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

ETH zürich

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

August 2020 (this 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

This version: August 7, 2020 (First version: June 29, 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 less fouls and yellow cards for the home team relative to the away team. During the ghost matches, these differences in fouls and cards changed so that home teams were treated less favorably than before. This effect is concentrated in matches where the relative share of supporters for the home team is large. 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 (Johnsallee 35, D-20148 Hamburg; marek.endrich@uni-hamburg.de)

[2]Center for Law & Economics ETH Zurich (IFW Building ETH Zurich, CH-8092 Zurich; tgesche@ethz.ch)

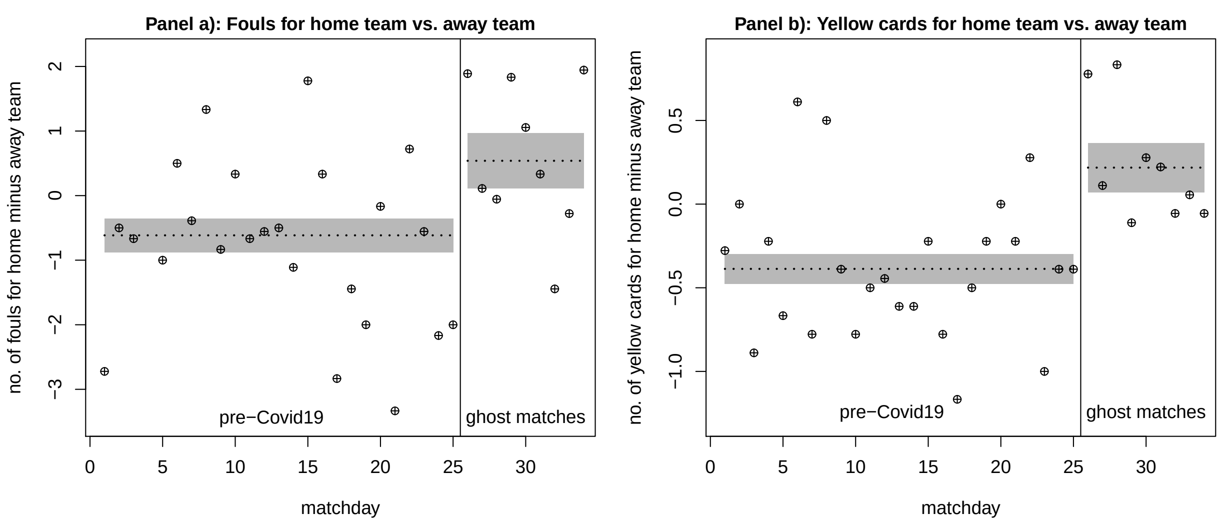
[3]We would like to thank Jerg Gutmann, Steve Heinke, Johannes Kunz, Max Mantei, Giuseppe Sorontini, and Stefan Zeisberger for helpful comments and input; we also acknowledge helpful comments by an anonymous referee. All remaining errors are, of course, our own.

1. Introduction

Working and 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 competing sides. Here, we provide evidence of such asymmetric audience effects on referee decisions. We exploit an 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 them to play in empty stadiums.

Our results provide evidence that referees favor the home team when there is an audience. First evidence for this is visualized in Figure 1a. It portrays the difference in the number of fouls whistled against the home team relative to the away team. On average, home teams received about 0.6 fouls less than away teams in the pre-Covid19 period, suggesting a referee bias towards the former. 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 *less* than the away team. This difference increases by about 0.6 *more* cards for the home team during the ghost games. In our main results section, we show that these effects are robust to controlling for referees and teams, teams' relative

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



Notes: Horizontal lines depict estimates with a dummy for ghost matches instead of the matchday fixed effects; grey rectangles depict the corresponding standard error of the mean. Points depict estimates from regressing the number of fouls (Panel a) or yellow cards (Panel b) on matchday fixed effects, a home team-dummy, and their interactions. The dotted horizontal lines depict estimates when a dummy for ghost matches and an intercept instead of the matchday fixed effects are used; grey rectangles depict the standard error of the mean.

strength (as measured by betting odds), and potential changes of in-match behavior. We also demonstrate that these effects are concentrated in matches where relative support by the audience for the home team is strong.

Our findings contribute to a literature on home bias in referee decisions such as timing, penalty, and calling goals (Garicano et al., 2005; Dohmen, 2008; Ponzo and Scoppa, 2016; for a survey see Dohmen and Sauermann 2015). Closest are Pettersson-Lidbom and Priks (2010), who study 21 Italian matches (3.6% of their observations) that were ordered to be without audience following hooligan violence. While in their setting, the affected teams and 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 a total of 165 matches (27.0% of our observations) in empty stadiums. We also use data on the number of avid supporters to trace out their effect on referee decision. Using this dataset and novel methods to measure the effect of in-match endogenous variables, we provide new evidence that home supporters can cause a referee bias that persists on several dimensions, even after years of improvements in refereeing techniques.[1]

2. Context, data, and model specification

Our investigation uses data from league 1 and 2 of German professional football during the 2019/20 season. From August 2019 on, 18 teams in each league played against every other team twice – once home, once away. Covid19 halted the season in mid-March and no matches took place for two months. Then, as society gradually re-opened, the league resumed and finished in June 2020. The following series of games were, for the first time ever, conducted as ghost matches without live audiences. In contrast, thousands watched players and referees before the break.

To estimate the effect of the ghost games on referee behavior, we use the following regression model:

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

In the above, $y_{i,m}$ measures our dependent variable of interest: either fouls (the referee calling a rule infringement; triggers a free kick for the other team) or yellow cards (given for repeated infringements or other misconducts).[2] Note that this specification means that each match m enters twice with i referring once to the home team and once to the away team (see, e.g., Garicano et al., 2005). To account for the resulting pair-wise correlation we cluster standard errors on matches.

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

[2] A second yellow card or a grave initial misconduct results in the player's dismissal (=red card).

The independent variables $Home_{i,m}$ and $GhostMatch_m$ indicate home- and ghost matches. Of primary interest is their interaction: the estimate for β_3 captures how referee decisions were affected by playing a ghost match at home as opposed to playing home pre-Covid19 with 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 a full set of fixed effects for each of the 2×18 teams and all referees. In addition, we control for the capacity of the hometeam's stadium and add weekday-dummies (see Krumer and Lechner, 2017).[3]

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.

3. Results

3.1. Fouls

Table 1 shows the results from fitting the model with our data. The dependent variable in Column 1a is the number of fouls received. The first coefficient represents the home effect pre-Covid19. When a live audience was present, the home teams received on average about 0.7 – or about 5.3% – fouls less than the away team (which got on average 12.3 fouls). Then, for ghost matches, the corresponding point estimate indicates that away teams got 0.5 more fouls than before but this change is not significant. The coefficient for the interaction term shows that for home teams, the effect of ghost matches is significantly larger by an additional 1.3 fouls.

This increase in the relative difference of fouls could potentially be due to omitted in-game behavior (such as player aggressiveness), which is affected differently by whether one plays home or away and whether an audience is present. We control for this directly by including a range of in-match behavior variables (Column 1b). Our main effect, the differential effect of ghost games 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. It contrasts the change of the diff-in-diff coefficient (Home×Ghost) upon inclusion of in-match controls with how well these controls capture behavior (as measured by the R-squared). We can quantify how much stronger the effect of *un*accounted behavior – relative to the effect of accounted in-game behavior – has to be to get a zero diff-in-diff estimate. Here, it needs to be 11.5 times larger and in the opposite direction.[4]

[3] Postponed matches count for the original matchday. The GhostMatch-dummy captures the audience at the actual date.

[4] Part of this estimation is to assume a R-squared from the (hypothetical) regression which includes unobserved in-match behavior. Following Oster (2019)'s recommendation, we multiply the R-squared from the regression with in-match controls by 1.3 (i.e., 1.3×0.372).

Table 1 Effect of ghost matches on referee decisions

	no. of fouls given		no. of yellow cards	
	(1a)	(1b)	(2a)	(2b)
Home	-0.650***	-0.505**	-0.311***	-0.278***
	(0.233)	(0.238)	(0.077)	(0.079)
GhostMatch	0.492	0.668**	-0.348***	-0.358***
	(0.313)	(0.303)	(0.111)	(0.113)
Home×GhostMatch	1.283***	1.343***	0.465***	0.477***
	(0.423)	(0.424)	(0.149)	(0.148)
Fouls			0.127***	0.115***
			(0.011)	(0.011)
In-match controls	no	yes	no	yes
R-squared	0.312	0.372	0.250	0.265
Observations	1,224	1,224	1,224	1,224

Notes: The dependent variable denotes the number of fouls (col. 1) or yellow cards (col. 2) given to the team under consideration. The independent variables indicate whether a match was a Ghost match and whether it was a home match; their interaction captures the effect of Ghost matches on the home effect. Control variables are standardized betting odds for the team under consideration, a full set of team and referee fixed effects, stadium capacity, and dummies for weekdays. In-match controls are the team's shots on goal, tackles, attempted and completed passes, ball possession, and running distance. Estimates are obtained by OLS; standard errors are in parentheses and clustered on the match level. $^{***}/^{**}/^{*}$ denotes significance at the 1/5/10%-level.

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 less than the 2.2 cards for away teams. The estimate for the GhostMatch-dummy implies that away teams got about another 0.3 yellow cards less when no audience was present. Relative to this decrease for away teams, the positive and significant interaction term shows that the ghost match-effect for home teams is 0.5 yellow cards higher.

It is important to note that this different effect of ghost matches on yellow cards for home teams is *conditional* on fouls given (where, all else equal, about every eighth foul results in a card). Converting a foul into a card is largely the referee's decision. Consistent with this notion, the inclusion of in-match controls leaves the model's coefficients and explanatory power almost unchanged (see Column 2b). According to Oster (2019)'s bounding procedure, the effect of unobservables would need to be negative and 8 times larger to get a zero interaction effect.[5]

[5] We get similar estimates if we include red cards, counting them as three yellow cards (see Ponzo and Scoppa, 2016). A placebo check with the corresponding match data from the 2018/2019-season with "pseudo ghost matches" yields home effects for fouls (-0.588) and cards (-0.375) but not for the "pseudo ghost match"-dummy and its interaction.

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

	no. of fouls given		no. of yellow cards	
	(1a)	(1b)	(2a)	(2b)
Home	-0.995**	-0.330	-0.298**	-0.206
	(0.402)	(0.380)	(0.140)	(0.131)
GhostMatch	0.128	1.086**	-0.675***	-0.070
	(0.389)	(0.477)	(0.158)	(0.181)
Home×GhostMatch	1.963***	0.816	0.814***	0.209
	(0.570)	(0.649)	(0.218)	(0.211)
Fouls			0.130***	0.097***
			(0.0157)	(0.016)
Away team support (rel.)	<median	≥median	<median	≥median
In-match controls	yes	yes	yes	yes
R-squared	0.461	0.400	0.333	0.304
Observations	612	612	612	612

Notes: The dependent variable denotes the number of fouls (col. 1) or yellow cards (col. 2) given to the team under consideration. The independent variables indicate whether a match was a Ghost match and whether it was a home match; their interaction captures the effect of Ghost matches on the home effect. Control variables are standardized betting odds for the team under consideration, a full set of team and referee fixed effects, stadium capacity, and dummies for weekdays. In-match controls are the team's shots on goal, tackles, attempted and completed passes, ball possession, and running distance. "Away team support (rel.)" is the number of fans that do, on average, travel with the guest team divided by the corresponding number for the home team and multiplied by stadium capacity. Estimates are obtained by OLS; standard errors are in parentheses and clustered on the match level. ***/**/* denotes significance at the 1/5/10%-level.

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 an alternative channel in which a change in playing style caused these results comes from including in-match controls, which do not change our estimates for the asymmetric home-effect of ghost matches on fouls. Further evidence against this channel comes from observing a similar pattern for yellow cards. As this result is conditional on fouls, it captures referee decisions *given* (perceived) player behavior.[6]

To directly test for the effect of avid supporters, we use the average number of fans who travel with a team to away matches ($= \bar{F}_i$). We then take the ratio of this number for the away team and the average number of away fans who travel with the home team (which we use as a proxy for avid supporters for the home team, as opposed to rather passive observers), i.e., we use $\bar{F}_{Away}/\bar{F}_{Home}$. To capture absolute supporter effects, this ratio is multiplied with the stadium's capacity. This yields a match-by-match measure of the number of avid supporters

[6]The sum of the Home-dummy and its interaction are positive, but not significant for cards (F-tests: $p \leq 0.073$ for columns 1a/1b and $p \geq 0.120$ for columns 2a/2b in Table 1). Given its mixed nature, this finding should be interpreted cautiously. 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 aggressive at home. If player did not adjust their behavior to a new equilibrium where the referee punished home teams equally, we would expect such a positive sum.

for the away team relative to avid supporters for the home team. We then use a median split of this measure to investigate how supporters influenced (or, for ghost matches, could have influenced) referee decisions.

Table 2 displays the results. Column 1a shows that when support for the away team relative to the home team was weak (below-median values for our measure), the home team received less fouls pre-Corona while this changed during the ghost matches. In contrast, there are no such effects for matches with relatively high support for the away team and low support for the home team (supporter measure above median, Column 1b). 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 relative support for the home team would have been particularly strong (or, conversely, particularly weak for the away team).

4. Conclusion

We show that referees punish teams less severely when they play at home and that this changed during ghost matches. We also show that this pattern is particularly pronounced when the home team's relative support is strong. Historically, previous seasons have also exhibited less fouls and yellow cards for the home team (on average, -1.139 and -0.238 over 2005/06 – 2018/19).[7] Our findings from ghost matches during the 2019/20-season suggest that pressure from the audience is behind these historical differences and that this channel still matters.

[7] This is for league 1, on which our historical data are limited. While the difference for fouls shrank by 0.099 per year (the -1.139 is the average difference over the years), it still persists as demonstrated in Figure 1. For yellow cards, the difference 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.

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.