Match Fixing and Sports Betting in Football: Empirical Evidence from the German Bundesliga

Christian Deutscher*
University of Bielefeld

Eugen Dimant
University of Pennsylvania

Brad R. Humphreys†
West Virginia University

Adam Nowak
West Virginia University

August 30, 2017

Abstract

Corruption in sports is a challenge to its integrity. Corruption can take many forms, including match fixing, which can be difficult to detect. We use a unique data set to analyze variation in bet volume on Betfair, an online betting exchange, for evidence of abnormal patterns associated with specific referees officiating matches. An analysis of 1,251 Bundesliga 1 football matches from five seasons reveals evidence of systematically higher bet volume for two referees relative to matches officiated by all other referees; results from a randomization experiment using Fischer exact p-values confirm this result. Our results are robust to alternative specifications.

Key words: corruption, betting exchange, match fixing, Fisher exact p-values

JEL Codes: D73, K42, L8, Z2

Introduction

Corruption in sports is a growing problem with new allegations of match fixing regularly appearing in the media. Since match fixers profit by placing bets on matches with pre-determined outcomes, evidence of match fixing usually presents itself as unusual patterns in aggregated data from betting markets (Forrest and Simmons, 2003). We extend the approach of Wolfers (2006) and investigate the idea that evidence of match fixing can be found in available data from a popular international betting exchange market, Betfair.

Sports betting is a growing industry and has become an integral tool for profiting from fixed matches. According to the European Gaming & Betting Association, regulated betting accounts

*We thank Michael Lechner, Victor Matheson and Amanda Ross as well as participants at the WEAI Portland 2016 and ESEA 2016 conference for comments on earlier versions of this paper.

†College of Business & Economics, 1601 University Ave., PO Box 6025, Morgantown, WV 26506-6025, USA; Email: brhumphreys@mail.wvu.edu.

for some \$58 billion yearly and is forecast to have reached \$70 billion in 2016; football (soccer in North America) accounts for about 70-85% of the bets placed (Foley-Train, 2014).

Economic models of match fixing predict that referees are prime candidates for corruption, since they can exert a strong influence on match outcomes and receive relatively low levels of compensation. Betting exchange markets provide convenient, highly liquid markets where match fixers can profit from influence on outcomes in sporting events. Previous match fixing scandals contain evidence of referee involvement.

Matches can be fixed in numerous ways and successful fixing usually involves many different individuals, including team managers, staff, players, and match officials. Matches can be fixed on numerous margins including outcome (home win, draw, away win), goals scored and other outcomes colloquially referred to as "proposition bets" in gambling markets (LaBrie et al., 2007). We focus on the role played by referees in conjunction with specific wagers on the total number of goals scored in football matches that can be linked to match fixing. This type of match fixing requires a small number of initiators, increasing the individual benefit for all parties involved. With their career at stake if detected, a referee takes on a huge risk when fixing a match.[1]

Match fixing by referees is an intentional form of referee bias. Dohmen and Sauermann (2016) survey the large literature on this topic that focuses on the idea of referees making advantageous decisions toward the home team as a result from social pressure by the home fans. In this context, referees can be understood as intermediaries who engage in corrupt behavior (Dusha, 2015).

The source of the exchange betting market data, Betfair, bills itself as the "worlds largest betting exchange" and, in 2015, reported 1.7 million active customers and a turnover of 475.6 million British pounds (approximately \$694 million). We capitalize on a unique data set to analyze variation in the volume of Betfair wagers on the total number of goals scored in German Bundesliga 1 games in the 2010/11 2014/15 seasons. We posit that match fixing, in terms of total goals scored, is more likely to occur than match fixing in terms of win/loss/draw outcomes. This manner of match fixing has a lower detection rate and, as a result, is less risky to the parties involved in the fixing process. In addition, unbiased referees should not affect the expected number of goals scored in a match.

Regression models with control variables capture specific match characteristics, including the identity of the referee for each match and unobservable home team-, away team-, matchday-, and season-level heterogeneity. Regression results indicate that Betfair betting volume on the proposition that games end with under 2.5 total goals scored was higher in games refereed by three specific Bundesliga referees over this period, even when controlling for team and match level observable and unobservable factors that might affect goal scoring. A randomization experiment holding match characteristics constant while randomly assigning referees to other matches generates randomization distributions and Fischer exact p-values for individual referee fixed effects parameter estimates. The results of these randomization experiments indicate that bet volume in the under 2.5 goals scored market on Betfair was unusually high for two referees under the null hypothesis

[1]Boeri and Severgnini (2011) emphasize the importance of expected future career earnings by referees in influencing their decisions to provide unbiased adjudication of play; they point out that referees involved in the *Calciopoli* match fixing scandal in Italy were coerced into influencing match outcomes rather than bribed.

that bet volume was identical for all referees in the sample. These results are consistent with the hypothesis that corruption might have influenced some Bundesliga 1 match outcomes over this period.

Literature Review and Context

Although the quality and extent of media coverage varies (Di Tella and Franceschelli, 2011), corruption in general, but particularly in sports, is a ubiquitous issue in both amateur and professional settings, especially in the form of match fixing. Corruption imposes adverse effects on both the society and economy. These impacts alone highlight the importance of gaining an understanding of the underlying mechanisms of match fixing in order to better inform policy makers (Rose-Ackerman and Palifka, 2016). For a comprehensive discussion of the mechanisms of corrupt behavior, as well as, the empirical findings on the causes and effects see Dimant and Schulte (2016) and Dimant and Tosato (2017). As a subset of corruption, match fixing represents a substantial threat to the integrity of sports (Dimant and Deutscher, 2014). It is a widespread phenomenon and has been uncovered in sports such as Basketball, Cricket, Football, Sumo Wrestling and Tennis. Such distortions create negative externalities not only at the individual level, but also at the aggregate level, such as loss of media interest.

Additionally, they erode the inherent principle of fair and competitive sports. Scholars have proposed a number of mechanisms to deal with match fixing, some of which have been implemented by policy makers with differing levels of success (Carpenter, 2012). Forrest et al. (2008) discuss the economic incentives that influence match fixing in sport from a more general view. They point out that the emergence of betting exchange markets like Betfair increases the incentives to fix matches since they provide enhanced opportunities to benefit financially from match fixing due to a quick and fairly anonymous exchange of money.

Substantial literature exists on the economics of match fixing in sports. Preston and Szymanski (2003) developed a game theoretic model of strategic interaction between bettors, bookmakers and participants in sporting events. The model assumes that participants in sporting events may be susceptible to corruption, given sufficient monetary incentives, and shows that the likelihood of corruption increases as the legal compensation of the participants decreases. This prediction implies that referees are prime candidates for corruption, because of their relatively low levels of compensation, especially in comparison to coaches and players (Forrest et al., 2008).

Pohlkamp (2014) reports compensation rates for Bundesliga 1 referees at €3,800 per match with no base compensation in 2009. In comparison, Premier League referees earned base salaries of €38,500 per season and an additional €1,170 per match in 2009. Referees in Bundesliga 1 pursue other jobs besides refereeing. Referees other professions range from being dentists to lawyers. Preston and Szymanski (2003) point out that a referee can have a larger impact on match outcomes than most players can, which also makes them prime candidates for corruption. Since the expected returns of wrongful behavior are negative if said behavior is uncovered, increasing referee compensation can be interpreted as reducing incentives to cheat. Premier League referees who switch from short-term contracts to salaried contracts show improved performance relative to those who do not (Bryson et al., 2011).

In line with this argument, Forrest and Simmons (2003) developed a model to explain match fixing in sports based on the expected costs and benefits of match fixing. This model also suggests that referees are likely candidates for corruption, in that the probability that an individual will take actions to affect the outcome of a sporting event increases with the likelihood that the actions succeed in affecting the outcome. The probability that the actions of an individual referee can influence outcomes exceeds the probability that coaches and almost all players can influence outcomes.

Forrest (2012) surveyed a number of recent match fixing scandals in sports, including the infamous 2011 “Bochum trial” in Germany where evidence of match fixing in more than 300 European football matches, including 53 in Germany, was introduced. Several of the match fixing scandals discussed by Forrest (2012) involved referees. Feltes (2013) discusses two key cases of match fixing in Germany involving referee corruption: the 2005 Hoyzer case and the 2011 “Bochum trial” involving Ante Sapina. Robert Hoyzer, a Bundesliga referee, was found guilty of fixing 23 Bundesliga matches and convicted of fraud. He was caught after a number of egregious calls, including awarding two penalties to SC Paderborn in a surprising 4-2 win after trailing 0-2 in a 2004 match against Hamburg, as well as, ejecting Hamburg’s star striker for misconduct. Hoyzer implicated Sapina as the source of his bribes, but Sapina was not prosecuted until years later.

Ante Sapina, the leader of a betting syndicate, was found guilty of fixing 32 football matches in Germany. Both referees and players were involved in the Sapina match fixing scandal. Two referees implicated in this scandal were banned for life by FIFA and UEFA, including Bosnian referee Novo Panic. Among the details emerging from the Sapina trial, one most relevant to our current work was his betting on the number of goals scored in matches among other “proposition” bets like the number of free kicks taken in a match.

Following the Hoyzer case, the committee of control for the Bundesliga (DFB) reacted by taking a number of actions, most prominently reducing the time between the designation of the referee for a match and the playing of the match. Aimed at minimizing the time available to fix football matches, the League proposed, but failed to implement, a planned two day notice before the match, because that process was ruled impractical. Instead, following the Hoyzer case, referees are assigned to matches according to the following schedule: referees for Friday games are announced on Wednesday and the referee assignments for games played between Saturday and Monday follows on Thursday; the announcement for Tuesday and Wednesday games happens on Fridays. This procedure features a time span of between two and five days between the referee assignment and the match, depending on the day of the match. In terms of the identity of the referee, the experience of the referee determines the chances a referee is assigned to a certain game; decisive games at the end of the season are assigned to experienced referees. To assure unpredictability of assignments, no clear referee assignment mechanism has been announced by the league. Possible limitations of this process for our analysis are described in the last section of the paper.

In another famous incident, Boeri and Severgnini (2011) analyze referee participation in the Calciopoli match fixing scandal in Italy in 2006. Using evidence about specific episodes of match fixing in Serie A uncovered through phone taps and other methods used in a criminal investigation of football match fixing, Boeri and Severgnini (2011) demonstrate how club officials used threats to adversely affect the career, and future earnings, of Italian football referees to Juventus's direct or indirect benefit. Referees actions included issuing red cards to key players in matches immediately before a team was scheduled to play Juventus (disqualifying said player from the next match), incorrectly ruling (or failing to rule) players offside or not ruling players offside and other subtle actions. The corrupted referees did not take overt actions like assessing red cards to opposing players in matches involving Juventus or awarding Juventus penalty kicks in important matches.

Various factors motivate our analysis of variation explicitly in bet volume in the Over 2.5 and Under 2.5 markets on Betfair. First, the media and the public are likely less sensitized to questionable calls that result in additional goals compared to dubious calls that change the outcome of the contest. Second, when comparing benefits and costs of match fixing, taking actions that affect the total number of goals scored reduces the chances of a referee getting caught, thus increasing the attractiveness of match fixing. For referees not being bribed, there should, ceteris paribus, be no systematic differences in the amount of money bet on the number of goals scored in a football match.[2]

Closely related is the case of Tim Donaghy and the 2007 NBA betting scandal in which the former referee bet money on the over proposition (the proposition that more than a specific number of points would be scored) in games he refereed (Lockwood, 2008).

Empirical Analysis

In order to develop evidence consistent with the presence of match fixing in the Bundesliga 1, we estimate reduced form models of the determination of betting volume on specific bets placed on Betfair. Match fixing occurs, because some individual or organization wants to profit from sporting event outcomes by influencing the outcome of these events in predictable ways. We assume that matches with an unusually large bet volume, or with unusual patterns in said bet volume, could potentially be fixed in some way, reflecting bets made by the match fixers.

Match fixing, and profits from match fixing, could take many forms. For example, in sporting events with prizes for winners, match fixing could involve payoffs to participants to guarantee specific outcomes; in a foot race with a first prize of $1,000 and a second prize of $500, the two fastest runners could agree before the race to split the sum of the first and second prizes equally. However, a simpler way to profit from match fixing is to bet on some match outcome that has been

[2]If non-bribed referees systematically differ in their evaluation of fouls, systematic differences in the money bet on over could occur. Imagine one referee is known to not call fouls and let the game continue in controversial situations. This knowledge could lead to statistically higher volume on over / under betting. This paper tries to control for this by using referee decision making as control variables (number of cards, penalties etc.)

determined in advance, or that some participant in the match has been paid or coerced to influence in a specific way. The Ante Sapina case discussed above revolved around profits earned by betting on fixed football matches in Germany.

An individual or organization attempting to profit from match fixing by betting on Betfair would need to place bets on an outcome that would be relatively straightforward to influence and relatively difficult to detect. The Appendix describes the types of betting markets available on Betfair. The largest Betfair football betting markets, in terms of bet volume, are the match odds markets (home win, draw, away win) and the over/under 2.5 goals markets. To profit on match odds betting, the match fixer would have to influence the outcome of the match, a more easily detected form of corruption due to the reasons described above. Exact match score markets have low volume and this outcome would be relatively difficult to fix. The over/under 0.5 and 1.5 goal markets also have low volume, so individual high volume bets would likely be identified as suspicious by market monitoring systems.

Based on these factors, the over/under 2.5 goal market appears to be a likely candidate for match fixers looking to earn profits of fixed matches to exploit. In the 1,530 Bundesliga 1 matches played in the 2010/11 through 2014/15 seasons, the average number of goals scored in a match was 2.92. Two or fewer goals were scored in 44% of the matches and three or more goals were scored in 56% of the matches. A player or referee would not need to influence scoring in a glaringly obvious way in order to influence scoring over/under 2.5 goals.

Bet volume can clearly be influenced by factors unrelated to match fixing. Bettors may prefer to wager on more popular teams, on teams with star strikers, on teams playing opponents with weak defenders, or simply prefer to wager in the Over 2.5 market, because they prefer matches with more scoring. In any event, if the Over 2.5 and Under 2.5 betting markets are weak form efficient, then all public information affecting match outcomes, including referee effects, should be reflected in betting odds.

Data

The data we used comes from Betfair, an on-line betting exchange founded in the UK in 1999. Betting exchanges like Betfair allow bettors to both back (bet that an athlete or team will win a sporting event, or bet that some event will occur) or lay (bet that an athlete or team will not win a sporting event, or an event will not occur) on any sporting event. Traditional betting with a bookmaker involves the bettor backing and the bookmaker laying on each transaction. On a betting exchange, each wager must be matched: at least one backer and one layer must agree to wager a specific amount of money at stated odds on a specific event. Betfair quickly matches backers and layers. Sometimes multiple backers and layers are matched at stated odds which allows for wagering both before and during (in-play) sporting events.

We obtained data on Betfair betting prior to football matches in Bundesliga 1, the top football league in Germany, over the 2010/11 through 2014/15 seasons at the match level. The data set contains 1,251 football matches. Here we only look at the bets made before the play started ("pre-play transactions"). The outcome variable of interest is the total volume of bets matched, in Pounds, for specific Betfair betting markets. We focus on two betting markets: bets that more than 2.5 goals will be scored in the football match (Over 2.5) and bets that fewer than 2.5 goals will be scored in the football match (Under 2.5).

We augmented the Betfair transactional data with information on match outcomes. We obtained the grade on a 1 to 6 scale (with 1 as the best performance grade and 6 as the worst) for the referee in each match from Kicker, a popular German football magazine. These grades represent assessments of the performance of each referee in each match; research suggests very good grades to increase nomination chances to succeeding games (Frick et al., 2008). Additionally, we obtained the name of the referee in each match and referee performance data (red and yellow cards given, penalty kicks given) from the German Football Association.

Table 1: Summary Statistics

	Mean	Standard Dev
Bet volume - over 2.5 goals (£)	33,103	36,904
Bet volume - under 2.5 goals (£)	22,456	38,22
Kicker Referee Grade	3.219	1.140
Home yellow cards	1.566	1.148
Home red cards	0.045	0.211
Home penalties	0.154	0.376
Home corners	5.421	2.900
Away yellow cards	1.967	1.230
Away red cards	0.054	0.230
Away penalties	0.104	0.328
Away corners	4.230	2.430

Table 1 contains summary statistics. The average volume of bets matched in the Over 2.5 market was about £33 thousand and the average volume of bets matched in the Under 2.5 market was about £22 thousand.

The football matches in the sample were officiated by 26 different referees. Table 4 shows the number of games officiated by each referee in the sample. The paper focuses on analyzing variation in bet volume in the Over 2.5 and Under 2.5 markets by referees. Figure 1 summarizes the variation of interest using box plots of bet volume in the Over 2.5 Goals (left panel) and Under 2.5 Goals (right panel) betting market for each referee. The box identifies the 75th and 25th percentile of the distribution for each referee, the interquartile range (IQR), and the lines identify the median values. The whiskers identify the smallest and largest values within 1.5 IQR of the nearest quartile, while the dots above the top whisker represent extreme values. The red line is the sample mean in each market.

As shown in Figure 1, quite a bit of variability exists in bet volume in both markets on Betfair across these 26 referees. The median value for each referee is lower than the average, indicating skew in the distribution for each referee. Figure 1 also shows a handful of very high volume

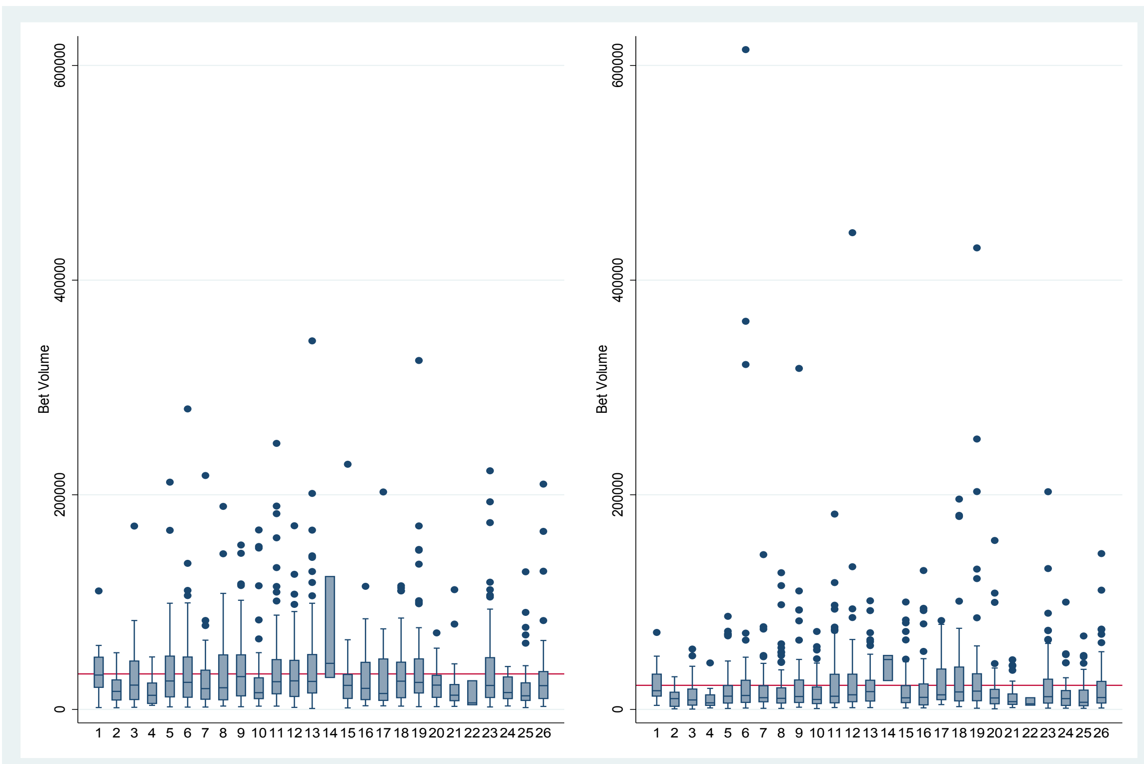

Figure 1: Bet Volume by Referee - Over (left) and Under (right) 2.5 Goals

football matches in the sample, including three matches in the Over 2.5 goal market with more than £400,000 in bet volume.

The average Kicker grade for a referee was 3.2 on the 1 to 6 German "school mark" grading scale. 1 is the best possible grade and 6 the worst possible grade. The Kicker grades are assessed in 0.5 unit increments. A referee received the highest grade (1.0 or 1.5) in only 4% of the matches. About 60% of the matches in the sample resulted in the referee getting a grade of between 2.0 and 4.0 ("good", "satisfactory" or "sufficient").

On average, home teams received about 1.5 yellow cards and away team received almost two yellow cards. Red cards were quite rare in the sample. Away teams also received more red cards on average. Penalty kicks were relatively rare, occurring in only about 1 in 10 matches. Home teams were awarded more penalty kicks on average (0.154 per game) than away teams (0.104 per game). This is in line with the literature on home bias in football (Dohmen and Sauermann, 2016).

Empirical Model

The aim of this study is to estimate the impact of referees on betting volume, controlling for game characteristics and visible referee statistics. The reduced form models of the determination of bet volume at the match level for the Over 2.5 and Under 2.5 markets on Betfair take the form

$$VOL_{ijrws} = \alpha_h HT_i + \alpha_v VT_j + \alpha_w MW_w + \alpha_s SEAS_s + \gamma_r REF_r + \beta X_{ijrws} + e_{ijrws} \tag{1}$$

where VOL_{ijrws} is the bet volume matched in one of the two exchange markets on a football match between home team i and visiting team j refereed by referee r in match week w of season s. Football matches occur in settings with substantial unobservable heterogeneity; teams face different constraints and incentives in terms of the effect of a draw or win on the teams expected success.

We control for unobservable match-level heterogeneity using a number of fixed effects in the regression model. HT_i is a vector of home-team specific indicator variables that reflect unobservable heterogeneity in home teams. These variables reflect team popularity, roster composition, pitch characteristics, fan characteristics and other unobservable home-team specific factors that affect bet volume on Betfair. VT_j is a vector of visiting-team specific indicator variables. MW_w is a vector of match week specific indicator variables. $SEAS_s$ is a vector of season-specific indicator variables. Boeri and Severgnini (2013) show that the probability of match fixing changes systematically over the course of a season. X_{ijrws} is a vector of match-specific characteristics, including variables reflecting observed referee performance.

e_{ijrws} is a random variable that captures all other factors that affect the matched bet volume in the Over 2.5 and Under 2.5 markets for each football match on Betfair. This random variable is assumed to be mean zero and possibly heteroscedastic.

Equation (1) also contains a vector of indicator variables for the referee in each match. Our sample includes 26 different referees. Two of these referees officiated only 6 Bundesliga 1 matches in our sample. The average referee in the sample officiated 48 matches and some refereed more than 130 matches. The vector of parameters on these referee-specific indicators, γ_r, are the parameters of interest here. These parameter estimates reflect the average effect of each referee on the volume of bets matched in the Over/Under 2.5 goal markets on Betfair. On average it would be expected that the betting volume on the proposition that more or less than 2.5 goals are scored in a football match would not be systematically related to the identity of the referee of a match, after controlling for other match-related factors, including unobservable heterogeneity related to teams, match weeks and seasons. If the estimated parameter on a referee-specific indicator is statistically different from zero, then the bet volume on matches worked by that referee are higher than average. This could reflect match fixing.

X_{ijrws} is a vector of match-specific variables. We include the Kicker grade for the referee and various variables reflecting referee performance in each match (including the number of red and yellow cards issued and the number of penalty kicks and corner kicks awarded) in this vector. The αs, γs and βs are unobservable parameters to be estimated.

Results

Table 2 shows results from estimating Equation (1) using OLS, with the estimated standard errors corrected for heteroscedasticity using the standard White/Huber "sandwich" correction. The dependent variable is bet volume in the Betfair Over 2.5 market in each match.

The baseline specification, Column (1), sets X_{ijrws} equal to zero and includes only home team, visiting team, match week, season, and referee indicator variables. This model explains 22.6% of the observed variation in bet volume in the Over 2.5 market. Two of the parameters on the referee-specific indicator variables are significantly different from zero at the 5% level in this model, suggesting that bet volume was above average in this market for matches refereed by these referees compared to the other referees in the sample.

Model specification (2) adds the Kicker referee grade variable to the model. This assumes that the Kicker grade reflects whatever actions each specific referee took during the match, and ultimately whether such actions were viewed as reasonable or unreasonable. Model specification (2), like specification (1), explains 21.6% of the observed variation in matched bet volume on Betfair in this market. The estimated parameter on the Kicker referee grade variable is not statistically different from zero. Again, two of the parameters on the referee-specific indicator variables are significantly different from zero at the 5% level in this model for the same two referees, suggesting that betting was above average in this market for the matches they refereed.

Model specification (3) replaces the Kicker referee grade variable with variables indicating the number of red and yellow cards issued, and the number of penalty kicks and corner kicks awarded, in each match. These variables reflect a different mechanism through which referee actions affect the outcome of football matches as players are potentially being sent off and, thus, directly affect effort and style of play of the affected team. Model specification (3) explains 22.1% of the observed variation in bet volume. None of the parameters on the card or penalty and corner kick variables

Table 2: Results - Over 2.5 Goals Scored Bet Volume

	(1)	(2)	(3)	(4)
Kicker Referee Grade	—	-1071	—	—
		-1.18		
Home yellow cards	—	—	-967	-1003
			-1.05	-1.09
Home red cards	—	—	-3349	-3522
			-0.68	-0.72
Home penalties	—	—	-1777	-1750
			-0.64	-0.63
Away yellow cards	—	—	-412	-412
			-0.47	-0.47
Away red cards	—	—	-6640	-6488
			-1.48	-1.44
Away penalties	—	—	-3604	-3599
			-1.14	-1.13
Home corner kicks	—	—	—	-301
				-0.81
Away corner kicks	—	—	—	-459
				-1.02
# Significant Referee Fixed Effects	2	2	3	3
Observations	1250	1250	1250	1250
R^2	0.216	0.216	0.221	0.222

Dependent variable: bet volume in over 2.5 goals scored market
Parameter estimates & t-stats. *: Significant at 5% level.

are statistically significant at conventional significance levels.

The parameter estimates on the two referee-specific indicator variables that were statistically different from zero in model specifications (1) and (2) are also statistically different from zero in model specification (3). In addition, a third parameter estimate on a referee-specific indicator variable is significantly different from zero (p-value 0.054) in this model specification. The p-value on the test of the null hypothesis (the null hypothesis being that the parameter estimate for this third referee is equal to zero) was 0.057 in model specification (1) and 0.071 in model specification (2). This represents weak evidence that betting volume in the Betfair Over 2.5 market was higher for this referee, as well.

Including both the Kicker grade and the card and penalty kick variables in the same model produces similar results. The model explains 22.1% of the observed variation in bet volume, none of the parameter estimates on the referee performance variables are statistically different from zero, and the parameter estimates on the same two referee-specific indicator variables are significantly different from zero. The p-value on the test of the null hypothesis that the third referee has a positive association with bet volume is 0.065 in this model.

Table 3 repeats the analysis above using bet volume in the Betfair Under 2.5 market as the dependent variable. Model specification (1) includes only the fixed effects variables to control for unobservable team, match week and season level heterogeneity. Model specification (2) includes the Kicker referee grade and model specification (3) includes the referee outcome variables (yellow and red cards and penalty kicks). As in Table 2, none of the parameter estimates on the referee performance variables are statistically different from zero, and the models explain about 22% of the observed variation in bet volume in this market.

Similar to Table 2, two of the parameter estimates on the referee-specific indicator variables are statistically different from zero at the 5% level. Bet volume in the Betfair Under 2.5 market was above average for matches officiated by these two referees, compared to the other referees in the sample. One of these referees is different from the three referees identified in Table 2. However, the second referee associated with higher bet volume in the Under 2.5 market is the referee identified as associated with a higher bet volume in the Over 2.5 market in model specification (3) on Table 2.

Note that this result is plausible. Suppose that an individual has fixed a football match by inducing the referee to make calls that would lead to 2 or fewer goals being scored in the match. On a betting exchange like Betfair, a match fixer could either back bets in the Under 2.5 market, or lay bets in the over 2.5 market to profit from this outcome. This would tend to increase bet volume in either market.

Actual match outcomes also differ systematically in a way consistent with the presence of match fixing. In the matches with the largest residuals from models (2) and (3) on Table 3, the 99th percentile of the distribution, the average number of goals scored was about 2.4; the overall average goals scored in the sample is 2.9. Fewer goals were scored on average in these high volume matches in the Under 2.5 goals market.

Table 3: Results - Under 2.5 Goal Bet Volume

	(1)	(2)	(3)	(4)
Kicker Referee Grade	—	-340	—	—
		-0.36		
Home yellow cards	—	—	467	451
			0.49	0.47
Home red cards	—	—	-4198	-4463
			-0.83	-0.88
Home penalties	—	—	1243	1112
			0.43	0.39
Away yellow cards	—	—	-1340	-1326
			-1.49	-1.47
Away red cards	—	—	-8361	-8115
			-1.80	-1.74
Away penalties	—	—	-3795	-3731
			-1.16	-1.14
Home corner kicks	—	—	—	-160
				-0.41
Away corner kicks	—	—	—	405
				0.87
# Significant Referee Fixed Effects	2	2	2	2
Observations	1251	1251	1251	1251
R^2	0.216	0.216	0.222	0.222

Dependent variable: bet volume in under 2.5 goals scored market
Parameter estimates & t-stats. *: Significant at 5% level.

The results on Tables 2 and 3 indicate that specific referees in Bundesliga 1 matches are associated with higher than average bet volume in the Betfair exchange markets for Over 2.5 and Under 2.5 goals. Since specific referees are identified as associated with higher bet volume in these markets, we next assess the extent to which these referees perform differently than the other referees in the sample in terms of quantifiable performance metrics.

Table 4 identifies referees associated with higher than average bet volume in the Betfair markets by ID number. The three referees consistently identified as associated with higher than average bet volume in the Over 2.5 market are Referee 6, Referee 11 and Referee 13.

Recall that Robert Hoyzer was identified as a referee fixing matches, and subsequently convicted of fraud after making a number of high-profile, egregious calls, including awarding penalty kicks and issuing red cards to star players during a match between Hamburger SV and SC Paderborn. However, matches fixed by Ante Sapina, and matches fixed during the Calciopoli match fixing scandal in Italy, were only discovered to have been fixed long after they occurred through sworn testimony during court cases or phone taps.

None of the referees associated with unusually high over/under 2.5 goal bet volumes had particularly good or bad average Kicker grades over the period, so an objective assessment of their performance indicates that none of the four performed unusually well or unusually poorly over the sample period. This is to be expected. Anyone attempting to influence match outcomes in sporting events must find ways to accomplish this goal without detection. The theoretical models of match fixing discussed above emphasize the important role played by the probability of detection in the decision to fix matches.

Table 4 also summarizes the number of yellow cards per match issued by each referee over the sample period. A yellow card is a caution issued by a referee to a player who has violated some football rule. A player receiving a yellow card can continue to play as long as he does not receive a second yellow card. None of the four referees associated with high bet volume issued an unusually large or small number of yellow cards per match.

Note that Table 4 does not show the exact number of games refereed by each official, only an indicator for refereeing more or less than 100 matches in the sample. This is to avoid identifying specific referees in the sample.

Table 5 summarizes the propensity of the referees to issue red cards to players and award penalty kicks and corner kicks. In no case were more than 2 red cards or penalty kicks awarded to any team in a match in the sample. Red cards result in player expulsion from the game. Penalties issued result in a free kick from 11 meters in front of the goal and typically result in a goal scored. Corner kicks are awarded to the attacking team when the ball leaves the field of play. Again, the four referees associated with unusually high bet volume in the over and under 2.5 goals scored markets did not issue an unusually large number of red cards, penalties, or corner kicks per match during the sample period.

Table 4: Individual Referee Results # of Significant Parameter Estimates

ID #	Matches	# significant	Grade	Home yellows	Away yellows
1	<50	0	worse (than mean)	below (mean)	below (mean)
2	<50	0	worse	below	above
3	50+	0	worse	above	above
4	<50	0	worse	above	below
5	50+	0	worse	above	below
6	50+	3 (+2.5)	worse	above	below
7	<50	0	worse	below	above
8	50+	0	better	above	above
9	50+	0	better	below	above
10	50+	0	worse	above	above
11	50+	3 (+2.5)	worse	above	below
12	50+	0	better	below	below
13	50+	3 (+2.5)	better	above	above
14	<50	0	worse	below	below
15	50+	0	better	below	below
16	50+	0	worse	below	below
17	<50	0	worse	below	above
18	50+	0	worse	above	below
19	50+	1 (+2.5), 3(-2.5)	better	below	below
20	50+	0	better	below	below
21	<50	0	worse	below	above
22	<50	0	worse	below	below
23	50+	0	better	below	above
24	<50	0	worse	below	below
25	<50	0	better	below	below
26	50+	0	better	above	above
Mean	48		3.22	1.57	1.97
+/- 1 SD			4.36/2.08	2.71/0.42	3.19/0.75

Table 5: Individual Referee Results - Red Cards and Penalties Called Per Match

ID	# Matches	Red away	Red home	Penalty away	Penalty home	Corners away	Corners home
1	<50	below (mean)	above (mean)	above (mean)	above (mean)	below (mean)	above (mean)
2	<50	below	above	above	below	above	above
3	50+	below	below	above	below	below	below
4	<50	above	below	above	below	above	below
5	50+	below	above	above	above	above	above
6	50+	above	above	above	below	below	below
7	<50	above	above	below	above	below	below
8	50+	below	above	below	below	above	below
9	50+	below	below	above	below	above	below
10	50+	above	below	above	below	below	above
11	50+	below	below	below	below	above	above
12	50+	below	below	above	below	above	below
13	50+	below	above	above	below	above	above
14	<50	below	below	above	below	below	above
15	50+	below	above	above	below	below	below
16	50+	above	above	above	below	below	above
17	<50	above	below	above	above	below	below
18	50+	below	above	above	above	above	below
19	50+	above	above	above	below	below	below
20	50+	above	above	above	above	above	above
21	<50	below	below	above	below	above	below
22	<50	below	below	below	below	below	above
23	50+	below	above	above	below	above	below
24	<50	above	above	below	below	below	above
25	<50	below	above	above	above	below	above
26	50+	above	above	above	below	above	below
Mean		0.05	0.04	0.10	0.15	4.12	5.55
+/- 1 SD		-0.17/0.28	-0.16/0.25	-0.22/0.43	-0.22/0.53	1.80/6.66	2.52/8.32

Analysis of Referee Fixed Effects: Fisher Exact P-Values

The empirical results above use estimated referee fixed effects parameters to identify specific referees associated with unusually large betting volume in the over 2.5 goals and under 2.5 goals betting markets on Betfair, conditional on other observable match-specific factors. While this is suggestive of potential match fixing, these referee fixed effects parameters capture all unobservable factors associated with each referee that affect bet volume. As a robustness check on the significance of these results, we use the *potential outcomes* approach to calculate exact p-values from a randomization distribution for these referee fixed effects.

This approach holds all match conditions constant and randomly assigns all referees in the sample to all other matches and calculates the exact distribution of the referee fixed effects parameter estimates under these conditions. Calculating the exact distribution from this randomization experiment allows us to make probabilistic statements about the likelihood of observing the specific referee fixed effects parameter estimates reported on Table ADD TABLE INFO and Table ADD TABLE INFO under the null assumption that no referees in the sample are involved in match fixing.

This "sharp" null hypothesis approach provides substantially stronger evidence of match fixing than a simple assessment of the statistical significance of the referee fixed effects parameters reported above, because it includes all possible combinations of referee assignment to football matches in the sample under the null hypothesis that no referees are involved in match fixing. Fisher (1925) and Rosenbaum (1984) developed the exact p-value approach; Athey et al. (2016) recently demonstrated the applicability of this approach to a similar economic setting.

In general, the *potential outcomes* approach is used when estimating *causal* effects. In order to understand this approach, it is useful to begin with the following definition of the *unit-level* causal effect for individual $i = 1, ..., N$

$$\tau_i = Y_i(1) - Y_i(0). \tag{2}$$

In 2, τ_i is the unit-level causal effect, $Y_i(1)$ is the outcome when an individual is given some treatment, and $Y_i(0)$ is the outcome when the individual is not given this treatment. You can think of the $0, 1$ arguments in Y_i as $Y_i(W_i = 1)$, where $W_i = 1$ if unit i receives the treatment, and $W_i = 0$ otherwise.

For example, we might be interested in the effect of a medicine that can reduce fevers in patients. To investigate this causal effect, we design a study where N_1 individuals with fevers are given the medicine and N_0 individuals with fevers are given a placebo. We then compare the body temperature of those who took the medicine, $Y_i(1)$, to those who did not take the medicine, $Y_i(0)$. The effect of the medicine on individual i is defined by 2.

Of course, we never observe both $Y_i(1)$ and $Y_i(0)$ in practice. That is, in the absence of a cloning technology or, unless the same $N = N_1 + N_0$ individuals come back with fevers at a later date, we cannot give the medicine to a patient and simultaneously not give that same patient the medicine.

As a result, we cannot compute τ_i for any individual in the study. For each i, define Y_i^{obs} as the outcome that is observed for each individual. Formally,

$$Y_i^{obs} = \begin{cases} Y_i(0) & W_i = 0 \\ Y_i(1) & W_i = 1 \end{cases} \tag{3}$$

We must, however, estimate the effect of the medicine on fevers. A natural estimator of the *average treatment effect* (ATE) is the difference in the Y_i outcome variables between each group

$$\hat{t} = \frac{1}{N_1} \sum W_i Y_i^{obs} + \frac{1}{N_0} \sum (1 - W_i) Y_i^{obs} = \bar{Y}_1 - \bar{Y}_0 \tag{4}$$

Other approaches exist. Fisher (1925) devised an ingenious assumption about the treatment effect: the null hypothesis that the medicine has *no effect* on the outcome for *any* individual.

$$H_0 : Y_i(1) = Y_i(0), \forall i. \tag{5}$$

This null hypothesis generates an alternative study using the same N individuals, a randomization study. If the null hypothesis in Equation (5) is true, we can randomly select another N_1 individuals (some of whom could have actually been given the medicine) and calculate what the ATE would be for this study

$$\hat{t}^* = \frac{1}{N_1} \sum W_i^* Y_i^{obs} + \frac{1}{N_0} \sum (1 - W_i^*) Y_i^{obs} \tag{6}$$

In Equation (6), the Y_i^{obs} have not been randomly re-assigned; rather, a sample of N_1 individuals have been randomly drawn and the W_i^* have been created based on the N_1 individuals selected.

There are $S = \frac{N!}{N_1! N_0!}$ total ways to assign N_1 individuals the treatment from the N individuals in the study. For each of these $s = 1, \ldots, S$ possible combinations, $\hat{t}_s^*$ can be calculated. Using these $\hat{t}_s^*$, we can then calculate the 2.5% and 97.5% quantiles of this distribution. Using these quantiles, we can then determine if the *actual* $\hat{t}$ is inside the 95% range for the $\hat{t}_s^*$.

The above procedure is similar to bootstrapping procedures, but differs slightly. When using bootstrap methods, the researcher creates a new sample of Y_i by simulation using a data-generating process, drawing from estimated error terms, or other means. In the Fisher method, we fix the Y_i and instead randomly redraw the W_i^*.

In the context of football referees, we will need to slightly modify the above approach. Specifically, there are $1 < K$ treatments that indicate the referee in charge of officiating football match i. If referee k is in charge of officiating match i, $W_{ik} = 1$ and $W_{ik'} = 0, k' \neq k$. Based on this, we write

$$\tau_k = Y_i(W_{ik} = 1) - Y_i(W_{ik} = 0) \tag{7}$$

with analogous null hypothesis

$$H_0 : Y_i(W_{ik} = 1) = Y_i(W_{ik} = 0), \forall i, k. \tag{8}$$

Like in the example above, we first calculate the $\hat{t}_k$ for each referee using

$$Y_i = x_i\beta + W_{ik}\tau_k + u_i \tag{9}$$

The x_i are game-specific controls, and $\hat{t}_k$ are the referee fixed-effects from a linear regression.

After estimating $\hat{t}_k$, we re-assign the referees to different matches, without replacement, and re-estimate Equation (9) using W^*_{ik}, calculating $\hat{t}^*_k$ for each $s = 1, ..., S$. In total there are $S = \frac{N!}{N_1!,...,N_K!}$ possible ways we could re-assign the referees to matches. The 95% confidence intervals for $\hat{t}_k$ are then given by the 95% range for the estimated $\hat{t}^*_k$s.

The results from the calculation of exact p-values include a randomized distribution generated by randomly assigning all referees to all football matches in the sample and estimating the referee specific fixed effect for each model, as well as, the 2.5% and 97.5% critical values of these distributions. One randomized distribution is generated for each referee in the sample. Under the null hypothesis that none of the referees is involved in match fixing (the treatment), the referee specific fixed effect would be close to the average effect of all referees in the sample. If a referee specific fixed effect is relatively large compared to the mean of the randomized distribution, then the null hypothesis of no treatment is unlikely to be true.

Figure 2 shows the randomized distributions in gold histograms, 2.5% and 97.5% critical values as black lines and actual referee fixed effects parameter estimates from Model (x) on Table 2 as a red line, for the three referees identified on Table 4 as associated with multiple referee fixed effects parameters statistically larger than zero based on observed variation in the over 2.5 goals scored market.

Based on the randomized distribution for each referee, while the referee fixed effects parameters are statistically larger than zero, these parameter estimates are not unusually large based on the randomization experiment described by Equations (7) - (9). If these three referees were replaced by all other referees in the sample, these parameter estimates would not lie outside the 95% confidence interval of the randomization distribution. In this sense, the estimated referee fixed effect parameters for these referees are not unusually large. The estimated referee specific fixed effects parameters for all other referees in the sample are not outside the 95% confidence intervals for their individual randomized distributions in the over 2.5 goals scored betting market.

Figure 3 shows the randomized distribution in gold histograms, 2.5% and 97.5% critical values as black lines, and actual parameter estimate from model (x) on Table 3 as red lines for two referees with exact p-values outside the 95% confidence interval for be volume in the under 2.5 goals scored market. The referee on the right, Referee #19 was identified on Table 4 as associated with unusually high bet volume in this market. The randomized distribution indicates that this parameter estimate is unusually large based on the randomization experiment.

Interestingly, the randomization experiment identifies a second referee, Referee #18, that had

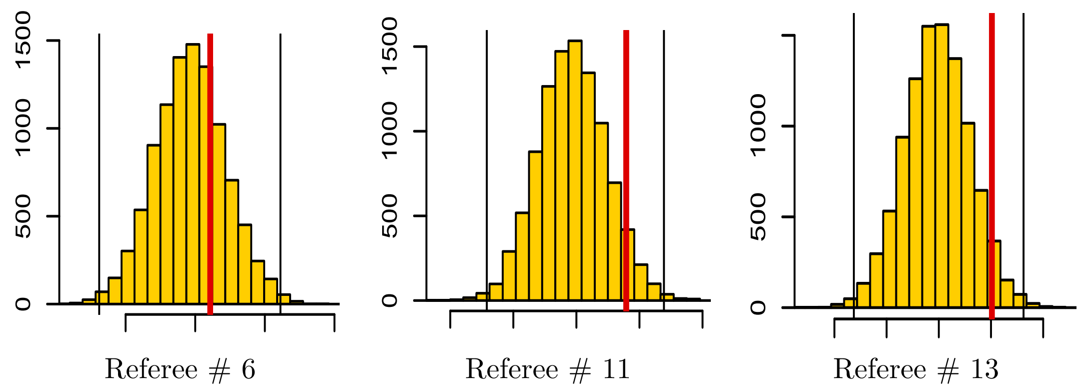

Figure 2: Fischer Exact P-Values, Over 2.5 Goals

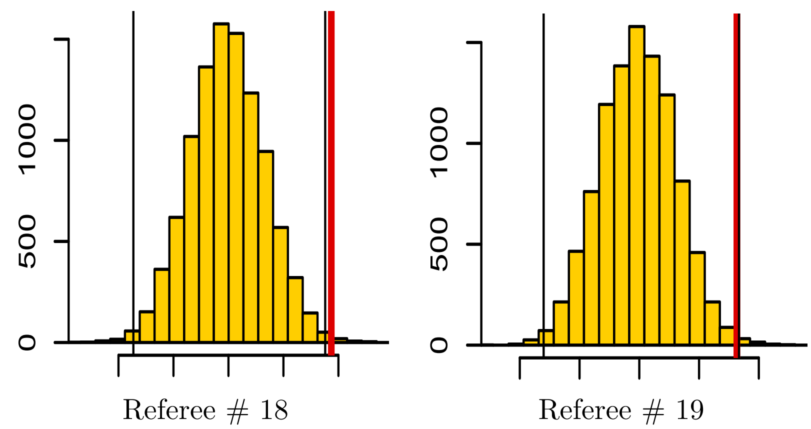

Figure 3: Fischer Exact P-Values, Under 2.5 Goals Scored

a large estimated referee fixed effect parameter. Referee #18 did not have any referee specific fixed effect parameter significantly different from zero at the 5% or 1% level for the regression models reported on Table 3. However, the p-values on Referee #18's fixed effects parameters are less than 0.10 for all four models, so the referee fixed effect parameters are borderline significant in all cases. Based on the randomized distribution for Referee #18, the estimated referee fixed effect for referee #18 appears to be unusually large in the context of the randomization experiment.

Additional Robustness Tests

The regression results above use individual referee fixed effects to identify referees in matches with unusually high bet volume in the markets for under/over 2.5 goals scored in matches. In each regression model, one referee is omitted to avoid perfect collinearity among the fixed effect parameters. These individual referee fixed effects reflect the average bet volume in matches refereed by each referee relative to the volume in the matches refereed by the omitted referee.

Since these parameters reflect relative match volumes, the referee fixed effect could be sensitive to the omitted referee. To assess the extent to which the results are driven by the omitted referee, we re-estimated Equation (1) with the full set of covariates 26 different times, systematically using a different omitted referee in each iteration. This results in $26x25 = 650$ different estimated referee fixed effects parameters.

Table 6 summarizes the results of this systematic re-estimation omitting different referees from each iteration. The parameter estimates are random variables, and some number of them are expected to be statistically different from zero by chance. At the 5% significance level, about 32 of the parameter estimates would be expected to differ significantly from zero. At the 1% significance level, about 6 parameter estimates would be expected to be statistically different from zero. If the parameter estimates were statistically different from zero by chance, then these significant parameter estimates would be expected to be randomly distributed across referees in the sample and not concentrated in a small number of referees.

Table 6: Robustness Checks Omitting Different Referees from Regression Models

	Over 2.5 Goals		Under 2.5 Goals	
	5% Level	1% Level	5% Level	1% Level
Referee ID	# Significant / 25	# Significant / 25	# Significant / 25	# Significant / 25
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	16	10
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	7	1	0	0
11	0	0	0	0
12	0	0	0	0
13	12	3	0	0
14	0	0	0	0
15	0	0	0	0
16	0	0	0	0
17	0	0	0	0
18	0	0	0	0
19	5	0	13	1
20	0	0	0	0
21	0	0	0	0
22	0	0	0	0
23	0	0	0	0
24	0	0	0	0
25	0	0	0	0
26	0	0	0	0

The first two columns of Table 6 show the results for the betting on over 2.5 goals scored markets. 24 of the estimated referee fixed effect parameters are statistically different from zero, slightly smaller than the number that would be expected to be significant by chance. However, these significant parameters occur for only three referees, Referees number 10, 13 and 19. These are the three referees identified on Table 4 as associated with several statistically significant individual referee fixed effects parameters.

Put another way, each referee in the sample has an individual fixed effect estimate from 25 different regression models. At the 5% level, each referee would be expected to have a statistically significant fixed effect parameter estimate in one of these models by chance. A referee having a statistically significant individual effect parameter 7 or 12 times out of 25 has substantially more significant parameter estimates than would be expected by chance. At the 1% level, only 0.2 out of 25 parameter estimates would be expected to be significantly different from zero by chance. From the second column of Table 6, having 3 statistically significant individual parameter estimates out of 25 is also more than would be expected by chance.

The second two columns on Table 6 show the results for betting on under 2.5 goals scored market. 29 individual estimated referee specific fixed effects parameters are statistically different from zero, about what would be expected. However, these statistically significant fixed effect parameters are for only two referees, Referees number 6 and 19. At the 1% level, 11 referee fixed effect parameters are statistically different from zero, well above the number that would be expected by chance. Ten of these statistically significant parameter estimates are for referee number 6. Again, both these referees were identified on Table 4 as refereeing matches with unusually high betting volume in the over/under 2.5 goals scored market.

Another explanation for the significant referee specific fixed effects parameters in the regression models explaining observed variation in bet volume in the over 2.5 goal and under 2.5 goal markets on Betfair could be based on observable referee styles, in terms of enforcement of rules, and bettor responses to these observable styles. Suppose that specific referees call games in a way that naturally leads to more goals scored in matches, or less goals scored in matches. If bettors observe these tendencies, then profit maximizing bettors would be more likely to place bets on matches officiated by these referees in the over 2.5 goal and under 2.5 goal markets on Betfair.

To assess the extent to which specific referees are associated with higher or lower goal scoring in matches they officiate, we estimate Equation (1) using the total goals scored in each match as the dependent variable. These regression models also include home and visiting team fixed effects, matchweek fixed effects and season fixed effects to capture unobserved heterogeneity in factors affecting goal scoring in Bundesliga matches over the sample period.

Table 6 shows the parameter estimates, estimated standard errors and other regression diagnostics for these models. Table 6 has the same format as Tables 2 and 2. The baseline specification, Column (1), sets X_{ijrws} equal to zero and includes only home team, visiting team, match week, season and referee indicator variables. The results in Column (2) add the Kicker referee score, and the results in Columns (3) and (4) add additional covariates reflecting actual referee decisions in each match.

The parameter estimate on the Kicker referee grade variable is not statistically different from zero. Goals scored in matches were not higher or lower in matches where the referee received higher or lower Kicker grades. The parameter estimates on the other covariates generally have the expected sign. Matches with more yellow cards issued to the home team have fewer goals scored. Matches with more penalties assessed on home and visiting teams, and more red cards given to visiting teams, have more goals scored. In general, these models explain between 11% and 16.5% of the observed variation in goals scored.

Table 7: Results - Total Goals Scored Models

	(1)	(2)	(3)	(4)
Kicker Referee Grade	—	-0.003	—	—
		-0.07		
Home yellow cards	—	—	-0.106*	-0.109*
			-2.38	-2.46
Home red cards	—	—	0.179	0.158
			0.76	0.67
Home penalties	—	—	0.612**	0.610**
			4.61	4.60
Away yellow cards	—	—	0.016	0.016
			0.37	0.38
Away red cards	—	—	0.655**	0.674**
			3.03	3.11
Away penalties	—	—	0.580**	0.582**
			3.81	3.83
Home corner kicks	—	—	—	-0.028
				-1.57
Away corner kicks	—	—	—	-0.023
				-1.07
# Significant Referee Fixed Effects	2	2	2	2
Observations	1251	1251	1251	1251
R^2	0.116	0.116	0.163	0.165

Dependent variable: goals scored in each match.
Parameter estimates & t-stats. */**: Significant at 5%/1% level.

From the results on Table 7, two referees are associated with games with statistically higher goals scored, even after controlling for observable game events like the number of penalty cards issued and the number of corner kicks awarded. These referees may have individual styles of enforcing rules that lead to systematically higher goal scoring. These two referees, Referee #9 and Referee #21, were not associated with higher bet volumes in either the over 2.5 goals scored market or under 2.5 goals scored market on Betfair, as reported on Table 2 and Table 3 above. So the unusually high bet volume in these markets does not reflect bettors realizing tendencies of specific referees and placing bets accordingly.

Taken together, the regression results suggest that specific referees officiated Bundesliga matches with unusually high volume in the under/over 2.5 goals scored market. The strongest evidence, based on bet volume in the under 2.5 goals scored betting market on Betfair, suggest that referees #18 and #19 officiated Bundesliga matches with unusually high volume in the under 2.5 goals scored betting market on Betfair. In order to profit from fixed matches, bets must be placed somewhere. Bundesliga matches with unusually high bet volume could potentially be fixed, if the match fixers use the Betfair platform to generate profits from fixed football matches.

The strongest evidence comes from the under 2.5 goals scored market; this may reflect a preference for bets backing under 2.5 goals scored bets when fixing matches. Backing a proposition is the only bet possible when betting with bookmakers, so individuals familiar with standard betting will be familiar with this type of bet.

It may also reflect the idea that a referee fixing a match would be better able to escape detection by suppressing scoring rather than enhancing scoring. Referee actions suppressing scoring include quick offsides calls, shading fouls in favor of the defending team at both ends of the pitch and technical violations on corner kicks in favor of the defending team at selected points in the game. Since Bundesliga matches had average total scores close to 2.5 goals in this sample, a referee fixing a match by reducing scoring would, on average, only have to eliminate one goal that would otherwise have been scored to generate a profitable betting opportunity for match fixers.

Conclusion

We performed an ex post analysis of betting volume in Betfair markets for over/under 2.5 goals scored in Bundesliga 1 football matches to determine if variation in this bet volume is consistent with the idea that some referees might have been engaged in match fixing. The empirical analysis identified a small number of referees that were associated with larger than average volume in the Under 2.5 markets. The Kicker referee grades, and observed average propensity of these referees to issue yellow and red cards and to award penalty kicks, did not appear to be different from other referees in the sample, suggesting no unusual behavior on the part of these referees.

While our results are compelling, we do not claim to identify match fixing in these Bundesliga 1 matches *per se*. Rather, we identify anomalous patterns in Betfair exchange betting volume that are consistent with the story of match fixing. From a theoretical point of view, a repeated engagement, rather than a unique involvement, in deviant behavior such as match fixing, is consistent with the slippery-slope literature (Gino and Bazerman, 2009) and anecdotal evidence from the uncovered cases of match fixing discussed previously. Wolfers (2006) undertook a similar analysis using point spread betting data from US college basketball games. Other explanations exist for the patterns in bet volume documented in these data. Alternative explanations for the results in Wolfers (2006) have been proposed (Bernhardt and Heston, 2010).

Since match fixing can be difficult to detect, especially match fixing by referees or team officials, it is important to develop methods for detecting match fixing based on observed data in betting markets. Some betting on fixed matches clearly takes place with traditional bookmakers. In general, only data on odds set by bookmakers on specific sporting events are publicly available; data on bet volume on specific sporting events at individual bookmakers are not easily available. This makes the Betfair data analyzed here of interest.

The analyses have some limitations. Referees are not totally randomly assigned to games. The Bundesliga reacted to the Hoyzer scandal by reducing the time between the announcements of referees and the actual games, reducing the time available to fix a game prior to all games in this sample. Veteran referees are assigned to decisive games more often than rookie referees. While our empirical work captures match week effects, there might still be some heterogeneity between the matches in our sample. Since betting volume cannot be disentangled on the day of each bet, this study includes overall betting volume on each possible outcome.

References

Athey, S., Eckles, D., and Imbens, G. W. (2016). Exact p-values for network interference. *Journal of the American Statistical Association*, (just-accepted).

Bernhardt, D. and Heston, S. (2010). Point shaving in college basketball: A cautionary tale for forensic economics. *Economic Inquiry*, 48(1):14–25.

Boeri, T. and Severgnini, B. (2011). Match rigging and the career concerns of referees. *Labour Economics*, 18(3):349–359.

Boeri, T. and Severgnini, B. (2013). Match rigging in italian professional soccer: the economic determinants of corruption. In *Match-Fixing in International Sports*, pages 101–112. Springer.

Bryson, A., Buraimo, B., and Simmons, R. (2011). Do salaries improve worker performance? *Labour Economics*, 18(4):424–433.

Carpenter, K. (2012). Match-fixing: The biggest threat to sport in the 21st century? *Internacional Sports Law Review*, 2:13–23.

Di Tella, R. and Franceschelli, I. (2011). Government advertising and media coverage of corruption scandals. *American Economic Journal: Applied Economics*, 3(4):119–151.

Dimant, E. and Deutscher, C. (2014). The economics of corruption in sports: The special case of doping. Working Papers in Economics and Management 17-2014, University of Bielefeld.

Dimant, E. and Schulte, T. (2016). The nature of corruption: An interdisciplinary perspective. *German Law Journal*, 17:53.

Dimant, E. and Tosato, G. (2017). Causes and effects of corruption: What has past decades empirical research taught us? *Journal of Economic Surveys*.

Dohmen, T. J. and Sauermann, J. (2016). Referee bias. *Journal of Economic Surveys*, 30(4):679–695.

Dusha, E. (2015). Intermediated corruption. *International Economic Review*, 56(3):997–1017.

Feltes, T. (2013). Match fixing in Western Europe. In Haberfeld, M. and Sheehan, D., editors, *Match-Fixing in International Sports: Existing Processes*, pages 15–30. Springer.

Fisher, R. A. (1925). *Statistical methods for research workers.* Genesis Publishing Pvt Ltd.

Foley-Train, J. (2014). Sports betting: Commercial and integrity issues. Report, European Gaming and Betting Association.

Forrest, D. (2012). The threat to football from betting-related corruption. *International Journal of Sport Finance*, 7(2):99–116.

Forrest, D., McHale, I., and McAuley, K. (2008). 'Say it ain't so': Betting-related malpractice in sport. *International Journal of Sport Finance*, 3(3):156–166.

Forrest, D. and Simmons, R. (2003). Sport and gambling. *Oxford Review of Economic Policy*, 19(4):598–611.

Frick, B., Gürtler, O., and Prinz, J. (2008). Men in black: monitoring and performance of german soccer referees. In Dietl, H., Franck, E., and Kempf, H., editors, *Soccer: economics of a passion*, pages 309–322. Hofmann.

Gino, F. and Bazerman, M. H. (2009). When misconduct goes unnoticed: The acceptability of gradual erosion in others unethical behavior. *Journal of Experimental Social Psychology*, 45(4):708–719.

LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., and Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of internet sports gambling behavior. *Journal of Gambling Studies*, 23(3):347–362.

Lockwood, R. I. (2008). The Best Interests of the League: Referee Betting Scandal Brings Commissioner Authority and Collective Bargaining Back to the Frontcourt in the NBA. *Sports Law Journal*, 15:137–172.

Pohlkamp, S. (2014). The impact of referees on match outcomes in professional sports: Evidence from the German Football Bundesliga. Working paper, University of Hamburg.

Preston, I. and Szymanski, S. (2003). Cheating in contests. *Oxford Review of Economic Policy*, 19(4):612–624.

Rose-Ackerman, S. and Palifka, B. J. (2016). *Corruption and government: Causes, consequences, and reform.* Cambridge University Press.

Rosenbaum, P. R. (1984). Conditional permutation tests and the propensity score in observational studies. *Journal of the American Statistical Association*, 79(387):565–574.

Wolfers, J. (2006). Point shaving: Corruption in NCAA basketball. *The American Economic Review*, 96(2):279–283.