntamente, sugerirem a existência de vantagem da equipe mandante em relação à equipe visitante, optou-se por explorar como essas diferenças se comportam quando comparamos diferentes divisões do Campeonato Brasileiro. Diante das diferenças entre as equipes que compõem cada competição, as

Tabela 1 – Estatísticas Mandante x Visitante - Dados Totais

	Média		Diferença	Testes de t (p-valor)
	Mandante	Visitante		
Gols	1.386	0.949	0.437	0.000***
Gols 1T	0.625	0.415	0.210	0.000***
Gols 2T	0.762	0.534	0.228	0.000***
Pontos	1.716	1.009	0.707	0.000***
Posse de Bola	51.80%	48.20%	0.036	0.000***
Posse de Bola 1T	52.60%	47.40%	0.052	0.000***
Posse de Bola 2T	50.90%	49.10%	0.018	0.000***
Chutes	14.265	10.902	3.363	0.000***
Chutes 1T	6.926	5.123	1.803	0.000***
Chutes 2T	7.744	6.083	1.661	0.000***
Escanteios	5.910	4.502	1.408	0.000***
Escanteios 1T	2.860	2.076	0.784	0.000***
Escanteios 2T	3.074	2.418	0.656	0.000***
Impedimentos	1.821	1.654	0.167	0.000***
Impedimentos 1T	0.880	0.798	0.082	0.000***
Impedimentos 2T	0.905	0.825	0.080	0.000***
Faltas Cometidas	15.581	15.530	0.051	0.567
Cartões Amarelos	2.257	2.564	-0.307	0.000***
Cartões Amarelos 1T	0.774	0.933	-0.159	0.000***
Cartões Amarelos 2T	1.507	1.642	-0.135	0.000***
Cartões Vermelhos	0.108	0.147	-0.039	0.000***
Cartões Vermelhos 1T	0.019	0.026	-0.007	0.017***
Cartões Vermelhos 2T	0.090	0.119	-0.029	0.000***
Pênaltis	0.153	0.100	0.053	0.000***

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Elaborado pelo autor.

diferentes condições dos estádios em que as partidas são jogadas, os diferentes árbitros que são elencados para apitar cada partida, os diferentes comportamentos das torcidas, é possível que as diversas variáveis que envolvem uma partida de futebol se comportem de diferentes formas em jogos da Série A e jogos da Série B.

A tabela 2 demonstra algumas dessas diferenças, principalmente ligadas à distribuição geográfica das equipes participantes, ao comportamento das torcidas e, consequentemente, à atmosfera criada dentro dos estádios. A média da distância percorrida pelos times

Tabela 2 – Outras Médias

	Total	Série A	Série B
Distância (km)	972.50	836.41	1,104.30
Público	11,982.86	17,705.83	6,223.15
Público Máx.	35,558.00	43,341.77	27,696.92
Ocupação	32.13%	41.64%	22.53%

Elaborado pelo autor.

visitantes para enfrentar os times mandantes se mostrou maior na Série B que na Série A, o que pode ser explicado por uma distribuição geográfica maior dos times participantes da segunda divisão do Campeonato Brasileiro. Os times participantes da Série A estão mais concentrados nas regiões Sul e Sudeste do país, enquanto na Série B, além dessas duas regiões, a região Nordeste também mostra uma participação relevante, como demonstra a tabela 3.

Tabela 3 – Distribuição Geográfica de Equipes e Partidas

	Série A		Série B	
	Equipes	Partidas	Equipes	Partidas
Norte	0	0	2	113
Nordeste	8	641	15	1,172
Centro-Oeste	3	247	5	454
Sudeste	14	2,164	23	1,270
Sul	11	1,118	15	1,156
Total	36	4,170	60	4,165

Elaborado pelo autor.

Além disso, a tabela 2 evidencia também um comportamento diferente em relação às torcidas nos estádios. Dados da Série A mostram que o público médio presente nas partidas representa quase o triplo do público médio em partidas da Série B. Mesmo que os estádios da Série A apresentem capacidade de receber um maior número de torcedores, o nível de ocupação das partidas da primeira divisão do futebol brasileiro representa, em média, quase o dobro do nível de ocupação das partidas na Série B. Este resultado pode ser justificado pelo fato de partidas da segunda divisão apresentarem um nível técnico inferior e menor apelo por parte dos torcedores.

Apesar das diferenças evidenciadas acima, não foi possível perceber um comportamento muito diferente em relação à proporção dos resultados das partidas do time mandante, como apresenta a tabela 4. O time que joga em casa tende a vencer, em média, grande parte das suas partidas e, quantitativamente, este resultado é bem próximo para

as duas divisões do Campeonato Brasileiro.

Tabela 4 – Proporção de Resultados Time Mandante

	Total	Série A	Série B
Vitória	48.04%	48.62%	47.49%
Empate	27.51%	27.05%	27.92%
Derrota	24.45%	24.33%	24.59%

Elaborado pelo autor.

Tabela 5 – Estatísticas Mandante x Visitante - Séries A e B

	Série A				Série B			
	Média		Diferença	Testes de t (p-valor)	Média		Diferença	Testes de t (p-valor)
	Mandante	Visitante			Mandante	Visitante		
Gols	1.412	0.946	0.466	0.000***	1.360	0.950	0.410	0.000***
Gols 1T	0.629	0.414	0.215	0.000***	0.620	0.415	0.205	0.000***
Gols 2T	0.783	0.532	0.251	0.000***	0.741	0.534	0.207	0.000***
Pontos	1.729	1.000	0.729	0.000***	1.704	1.017	0.687	0.000***
Posse de Bola	51.60%	48.40%	0.032	0.000***	52.10%	47.90%	0.042	0.000***
Posse de Bola 1T	52.20%	47.80%	0.044	0.000***	53.10%	46.90%	0.062	0.000***
Posse de Bola 2T	50.80%	49.20%	0.016	0.000***	51.20%	48.80%	0.024	0.000***
Chutes	13.908	10.658	3.250	0.000***	14.816	11.271	3.545	0.000***
Chutes 1T	6.854	5.086	1.768	0.000***	7.023	5.172	1.851	0.000***
Chutes 2T	7.709	6.068	1.641	0.000***	7.793	6.099	1.694	0.000***
Escanteios	5.902	4.528	1.374	0.000***	5.926	4.462	1.464	0.000***
Escanteios 1T	2.871	2.095	0.776	0.000***	2.847	2.049	0.798	0.000***
Escanteios 2T	3.072	2.421	0.651	0.000***	3.078	2.414	0.664	0.000***
Impedimentos	1.830	1.662	0.168	0.000***	1.806	1.643	0.163	0.000***
Impedimentos 1T	0.861	0.778	0.083	0.001***	0.902	0.821	0.081	0.007***
Impedimentos 2T	0.905	0.828	0.077	0.004***	0.905	0.822	0.083	0.005***
Faltas Cometidas	14.993	15.065	-0.072	0.525	16.439	16.217	0.222	0.118
Cartões Amarelos	2.231	2.537	-0.306	0.000***	2.293	2.602	-0.309	0.000***
Cartões Amarelos 1T	0.768	0.928	-0.160	0.000***	0.779	0.937	-0.158	0.000***
Cartões Amarelos 2T	1.502	1.618	-0.116	0.000***	1.514	1.665	-0.151	0.000***
Cartões Vermelhos	0.103	0.151	-0.048	0.000***	0.114	0.142	-0.028	0.004***
Cartões Vermelhos 1T	0.021	0.027	-0.006	0.145	0.018	0.026	-0.008	0.053*
Cartões Vermelhos 2T	0.083	0.122	-0.039	0.000***	0.096	0.116	-0.020	0.021**
Pênaltis	0.159	0.104	0.055	0.000***	0.144	0.094	0.050	0.000***

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Elaborado pelo autor.

Em relação ao comportamento das variáveis de uma partida individualizado para as Séries A e B (Tabela 5), é possível perceber que, apesar de ocorrer em diferentes níveis para cada divisão, o time mandante continua demonstrando maior imposição, em média, sobre o time visitante. Dados da Série A apresentaram um maior diferencial de gols, conquista de pontos e marcação de impedimentos, em média. Dados da Série B, por outro lado, apresentaram um maior diferencial de posse de bola, chutes e escanteios.

Diferentemente da tabela 1, que agrega os dados, o diferencial de faltas cometidas na Série B, equivalente a 0,222, em média, quase apresentou relevância estatística ao nível

de significância de 10%. Na Série A, o diferencial de faltas manteve o comportamento da tabela anterior, sem qualquer relevância estatística. Mesmo com a possibilidade de times mandantes cometerem um maior número de faltas em comparação aos times visitantes, o comportamento do diferencial de cartões ainda se manteve favorável aos times mandantes, ou seja, estes foram menos penalizados se comparados aos seus adversários em ambas as divisões. Ainda, o diferencial de cartões vermelhos se mostrou maior na primeira divisão que na segunda divisão.

Tabela 6 – Diferencial Mandante x Visitante - Covid-19

	Covid		Testes de t (p-valor)
	Não	Sim	
Gols	0.460	0.308	0.001***
Gols 1T	0.213	0.189	0.428
Gols 2T	0.248	0.118	0.000***
Pontos	0.743	0.501	0.002***
Posse de Bola	3.70%	3.00%	0.238
Posse de Bola 1T	5.50%	4.20%	0.049**
Posse de Bola 2T	1.90%	1.80%	0.974
Chutes	3.613	2.461	0.000***
Chutes 1T	1.958	1.310	0.000***
Chutes 2T	1.823	1.151	0.000***
Escanteios	1.573	0.811	0.000***
Escanteios 1T	0.878	0.489	0.000***
Escanteios 2T	0.762	0.322	0.000***
Impedimentos	0.191	0.077	0.097*
Impedimentos 1T	0.093	0.046	0.309
Impedimentos 2T	0.096	0.031	0.178
Faltas Cometidas	-0.104	0.594	0.001***
Cartões Amarelos	-0.365	-0.030	0.000***
Cartões Amarelos 1T	-0.191	-0.036	0.000***
Cartões Amarelos 2T	-0.172	0.007	0.001***
Cartões Vermelhos	-0.046	-0.010	0.016**
Cartões Vermelhos 1T	-0.006	-0.010	0.559
Cartões Vermelhos 2T	-0.037	0.001	0.005**
Pênaltis	0.068	-0.001	0.000***

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Elaborado pelo autor.

Como mencionado anteriormente, a pandemia do Covid-19 gerou um ambiente de rara presença no futebol: partidas com ausência total do público. Naturalmente, é possível imaginar que esta atmosfera diferente nas partidas jogadas mude também o comportamento das equipes. Por este motivo, criou-se uma tabela parecida com aquelas apresentadas anteriormente, em que se analisa o diferencial de cada variável entre equipes mandantes e visitantes. Contudo, ao contrário das tabelas anteriores, em que se comparou diretamente as médias dos times mandantes e times visitantes, na tabela 6, comparou-se as médias dos diferenciais entre mandantes e visitantes para partidas que não ocorreram durante o período do Covid-19 e partidas que foram jogadas durante a pandemia (aquelas ocorridas desde o início de 2020 até meados de outubro de 2021). Além disso, a tabela considera toda a base de dados e não distingue os números para cada divisão do campeonato.

Os resultados obtidos mostram que, aparentemente, os times mandantes impuseram menos controle sobre a partida comparado ao que se observava antes. Isto fica evidente com uma diminuição do diferencial das variáveis ligadas ao comportamento ofensivo das equipes, como gols, chutes, escanteios e impedimentos, além do diferencial do número de pontos conquistados que, em média, também diminuiu. É possível observar um impacto negativo de 30% a 60% sobre essas variáveis, seja por uma redução do ímpeto ofensivo da equipe mandante ou por um aumento deste ímpeto por conta da equipe visitante.

É notável a mudança nas variáveis ligadas à arbitragem também: uma inversão do sinal de negativo para positivo do diferencial de faltas cometidas, acompanhada do aumento considerável dos diferenciais de cartões amarelos e vermelhos. Em relação aos cartões amarelos, os jogos ocorridos durante a pandemia apresentaram, em média, um diferencial mais de dez vezes menor, em módulo, que em jogos em outros períodos, o que trouxe o indicador quase a zero. O comportamento em relação aos cartões vermelhos foi parecido, com redução próxima dos 80%, em módulo. Como esse comportamento é acompanhado do aumento no diferencial de faltas cometidas, seja pelo aumento das faltas cometidas pelo time mandante ou redução de faltas cometidas pelo time visitante, é possível que as mudanças nos números das penalidades (cartões) aplicadas a cada time sejam explicadas parcialmente por isso.

Ademais, o período da pandemia parece ter afetado também a dinâmica dos pênaltis marcados a favor de uma ou outra equipe, já que o diferencial desta variável se tornou quase zero durante este período, mais um indicador de redução da vantagem antes percebida para o time mandante.

Além da pandemia do Covid-19, outro evento com potencial de alterar a dinâmica e o comportamento de algumas métricas que envolvem uma partida de futebol também ocorreu durante o período estudado: a introdução do VAR ao Campeonato Brasileiro. Da mesma forma que foi feito para o caso particular da pandemia, também se analisou o comportamento da média dos diferenciais ao se comparar os dados de partidas jogadas

sem a presença e com a presença do VAR. Importante pontuar que, na Série A, o VAR foi implementado desde o início do campeonato de 2019 e, na Série B, a implementação ocorreu de forma mais tardia, a partir da vigésima rodada do campeonato de 2021. Diferentemente do caso anterior, entretanto, optou-se por direcionar a análise a variáveis relacionadas aos resultados das partidas (gols e pontos conquistados) e, principalmente, às variáveis de arbitragem (impedimentos, faltas, cartões e pênaltis).

Na tabela 7, é possível observar diferenças significativas em poucas variáveis, com maior destaque para os diferenciais de faltas cometidas e de cartões amarelos, que aumentaram. Contudo, algumas variáveis que são afetadas diretamente por consultas ao VAR, como a aplicação de cartões vermelhos e a marcação de pênaltis, não apresentaram mudanças estatisticamente significantes.

Tabela 7 – Diferencial Mandante x Visitante - VAR

	VAR		Testes de t (p-valor)
	Não	Sim	
Gols	0.445	0.417	0.456
Gols 1T	0.202	0.232	0.240
Gols 2T	0.243	0.186	0.044**
Pontos	0.726	0.653	0.235
Impedimentos	0.174	0.153	0.734
Impedimentos 1T	0.089	0.071	0.645
Impedimentos 2T	0.079	0.082	0.927
Faltas Cometidas	-0.165	0.404	0.001***
Cartões Amarelos	-0.377	-0.138	0.000***
Cartões Amarelos 1T	-0.203	-0.080	0.000***
Cartões Amarelos 2T	-0.178	-0.058	0.006***
Cartões Vermelhos	-0.043	-0.031	0.337
Cartões Vermelhos 1T	-0.008	-0.004	0.459
Cartões Vermelhos 2T	-0.031	-0.026	0.690
Pênaltis	0.054	0.047	0.639

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Elaborado pelo autor.

A partir dos resultados apresentados nas tabelas desta seção, é possível observar que, apesar de suas diferenças na distribuição geográfica das equipes e na assiduidade do público, evidências de VC existem em ambas as divisões do Campeonato Brasileiro. A pandemia do Covid-19 impactou, reduzindo a VC, não só o rendimento das equipes mandantes, que parecem ter diminuído o ímpeto ofensivo, como também o comportamento

da arbitragem, que penalizou relativamente mais as equipes da casa. Como era de se esperar, a implementação do VAR trouxe impacto sobre alguns comportamentos da arbitragem, como a marcação de faltas e, conseqüentemente, a aplicação de cartões amarelos. Contudo, não parece ter sido determinante para os resultados das partidas.

4 Metodologia

4.1 Método Ajustado de Pollard

Há diversas formas de se quantificar a VC e as mais presentes na literatura são aquelas que utilizam o número de pontos conquistados e o número de gols marcados. Para isto, um dos métodos que será utilizado para identificar indícios de VC será uma versão redimensionada daquele apresentado por Pollard (1986), que visa contornar algumas críticas sofridas pela versão inicial, em que o valor de referência para ausência de VC era de 50%. Em sua nova versão, o novo valor de referência passa a ser de 0% e a vantagem e desvantagem de jogar em casa passam a ser representadas, respectivamente, por valores estritamente positivos e negativos. Este modelo redimensionado, apresentado por Matos, Amaro e Pollard (2020), segue a seguinte fórmula:

$$VC = \left[\frac{(Casa - Fora)}{(Casa + Fora)} \right] \times 100\% \quad (4.1)$$

em que *Casa* indica as variáveis calculadas para partidas jogadas em casa e *Fora* indica as variáveis calculadas para partidas jogadas fora de casa. Com o objetivo de complementar a análise anterior, calculou-se o valor de VC para cada ano, de 2012 a 2022, a fim de que se possa ter uma percepção de seu comportamento ao longo do tempo.

4.2 Modelo Logístico Binário

Para analisar o impacto de algumas variáveis sobre os resultados das partidas, primeiramente optou-se por utilizar um modelo *logit* binário, que é utilizado em pesquisas para estimar a probabilidade de ocorrência e um evento binário, representado por uma variável dependente Y , que assume valores 0 ou 1, indicando a ausência ou presença do evento de interesse, respectivamente. O modelo *logit* binário pode ser expresso pela seguinte função de ligação logística:

$$\ln \left(\frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n, \quad (4.2)$$

em que:

- p é a probabilidade de $Y = 1$, ou seja, a probabilidade de ocorrência do evento de interesse;

- $\frac{p}{1-p}$ é a razão de chances (*odds ratio*) de p , representando a chance de ocorrência do evento em comparação à chance de não ocorrência;
- β_0 é o intercepto, que representa o logaritmo das chances de $Y = 1$ quando todas as variáveis independentes (X_i) são iguais a zero;
- $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes preditores, que medem a mudança no logaritmo das chances de $Y = 1$ para uma unidade de mudança na variável preditora correspondente, mantendo todas as outras variáveis constantes.

Os coeficientes (β_i) do modelo *logit* são interpretados em termos de logaritmo das chances. Um coeficiente positivo ($\beta_i > 0$) indica que um aumento na variável preditora está associado ao aumento do logaritmo das chances de ocorrência do evento e o contrário para um coeficiente negativo.

A modelagem escolhida segue o formato a seguir:

$$Y_i = \beta_0 + \beta_1 \times \text{ocupação}_i + \beta_2 \times \text{distância100}_i + \beta_3 \times \text{covid}_i + \beta_4 \times \text{rankdif}_i + \varepsilon_i, \quad (4.3)$$

em que Y assume valor 1 para vitória do mandante, que é o evento de interesse, e 0 para empate ou derrota, que representam ausência do evento de interesse. β_0 é o intercepto e *ocupação* representa o nível de ocupação do estádio em que a partida foi jogada, ou seja, a razão entre o número de torcedores presentes e o número máximo de torcedores suportados pelo estádio. Ademais, *distância100* representa a distância, a cada 100 Km, entre o estádio no qual o time visitante mais joga os seus jogos como mandante e o estádio no qual a partida observada foi jogada, com o intuito de simular a distância percorrida pelo time visitante para realizar o jogo contra o time mandante. Ainda, temos uma variável *dummy* (*covid*), que assume valor igual a um para partidas jogadas durante o período da pandemia e sem público e zero caso contrário, e uma variável de controle de força relativa entre os times, representada pela diferença entre as colocações dos times mandante e visitante no campeonato quando a partida é jogada (*rankdif*), ou seja, uma diferença com valor negativo indica que o time mandante está mais bem colocado que o time visitante e vice-versa. Por fim, temos o termo de erro, representado por ε . O subscrito i indica que as variáveis são consideradas para cada uma das observações (partidas).

Após a estimação do modelo *logit* binário, a interpretação dos coeficientes β fornece uma visão inicial sobre a direção do relacionamento entre as variáveis preditoras e a probabilidade de vitória do mandante. Contudo, para uma compreensão completa do impacto dessas variáveis, calculou-se também os efeitos marginais médios (EMMs) de cada uma dessas variáveis sobre esta probabilidade. Analisar os EMMs pode trazer maior intuição aos resultados, já que representam a mudança na probabilidade predita da variável

$Y = 1$ (vitória da equipe mandante), em média, associada a uma mudança unitária em determinada variável independente.

4.3 Modelos por Mínimos Quadrados Ordinários

Além do modelo *logit*, que foi utilizado para estimar o impacto sobre a probabilidade de vitória do time mandante, também se utilizou uma regressão por Mínimos Quadrados Ordinários (MQO) para investigar o impacto das mesmas variáveis sobre a diferença entre os gols marcados pelo time mandante e os gols marcados pelo time visitante em uma mesma partida. A modelagem seguiu o mesmo formato da regressão *logit*, com a única diferença de que, neste caso, Y_i representa o diferencial de gols entre as equipes e não mais uma variável binária.

Para analisar o impacto de algumas variáveis sobre decisões de arbitragem, também se optou por utilizar uma regressão por MQO. Primeiramente, ao considerar o diferencial de cartões amarelos e vermelhos como variáveis dependentes, a modelagem assumiu o formato a seguir:

$$Y_i = \beta_0 + \beta_1 \times ocupação_i + \beta_2 \times covid_i + \beta_3 \times VAR_i + \beta_4 \times faltasdif_i + \beta_5 \times (ocupação \times VAR)_i + \text{árbitros}_i + \varepsilon_i, \quad (4.4)$$

em que Y_i representa as variáveis dependentes (cartões amarelos e vermelhos) e $ocupação_i$ e $covid_i$ representam as mesmas variáveis utilizadas no modelo *logit*. VAR é uma novidade, representativa de uma variável *dummy*, que assume valor um para partidas jogadas com a presença do VAR e zero caso contrário. $faltasdif_i$, representada pela diferença entre as faltas cometidas pelo time mandante e pelo time visitante, é uma variável de controle, já que se espera que uma equipe seja punida com mais cartões à medida que cometa um maior número de faltas. Por fim, temos uma interação entre as variáveis $ocupação$ e VAR , a fim de identificar se a introdução do VAR trouxe algum efeito adicional da ocupação nos estádios sobre o diferencial de cartões, um vetor de efeitos fixos (*árbitros*) para cada árbitro, isto é, uma *dummy* para cada um dos 216 árbitros identificados, e um termo de erro ε_i .

Uma segunda modelagem foi escolhida para analisar os impactos sobre impedimentos e pênaltis:

$$Y_i = \beta_0 + \beta_1 \times ocupação_i + \beta_2 \times covid_i + \beta_3 \times VAR_i + \beta_4 \times (ocupação \times VAR)_i + \beta_5 \times (VAR \times covid)_i + \text{árbitros}_i + \varepsilon_i, \quad (4.5)$$

em que Y representa as variáveis dependentes (impedimentos e pênaltis), incluiu-se uma interação entre as variáveis *dummy covid* e *VAR* e retirou-se a variável controle de diferencial de faltas cometidas. Além dessas diferenças, as variáveis independentes adotadas neste modelo seguem as mesmas especificações do modelo anterior. Importante salientar que, a fim de evitar um problema de multicolinearidade perfeita, uma das dummies de árbitros foi retirada das regressões.

5 Resultados e Discussão

5.1 Método Ajustado de Pollard

A tabela 8 apresenta os resultados obtidos ao aplicar o método ajustado de Pollard para o número de gols marcados e o número de pontos conquistados. Importante lembrar que a referência para ausência de VC é o nível de 0%. Em todas as condições analisadas, mesmo em diferentes divisões do Campeonato Brasileiro ou no período da pandemia, o time que joga em casa apresentou vantagem em relação ao time que joga fora de casa.

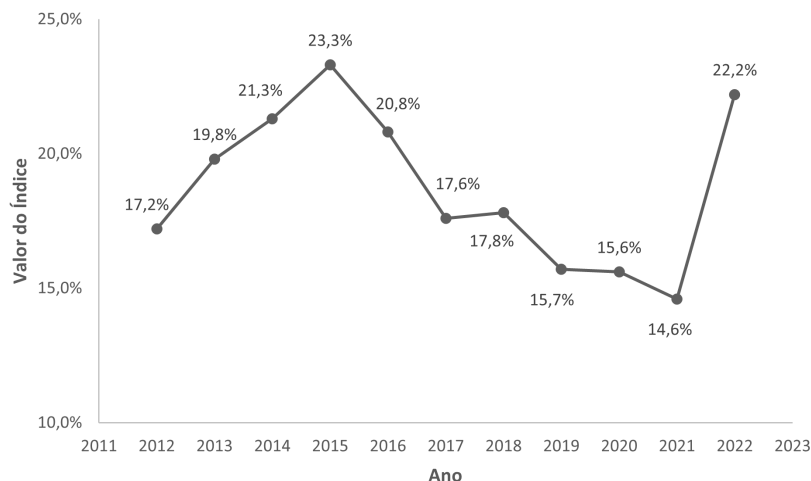
Tabela 8 – Método Ajustado de Pollard

	Total	Série A	Série B	Covid-19
Gols	18.77%	19.74%	17.77%	13.74%
Pontos	25.97%	26.69%	25.25%	18.63%

Elaborado pelo autor.

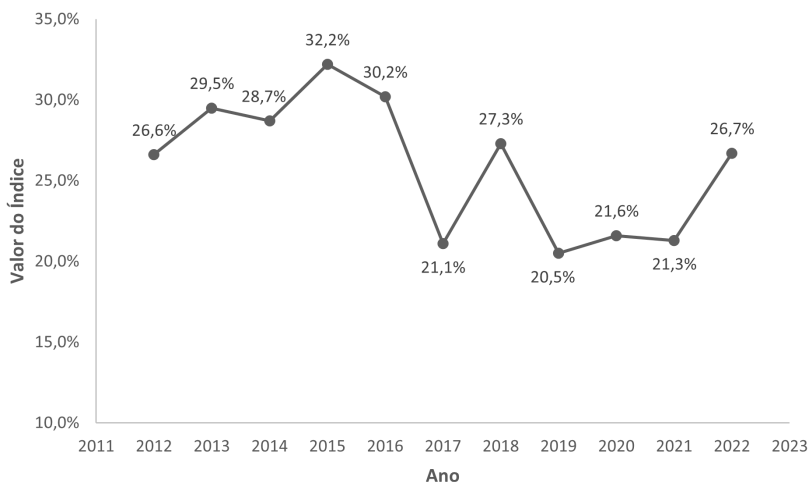
Entre as divisões do campeonato, é possível observar uma ligeira diferença a favor dos times mandantes da Série A em relação aos da Série B tanto para as métricas de gols quanto para pontos. O período da Covid-19, mesmo ainda demonstrando existência de VC, mostrou que o desempenho dos mandantes piorou, refletido pela queda no índice de VC calculado. Para complementar esta análise e trazer uma percepção maior do dinamismo desses índices, calculou-se os valores para cada ano desde 2012 a 2022, representados nos gráficos abaixo.

Figura 1 – Método Ajustado de Pollard (2012/2022) - Gols



Elaborado pelo autor.

Figura 2 – Método Ajustado de Pollard (2012/2022) - Pontos



Elaborado pelo autor.

Como é possível perceber a partir dos resultados para gols marcados, o índice já apresentava, aparentemente desde 2015, uma tendência de queda na VC. Em 2020 e 2021, anos em que houve partidas jogadas com ausência de público devido à pandemia, o índice atingiu os menores valores desta série histórica. Ainda, o índice aumentou de forma significativa no ano seguinte, em que alcançou um nível próximo ao observado em 2015, sete anos antes. Em relação à variável pontos conquistados, observa-se um comportamento mais volátil do índice, apesar de 2020 e 2021 não representarem os anos em que este índice atingiu os seus menores níveis. O período da pandemia mostra sinais de que trouxe dificuldades para as equipes mandantes em relação aos visitantes, mas este não parece ter sido um comportamento exclusivo deste período.

5.2 Resultados de Desempenho da Equipe Mandante

Os resultados das regressões logísticas estão representados na tabela 9, que incluem os valores para a base de dados total e para as partidas da Série A e partidas da Série B. Esta foi uma escolha a fim de entender se alguns fatores específicos influenciam mais os resultados de uma divisão do campeonato que outros. Abaixo dos valores dos coeficientes, os erros-padrão individuais são apresentados entre parênteses. Outros resultados foram obtidos a partir da substituição da variável independente de ocupação pela quantidade absoluta de torcedores presentes e, pelo fato de não terem apresentado diferenças significativas em relação aos primeiros, foram incluídos e podem ser consultados no apêndice.

Primeiramente, regrediu-se cada variável explicativa individualmente (colunas 1, 2 e 3) e, sob uma análise qualitativa dos resultados obtidos, os sinais dos coeficientes de regressão de cada variável estão em conformidade com o que se esperava: uma maior ocupação nos estádios e uma maior distância percorrida pela equipe visitante aumentam

a probabilidade de vitória do time que joga em casa, enquanto partidas jogadas sob as condições do período da pandemia trazem o efeito contrário, representado pelo sinal negativo da variável *dummy* de Covid. Exceto o coeficiente da distância percorrida pelo time visitante na Série A, todos os outros coeficientes apresentaram valores estatisticamente significantes nos diferentes grupos de dados.

Tabela 9 – Resultados Logit

Total	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ocupação	0.484*** (0.089)			0.251*** (0.095)			0.311*** (0.097)	0.165 (0.110)
Distância em 100 km		0.011*** (0.003)			0.010*** (0.003)		0.011*** (0.003)	0.011*** (0.004)
Covid			-0.223*** (0.063)			-0.249*** (0.066)		-0.215*** (0.075)
Diferencial de Colocação				-0.083*** (0.003)	-0.084*** (0.003)	-0.083*** (0.003)	-0.083*** (0.003)	-0.084*** (0.003)
Constante	-0.211*** (0.033)	-0.173*** (0.039)	-0.045* (0.024)	-0.165*** (0.035)	-0.182*** (0.041)	-0.060** (0.025)	-0.281*** (0.051)	-0.208*** (0.057)
Observações	8,299	8,063	8,299	8,299	8,063	8,299	8,063	8,063
Pseudo-R ² (McFadden)	0.002	0.001	0.001	0.080	0.081	0.081	0.082	0.083
Akaike Information Criterion (AIC)	11,466.80	11,161.26	11,483.55	10,580.24	10,266.80	10,573.06	10,258.60	10,252.44
Série A								
Ocupação	0.374*** (0.120)			0.245*** (0.125)			0.293** (0.129)	0.0260 (0.157)
Distância em 100 km		0.001 (0.005)			0.0002 (0.005)		0.002 (0.005)	0.002 (0.005)
Covid			-0.298*** (0.089)			-0.328*** (0.092)		-0.339*** (0.114)
Diferencial de Colocação				-0.062*** (0.004)	-0.063*** (0.004)	-0.063*** (0.004)	-0.063*** (0.004)	-0.063*** (0.004)
Constante	-0.189*** (0.053)	-0.050 (0.053)	-0.013 (0.034)	-0.135** (0.055)	-0.038 (0.055)	-0.001 (0.035)	-0.156** (0.075)	-0.012 (0.090)
Observações	4,155	3,967	4,155	4,155	3,967	4,155	3,967	3,967
Pseudo-R ² (McFadden)	0.002	0.000	0.002	0.047	0.047	0.049	0.048	0.050
Akaike Information Criterion (AIC)	5,751.20	5,501.42	5,749.65	5,489.55	5,243.81	5,480.71	5,240.63	5,233.67
Série B								
Ocupação	0.735*** (0.155)			0.070 (0.169)			0.032 (0.171)	-0.104 (0.187)
Distância em 100 km		0.020*** (0.004)			0.021*** (0.005)		0.021*** (0.005)	0.021*** (0.005)
Covid			-0.148* (0.088)			-0.158* (0.096)		-0.190* (0.105)
Diferencial de Colocação				-0.106*** (0.004)	-0.107*** (0.004)	-0.106*** (0.004)	-0.107*** (0.005)	-0.107*** (0.005)
Constante	-0.241*** (0.043)	-0.320*** (0.058)	-0.078** (0.034)	-0.175*** (0.047)	-0.395*** (0.064)	-0.139*** (0.037)	-0.401*** (0.071)	-0.349*** (0.077)
Observações	4,144	4,096	4,144	4,144	4,096	4,144	4,096	4,096
Pseudo-R ² (McFadden)	0.004	0.004	0.001	0.122	0.126	0.122	0.126	0.127
Akaike Information Criterion (AIC)	5,715.49	5,652.05	5,735.48	5,042.29	4,959.30	5,039.72	4,961.27	4,959.99

Erros-padrões IID entre parênteses

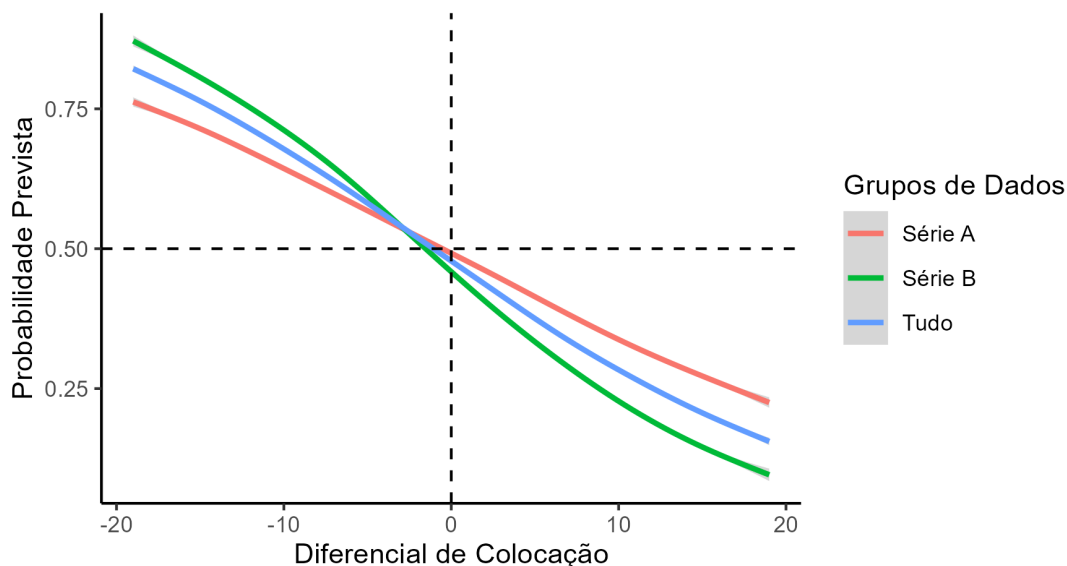
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Elaborado pelo autor.

Depois, regrediu-se cada uma dessas variáveis mais uma vez, com a inclusão de uma variável de controle de força relativa entre os times (colunas 4, 5 e 6), representada pelo diferencial de colocação das equipes na competição no momento da partida, já que é natural pressupor que uma equipe mais bem colocada na tabela se encontra em melhor fase que uma equipe em pior colocação e possui, por este motivo, uma probabilidade maior de vitória. Após a inclusão desta variável de controle, é possível observar um aumento considerável, em todos os casos, da medida de ajuste dos modelos, representada pelo pseudo- R^2 de McFadden. Além disso, o valor do coeficiente da variável de ocupação mostrou ser o mais afetado, em todos os conjuntos de dados, perdendo relevância e, particularmente na Série B, perdendo até mesmo significância estatística. Em relação às outras variáveis (distância e *dummy* de Covid), os respectivos coeficientes não sofreram grandes mudanças.

É possível observar que, ao longo das regressões, o coeficiente da variável de diferencial de colocação mantém a sua magnitude, o que pode ser um indicativo de que está adequada ao modelo proposto. Contudo, essa magnitude varia entre as divisões do campeonato e parece mostrar maior relevância sobre a probabilidade de vitória na Série B que na Série A, ou seja, em confrontos entre equipes bem colocadas e equipes mal colocadas, a diferença de força entre os times é ainda mais evidente na segunda divisão, o que leva a um desequilíbrio maior desta competição.

Figura 3 – Probabilidades Previstas - Diferencial de Colocação



Elaborado pelo autor.

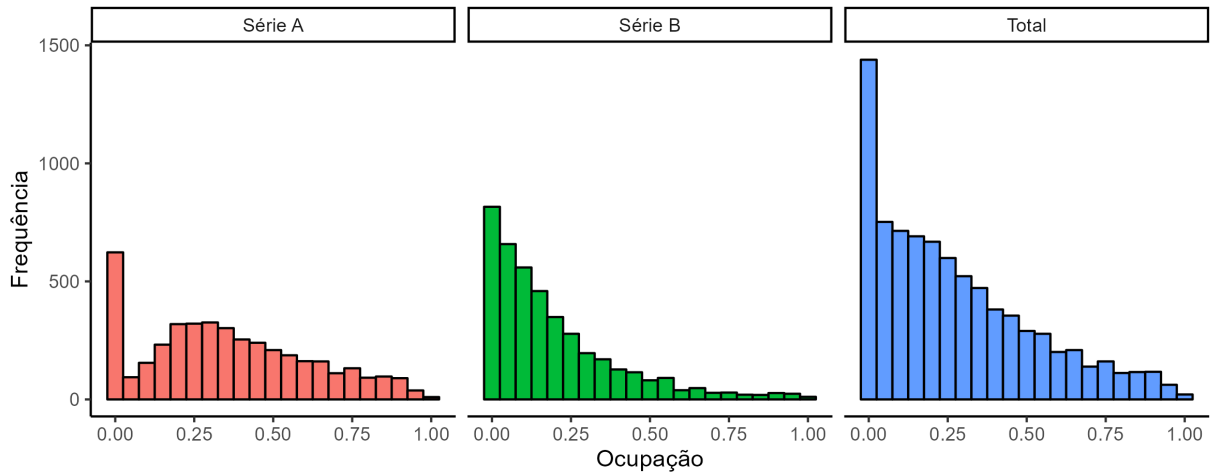
O gráfico acima representa as probabilidades de vitória do time mandante, calculadas para cada observação e organizadas de acordo com o diferencial de colocação presente em cada uma delas. Quando as equipes estão próximas na tabela, a probabilidade prevista de vitória do time mandante se situa perto dos 50%. À medida que se afasta do centro do gráfico e se encontra situações extremas, em que o diferencial de colocação é maior, a

probabilidade de vitória da equipe mandante se mostra mais sensível na Série B.

Quando presentes na mesma regressão (coluna 7), as variáveis de ocupação e distância não parecem interferir uma sobre o coeficiente da outra. O grupo de dados da Série A demonstrou a ocupação como um fator mais determinante para a vitória do time mandante, enquanto o grupo de dados da Série B responsabilizou a distância percorrida pela equipe visitante. Na última regressão (coluna 8), a inclusão da variável *dummy* de Covid é capaz de retirar a significância estatística dos coeficientes de ocupação do grupo de dados total e da Série A, já inexistente na Série B.

Como os jogos que ocorreram durante a pandemia apresentaram ausência total de público nos estádios, uma possível explicação para este resultado é o fato de que há, na amostra coletada, além das partidas com público zerado, uma concentração de jogos que não possui um nível de ocupação alto, como se pode perceber a partir dos histogramas abaixo. Na Série A, há uma concentração de jogos ao redor do nível de ocupação de 25% e observa-se cada vez menos partidas com níveis de ocupação superiores a este. Na Série B, encontrar partidas com níveis de ocupação maiores que 25% torna-se um desafio ainda maior. Importante destacar o peso do número de partidas com ausência total de público em todos os grupos de dados observados.

Figura 4 – Histogramas por Grupo de Dados



Elaborado pelo autor.

As diferentes divisões do Campeonato Brasileiro apresentaram diferentes respostas com relação ao período da pandemia do Covid-19. Assim como identificaram Ribeiro et al. (2022), houve redução de VC da Série A em razão de efeitos do evento, representada pelo coeficiente da variável *dummy* na tabela 9. Contudo, enquanto os mesmos autores identificaram ausência de efeitos na Série B, o presente estudo identificou uma redução do fenômeno na segunda divisão, mas em menor grau se comparada à primeira. Logo, os resultados para ambas as divisões se alinharam com as expectativas de que a ausência de torcida reduziria a vantagem de jogar em casa (INAN, 2020; TILP; THALLER, 2020).

Em contrapartida, os resultados de Matos et al. (2021) parecem discordar, já que não reportaram reduções de VC sob condições similares no campeonato português.

Também existem estudos que investigam o impacto do público sobre a VC no futebol brasileiro fora do período da pandemia, mas que não encontram evidências de efeitos significativos deste fator sobre os resultados das partidas (BRAGA; GUILLEN, 2012; BELCHIOR, 2020). É possível imaginar, portanto, que a vantagem do time mandante responda à ausência total do público, mas não ao número de torcedores presentes, o que pode justificar a perda de significância do coeficiente da variável de ocupação na tabela de resultados em todos os casos testados.

Além disso, é possível que a presença de torcedores seja menos decisiva na Série B que na Série A no futebol brasileiro. Uma lógica semelhante, que pode ajudar a entender o menor impacto da pandemia sobre a VC na segunda divisão do Campeonato Brasileiro, foi proposta por Fischer e Haucap (2020). Os autores mostraram que os times alemães que estavam mais habituados com estádios mais preenchidos mostraram uma redução mais drástica de VC se comparados a times acostumados com estádios mais vazios. Logo, é possível que essa diferença tenha ocorrido pela baixa média de público apresentada pela Série B, conforme evidencia a tabela 2.

Um fato também a ser investigado é a interpretação das constantes obtidas nas regressões logísticas binárias. A constante (intercepto) representa o logaritmo das chances quando todas as variáveis independentes no modelo são iguais a zero. Ela representa um componente crítico do modelo, mas não possui efeito marginal pelo fato de não apresentar variação. Ao considerar os valores de constantes obtidos na regressão da coluna 7 da tabela 9 para cada grupo de dados, por exemplo, e atribuir valor zero às variáveis independentes do modelo, será possível calcular a probabilidade predita de a equipe mandante vencer a partida em um cenário em que há ausência completa de público, as equipes são da mesma cidade (um clássico local, por exemplo) e elas estão posicionadas muito próximas na tabela de classificação.

Com o intuito de trazer essa perspectiva quantitativa dos resultados da regressão logística, elaborou-se uma tabela de Efeitos Marginais Médios (EMMs) para cada uma das variáveis independentes presentes na modelagem da regressão da coluna 7 da tabela 9, além da probabilidade prevista calculada a partir do valor da constante, conforme apresentado acima. Os resultados são apresentados na tabela 10.

Como é possível observar, se todas as variáveis independentes assumirem valor nulo, as probabilidades preditas de vitória do time mandante, originadas da constante, possuem todas valores abaixo de 50%. Apesar de o nível de ocupação e a distância percorrida pelo time visitante se mostrarem motores importantes para o aumento dessas probabilidades nas Séries A e B, respectivamente, o diferencial de colocação mostra ser o principal potencial impulsionador dessa probabilidade de vitória da equipe mandante. A título de exemplo,

o efeito positivo sobre a probabilidade observado de um estádio com nível de ocupação máximo é completamente anulado se a equipe visitante estiver 4 posições à frente do time mandante na tabela.

Tabela 10 – Efeitos Marginais Médios e Probabilidade Preditada da Constante

	Constante	Ocupação	Distância (100 km)	Diferencial de Colocação
Total	43.03%	6.92 p. p.	0.24 p. p.	-1.85 p. p.
Série A	46.10%	6.84 p. p.	0.04 p. p.	-1.46 p. p.
Série B	40.11%	0.67 p. p.	0.44 p. p.	-2.23 p. p.

Pontos percentuais: p. p.

Elaborado pelo autor.

Os impactos das mesmas variáveis independentes utilizadas no modelo logístico sobre o diferencial de gols em cada partida também foram investigados e os resultados são apresentados na tabela 11.

Diferentemente da regressão logística binária, os coeficientes da regressão por MQO trazem uma abordagem não só qualitativa como também quantitativa dos efeitos das variáveis independentes sobre o diferencial de gols entre mandante e visitante. Qualitativamente, contudo, os resultados são semelhantes aos obtidos na regressão logística, até mesmo nos impactos causados sobre os coeficientes devido à inclusão de novas variáveis ao modelo, como é o caso de perda de significância estatística do coeficiente da variável de ocupação após a inclusão da variável *dummy* do Covid. Este comportamento qualitativo não surpreende pelo fato de que um diferencial de gols positivo está intimamente ligado à vitória da equipe mandante e, naturalmente, à probabilidade deste evento ocorrer. Assim como no caso da regressão logística, resultados adicionais foram obtidos a partir da substituição da variável independente de ocupação pela quantidade absoluta de torcedores e estão disponíveis para consulta no apêndice.

Diante do exposto acima, o comportamento do coeficiente relacionado à distância percorrida pela equipe visitante é mantido, com destaque deste efeito para a Série B. Pollard, Silva e Medeiros (2008) identificaram que a VC é maior para equipes que se localizam em regiões mais distantes do centro do país, como as regiões Sul e Nordeste, e calculam um impacto estimado de 0,0115 sobre o diferencial de gols para cada 100 km percorridos pela equipe visitante, resultado próximo do apresentado na tabela 10. Na tabela 3, é possível verificar que a distribuição geográfica das equipes da segunda divisão é mais distribuída e possui maior peso nas regiões Sul e Nordeste que a primeira divisão, o que pode explicar a relevância do coeficiente obtido. Logo, um efeito mais relevante dessa variável na Série B poderia ser explicada parcialmente pelo fato de haver maiores deslocamentos e, ainda, piores condições de viagens em relação aos times da Série A (ALMEIDA; OLIVEIRA; SILVA, 2011).

Tabela 11 – Resultados MQO - Diferencial de Gols

Total	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ocupação	0.337*** (0.065)			0.123*** (0.061)		0.152** (0.062)	0.058 (0.070)	
Distância em 100 km		0.007*** (0.002)			0.005*** (0.002)		0.006*** (0.002)	0.006*** (0.002)
Covid			-0.153*** (0.045)			-0.151*** (0.042)		-0.137*** (0.048)
Diferencial de Colocação				-0.062*** (0.002)	-0.062*** (0.002)	-0.062*** (0.002)	-0.062*** (0.002)	-0.062*** (0.002)
Constante	0.346*** (0.024)	0.372*** (0.028)	0.460*** (0.017)	0.397*** (0.022)	0.379*** (0.026)	0.453*** (0.016)	0.331*** (0.033)	0.378*** (0.037)
Observações	8,299	8,063	8,299	8,299	8,063	8,299	8,063	8,063
R ²	0.003	0.001	0.001	0.133	0.134	0.134	0.134	0.135
Estatística F	26.894***	8.850***	11.389***	636.478***	622.025***	641.573***	416.921***	315.026***
Série A								
Ocupação	0.261*** (0.089)			0.135 (0.086)		0.153* (0.088)	-0.054 (0.107)	
Distância em 100 km		0.0002 (0.004)			-0.0001 (0.004)		0.001 (0.004)	0.001 (0.004)
Covid			-0.226*** (0.066)			-0.233*** (0.063)		-0.262*** (0.077)
Diferencial de Colocação				-0.052*** (0.003)	-0.052*** (0.003)	-0.052*** (0.003)	-0.051*** (0.003)	-0.052*** (0.003)
Constante	0.373*** (0.039)	0.468*** (0.039)	0.498*** (0.025)	0.426*** (0.038)	0.480*** (0.038)	0.507*** (0.024)	0.418*** (0.052)	0.530*** (0.062)
Observações	4,155	3,967	4,155	4,155	3,967	4,155	3,967	3,967
R ²	0.002	0.000	0.003	0.087	0.086	0.089	0.087	0.089
Estatística F	8.536***	0.000	11.841***	197.513***	186.629***	203.679***	125.491***	97.231***
Série B								
Ocupação	0.461*** (0.109)			-0.076 (0.100)		-0.090 (0.100)	-0.165 (0.110)	
Distância em 100 km		0.014*** (0.003)			0.012*** (0.003)		0.012*** (0.003)	0.013*** (0.003)
Covid			-0.081 (0.062)			-0.067 (0.056)		-0.104* (0.061)
Diferencial de Colocação				-0.073*** (0.002)	-0.073*** (0.002)	-0.073*** (0.002)	-0.073*** (0.002)	-0.073*** (0.002)
Constante	0.322*** (0.031)	0.253*** (0.041)	0.423*** (0.024)	0.396*** (0.028)	0.245*** (0.037)	0.392*** (0.022)	0.262*** (0.042)	0.290*** (0.045)
Observações	4,144	4,096	4,144	4,144	4,096	4,144	4,096	4,096
R ²	0.004	0.005	0.0004	0.192	0.195	0.192	0.195	0.196
Estatística F	17.751***	20.686***	1.669	491.379***	496.289***	491.918***	331.113***	249.168***

Erros-padrões IID entre parênteses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Elaborado pelo autor.

Ainda, caso se atribua valor zero a todas as variáveis independentes, pelo fato de as constantes obtidas em todas as regressões realizadas possuírem sinal positivo, o diferencial

de gols entre o time mandante e visitante também assumiria um valor positivo, em média, o que representa uma vantagem existente para o time que joga em casa. Mesmo sob a presença da pandemia (*dummy* de Covid igual a 1), essa vantagem do time mandante é atenuada, mas não desaparece completamente.

Os resultados obtidos em relação ao desempenho da equipe mandante, representados pela probabilidade de vitória e pelo diferencial de gols, mostram que a presença do público contribui na construção da VC existente na Série A do Campeonato Brasileiro. A ausência do público, contudo, presente nas partidas jogadas durante a pandemia, parece demonstrar um impacto ainda mais significativo, alcançando as primeira e segunda divisões. A distância percorrida pelo time visitante também mostra participação sobre o efeito de VC na Série B, mas, dentre os fatores escolhidos, a força relativa entre os times mostrou um maior poder de influência sobre o fenômeno em ambas as divisões.

5.3 Resultados de Arbitragem

Além da análise de probabilidades de vitória ou de variáveis intimamente ligadas ao resultado das partidas, como o diferencial de gols ou de pontos conquistados, o comportamento da arbitragem também foi investigado. Assim como foi feito na regressão do diferencial de gols, os coeficientes das regressões das variáveis de arbitragem (cartões, impedimentos e pênaltis) também foram obtidos por MQO. Os resultados foram divididos em duas tabelas, pois há uma diferença nas variáveis de controle utilizadas em cada grupo de regressões. O diferencial de faltas cometidas foi utilizado como variável de controle no primeiro grupo, das regressões que envolvem o diferencial de cartões como variável dependente, o que não acontece nas regressões de impedimentos e pênaltis (segundo grupo). Como a penalidade de aplicação de um cartão amarelo ou vermelho se origina, normalmente, a partir de uma falta cometida, entendeu-se que seria uma variável relevante para a captura de parte deste efeito de penalidade dos árbitros. Em contrapartida, no segundo grupo incluiu-se, como variável independente, uma interação entre a variável *dummy* da pandemia e a variável *dummy* do VAR.

5.3.1 Grupo 1 - Cartões:

Os resultados das regressões do primeiro grupo são apresentados na tabela 12. Antes de qualquer análise, é importante mencionar que, apesar de as regressões da coluna 6 apresentarem as mesmas variáveis explicativas das regressões da coluna 8, as últimas incluem o controle por efeitos fixos da arbitragem, em que se utilizou uma variável *dummy* para cada árbitro que apitou as partidas da amostra. Na tabela, é possível observar o impacto negativo do nível de ocupação tanto sobre o diferencial de cartões amarelos quanto sobre o diferencial de cartões vermelhos, ou seja, o árbitro aplicaria menos cartões ao

time mandante e/ou um número maior de cartões ao time visitante se o estádio em que a partida fosse jogada estivesse com uma presença maior de torcida. Esse comportamento é observado, em diferentes níveis, em todas as regressões realizadas, mesmo após o controle pelo diferencial de faltas cometidas, que não altera significativamente o coeficiente da variável de ocupação (coluna 4).

Assim como o diferencial de colocação entre os times, variável de controle utilizada nas regressões anteriores, o diferencial de faltas cometidas também não demonstrou alterações em seus coeficientes estimados ao longo dos diferentes modelos de regressão utilizados para o diferencial de cartões, especialmente no caso dos cartões amarelos, em que apresentou significância estatística. Além disso, a medida do R^2 indicou que houve um aumento da adequação do modelo de cartões amarelos após a inclusão desta variável. A ausência deste comportamento no caso dos cartões vermelhos pode ser justificada pelo fato de a aplicação desta penalidade mais gravosa ser de rara ocorrência nos jogos.

A *dummy* de Covid, isoladamente, apresenta impacto positivo sobre o diferencial de cartões, ou seja, maior penalidade para a equipe mandante e/ou menor penalidade para o time visitante. Resultados semelhantes podem ser observados nos trabalhos de Fischer e Haucap (2020) e Endrich e Gesche (2020). Como esta variável representa as partidas jogadas durante a pandemia que não tiveram a presença do público, este coeficiente positivo mantém relação com o coeficiente negativo obtido para a variável do nível de ocupação. Logo, em partidas nas quais o time mandante possui o apoio de sua torcida, o viés de arbitragem (maior diferencial de cartões) parece ser potencializado justamente pela presença da torcida. Em regressões que contam com a presença da variável de ocupação e da *dummy* do Covid, o coeficiente desta última perde significância estatística nos cenários de ambos os cartões, o que contraria os achados de Bryson et al. (2021), que identifica que a ausência total de torcida possui impacto mais representativo que a variação normal do número de torcedores presentes. . A tentativa de identificar se a presença do VAR potencializou o efeito da ocupação dos estádios sobre o diferencial de cartões não mostrou resultados significativos (coluna 7).

Apesar disso, a variável *dummy* que representa as partidas disputadas com a presença do VAR apresentou coeficientes significativos e de forma a combater o viés de arbitragem (sinal positivo) apenas no caso dos cartões amarelos. Como foi exposto na seção de estatísticas descritivas, no período em que o VAR esteve presente, o diferencial de faltas cometidas aumentou ao ponto de se tornar positivo (equipes mandantes cometendo mais faltas que equipes visitantes), fato que poderia explicar a redução no diferencial de cartões amarelos percebida após a implementação da nova tecnologia de arbitragem. Contudo, mesmo ao controlar pelo diferencial de faltas cometidas, o coeficiente da *dummy* de VAR ainda mostrou exercer algum efeito sobre a aplicação da penalidade menos gravosa (colunas 5 a 8). Quantitativamente, na coluna 7, o efeito da presença do VAR seria equivalente ao

Tabela 12 – Resultados MQO - Cartões

Cartões Amarelos	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ocupação	-0.447*** (0.087)			-0.437*** (0.092)	-0.431*** (0.092)	-0.339*** (0.113)	-0.289** (0.136)	-0.424*** (0.119)
Covid		0.334*** (0.058)				0.102 (0.072)	0.097 (0.073)	0.116 (0.076)
VAR			0.238*** (0.048)		0.165*** (0.050)	0.152*** (0.051)	0.187*** (0.074)	0.138** (0.059)
Diferencial de Faltas Cometidas				0.088*** (0.004)	0.087*** (0.004)	0.087*** (0.004)	0.087*** (0.004)	0.087*** (0.004)
Ocupação × VAR							-0.124 (0.0187)	
Constante	-0.180*** (0.033)	-0.364*** (0.024)	-0.376*** (0.026)	-0.155*** (0.036)	-0.220*** (0.041)	-0.263*** (0.051)	-0.277*** (0.055)	0.120 (0.402)
Observações	7,032	7,032	7,032	5,424	5,424	5,424	5,424	5,424
R ²	0.004	0.005	0.003	0.0086	0.088	0.088	0.089	0.117
Estatística F	26.509***	32.686***	24.256***	255.995***	174.580***	131.458***	105.244***	4.554***
Cartões Vermelhos								
Ocupação	-0.086*** (0.022)			-0.098*** (0.025)	-0.098*** (0.033)	-0.104*** (0.030)	-0.103** (0.036)	-0.116*** (0.032)
Covid		0.036** (0.015)				-0.007 (0.019)	-0.007 (0.019)	0.0003 (0.020)
VAR			0.012 (0.048)		0.006 (0.050)	0.006 (0.051)	0.007 (0.074)	-0.007 (0.059)
Diferencial de Faltas Cometidas				-0.0005 (0.001)	-0.0005 (0.001)	-0.0005 (0.001)	-0.0005 (0.001)	-0.0001 (0.001)
Ocupação × VAR							-0.001 (0.050)	
Constante	-0.015* (0.009)	-0.045*** (0.006)	-0.043*** (0.007)	-0.007 (0.009)	-0.009 (0.011)	-0.006 (0.014)	-0.007 (0.015)	0.005 (0.107)
Observações	7,032	7,032	7,032	5,424	5,424	5,424	5,424	5,424
R ²	0.002	0.001	0.0001	0.003	0.003	0.003	0.003	0.030
Estatística F	14.691***	5.551**	0.985	7.974***	5.372***	4.058***	3.246***	1.049

Erros-padrões IID entre parênteses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Nota: A coluna (8) representa o modelo com o vetor de variáveis dummy dos árbitros e os coeficientes foram ocultos para melhor apresentação.

Elaborado pelo autor.

efeito trazido por uma diferença de duas faltas cometidas entre o mandante e o visitante.

Diferentemente de trabalhos como os de Han et al. (2020), Lago-Peñas, Rey e Kalén (2019) e Kubayi, Larkin e Toriola (2021), que identificaram mudanças no número absoluto de decisões de arbitragem após a introdução do VAR em diferentes ligas de futebol, este estudo parte de uma abordagem diferente, em que procura identificar se houve mudança no diferencial, entre mandante e visitante, dessas mesmas decisões de arbitragem. O uso do VAR proporciona aos árbitros uma nova oportunidade de analisar situações cruciais de jogo com mais tempo disponível e de uma forma coletiva, além de proporcionar a visualização do lance a partir de diversos ângulos, o que favorece a diminuição de erros em algumas decisões e diminui o impacto da pressão social causada pela torcida (RIBEIRO et

al., 2022; UNKELBACH; MEMMERT, 2010). Isso pode explicar os resultados obtidos a partir das variáveis de ocupação e *dummy* de VAR, principalmente para o caso dos cartões amarelos.

A última coluna de regressões, que inclui o controle de efeitos fixos de cada um dos 216 árbitros que apitaram as partidas da amostra, não parece apresentar resultados muito diferentes, quantitativamente, dos resultados presentes nas regressões das colunas 6 e 7. Para ambos os tipos de cartões, o coeficiente de ocupação ganha até mesmo maior relevância como variável explicativa após a introdução desse controle. A constante, no caso de cartões amarelos, que se mostrou consistentemente negativa nas regressões anteriores, perde significância estatística. Qualitativamente, o restante dos coeficientes mantém o comportamento.

O fato de a aplicação de cartões vermelhos ser um evento de mais rara ocorrência gera uma amostra com grande concentração de valores iguais a zero, o que pode representar um obstáculo para obter resultados mais conclusivos em relação ao comportamento desta variável. Recomenda-se, para estudos futuros, testar outras formas de modelagem que contornem este empecilho.

5.3.2 Grupo 2 - Impedimentos e Pênaltis:

Os resultados das regressões para o diferencial de impedimentos e pênaltis, também obtidos por MQO, estão dispostos na tabela 13. A quantidade de observações para cada uma das variáveis de arbitragem é menor que a quantidade observada nas regressões de cartões, pois há ausência de dados para alguns períodos da amostra: não há dados de impedimento de 2012 e 2013 (Série A) e de 2012 a 2016 (Série B), além de os dados de pênalti cobrirem apenas as partidas jogadas entre 2014 e 2021 (Série A) e entre 2016 e 2021 (Série B). Ainda, da mesma forma que foi apresentada na tabela voltada aos cartões, a coluna 8 apresenta os resultados com o controle de efeitos fixos dos árbitros que apitaram as partidas.

Como se pode observar, os coeficientes obtidos nas regressões de impedimentos não apresentaram qualquer significância estatística ao nível de 10%, com exceção da interação entre a variável de ocupação e a variável *dummy* do VAR na coluna 5, que não traz, em um contexto mais amplo, uma interpretação clara. A variável de impedimentos, ao contrário das outras variáveis de arbitragem incluídas neste estudo, tende a estar mais relacionada à estratégia adotada por cada equipe na partida (ofensiva ou defensiva) do que a uma atuação pura de arbitragem. Isso é razoável supor, pois equipes que tomam maior iniciativa nas partidas terão uma frequência de ataques maior e, conseqüentemente, estarão sujeitas a ter um maior número de jogadas com posição irregular. Além disso, a decisão do árbitro por marcar um impedimento é menos subjetiva, pois carece de interpretação, o que não é o caso na marcação de uma falta ou aplicação de um cartão. Contudo, como o

impedimento é um fator de análise em decisões, pelo VAR, de validação de alguns gols com posicionamento duvidoso, ele foi incluído neste estudo.

Tabela 13 – Resultados MQO - Impedimentos e Pênaltis

Impedimentos	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ocupação	0.121 (0.113)			0.120 (0.113)	-0.048 (0.149)	-0.111 (0.164)	-0.142 (0.174)	-0.020 (0.146)
Covid		-0.113 (0.072)				-0.084 (0.091)	-0.125 (0.119)	-0.099 (0.095)
VAR			-0.020 (0.062)	-0.018 (0.062)	-0.129 (0.089)	-0.110 (0.092)	-0.161 (0.131)	0.009 (0.073)
Ocupação × VAR					0.395* (0.229)	0.369 (0.231)	0.461 (0.287)	
VAR × Covid							0.099 (0.184)	
Constante	0.131*** (0.044)	0.190*** (0.034)	0.173*** (0.037)	0.138*** (0.050)	0.187*** (0.058)	0.219*** (0.067)	0.235*** (0.073)	-0.408 (0.501)
Observações	5,283	5,283	5,283	5,283	5,283	5,283	5,283	5,283
R ²	0.0002	0.0005	0.00002	0.0002	0.001	0.001	0.001	0.031
Estatística F	1.148	2.450	0.102	0.614	1.399	1.263	1.069	1.078
Pênaltis								
Ocupação	0.072*** (0.025)			0.072*** (0.025)	0.036 (0.030)	-0.008 (0.032)	-0.011 (0.034)	0.030 (0.032)
Covid		-0.069*** (0.015)				-0.066*** (0.020)	-0.071*** (0.023)	-0.063*** (0.021)
VAR			-0.007 (0.015)	-0.001 (0.015)	-0.029 (0.019)	-0.001 (0.021)	-0.010 (0.034)	0.030* (0.017)
Ocupação × VAR					0.127** (0.056)	0.081 (0.058)	0.099 (0.013)	
VAR × Covid							0.015 (0.013)	
Constante	0.034*** (0.009)	0.068*** (0.007)	0.054*** (0.007)	0.034*** (0.010)	0.044*** (0.011)	0.066*** (0.013)	0.067*** (0.013)	-0.134 (0.112)
Observações	5,291	5,291	5,291	5,291	5,291	5,291	3,967	5,291
R ²	0.002	0.004	0.00005	0.002	0.003	0.005	0.087	0.040
Estatística F	8.411***	21.310***	0.249	4.208**	4.514***	6.219***	125.491***	1.325***

Erros-padrões IID entre parênteses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Nota: A coluna (8) representa o modelo com o vetor de variáveis dummy dos árbitros e os coeficientes foram ocultos para melhor apresentação.

Elaborado pelo autor.

Semelhante ao que foi apresentado em relação ao diferencial de cartões vermelhos, por se caracterizar como um evento de rara ocorrência, a amostra de pênaltis enfrenta o mesmo empecilho, diversos valores iguais a zero. A variável *dummy* de Covid parece mostrar

o resultado mais significativo no comportamento do diferencial de pênaltis marcados, o que já havia sido identificado na seção de estatísticas descritivas quando o período sem pandemia foi comparado ao período durante a pandemia, em que o diferencial de pênaltis assumiu valor zero. A perda de significância estatística do coeficiente da variável de ocupação, que ocorre ao longo das regressões após a inclusão da variável *dummy* de Covid, pode ser explicado por motivos semelhantes ao que foi observado na regressão logística, a partir dos histogramas do gráfico 2. A existência de muitos jogos com ausência completa de público oferece um peso maior a esta *dummy*, que só engloba jogos com esta característica. Nas regressões das colunas 6 e 7, é possível observar um coeficiente da variável *dummy* da Covid que anula completamente o efeito do intercepto, o que pode representar ausência de viés de arbitragem na marcação de pênaltis quando não há presença de público nos estádios.

Em contrapartida, a variável *dummy* do VAR parece não demonstrar, isoladamente, um efeito significativo sobre o diferencial de pênaltis. Como a avaliação da marcação adequada de um pênalti é um dos motivos para o uso da tecnologia de vídeo durante a partida, esperava-se, pela íntima relação entre os acontecimentos, algum resultado diferente. Contudo, a presença do VAR parece ter aumentado o efeito que a variável de ocupação possui sobre o diferencial de pênaltis das partidas (coluna 5). Talvez, pelo simples fato de a tecnologia existir e haver possibilidade de revisão do lance, o árbitro se sinta mais confrontado pela torcida a checar, no vídeo, algum lance duvidoso da partida.

Após o controle por efeitos fixos dos árbitros (coluna 8), é possível perceber que a variável *dummy* de Covid mantém a sua relevância como variável explicativa e a *dummy* de VAR conquista relevância individual, mas contribui para o aumento do diferencial de pênaltis.

Os resultados obtidos em relação às variáveis de arbitragem mostram uma contribuição da torcida sobre a decisão de aplicação de cartões no sentido de favorecer a equipe mandante. O período da pandemia, marcado pela ausência das torcidas nos estádios, mostrou efeito mais significativo sobre o diferencial de pênaltis, reduzindo o efeito do fenômeno da VC. Ainda, a introdução do VAR apresentou efeitos mistos sobre a VC, a partir do aumento do diferencial de pênaltis marcados a favor e o aumento do diferencial de cartões amarelos aplicados.

6 Conclusão

A presente dissertação investigou a hipótese da existência da vantagem de jogar em casa nas principais divisões do Campeonato Brasileiro de futebol, com o objetivo de identificar a relevância de alguns fatores, como ocupação dos estádios, distância viajada pelo time visitante e viés de arbitragem, ao explicar este fenômeno. Apesar de apresentar variações, a vantagem de jogar em casa, calculada sob a ótica dos pontos conquistados e dos gols marcados a partir do método modificado de Pollard, se mostrou presente em todos os anos da base de dados utilizada. Além disso, as estatísticas descritivas, de forma consolidada, também mostraram um domínio maior da partida por parte da equipe mandante, que teve mais iniciativa ofensiva. Este comportamento é observado nas Séries A e B, assim como, em menor ou maior grau, durante períodos com ausência de público ou presença do VAR.

Após identificar a presença da vantagem, procurou-se identificar a influência de alguns fatores sobre o fenômeno da VC. A partir de alguns modelos de regressão, identificou-se que o fenômeno possui uma dinâmica diferente a depender da divisão do campeonato analisada. Na Série A, o nível de ocupação dos estádios mostrou maior relevância que a Série B na explicação do fenômeno. Já na segunda divisão, entretanto, foi a vez da distância percorrida pelo time visitante se destacar quando comparada à primeira divisão. Naturalmente, como o período da pandemia do Covid-19 foi marcado pela ausência de público, o impacto do evento sobre a Série A do Campeonato Brasileiro também foi maior. De qualquer forma, apesar de potencialmente contribuírem para aumentar o diferencial de gols e, conseqüentemente, a probabilidade de vitória da equipe mandante, a diferença de força relativa entre os times, representada pelo diferencial de colocação na tabela do campeonato no momento da partida, mostrou ser a variável com o maior poder de explicação dessa vantagem.

O viés de arbitragem também foi investigado e, apesar da introdução da tecnologia do árbitro de vídeo (VAR) ao futebol brasileiro, o seu impacto sobre algumas decisões de arbitragem mais relevantes e que possuem a capacidade de alterar o rumo de uma partida, como cartões vermelhos ou pênaltis, não se mostrou relevante. Em contrapartida, uma decisão que é menos gravosa e ocorre com maior frequência nos jogos, como a aplicação de cartões amarelos, demonstrou responder favoravelmente ao time mandante em estádios mais cheios e favoravelmente ao time visitante, com menor efeito, na presença do VAR. Ainda, o diferencial de cartões vermelhos aplicados ao time mandante e visitante e o diferencial de pênaltis marcados a favor de cada equipe, mostraram ser afetados negativamente, respectivamente, pela presença da torcida e ausência de torcida (jogos durante a Covid-19).

Em suma, o fenômeno da VC se mostra complexo mesmo no caso brasileiro, sugerindo que depende de diversos fatores em diferentes níveis. Além de contribuir para a literatura de vantagem de jogar em casa no Brasil, que já não é vasta, os resultados obtidos neste trabalho sugerem um ponto de partida para futuras investigações, como o impacto de variáveis psicológicas e físicas, intimamente ligadas aos fatores de presença de público e deslocamento em viagens, além da possibilidade de aprofundar na análise do impacto do VAR, a partir das revisões de lances individuais já realizadas pelos árbitros desde a sua implementação.

Referências

- AGUILAR, G.; GARCÍA, M. S.; ROMERO, J. J. F. Measuring Home Advantage in Spanish Handball. *Perceptual and Motor Skills*, v. 114, p. 329–338, 2012. Citado na página 20.
- ALMEIDA, C. H.; LEITE, W. Professional football in times of COVID-19: did the home advantage effect disappear in European domestic leagues? *Biology of Sport*, v. 38, p. 693–701, 2021. Citado na página 23.
- ALMEIDA, L. G. d.; OLIVEIRA, M. L. d.; SILVA, C. D. d. Uma análise da vantagem de jogar em casa nas duas principais divisões do futebol profissional brasileiro. *Revista Brasileira De Educação Física E Esporte*, v. 25, p. 49–54, 2011. Citado 2 vezes nas páginas 26 e 47.
- ANDERSON, M. N. et al. Perspectives on the home advantage: A comparison of football players, fans and referees. *Psychology of Sport and Exercise*, v. 13, p. 311–316, 2012. Citado na página 21.
- BELCHIOR, C. A. Fans and Match Results: Evidence from a Natural Experiment in Brazil. *Journal of Sports Economics*, v. 21, p. 663–687, 2020. Citado 2 vezes nas páginas 27 e 46.
- BRAGA, B.; GUILLEN, D. Working under pressure: Evidence from the impacts of soccer fans on players' performance. *Economics Letters*, v. 114, p. 212–215, 2012. Citado 2 vezes nas páginas 27 e 46.
- BROCHERIE, F. et al. Influence of Weather, Rank, and Home Advantage on Football Outcomes in the Gulf Region. *Medicine Amp; Science in Sports Amp; Exercise*, v. 47, p. 401–410, 2015. Citado na página 20.
- BRYSON, A. et al. Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters*, v. 198, p. 109664, 2021. Citado 3 vezes nas páginas 22, 23 e 50.
- CARMICHAEL, F.; THOMAS, D. Home-Field Effect and Team Performance. *Journal of Sports Economics*, v. 6, p. 264–281, 2005. Citado na página 21.
- CORREIA-OLIVEIRA, C. R.; ANDRADE-SOUZA, V. A. Home advantage in soccer after the break due to COVID-19 pandemic: does crowd support matter? *International Journal of Sport and Exercise Psychology*, v. 20, p. 1245–1256, 2021. Citado na página 23.
- DAMME, N. v.; BAERT, S. Home Advantage in European International Soccer: Which Dimension of Distance Matters? *SSRN Electronic Journal*, 2019. Citado na página 25.
- DEWENTER, R.; NAMINI, J. E. How to Make Soccer More Attractive? Rewards for a Victory, the Teams' Offensiveness, and the Home Bias. *Journal of Sports Economics*, v. 14, p. 65–86, 2011. Citado na página 21.
- DOHMEN, T.; SAUERMAN, J. Referee Bias. *Journal of Economic Surveys*, v. 30, p. 679–695, 2015. Citado na página 24.

- ENDRICH, M.; GESCHE, T. Home-Bias in Referee Decisions: Evidence from 'Ghost Matches' During the COVID-19 Pandemic. *SSRN Electronic Journal*, 2020. Citado 2 vezes nas páginas 23 e 50.
- FISCHER, K.; HAUCAP, J. Does Crowd Support Drive the Home Advantage in Professional Soccer? Evidence from German Ghost Games During the Covid-19 Pandemic. *SSRN Electronic Journal*, 2020. Citado 3 vezes nas páginas 23, 46 e 50.
- FRANCHINI, E.; TAKITO, M. Y. Home advantage in combat sports during the Olympic games. *Sport Sciences for Health*, v. 12, p. 287–290, 2016. Citado na página 20.
- GARCÍA, M. S. et al. Calculating Home Advantage in the First Decade of the 21th Century UEFA Soccer Leagues. *Journal of Human Kinetics*, v. 38, p. 141–150, 2013. Citado na página 21.
- GILBERT, E. J.; DIXON, J. C.; LOUGHEAD, T. M. An Examination of Travel Effects on Performance Outcomes in Major League Soccer. *Journal of Applied Sport Management*, 2020. Citado na página 25.
- GOUMAS, C. Home advantage and referee bias in European football. *European Journal of Sport Science*, v. 14, 2012. Citado na página 24.
- GOUMAS, C. Home advantage in Australian soccer. *Journal of Science and Medicine in Sport*, v. 17, p. 119–123, 2014. Citado na página 26.
- HAN, B. et al. The influence of the video assistant referee on the Chinese Super League. *International Journal of Sports Science Amp; Coaching*, v. 15, p. 662–668, 2020. Citado 3 vezes nas páginas 24, 25 e 51.
- INAN, T. Analyzing the Home-Field advantage in Major European Football Leagues. *The International Journal of Environmental and Science Education*, v. 13, p. 113–124, 2018. Citado na página 20.
- INAN, T. The Effect of Crowd Support on Home-Field Advantage: Evidence from European Football. *Annals of Applied Sport Science*, v. 8, 2020. Citado 2 vezes nas páginas 22 e 45.
- KUBAYI, A.; LARKIN, P.; TORIOLA, A. The impact of video assistant referee (VAR) on match performance variables at men's FIFA World Cup tournaments. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, v. 236, p. 187–191, 2021. Citado 2 vezes nas páginas 25 e 51.
- LAGO-PEÑAS, C.; REY, E.; KALÉN, A. How does Video Assistant Referee (VAR) modify the game in elite soccer? *International Journal of Performance Analysis in Sport*, v. 19, p. 646–653, 2019. Citado 2 vezes nas páginas 25 e 51.
- LAGO-PEÑAS, C.; SÁNCHEZ-GÓMEZ, M. C.; POLLARD, R. The effect of the Video Assistant Referee on referee's decisions in the Spanish LaLiga. *International Journal of Sports Science Amp; Coaching*, v. 16, p. 824–829, 2020. Citado na página 25.
- LEITE, W. Home Advantage: Comparison between the Major European Football Leagues. *Athens Journal of Sports*, v. 4, p. 65–74, 2017. Citado 2 vezes nas páginas 20 e 22.

LEITNER, M. C. et al. The cauldron has cooled down: a systematic literature review on home advantage in football during the COVID-19 pandemic from a socio-economic and psychological perspective. *Management Review Quarterly*, v. 73, p. 605–633, 2022. Citado na página 23.

MATOS, R. et al. Home-Advantage during COVID-19: An Analysis in Portuguese Football League. *International Journal of Environmental Research and Public Health*, v. 18, p. 3761, 2021. Citado 2 vezes nas páginas 23 e 46.

MCCARRICK, D. et al. Home advantage during the COVID-19 pandemic: Analyses of European football leagues. *Psychology of Sport and Exercise*, v. 56, p. 102013, 2021. Citado na página 23.

MCHILL, A. W.; CHINOY, E. D. Utilizing the national basketball association's COVID-19 restart "bubble" to uncover the impact of travel and circadian disruption on athletic performance. *Scientific Reports*, v. 10, 2020. Citado na página 20.

PIC, M. Performance and Home Advantage in Handball. *Journal of Human Kinetics*, v. 63, p. 61–71, 2018. Citado na página 20.

POLLARD, R. Home advantage in soccer: a retrospective analysis. *Journal of Sports Sciences*, v. 4, p. 237–248, 1986. Citado 4 vezes nas páginas 19, 20, 21 e 22.

POLLARD, R. Worldwide regional variations in home advantage in association football. *Journal of Sports Sciences*, v. 24, p. 231–240, 2006. Citado na página 21.

POLLARD, R.; ARMATAS, V. Factors affecting home advantage in football World Cup qualification. *International Journal of Performance Analysis in Sport*, v. 17, p. 121–135, 2017. Citado na página 22.

POLLARD, R.; POLLARD, G. Long-term trends in home advantage in professional team sports in North America and England (1876–2003). *Journal of Sports Sciences*, v. 23, p. 337–350, 2005. Citado na página 20.

POLLARD, R.; PRIETO, J.; GÓMEZ, M. Global differences in home advantage by country, sport and sex. *International Journal of Performance Analysis in Sport*, v. 17, p. 586–599, 2017. Citado na página 19.

POLLARD, R.; SILVA, C. D.; MEDEIROS, N. Home advantage in football in Brazil: Differences between teams and the effects of distance traveled. *Brazilian Journal of Soccer Science*, v. 1, p. 3–10, 2008. Citado 2 vezes nas páginas 26 e 47.

PONZO, M.; SCOPPA, V. Does the Home Advantage Depend on Crowd Support? Evidence From Same-Stadium Derbies. *Journal of Sports Economics*, v. 19, p. 562–582, 2016. Citado na página 22.

RIBEIRO, L. d. C. et al. Did the Absence of Crowd Support During the Covid-19 Pandemic Affect the Home Advantage in Brazilian Elite Soccer? *Journal of Human Kinetics*, v. 81, p. 251–258, 2022. Citado 3 vezes nas páginas 26, 45 e 52.

SCHWARTZ, B.; BARSKY, S. F. The Home Advantage. *Social Forces*, v. 55, p. 641, 1977. Citado na página 19.

- SILVA, C. D. d.; MEDEIROS, N. C.; SILVA, A. C. D. d. Vantagem em casa no campeonato brasileiro de futebol: efeito do local do jogo e da qualidade dos times. *Revista Brasileira De Cineantropometria E Desempenho Humano*, v. 12, 2011. Citado na página 26.
- SORS, F. et al. Pressing crowd noise impairs the ability of anxious basketball referees to discriminate fouls. *Frontiers in Psychology*, v. 10, 2019. Citado na página 20.
- SPITZ, J. et al. Video assistant referees (VAR): The impact of technology on decision making in association football referees. *Journal of Sports Sciences*, v. 39, p. 147–153, 2020. Citado na página 25.
- TABASSUM, Y. et al. Effect of playing venue on pre-competition cortisol level and competitive state anxiety in university basketball players. *Journal of Pharmaceutical Research International*, p. 248–255, 2021. Citado na página 20.
- TILP, M.; THALLER, S. Covid-19 Has Turned Home Advantage Into Home Disadvantage in the German Soccer Bundesliga. *Frontiers in Sports and Active Living*, v. 2, 2020. Citado 2 vezes nas páginas 23 e 45.
- UNKELBACH, C.; MEMMERT, D. Crowd Noise as a Cue in Referee Decisions Contributes to the Home Advantage. *Journal of Sport and Exercise Psychology*, v. 32, p. 483–498, 2010. Citado na página 52.
- WUNDERLICH, F. et al. 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*, v. 16, p. e0248590, 2021. Citado na página 23.

Apêndices

APÊNDICE A – Tabelas Adicionais

Tabela 14 – Resultados Logit - Público Absoluto

Total	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Público em 10,000 torcedores	0.117*** (0.018)			0.042** (0.020)			0.058*** (0.020)	0.033 (0.021)
Distância em 100 km		0.011*** (0.003)			0.010*** (0.003)		0.011*** (0.003)	0.011*** (0.004)
Covid			-0.223*** (0.063)			-0.249*** (0.066)		-0.227*** (0.071)
Diferencial de Colocação				-0.082*** (0.003)	-0.084*** (0.003)	-0.083*** (0.003)	-0.083*** (0.003)	-0.083*** (0.003)
Constante	-0.197*** (0.029)	-0.173*** (0.039)	-0.045* (0.024)	-0.139*** (0.031)	-0.182*** (0.041)	-0.060** (0.025)	-0.254*** (0.048)	-0.195*** (0.051)
Observações	8,299	8,299	8,299	8,063	8,299	8,063	8,063	
Pseudo-R ² (McFadden)	0.004	0.0009	0.001	0.080	0.081	0.080	0.082	0.083
Akaike Information Criterion (AIC)	11,455.47	11,161.26	11,483.55	10,582.50	10,266.80	10,573.06	10,260.45	10,252.27
Série A								
Público em 10,000 torcedores	0.103*** (0.023)			0.045* (0.024)			0.059** (0.025)	0.019 (0.028)
Distância em 100 km		0.001 (0.005)			0.0002 (0.005)		0.002 (0.005)	0.002 (0.005)
Covid			-0.298*** (0.089)			-0.328*** (0.092)		-0.316*** (0.106)
Diferencial de Colocação				-0.062*** (0.004)	-0.063*** (0.004)	-0.063*** (0.004)	-0.061*** (0.004)	-0.063*** (0.004)
Constante	-0.211*** (0.047)	-0.050 (0.053)	-0.013 (0.034)	-0.117** (0.048)	-0.038 (0.055)	-0.001 (0.035)	-0.140** (0.070)	-0.037 (0.078)
Observações	4,155	4,155	4,155	3,967	4,155	3,967	3,967	
Pseudo-R ² (McFadden)	0.004	0.000	0.002	0.047	0.047	0.049	0.048	0.049
Akaike Information Criterion (AIC)	5,740.67	5,501.42	5,749.64	5,489.82	5,243.80	5,480.70	5,240.18	5,233.24
Série B								
Público em 10,000 torcedores	0.226*** (0.043)			-0.007 (0.046)			-0.024 (0.047)	-0.053 (0.049)
Distância em 100 km		0.020*** (0.004)			0.021*** (0.005)		0.021*** (0.005)	0.022*** (0.005)
Covid			-0.148* (0.088)			-0.158* (0.096)		-0.200** (0.101)
Diferencial de Colocação				-0.106*** (0.005)	-0.107*** (0.004)	-0.106*** (0.004)	-0.107*** (0.005)	-0.108*** (0.005)
Constante	-0.220*** (0.038)	-0.320*** (0.058)	-0.078** (0.034)	-0.159*** (0.041)	-0.395*** (0.064)	-0.139*** (0.037)	-0.384*** (0.067)	-0.343*** (0.070)
Observações	4,144	4,096	4,144	4,144	4,096	4,144	4,096	4,096
Pseudo-R ² (McFadden)	0.005	0.004	0.0005	0.122	0.126	0.122	0.126	0.127
Akaike Information Criterion (AIC)	5,709.20	5,652.04	5,735.47	5,042.44	4,959.30	5,039.71	4,961.03	4,959.10

Erros-padrões IID entre parênteses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Elaborado pelo autor.

Tabela 15 – Resultados MQO - Diferencial de Gols - Público Absoluto

Total	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Público em 10,000 torcedores	0.099*** (0.013)			0.035*** (0.012)			0.043*** (0.013)	0.031** (0.014)
Distância em 100 km		0.007*** (0.002)			0.005** (0.002)		0.006*** (0.002)	0.006*** (0.002)
Covid			-0.153*** (0.045)			-0.151*** (0.042)		-0.119*** (0.045)
Diferencial de Colocação				-0.062*** (0.002)	-0.062*** (0.002)	-0.062*** (0.002)	-0.061*** (0.002)	-0.062*** (0.002)
Constante	0.337*** (0.021)	0.372*** (0.028)	0.460*** (0.017)	0.395*** (0.020)	0.379*** (0.026)	0.453*** (0.016)	0.326*** (0.031)	0.357*** (0.033)
Observações	8,299	8,063	8,299	8,299	8,063	8,299	8,063	8,063
R ²	0.007	0.001	0.001	0.133	0.134	0.134	0.135	0.136
Estatística F	55.77***	8.85***	11.38***	638.60***	622.02***	641.57***	419.07***	316.27***
Série A								
Público em 10,000 torcedores	0.088*** (0.017)			0.036** (0.016)			0.044*** (0.017)	0.017 (0.019)
Distância em 100 km		0.0002 (0.004)			-0.0001 (0.004)		0.001 (0.004)	0.001 (0.004)
Covid			-0.226*** (0.066)			-0.233*** (0.063)		-0.208*** (0.072)
Diferencial de Colocação				-0.051*** (0.003)	-0.052*** (0.003)	-0.052*** (0.003)	-0.050*** (0.003)	-0.051*** (0.003)
Constante	0.332*** (0.034)	0.468*** (0.039)	0.498*** (0.025)	0.420*** (0.033)	0.480*** (0.038)	0.507*** (0.024)	0.404*** (0.048)	0.472*** (0.053)
Observações	4,155	3,967	4,155	4,155	3,967	4,155	3,967	3,967
R ²	0.006	0.00000	0.003	0.087	0.086	0.089	0.088	0.090
Estatística F	27.06***	0.003	11.84***	198.75***	186.62***	203.67***	126.82***	97.38***
Série B								
Público em 10,000 torcedores	0.171*** (0.029)			-0.007 (0.027)			-0.017 (0.027)	-0.030 (0.028)
Distância em 100 km		0.014*** (0.003)			0.012*** (0.003)		0.012*** (0.003)	0.013*** (0.003)
Covid			-0.081 (0.062)			-0.067 (0.056)		-0.085 (0.059)
Diferencial de Colocação				-0.073*** (0.002)	-0.073*** (0.002)	-0.073*** (0.002)	-0.073*** (0.002)	-0.073*** (0.002)
Constante	0.320*** (0.027)	0.253*** (0.041)	0.423*** (0.024)	0.386*** (0.024)	0.245*** (0.037)	0.392*** (0.022)	0.253*** (0.039)	0.271*** (0.041)
Observações	4,144	4,096	4,144	4,144	4,096	4,144	4,096	4,096
R ²	0.008	0.005	0.0004	0.192	0.195	0.192	0.195	0.196
Estatística F	34.06***	20.68***	1.66	491.07***	496.28***	491.91***	330.94***	248.79***

Erros-padrões IID entre parênteses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Elaborado pelo autor.