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

June 2020

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

Marek Endrich

Unversity of Hamburg

Tobias Gesche

ETH Zurich

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 relative to the pre-Covid19 period, the difference between home and away teams in fouls and the number of cards increases. 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. Introduction

Working and competing in front of an audience can be a double-edged sword: On the one hand, being in the focus of others can spur effort, increase competition, and yield better prices or quality. On the other hand, 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. In this paper, 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 incduced 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 exogenous and sudden change forced them to play in empty stadiums.

Our results provide evidence of referees systematically favoring the home team when playing in front of an audience. The first part of this evidence is visualized in Figure 1, which 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 the period of the ghost matches, however, the relative number of fouls for the home team increased by about one. Figure 1b shows a similar pattern over time for the number of cards that referees gave. Before Covid19, the average home team received 0.4 yellow cards *less* than the away team. This difference then increased by about 0.6 *more* yellow cards for the home team during the ghost games. In our main results section, we show that these effects remain

Figure 1: Difference in fouls and yellow cards over the season
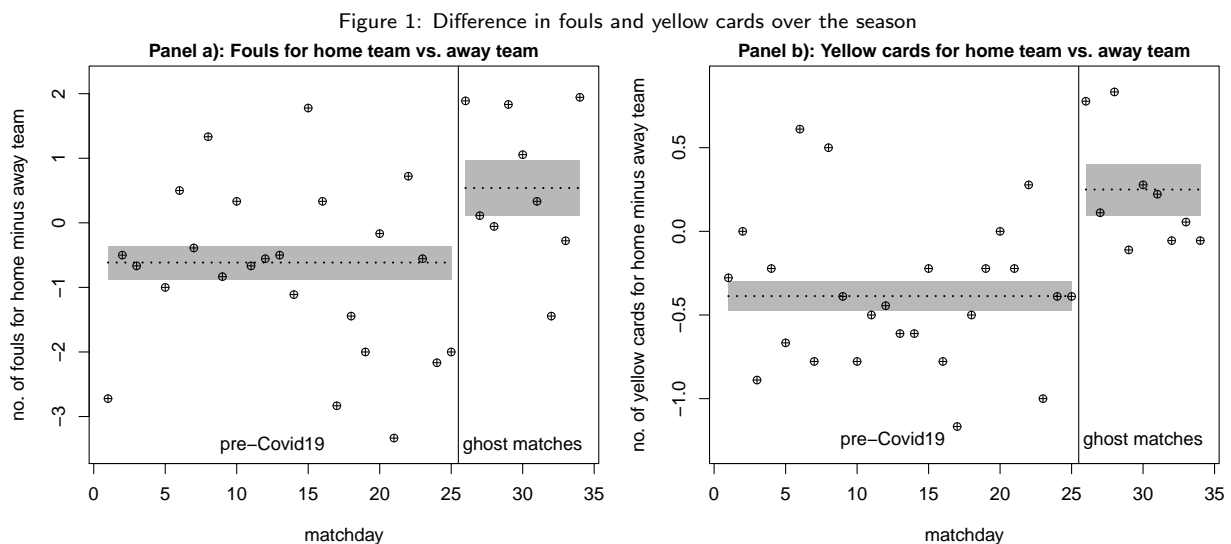
Notes: Points depict estimates from regressing the number of fouls (Panel a) or yellow cards (Panel b) on matchday fixed effects, a home team-dummy, and their interactions. The dotted horizontal lines depict estimates when a dummy for ghost matches and an intercept instead of the matchday fixed effects are used; grey rectangles depict the standard error of the mean.

stable and significant if we control for referee and team fixed effects, the relative strength of teams (as measured by betting odds), audience sizes, and variables controlling for potential changes of in-match behavior by players.

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). Pettersson-Lidbom and Priks (2010), who studied the effect of ghost games in Italy that were a result of an eruption of hooligan violence, are closest to our study. They studied 21 matches (that might have been non-randomly determined), whereas in our data, *all* competing teams had to play a quarter of the season (145 matches) in empty stadiums.[1] Using this encompassing dataset and new methods to measure potential changes from in-match endogenous variables, we provide new evidence that a referee home bias persists on several dimensions, even after years of improvements in refereeing techniques.

2. Data and results

2.1. Context, data, and model specification

Our investigation uses data from the top two divisions of Germany's professional football league ("1. Bundesliga" and "2. Bundesliga") during the 2019/20 season. Both divisions use a double all-play-all format, in which each team plays against every other team twice – once home, once away. Each division has 18 teams with 612 matches between August 2019 and June 2020. Covid19 halted the season in mid-March and no matches took place for two months. Then, as society gradually re-opened, the league resumed. Crucially, these matches had to occur without live audiences, whereas previously, tens of thousands watched the players and referees.

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 (e.g., the number of fouls). Note that this specification means that each match m enters twice; once where i refers to the home team and once where i refers to the away team (see, e.g., Garicano et al., 2005, Ponzo and Scoppa, 2016, Belchior, 2020). To account for the resulting pair-wise correlation we cluster standard errors on matches (see Abadie et al., 2017).

The independent variables $Home_{i,m}$, $GhostMatch_m$ indicate home- and ghost matches. The main independent variable of interest is the interaction of these two dummies: The estimate for β_3 captures how referee

---

[1]Whereas due to Covid19, the duration of other leagues exceeded many players' contract ending in June (creating incentive- and hold-up problems), Bundesliga ended within that month.

decisions were affected by playing a ghost match at home as opposed to playing home pre-Covid19 with an audience (predominantly made up the home team's supporters).[2]

The $Controls_{i,m}$-vector collects additional independent variables. Firstly, we add an index of betting odds for team i winning to control for relative team strength and winning incentives at the season's end. We also add a full set of referee and team fixed effects. Finally, we control for the pre-Covid19 average audience in the hometeam's stadium and weekday-dummies (see Krumer and Lechner, 2017).[3]

In addition, we add a comprehensive set of controls for in-match player behavior. These are the team's shots on goal, tackles, attempted and completed passes, ball possession, and running distance.

3. Results

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 12.3 fouls for away teams. Then, for ghost matches, the corresponding point estimate indicates 0.5 more fouls for the away team but 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 an omitted variable (such as player aggressiveness) that is related to playing home or away and to whether an audience is present. We control for this by including a range of in-match controls. Our main effect, the differential effect of ghost games on home relative to away teams, remains positive and significant.

To pin down the effect of changes in in-match behavior, we use a bounding method by Oster (2019). It contrasts the change in the diff-in-diff estimate upon inclusion of the in-match controls with how well these controls capture behavior (as measured by the resulting change of the R-squared). We find that in order to drive this estimate down to zero, the relationship between outcomes and unobserved in-game behavior would need to be 11.5 times larger (and negative) than what we capture.[4] We summarize these findings as follows:

Result 1. *The (initially negative) difference in the number of fouls given to the home team relative to away teams increases during ghost matches.*

---

[2]With the resumption after the Covid19-pause, a few new game rules (5 instead of 3 player substitutions) were introduced. However, this affected all teams and should therefore not affect the estimates for the Home-dummy and/or its interaction.

[3]Postponed matches are counted for the originally scheduled matchday. The GhostMatch-dummy captures the audience-situation when these matches actually occurred.

[4]We follow Oster (2019) and multiply the R-squared from the regressions with in-match controls by 1.3 to get an estimate for the R-squared from the (hypothetical) regressions that also includes unobserved in-match behavior.

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	0.371	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, average audience size in the stadium where the match was played and dummies for weekdays. In-match controls are the team's 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.

In Column 2a, we present the results when the dependent variable is the number of 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 following estimate for the GhostMatch-dummy implies that away teams then 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 the ghost match-effect for home teams is 0.5 yellow cards higher.

It is important to note that this differential effect of ghost matches on yellow cards for home and away teams is *conditional* on fouls given (where we find that, 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 for player behavior leads to only limited changes in the coefficients estimated and the model's explanatory power (see Column 2b). According to Oster (2019)'s bounding procedure, the effect of unobservables would need to be negative and 8 times larger to drive the interaction effect to zero. Together, these finding lead us to state the following main result:[5]

Result 2. *Controlling for the referee decision to give a foul, the (initially negative) difference in the number of yellow cards given to the home team relative to away teams increases during ghost matches.*

---

[5]We get similar estimates if we follow Ponzo and Scoppa (2016) and use the number of red cards plus three times yellow cards. In addition, we performed a placebo check on the corresponding match data from the 2018/2019-season with "pseudo ghost matches". We find significantly negative home effects for fouls (-0.586) and yellow cards (-0.377) but not for the dummy indicating pseudo ghost matches or its interaction.

4. Discussion and Conclusion

We find that ghost matches lead referees to give more fouls (Result 1) and yellow cards (Result 2) for the home team. An explanation for *both* effects is that referees are affected by social pressure from the rank.

In principle, a different playing style in ghost matches could also have affected referee decisions (e.g., more aggressive player behavior). However, in order to explain our Result 1, this change in player behavior would need to be particularly pronounced for home matches but not when playing away. Further evidence comes from our controls for potential changes in player behavior through a set of comprehensive in-match controls. While these controls do capture actual behavior, they do not substantially change our estimates for the asymmetric effect of ghost matches on referee decisions. Finally, the observation of a similar asymmetric effect for yellow cards stands against player behavior as a unifying explanation. As Result 2 is conditional on fouls given, it captures referee behavior *given* (perceived) player behavior.

We therefore interpret our Result 2 as providing evidence towards a referee bias on the intensive margin (=punishing fouls by the home team less harshly through less yellow cards) that changed during ghost games. In line with this, Result 1 supports the notion of a referee bias on the extensive margin (=calling less fouls for the home team). Together, our findings highlight how audience creates pressure that affects referee decisions on multiple dimensions.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When Should You Adjust Standard Errors for Clustering? Working Paper 24003, National Bureau of Economic Research.

Belchior, C. A. (2020). Fans and Match Results: Evidence From a Natural Experiment in Brazil. *Journal of Sports Economics*.

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.