Home-Bias in Referee Decisions: Evidence from “Ghost Matches” during the COVID19-Pandemic

Center for Law & Economics
Working Paper Series

Number 09/2020

Home-Bias in Referee Decisions: Evidence from "Ghost Matches" during the Covid19-Pandemic

Marek Endrich
Tobias Gesche

October 2020 (version accepted by Economics Letters)
August 2020 (updated version)
June 2020 (first version)

Home-Bias in Referee Decisions: Evidence from "Ghost Matches" during the Covid19-Pandemic

Marek Endrich

University of Hamburg

Tobias Gesche

ETH Zurich

October 12, 2020

Abstract

We use ghost matches induced by Covid19 in the Bundesliga, Germany's top two football (soccer) divisions, to investigate whether audiences affect referees. We find that pre-Covid19 referees gave fewer fouls and yellow cards for the home team relative to the away team. These differences in fouls and cards changed during the ghost matches so that home teams were treated less favorably than before. This effect is concentrated in matches where support for the away team is particularly weak. The results provide evidence for a home bias in referee decisions through social pressure.

JEL codes: D8, K4

Keywords: referee bias, football, natural experiment

[1]Institute of Law and Economics, University of Hamburg; marek.endrich@uni-hamburg.de

[2]Center for Law & Economics, ETH Zurich; tgesche@ethz.ch

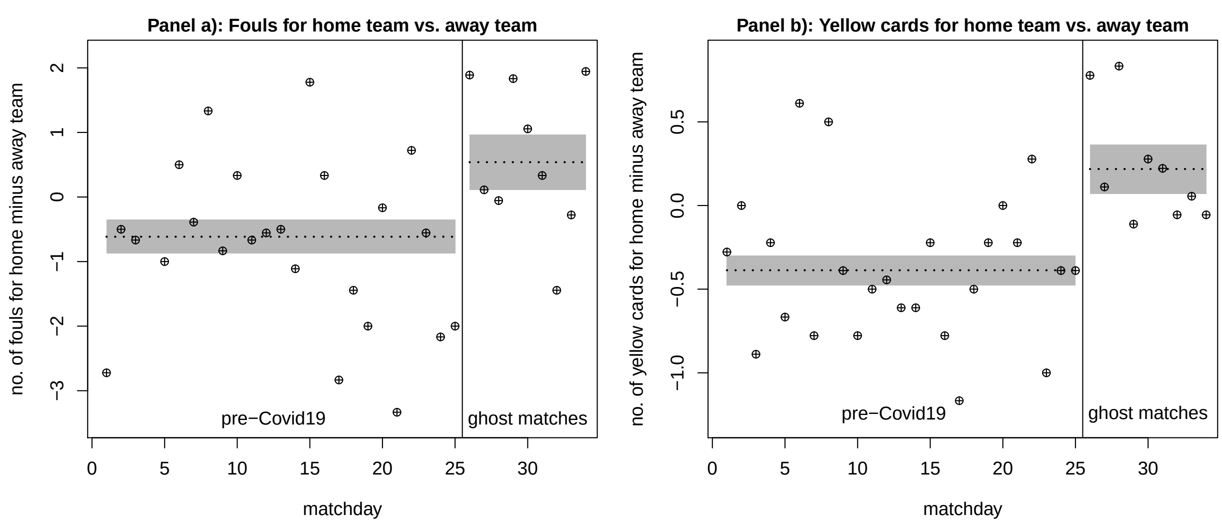
[3]We would like to thank Jerg Gutmann, Steve Heinke, Johannes Kunz, Max Mantei, Giuseppe Sorrenti and Stefan Zeisberger for helpful comments and input; we also acknowledge helpful comments by an anonymous referee. All remaining errors are, of course, our own.

1. Introduction

Competing in front of an audience can be a double-edged sword: Being exposed can spur effort, increase competition and yield better prices or quality. However, higher publicity puts the arbiters of such increased competition (e.g. regulators, supervisors or judges) under increased pressure. This applies particularly if the public favors one of the competitors. Here, we provide evidence of such asymmetric audience effects on referee decisions. We exploit a hitherto unprecedented, exogenous shock in the context of a highly competitive tournament. Specifically, we look at the effect of ghost matches induced by Covid19 in the Bundesliga, Germany's professional football (soccer) league and a EUR 4.8bn market. Whereas matches in the top two divisions normally attract an average crowd of 41,000 and 20,300, respectively, this sudden change forced the teams to play in empty stadiums.

We show evidence that referees favor the home team when there is an audience. A first piece of such evidence is visualized in Figure 1a. It portrays the difference in the number of fouls whistled against the home team relative to the away team. Home teams received, on average, about 0.6 fouls fewer than away teams in the pre-Covid19 period, suggesting a referee bias for home teams. During ghost matches, however, the relative number of fouls for the home team increased by about one. Figure 1b shows a similar pattern for the number of yellow cards that referees gave. Before Covid19, the average home team received 0.4 cards fewer than the away team. This difference changes to about 0.2 *more* cards for the home team during the ghost matches. We show in our main results section that these effects are robust to controlling for referees and teams, teams' relative strength (as measured by betting odds), audience and potential changes of in-match behavior. We also demonstrate that the effects are concentrated in matches where support by the audience for the away team is weak.

Figure 1: Difference in fouls and yellow cards over the season



Notes: Points depict estimates from regressing fouls/yellow cards on matchday fixed effects, a home team-dummy and their interactions. Horizontal lines depict estimates with a dummy for ghost matches instead of the matchday fixed effects, grey rectangles depict the corresponding SEM.

Our findings contribute to a literature documenting a bias in referee decisions such as in timing, penalties, and calling goals (Garicano et al., 2005; Dohmen, 2008; Ponzo and Scoppa, 2016; for a survey see Dohmen and Sauermann 2015). Closely related are Pettersson-Lidbom and Priks (2010), who study 21 Italian matches (3.6% of their observations) that were ordered to be without an audience following hooligan violence. While these matches might have been selected non-randomly and the small share of ghost matches limits statistical power, we can exploit a natural experiment that forced *all* teams to play 165 matches (27.0% of our observations) in empty stadiums. Using novel methods to measure the effect of in-match endogenous variables and additional data on fans travelling to away matches, we provide new evidence that home supporters can cause referee bias, even after years of improvements in refereeing techniques.[1]

2. Context, data and model specification

Our investigation uses data from the first two divisions of German professional football during the 2019/20-season. From August 2019 on, 18 teams in each division played against all other 17 teams twice – once at home, once away. Covid19 halted the season in mid-March and no matches took place for two months. Then, as society gradually reopened, the league resumed and finished in June 2020. The following matches were, for the first time ever, "ghost matches" without live audiences. By contrast, thousands watched players and referees before.

We use the following regression model to estimate the effect of the ghost matches on referee behavior by OLS:

$$y_{i,m} = \alpha + \beta_1\, Home_{i,m} + \beta_2\, GhostMatch_m + \beta_3\, Home_{i,m} \times GhostMatch_m + \gamma\, Controls_{i,m} + \epsilon_{i,m}$$

In the above, $y_{i,m}$ measures our dependent variable: Either fouls (the referee calling an offence by a player in team i, resulting in a free kick for the other team) or yellow cards (the referee judging an offence as "cautionable" – a second yellow card leads to a red card and a player's sending-off; see FIFA, 2019, Section 12). This specification counts each match m twice, once with i referring to the home team and once to the away team (see, e.g., Garicano et al., 2005, Ponzo and Scoppa, 2016), which is why we cluster standard errors on the match level.

The independent variables $Home_{i,m}$ and $GhostMatch_m$ indicate home and ghost matches. Their interaction is of primary interest: The estimate for β_3 captures how referee decisions were affected by playing a ghost match at home as opposed to playing at home pre-Covid19 in front of an audience.

The $Controls_{i,m}$-vector collects additional independent variables: Firstly, we include an index of betting odds to control for team i's relative strength (and winning incentives at the season's end). We also include fixed effects

[1]For example, the goal-line technology and video assistant referees were introduced in 2015 and 2017. Referees came also under increased scrutiny following a match-fixing scandal in 2005.

Table 1 Effect of ghost matches on referee decisions

	no. of fouls given		no. of yellow cards	
	(1a)	(1b)	(2a)	(2b)
Home	-0.651***	-0.509**	-0.310***	-0.277***
	(0.233)	(0.238)	(0.077)	(0.079)
GhostMatch	0.489	0.669**	-0.348***	-0.359***
	(0.314)	(0.303)	(0.110)	(0.113)
Home×GhostMatch	1.288***	1.350***	0.464***	0.476***
	(0.424)	(0.425)	(0.149)	(0.148)
Fouls			0.127***	0.115***
			(0.011)	(0.011)
In-match controls	no	yes	no	yes
R^2	0.312	0.372	0.250	0.265
Observations	1,224	1,224	1,224	1,224

Notes: OLS standard errors clustered on the match level in parentheses. ***/**/*: significance at the 1/5/10% level.

for each team i (i.e. 36 dummies for the 18 teams playing in each of the two divisions) and all 47 referees. In addition, we control for a stadium's average audience and add weekday-dummies (see Krumer and Lechner, 2017). Finally, we add a comprehensive set of controls for in-match player behavior. These are shots on goal, tackles, attempted and completed passes, ball possession, and running distance.[2]

3. Results

3.1. Fouls

Column 1a in Table 1 presents the results for the number of fouls received. The first coefficient represents the home effect pre-Covid19. With a live audience, the home teams received on average about 0.7, or 5.3%, fouls fewer than the away team (which got on average 12.3 fouls). The estimate for ghost matches means that away teams got 0.5 more fouls without audiences, but this change is not significant. The interaction term then shows that for home teams the ghost match-effect is significantly larger, by an additional 1.3 fouls.

This increase in the relative difference of fouls could be due to a change in players' in-match behavior, which is affected differently by whether one plays at home or away and whether an audience is present. We control for such an explanation in Column 1b by including a range of in-match behavior variables (described in Section 2). Our main effect, the differential effect of ghost matches for home teams, remains positive and significant.

Controlling for in-match behavior also allows us to estimate the effect of *un*observed behavior, using Oster (2019)'s bounding method. This method contrasts the change of the diff-in-diff coefficient (Home×Ghost) upon

[2]Data is obtained from www.football-data.co.uk (results and betting odds), www.kicker.de and www.sport.de (in-match statistics) and www.fussballmafia.de (away fans, see Subsection 3.3). The codebook in the accompanying data set provides more details.

inclusion of in-match controls with how well these controls capture behavior (as measured by the R^2). We can quantify how much stronger the effect of *un*accounted behavior – relative to the effect of accounted in-match behavior – has to be to get a zero diff-in-diff estimate. Using her procedure, we find that it would need to be 11.5 times larger and in the opposite direction.[3]

3.2. Yellow cards

Column 2a in Table 1 presents the results for yellow cards. Pre-Covid19, home teams received on average 0.3, or 14.1%, yellow cards fewer than the 2.2 cards for away teams. The GhostMatch-estimate implies that during ghost matches, away teams got 0.3 yellow cards fewer compared to when audiences were present. The diff-in-diff estimate shows that this (negative) difference increases by 0.5 yellow cards for home teams.[4]

Importantly, our findings for yellow cards do not just mirror those for fouls. This is because they are *conditional* on fouls given (where, all else equal, about every eighth foul results in a card). Since a foul can – but does not have to – result in a card, these findings indicate a change in how harshly referees punished given fouls. Consistent with this notion, the inclusion of in-match controls as presented in Column 2b leaves the model's coefficients and explanatory power almost unchanged. According to Oster (2019), the effect of unobservables would need to be in the other direction and 8 times larger than the effect of observables to get a zero interaction effect.[5]

3.3. Supporter effects

Our results are in line with the argument that referees are affected by social pressure from the teams' supporters. Evidence against a change in playing style as an alternative explanation comes from the robustness of our results to including in-match controls.

To test directly for pressure by supporters, we use the average number of fans who travel with the visiting team. Based on a median split of this measure, we then repeat our preceding analysis. The underlying idea is that these fans act as a counteracting force to supporters of the home team. Thus, in matches with few away fans (i.e. below-median values), support for the home team and, therefore, pressure on the referee to favor the home team can be expected to be stronger. But during ghost matches such pressure by the supporters cannot materialize,

[3]That is, using Oster (2019)'s language, we find that $\delta = -11.5$. Part of this estimation is to assume a R^2_{max} from the (hypothetical) regression which includes unobserved in-match behavior. Following Oster (2019), p.189's recommendation, we compute it by multiplying the R^2 from the regression with observable in-match controls by 1.3 (i.e. we use the R^2 =0.372 from Column 1b in Table 1 to compute R^2_{max} =1.3×0.372).

[4]We get similar estimates if we include red cards and count them as three yellow cards (see Ponzo and Scoppa, 2016). A placebo check with the corresponding match data from the 2018/2019-season and "pseudo ghost matches" yields home effects for fouls (-0.558) and cards (-0.375) but not for the "pseudo ghost match"-dummy and its interaction.

[5]The sum of the Home-dummy and its interaction are positive but not significant for cards (F-tests: p≤0.073 for columns 1a/1b and p≥0.124 for columns 2a/2b in Table 1). This finding should be interpreted cautiously given its mixed nature. It is consistent, though, with the notion that pre-Corona, players could have been (implicitly) aware of the referees' home bias so that they played relatively aggressively at home. If players did not adjust their behavior to a new equilibrium where the referee punished home teams equally, we would expect such a positive sum.

Table 2 Effect of ghost matches on referee decisions; median-split by support for the away team

	no. of fouls given		no. of yellow cards	
	(1a)	(1b)	(2a)	(2b)
Home	-1.002**	-0.632	-0.375**	-0.302**
	(0.433)	(0.404)	(0.150)	(0.130)
GhostMatch	0.236	0.925**	-0.471***	-0.243
	(0.446)	(0.464)	(0.168)	(0.169)
Home×GhostMatch	2.008***	0.929	0.711***	0.322
	(0.580)	(0.672)	(0.225)	(0.214)
Fouls			0.120***	0.102***
			(0.017)	(0.015)
Away team support	<median	≥median	<median	≥median
In-match controls	yes	yes	yes	yes
R^2	0.432	0.400	0.337	0.311
Observations	612	612	612	612

Notes: OLS standard errors clustered on the match level in parentheses. ***/**/*: significant at 1/5/10%.

which is then captured by our diff-in-diff estimate. We therefore expect this estimate to be particularly strong for below-median matches (where pressure from the home team's supporters would have been pronounced), in contrast to matches with above-median support by away fans.

Table 2 displays the results. Column 1a shows that for below-median matches, the home team received fewer fouls pre-Corona. This changed during the ghost matches, as indicated by the significant diff-in-diff estimate. By contrast, Column 1b shows that there is no such effect for matches with above-median support for the away team. We also observe a very similar pattern for yellow cards in columns 2a and 2b. Thus, the shift in the home bias is concentrated in those matches where support for the away team would have been particularly weak (or, conversely, particularly strong for the home team).

4. Conclusion

We show that referees punish teams less severely when they play at home and this difference in the teams' treatment changed during ghost matches. We then demonstrate that this pattern is concentrated in matches where the away team's support is weak. A historical analysis, which we report in the appendix (see Table A.1), documents that from the 2005/06 to the 2018/19-season, home teams received, on average, 1.1 and 0.2 fewer fouls and yellow cards, respectively.[6] Our results from the Covid19-induced ghost matches in the 2019/20-season show that these historical differences could well have been caused by pressure from the audience and that a home bias still matters.

[6]While the difference for fouls shrank by 0.1 per year (the 1.1 is the average difference over the historical sample's 14 years), it still persists, as demonstrated in Figure 1a. For yellow cards, the home-effect has remained stable.

References

Dohmen, T. and J. Sauermann (2015). Referee Bias. *Journal of Economic Surveys 30*(4), 679–695.

Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization 65*(3-4), 636–653.

FIFA (2019). International Football Association Board - The Laws of the Game 2019/20.

Garicano, L., I. Palacios-Huerta, and C. Prendergast (2005). Favoritism Under Social Pressure. *The Review of Economics and Statistics 87*(2), 208–216.

Krumer, A. and M. Lechner (2017). Midweek Effect on Performance: Evidence from the German Soccer Bundesliga. *Economic Inquiry 56*(1), 193–207.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Pettersson-Lidbom, P. and M. Priks (2010). Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters 108*(2), 212–214.

Ponzo, M. and V. Scoppa (2016). Does the Home Advantage Depend on Crowd Support? Evidence From Same-Stadium Derbies. *Journal of Sports Economics 19*(4), 562–582.

For Online Publication

Home-Bias in Referee Decisions: Evidence from "Ghost Matches" during the Covid19-Pandemic

Appendix

Marek Endrich[1] & Tobias Gesche[2]

[1]marek.endrich@uni-hamburg.de, Institute of Law and Economics, University of Hamburg
[2]tgesche@ethz.ch, Center for Law & Economics, ETH Zurich

Historical analysis

In this appendix, we provide the analysis of historical data referred to in the conclusion. It is based on 4,284 matches in the German first division, which took place between the 2005/06 and the 2018/19-seasons. The estimates are obtained by fitting the following regression equation, similar to the one in the main text:

$$y_{i,m} = \alpha + \beta_1\, Home_{i,m} + \beta_2\, TimeTrend_m + \beta_3\, Home_{i,m} \times TimeTrend_m + \gamma\, TeamFE_i + \epsilon_{i,m}$$

In the above, $y_{i,m}$ and $Home_{i,m}$ are, as before, the dependent variable (fouls given or yellow cards received) and a dummy indicating whether team i in match m was the home team. $TimeTrend_m$ is a linear time trend and indicates the difference (in years) of the season in which match m is played to the 2005/06-season, the first in our historical sample. $TeamFE_i$ is a vector of team fixed effects.

Column 1a in Table A.1 presents the results when the number of cards is the dependent variable and the time trend is not included. The estimate for the $Home_{i,m}$-dummy indicates that the home team received, on average, about 1.1 fouls fewer than the away team over historical sample's 14 years. In Column 1b, we add the time trend and its interaction with the $Home_{i,m}$-dummy. While the non-interacted coefficient for the time trend shows that the number of fouls given to the away team decreased by about 0.6 per year, the interacted term shows that this decrease over time was significantly less pronounced, by about 1.0 fouls per year, for the home team.

Column 2a and 2b repeat the analysis with the number of yellow cards as the dependent variable, controlling for number of fouls. While Column 1a shows that historically, the home team also gets fewer yellow cards (about 0.2 cards fewer) and Column 2b shows that the number of yellow cards increased over time (by about 0.04 per year), this change in time was no different for the home than for the away team (insignificant, close to a zero interaction coefficient).

Table A.1 Decisions for the home team from a historical perspective (seasons 2005/06 – 2018/19; first division)

	no. of fouls given		no. of yellow cards	
	(1a)	(1b)	(2a)	(2b)
Home	-1.139***	-1.782***	-0.241***	-0.196***
	(0.088)	(0.175)	(0.024)	(0.046)
TimeTrend (years)		-0.586***		0.042***
		(0.018)		(0.005)
Home×TimeTrend (years)		0.099***		-0.005
		(0.021)		(0.006)
Fouls			0.082***	0.094***
			(0.003)	(0.003)
R^2	0.015	0.245	0.132	0.157
Observations	8,568	8,568	8,568	8,568

Notes: OLS standard errors clustered on the match level in parentheses. ***/**/*: significant at 1/5/10%.