ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Strategy Economics

The Impact of the VAR on Match Characteristics and Predictability in European Football

Name student: Arvid Oostrom

Student ID number: 543290

Supervisor: Thomas Peeters

Second assessor: Jan van Ours

Date final version: 02-08-24

ABSTRACT

This study demonstrates that the introduction of the Video Assistant Referee (VAR) has a mixed impact on match characteristics and the predictability of match outcomes in European football leagues. It contributes to economic theory by employing a staggered Difference-in-Differences (DiD) method and finding differing results from standard Two-Way Fixed Effects models, highlighting the importance of selecting an appropriate model. The staggered DiD model proves suitable for sports contexts with varying treatment timings. The study also shows that refereeing oversight influences decision-making, as evidenced by a significant increase in yellow cards issued and a significant decline in average goals scored per match. These findings suggest behavioral changes among players and referees or adjustments in offensive strategies due to VAR oversight. However, no significant effects are found in the number of fouls, red cards, shots, shots on target, or corners. Additionally, the VAR's effect on the predictability of match outcomes is not significant.

Keywords: Video Assistant Referee (VAR), staggered Difference-in-Differences (DiD), match characteristics, predictability of match outcomes, referee decision-making, changing player behavior, and technological advancement.

Table of Contents

1. Introduction

The roar of the crowd erupts as the ball crosses the line – goal! But celebrations turn into frustration due to controversial referee decisions, such as Lampard's "ghost goal" in the 2010 World Cup, Scholes' disallowed goal against Porto in 2004, and the three denied clear penalties for Chelsea against Barcelona in 2009. Such incidents, with their potential to dramatically affect match outcomes and have significant financial consequences for clubs, highlight the critical role of accurate officiating in football.

The introduction of VAR represents a significant shift in football officiating, utilizing video technology to assist referees with critical on-field decisions. Specifically, VAR reviews incidents such as goals, penalties, direct red cards, and cases of mistaken identity (Inside FIFA, 2023). This technological advancement aims to minimize human error, thus ensuring that match outcomes more accurately reflect team performances. Considering this, this study seeks to investigate the impact of VAR on various match characteristics and the predictability of match outcomes in European football leagues from 2017 to 2020.

The introduction of VAR sparked ongoing debate among players, coaches, and stakeholders about its impact and effectiveness. However, the VAR debate extends beyond players and managers expressing their views on its impact on various aspects of the game. Prior research highlights the significant impact of referees on match characteristics and outcomes. For instance, Fontenla et al. (2018) demonstrated that individual referees significantly affect game scores in Argentina's first division, even when controlling for various factors. Albanese et al. (2020) found similar results in Italian Serie A. These findings underscore the necessity of improving officiating accuracy, a gap that VAR aims to fill.

Despite these findings, there is a lack of comprehensive analysis of the VAR's effects across multiple seasons and leagues. This study fills that gap using a staggered difference-in-differences approach, aiming to provide a comprehensive analysis of how the VAR influences football matches. This research aims to answer the following research question:

- How does the implementation of the Video Assistant Referee (VAR) system affect professional football matches in men's European football leagues from 2017 to 2020?

This paper will contribute scientific value by employing causal inference methods with multiple treatment periods to test the effects. Notably, it will utilize the new "staggered Difference-in-Differences" method developed by Callaway and Sant'Anna (2021). The use of this relatively new method adds scientific relevance, showing that this method can be applied in this research area as well. Understanding the effectiveness and applicability of this method is crucial because it provides more accurate and robust estimates in settings with varying treatment timings, which is common in many economic and social policy evaluations. By validating and demonstrating the utility of this method in the context of VAR in football, the toolkit available for researchers can be improved, leading to more precise policy recommendations and interventions in various fields beyond sports.

Furthermore, this paper contributes to economic theory by examining how the implementation of the Video Assistant Referee (VAR) system influences various match characteristics and outcomes in football. Similar to the judicial system, where stricter oversight by higher courts limits district judges' flexibility in sentencing decisions (Fischman & Schanzenbach, 2011), this study explores how oversight mechanisms influence match characteristics and outcomes, potentially as a result of influenced decision-making of referees and players.

Additionally, this research holds relevance given that football is the world's most popular sport, and the introduction of VAR represents a major change in the rules of the game, affecting the debate and the entertainment value for the fans. While media and fans used to frequently debate referees' decisions, they now find themselves continually discussing the VAR.

The remainder of this paper is structured as follows. The next section covers the literature review, analyzing how the VAR works and previous research on its impact on match characteristics and predictability of matches followed by my hypotheses. After this, a description of the data is provided, followed by the methods used for the models, the results, and a detailed discussion and conclusion.

2. Theoretical Framework

In this section, mechanisms will be explained behind how the VAR works in general, how it can affect match characteristics, and how it possibly can affect the predictability of match outcomes. In between, the hypotheses to answer the research question will be demonstrated.

2.1 Rules of the VAR

The Video Assistant Referee (VAR) system involves a team of officials at a neutral location who review live incidents using multiple camera angles. They intervene in situations involving goals, penalties, direct red cards, and mistaken identity, aiming to correct clear errors and enhance decision accuracy. When the referee on the field makes a 100% factual wrong decision, the VAR can communicate with him and tell him to change his decision (offside or ball out of play leading up to goal/penalty, foul committed inside or outside the penalty area for a penalty decision, mistaken identity). When the decision is by interpretation of the VAR 100% wrong (foul committed by an attacking player or offside interference leading up to goal/penalty, and direct red card incidents), the VAR can communicate with the on-field referee to review his decision on the monitor standing next to the field and have the option to change his decision afterward (INSIDE FIFA, 2023). Thus, if the VAR team sees that the on-field referee did make a 100% wrong decision, the on-field referee gets the chance to recover his mistake. From now on the VAR team will be described as the VAR.

2.2 VAR affecting match characteristics

Recent literature already provides some valuable insights into the influence of the Video Assistant Referee (VAR) system in professional football and the various effects of this technological advancement on match characteristics. Studies by Kubayi et al. (2021) have shown that VAR implementation during Men's FIFA World Cup matches led to more penalties, increased playing time, and a decrease in offsides. Possible reasons for these increases and decreases may be that penalty situations sometimes were missed before the introduction of the VAR, and the match is now stopped more often to review certain situations, resulting in increased playing time. Earlier, Lago-Peñas et al. (2019) addressed that in Serie A and German Bundesliga, there was a decrease in fouls, yellow cards, and offsides following the introduction of VAR, all indirect effects of the VAR. But the number of offsides, for instance, may have been impacted by the VAR by the change in the protocol, which now recommends correcting factual errors in-field, but if in doubt continue playing in onside/offside situations because goals will be reviewed. The decrease in fouls and yellow cards could only come from a behavioral change of the players or on-field referees because yellow cards are not reviewed by the VAR. Moreover, research by Spitz et al. (2020) revealed that VAR usage across 13 different leagues was associated with an increase in red cards, penalties, and goals scored. Where red cards and penalties are direct effects of interference with the VAR, the number of goals is subject to the context of the match and the number of penalties which is a direct effect of the VAR. All these findings suggest a unique impact on match characteristics, indicating a shift in disciplinary actions and goal-scoring patterns influenced by VAR interventions or behavioral changes.

- Hypothesis 1: The introduction of VAR will influence match characteristics, increasing the number of fouls, yellow cards, and red cards, while decreasing the number of goals, shots, and shots on target.

2.3 VAR affecting the predictability of match outcomes

Research by Spitz et al. (2020) shows an increase in the predictive odds for referees of making the right decision from 92.1% to 98.3% suggesting an accuracy increase in refereeing decision-making. Eventually, this could result in matches becoming more predictable if this increasing accuracy would lead to better decisions for the predicted winners. However, the refereeing inaccuracy was not unbiased before. Dohmen and Sauermann (2015) show that referees are biased in their decision-making due to social payoffs and material payoffs. Referees are, for instance, subject to home team bias explained by Boyko et al. (2007) and Nevill et al. (2002) and/or the successful team bias by Erikstad and Johansen (2020). Adding to this research, Holder et al. (2021) studied whether the VAR has changed the decision-making behavior of the referees in Serie A and Bundesliga. They showed that referees did change their behavior after the implementation of the VAR, where referees had a home-team preference in their decision-making, the home-team advantage partially decreased after the introduction of the VAR. The home team advantage that remains seems to emerge from the advantages of playing in one's local surroundings, but the referee home bias seems to disappear. Just like papers from Han et al. (2020) and Dufner et al. (2023), who both found the same significant decrease in home advantage. It could be the case that when the referee is biased toward the team that is predicted to win (home team or most successful team) and this bias is diminished by the VAR, matches become less predictable.

Adding more weight to this theory, Johnson and Fort (2022) found that fans prefer more uncertain matches, suggesting that the predictability of match outcomes can significantly influence fan satisfaction and engagement. Similarly, Winand et al. (2021) emphasized that the ongoing debate among fans plays a crucial role in their overall satisfaction with the game. Thus, finding a decrease in the predictability of the game would increase the debate, fan satisfaction, and engagement with the game. This paper will, therefore, investigate whether matches become less predictable following the introduction of VAR.

- Hypothesis 2: VAR implementation will lead to a significant decrease in the alignment of the predicted and real match outcomes, measured by the alignment of full-time results with pre-match bookmaker odds.

These interconnected hypotheses highlight the complex and multifaceted nature of VAR's impact on the beautiful game. This study aims to generate average results of the effects of the VAR in top European leagues from the first implementation in Italy and Germany to three years later, in which other competitions also implemented VAR in between. The findings will be both scientific and social, contributing to the ongoing debate about the VAR in football.

3. Data

3.1 Data collection and preparation

To test the hypotheses, multiple panel datasets are used. Each hypothesis is modeled with a different dataset, incorporating as many data points and leagues as possible. This approach is necessary because match characteristics data like yellow cards, fouls, or shots are not available for all leagues. All data is gathered from one database: football.data.co.uk. From this database, match characteristics such as match date, home and away team, referee, number of shots, shots on target, goals, fouls, yellows, or reds for the home and away team, but also multiple pre-match bookmaker odds are obtained. As shown in Table 1, for the first hypothesis, a panel dataset with the German Bundesliga, Italian Serie A, France Ligue 1, Spanish La Liga, English Premier League, English Championship, and Scottish Premier League is used from the season 2012/13 until March in the season 2019/20, when Covid-19 interrupted all leagues and societies. Later matches and seasons were not included in this analysis to avoid changes in referee biases arising from the Covid-19 pandemic and the empty stadiums that came with it. Sors et al. (2020) found that home advantage and referee bias in football decreased during matches played without spectators due to the COVID-19 pandemic.

For the second hypothesis, the panel dataset again consists of the German Bundesliga, Italian Serie A, France Ligue 1, Spanish La Liga, English Premier League, English Championship, Scottish Premier League, but also the Dutch Eredivisie, Belgium Jupiler Pro League, Greek Super League, and Turkish Super Lig. This dataset is also used to estimate the effect of VAR on the total number of goals per match, as part of the first hypothesis. This match characteristic is the only match characteristic of hypothesis 1 that is available for all leagues included in the second dataset, ensuring the results are based on the maximum number of data points possible. In Table 1 the data coverage per dataset is shown.

Table 1

Data Coverage

Dataset 1	Obs.	Dataset 2	Obs.
Season		Season	
2012-13	56	2012-13	96
2013-14	56	2013-14	96
2014-15	56	2014-15	96
2015-16	56	2015-16	96
2016-17	56	2016-17	96
2017-18	56	2017-18	96
2018-19	56	2018-19	96
2019-20	56	2019-20	96
Div		Div	
Bundesliga	88	Bundesliga	88
English PL	72	English PL	72
Ligue 1	80	Ligue 1	80
Serie A	96	Serie A	96
Scottish PL	40	Scottish PL	40
La Liga	72	La Liga	72
		Jupiler Pro	72
		Greek Super.	56
		Turkish Lig.	40
		Eredivisie	88
		Liga NOS	64

Note. All observations represent averages per home team per season. Teams that did not consistently play in the first league across all seasons have been excluded as home teams.

As mentioned before, this paper will make use of a staggered Difference-in-Differences method developed by Callaway and Sant'Anna (2021). This method will be used due to the different treatment timings of all leagues. In Table 2, the different treatment timings are displayed. The Scottish Premier League is never treated. The Bundesliga, Serie A, Liga NOS, and the Jupiler Pro League implemented the VAR at the start of the 2017/18 season, the Ligue 1, La Liga, Eredivisie, and Turkish Super Lig from 2018/19, and the English Premier League and the Greek Super League from 2019/20.

Table 2

Treatment timing

Never treated	2017-18	2018-19	2019-20
Scottish PL	Bundesliga	Ligue 1	English PL
	Serie A	La Liga	Greek Super.
	Liga NOS	Eredivisie	
	Jupiler Pro	Turkish Lig.	

Note. All treatment timings are at the beginning of the season.

In the datasets, new variables are created like the total number of shots, goals, fouls, yellows, and reds per match (home + away) but also the differences between the home and away teams in the number of shots, goals, fouls, yellows, and reds.

For the second hypothesis, a new variable was created as the difference in expected outcome minus actual outcome. The variable shows the difference between 1 and the percentage of the pre-match chance of the actual outcome (home win/draw/away win).

$$Difference\ in\ Result = 1 - pre_{match}\ chance\ of\ the\ actual\ outcome$$

The chance of the actual outcome is calculated by dividing 1 by the win, draw, or away odds from bet365.

$$pre_{match}\ chance\ of\ the\ actual\ outcome = \frac{1}{win, draw\ or\ away\ odd}$$

But because the odds from bookmakers have a bookmaker margin incorporated to win money for the bookmakers, this margin needs to be subtracted from the odds to get the real pre-match chance of a home win, draw, or away win. To get the bookmaker margin, all chances are added up to each other and subtracted by 1.

$$Bookmaker\ margin = \frac{1}{\text{home win odd}} + \frac{1}{\text{draw odd}} + \frac{1}{\text{away win odd}} - 1$$

Then the real chances are found by multiplying the odds with the bookmaker margin added by 1. Thus, when the home team wins, the difference is calculated as 1 minus the chance of the home team winning the match regarding the pre-match bet365 win odds.

$$DiffinResult\ (outcome = home_{win}) = 1 - \frac{1}{\mathrm{home_{win}\ odd} * (1 + \mathrm{Bookmaker\ margin})}$$

When the away team wins, the difference is calculated as 1 minus the chance of the away team winning. When the real outcome is a draw, the difference is calculated as 1 minus the chance of the match ending in a draw. A number closer to 1 indicates a greater difference between the actual and predicted outcomes.

For all leagues, a treatment dummy (VAR_use) is created which is "1" from the moment when a league implements the VAR onwards. Lastly, a variable (treatment_moment) is created that always shows the year of implementation of the VAR for the specific league.

To run a staggered DiD model the panel needs to be balanced. Therefore, all periods are required to have the same number of observations. Due to leagues having different numbers of teams and matches per season, all statistics are averaged per home team per season. The averages do not only reflect the home team's match performance; rather, they are based on the sum of the home and away teams' match performance. So, the total match characteristics (eg. home team yellows + away team yellows) per match are averaged per home team per season. All teams that did not participate in all sample years were dropped as home teams (but kept in as away teams). By doing so, balanced datasets are created with an evenly distributed number of characteristics per unit, per period. This approach ensures the retention of the maximum number of data points while minimizing the potential for bias due to sample selection. In the panel datasets, the units are the home teams, and the periods are the seasons. However, this approach introduces a limitation of the method used. By excluding these teams as home teams, the analysis may not fully capture the variability and dynamics of all teams in the league, potentially leading to biased estimates or loss of generalizability.

3.2 Descriptive Statistics

Table 3 shows the descriptive statistics of the first and the second dataset. The table provides descriptive statistics of various match characteristics and the average difference in results for each dataset, split by whether the VAR was used or not. The descriptive statistics include the number of observations, mean, standard deviation, minimum, and maximum values for each variable. Lastly, the table also contains information on the treatment moments in both datasets.

As displayed in Table 3, matches with VAR have slightly higher average yellow cards but slightly lower average red cards and fouls compared to matches without VAR. This result suggests VAR could lead to an increase in yellow cards but a decrease in red cards and fouls. Matches with VAR have again a slightly higher number of shots on target but a slightly lower number of shots, goals, and corners. This result suggests VAR could lead to an increase in shots on target but a decrease in shots, goals, and corners. The average difference in predicted and actual outcomes is nearly identical between matches with and without VAR, suggesting that the introduction of VAR does not significantly impact this metric. Furthermore, the standard deviations are relatively consistent across all variables, indicating a stable spread of values. Finally, the second dataset contains a larger number of observations for all periods of VAR implementation compared to the first dataset, except for the group that was never treated. This increase in data points is attributable to the inclusion of additional leagues in the second dataset, as shown in Table 1, from all three different treatment timings, as shown in Table 2.

Table 3

Descriptive Statistics

Variable	Obs.	Mean	St. Dev	Min	Max
VAR_use = 0					
avg_yellows	332	3.856	0.927	1.895	6.263
avg_reds	332	0.213	0.125	0	0.632
avg_fouls	332	26.257	4.052	16.632	36
avg_shots	332	25.005	2.863	17.111	32.947
avg_SOT	332	9.173	1.539	6.105	16.737
avg_corners	332	10.284	1.133	7.353	13.263
VAR_use = 1					
avg_yellows	116	4.151	0.946	1.917	6.75
avg_reds	116	0.202	0.130	0	0.643
avg_fouls	116	24.826	3.164	17.2	32.167
avg_shots	116	24.652	2.770	18.421	31.471
avg_SOT	116	9.523	1.433	7.214	13.5
avg_corners	116	10.202	1.131	6.941	12.933
treatment_moment					
0	40				
2017-18	184				
2018-19	152				
2019-20	72				
VAR_use = 0					
avg_goals	562	2.844	0.529	1.895	4.632
avg_DiffResult	562	0.538	0.109	0.198	0.707
VAR_use = 1					
avg_goals	206	2.919	0.636	1.5	4.882
avg_DiffResult	206	0.539	0.107	0.215	0.696
treatment_moment					
0	40				
2017-18	320				
2018-19	280				
2019-20	128				

Note. All values have been rounded to three decimal places. The descriptive statistics table is divided into two parts based on the different datasets for the different hypotheses. The data descriptives of the first dataset first and the data descriptives of the second dataset afterward. All match characteristic values and the difference in result values are calculated from the subsamples non-treated when VAR_use = 0 and treated when VAR_use = 1. All match characteristic averages (e.g., avg_yellows, avg_reds, avg_fouls) represent the averages per game, per home team, per season. The average difference in result (avg_DiffResult) is the average difference between predicted and actual outcomes, calculated per home team, per season. The treatment moment indicates the season from which the teams started using VAR. The number of observations demonstrates the number of teams treated since the specified year, multiplied by the number of periods in the sample.

4. Methodology

4.1 Staggered Difference-in-Differences Models

To test the hypotheses, the staggered Difference-in-Differences (DiD) method developed by Callaway and Sant'Anna (2021) is employed. This approach is suitable due to the staggered implementation of VAR across different European leagues, as shown in Table 2, allowing us to estimate treatment effects for each specific implementation timing and average these effects for overall impact analysis.

The staggered DiD method estimates treatment effects for each specific timing of VAR implementation. For each point in time when a new group receives the treatment, the model compares the treated group to the groups that are not yet treated at that moment. Consequently, each group has a distinct control group. After estimating these effects, the model aggregates the effects to capture the overall impact of VAR implementation across all leagues. This approach allows us to estimate the total and real effect of VAR, rather than a partial effect based on a subsample.

In the models, the dependent variables include match characteristics or measures such as the average total number of yellow cards per match per season per home team, while the independent variable is the VAR implementation. Each match characteristic or measure is tested using different models and dependent variables. The home team serves as the panel identifier, the season as the time identifier, and the treatment moment identifies from which moment groups receive the treatment.

The estimation equation for the staggered DiD method is as follows:

$$Y_{it} = \alpha_i + \lambda_t + \beta \cdot T_{it} + \epsilon_{it}$$

where:

- Y_{it} is the dependent variable, which varies across models (e.g., average yellow cards, average difference in results)
- α_i represents unit fixed effects (home teams)
- λ_t represents time fixed effects (seasons)

- T_{it} is the treatment indicator (whether VAR is implemented)
- β is the treatment effect (impact of VAR implementation)
- ϵ_{it} is the error term, capturing unobserved factors affecting the dependent variable not explained by the independent variables or fixed effects

4.2 Robustness checks

Because the DiD models account for the unit and time-fixed effects, no other controls are needed to ensure exogenous treatment effects. The only assumption that needs to be held is that the control group needs to be a valid counterfactual for the treatment in the absence of the intervention to obtain causal effects. This can be divided into two other assumptions: the parallel trends assumption and the Stable Unit Treatment Value Assumption (SUTVA). In the staggered DiD setup, these assumptions differ a little bit. Due to the treatment being introduced at different times for different units, the context is more complex than in a standard DiD where all treated units receive the treatment at the same time.

Therefore the “conditional” parallel trends assumption is introduced to address this complexity and ensure the validity of the estimated treatment effects. This staggered DiD complexity comes from the fact that the different units, and home teams, receive their treatment at different points in time. The effect of the treatment can thus vary not only across units but also over time. The treatment effect might differ depending on when a unit receives the treatment and the specific context at that time. Additionally, units that receive the treatment at different times might also differ systematically in ways that affect the outcome. These differences can be due to time-varying factors that correlate with both the timing of treatment and the outcome of interest. To address these complexities, the “conditional” parallel trends assumption is used. By conditioning on observed covariates, the assumption allows for the possibility that units differ in ways that affect their outcome trajectories. This helps in isolating the effect of the treatment from other confounding factors. Furthermore, the conditional assumption ensures that after accounting for covariates, the remaining differences in trends between treated and control units are solely due to the treatment effect. This is crucial when treatment timing is staggered. By this theory, the conditional parallel trends assumption is a refinement of the standard parallel trends' assumption. It now also accounts for the additional complexity introduced by varying treatment timings and heterogeneous treatment effects, ensuring that the estimated treatment effects are valid and unbiased. The conditional parallel trends assumption assumes that, after controlling for observed covariates, the trends in the pre-intervention period between the treatment and control units are equal or parallel.

To evaluate the parallel trends assumption, this study employs the chi-squared test using the estat pretrend function. This test assesses whether pre-treatment trends in the outcome variable differ significantly between treated and control groups. The test examines the coefficients of the Time x Treatment interaction terms, which capture differential changes in the outcome variable over time between groups. The null hypothesis claims that all Time x Treatment interaction coefficients are zero, indicating parallel trends. A low p-value, below 0.05, suggests the trends differed even before the treatment. This outcome implies that the treated and control groups are not initially comparable regarding the outcome variable. A high p-value, above 0.05, suggests that trends are parallel before the treatment, implying that treated and control groups are comparable. The test will be executed with 15 degrees of freedom. The degrees of freedom are related to the number of periods before the treatment is applied and the number of distinct groups receiving treatment at different times. Due to the 5 pre-treatment periods and the 4 treatment/control groups, 15 degrees of freedom are used.

The SUTVA requires that the treatment effect is due solely to the treatment itself, without interactions between treatment and control units. This assumption must hold in both standard and staggered DiD settings. In staggered DiD, SUTVA ensures no interference between units, meaning the treatment status of one unit does not affect another's outcome. For instance, the implementation of the VAR in one league should not impact neighboring leagues. Additionally, SUTVA requires consistent treatment application across units and time. Any variations in treatment implementation must be controlled for to avoid biased estimates, ensuring the measured effects are attributable to the treatment alone.

To test these assumptions, the trends are plotted, and a statistical pre-treatment trend test is estimated with the staggered DiD results to test the conditional parallel trends assumption. To ensure that the SUTVA holds, it will be carefully discussed if there is a possibility of spillover effects or interference between units. To ensure that the treatment does not vary across all units and times the VAR treatments will be compared.

Lastly, in addition to employing the staggered DiD method, this study includes a robustness check using a standard Two-Way Fixed Effects (TWFE) model. By comparing the results from the staggered DiD method and the TWFE model, the added value of using the staggered DiD approach is highlighted. The TWFE model is an extension of the fixed effects model that controls for unobserved heterogeneity by including unit- and time-fixed effects. The staggered DiD method is expected to provide more accurate and reliable estimates, as it better handles variations in treatment effects over time and across different groups. This comparison underscores the robustness of the findings and demonstrates the methodological importance of employing the staggered DiD approach in this context.

5. Results

5.1 Regression Results

5.1.1 Hypothesis 1: Impact on Match Characteristics

All results from the Callaway and Sant' Anna Difference-in-Differences models are shown in columns 2 to 8 of Table 3. Furthermore, figures 1 to 7 zoom in on the group treatment effects for each match characteristic and its 95% confidence intervals.

Table 3

Staggered Callaway and Sant'Anna Difference-in-Differences Effects

	(1) avg_yellows	(2) avg_reds	(3) avg_fouls	(4) avg_shots	(5) avg_SOT	(6) avg_goals	(7) avg_corners
VAR_use	0.516***	0.026	-0.575	-0.529	-0.106	-0.224*	0.185
	(0.172)	(0.032)	(0.525)	(0.546)	(0.309)	(0.118)	(0.212)
Observations	448	448	448	448	448	768	448

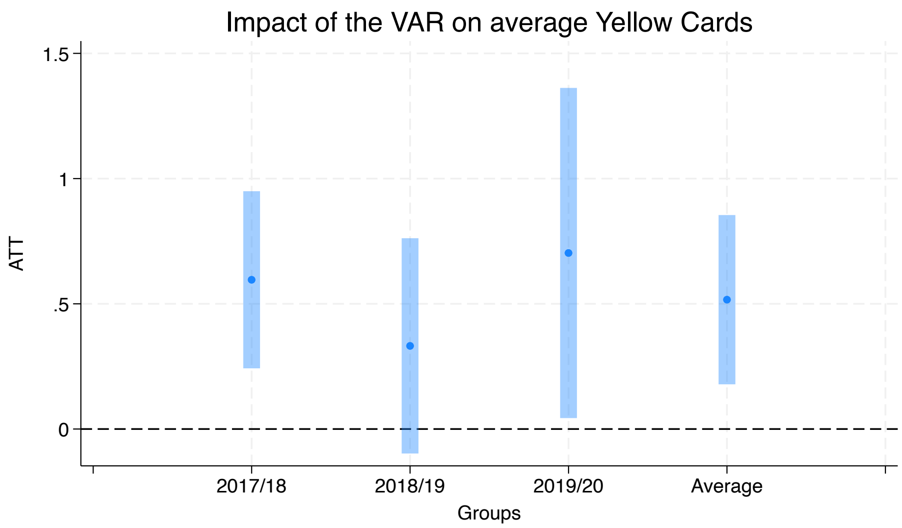
Note. All values have been rounded to three decimal places. VAR_use is the aggregated effect of the use of the VAR in all different treatment timings and groups on the dependent variables. Standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Contrary to recent literature from Lago-Peñas et al. (2019), the results show an increase in the number of yellow cards. On average, being treated with the use of the VAR increases the average number of yellow cards per match per season by 0.516 cards compared to being untreated, ceteris paribus, which is significant at a 1% significance level. This effect is quite large as the increase is 13.4% of the mean of average yellow cards for teams that were not yet treated with the VAR (VAR_use = 0).

This effect represents the average of the three treatment effects at different treatment timings, as illustrated in Figure 1. Figure 1 illustrates the impact of VAR on the average number of yellow cards. The x-axis represents different groups based on the season they started using VAR (2017/18, 2018/19, 2019/20) and the overall average. The y-axis shows the average treatment effect (ATT) of VAR on the number of yellow cards. The blue dots indicate the estimated average treatment effect for each group and the overall average. The vertical blue lines around the dots represent the 95% confidence intervals for each estimate.

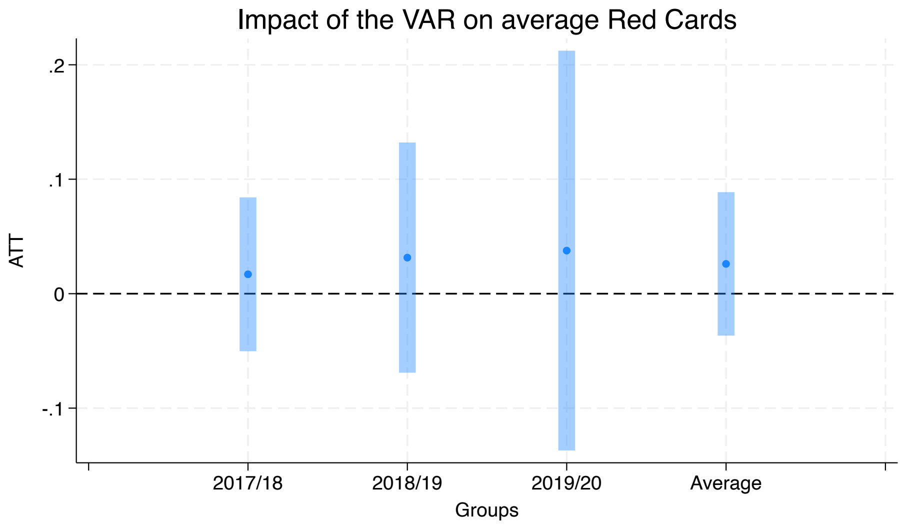
The plot shows that for the 2017/18 group, the effect of VAR on yellow cards indicates a positive and statistically significant increase, as the confidence interval does not cross the x-axis. The 2018/19 group's estimated effect is also positive, but its confidence interval crosses the x-axis, indicating that this effect is not statistically significant at the 5% level. The effect for the 2019/20 group is positive and statistically significant, with its confidence interval not intersecting the x-axis. The overall average effect, combining all groups, shows a positive and statistically significant increase in yellow cards due to VAR. Concluding, this figure illustrates that all treatment groups experienced an increase in yellow cards with the introduction of VAR. However, the 2018/19 group's effect is not statistically significant, as shown by its confidence interval intersecting the x-axis. Despite this, the other groups and the overall average effect remain statistically significant, reinforcing the conclusion that VAR generally leads to more yellow cards being issued.

Figure 1



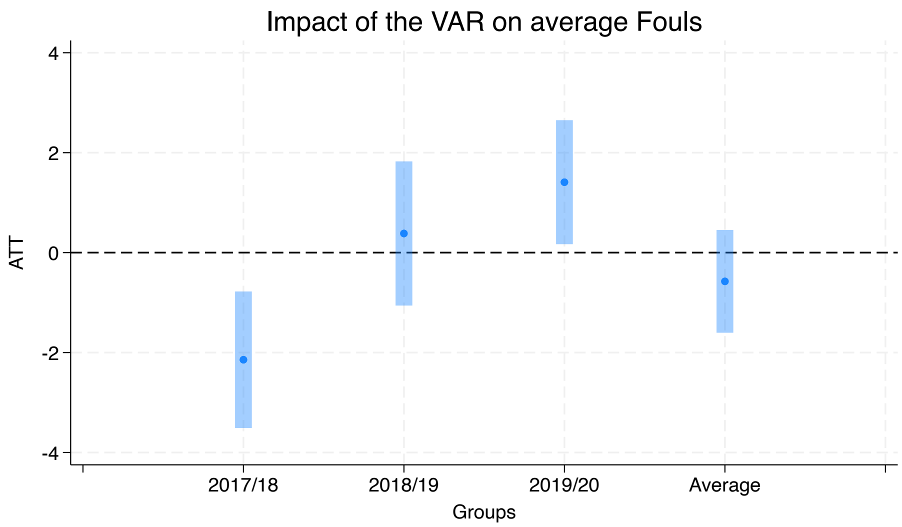
Next, on average, the average number of red cards per match per season increases by 0.026 cards when being treated with the use of the VAR compared to being untreated, ceteris paribus, but this is not significant at a 10% significance level. This result is in line with the findings of Spitz et al. (2020). Figure 2 also shows that, despite all effects being positive, none of the treatment groups has a statistically significant effect at the 5% significance level. This illustration strengthens the suggestion that the VAR does not have a statistically significant effect on red cards.

Figure 2



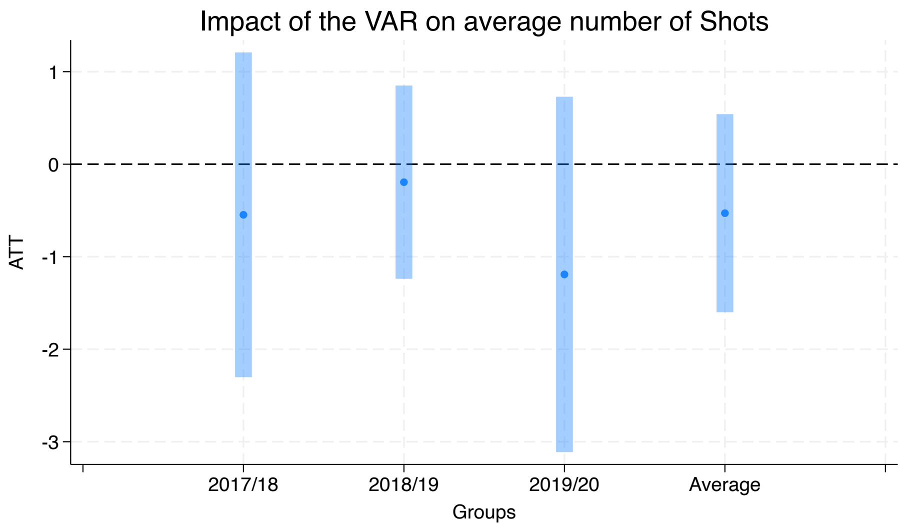
Next, the results of the average number of fouls are in line with the findings of Lago-Peñas et al. (2019). Namely, being treated with the use of the VAR decreases the average number of fouls per match per season by 0.575 fouls compared to being untreated, ceteris paribus, which is not significant at a 10% significance level. When zooming in on the group treatment effects in Figure 3, a negative and significant treatment effect for the first group is found, the second is positive and insignificant and the last is positive and significant. As all effects are different, there is no statistically significant average effect.

Figure 3



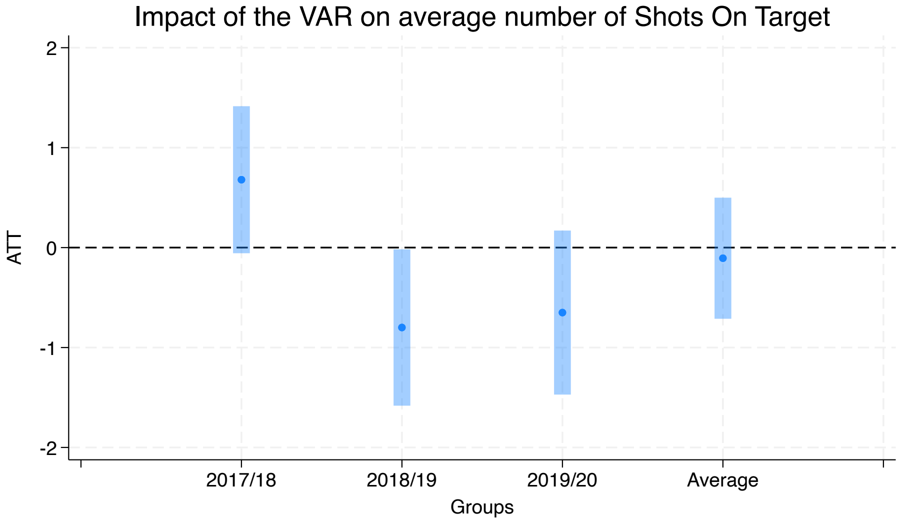
The next figure shows, that on average, being treated with the use of the VAR decreases the average number of shots per match per season by 0.529 shots compared to being untreated, ceteris paribus, which is not significant at a 10% significance level. Figure 4 is in line with these findings, illustrating that all group effects are negative but insignificant at the 5% significance level.

Figure 4



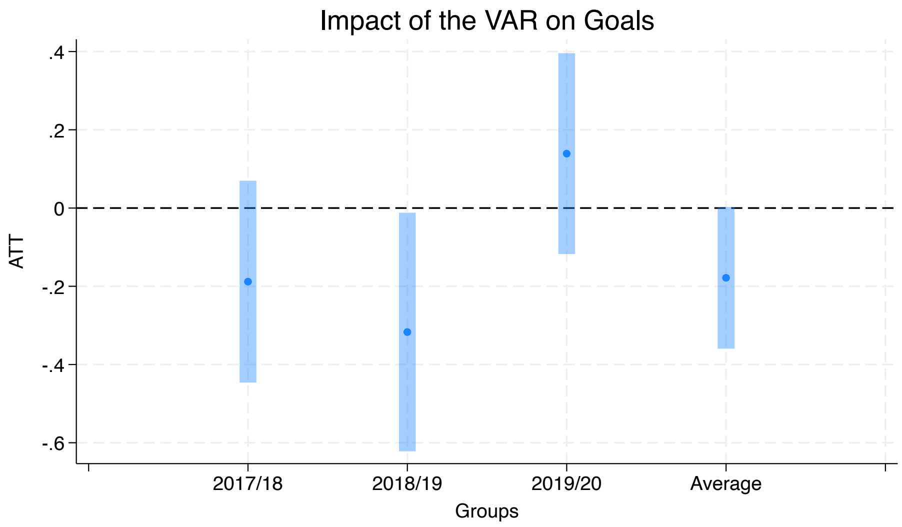
Reversibly, the average shots on target per match per season on average decreased by 0.106 shots on target compared to being untreated, ceteris paribus, which again is not significant at a 10% significance level. This effect is again supported by the illustration in Figure 5. The figure shows that the treatment effect for the first group is positive and insignificant, the second is negative and significant and the last is negative and insignificant. From this, the conclusion can be drawn that there is no consistent and significant effect of the VAR on shots on target, as each group has a different treatment effect in sign and significance.

Figure 5



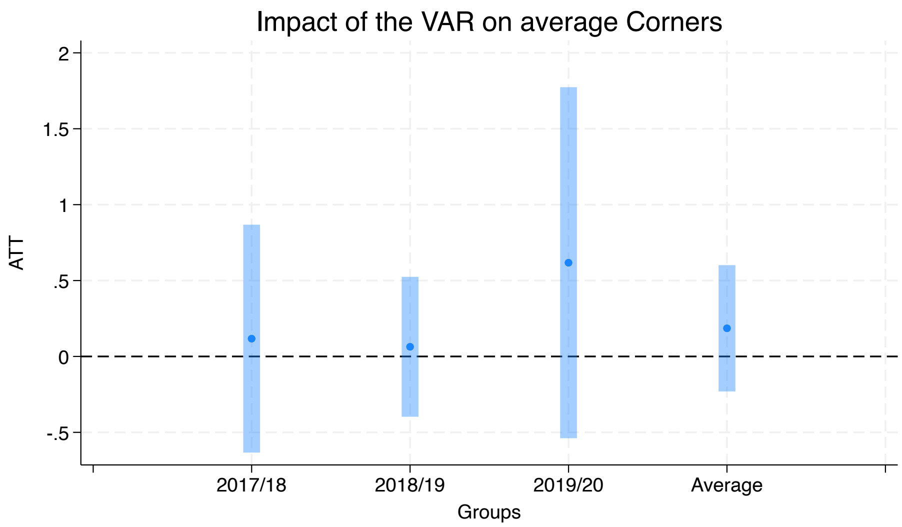
While the effects of the shots and shots on target were not researched yet, the goals scored were researched by Spitz et al. (2020). Contrary to their results, these findings show that on average, being treated with the use of the VAR decreases the average number of goals scored per match per season by 0.178 goals compared to being untreated, ceteris paribus, which is significant at a 10% significance level. This effect is relatively small compared to the effect on yellow cards, as it is only 6,2% of the mean of average goals scored for teams that were not treated yet with the VAR (VAR_use = 0). However, Figure 6 shows interestingly that for the last treated group, the effect is positive but insignificant. Illustrating that not all group effects are negative and only the 2018/19 effect is significant at the 5% significance level, this average negative significant effect should be interpreted with caution.

Figure 6



Lastly, the results show that on average, being treated with the use of the VAR increases the average number of corners per match per season by 0.185 corners compared to being untreated, ceteris paribus, which is not significant at a 10% significance level. Figure 7 is again in line with these findings, illustrating that all group effects are positive and insignificant at the 5% significance level.

Figure 7



In summary, as hypothesized, VAR implementation has led to a significant increase in yellow cards and a decrease in goals, but with no significant impact on fouls, red cards, shots, shots on target, or corners.

5.1.2 Hypothesis 2: Impact on Predictability of Match Outcome

For the second hypothesis, the results show an increase in the difference between the predicted match outcome and the actual match outcome. On average, being treated with the use of the VAR increases the average difference between the predicted match outcome and the actual match outcome per match per season by 2.6 percentage points compared to being untreated, ceteris paribus, which is not significant at a 10% significance level. This result can be seen in column 2 of Table 4.

Table 4

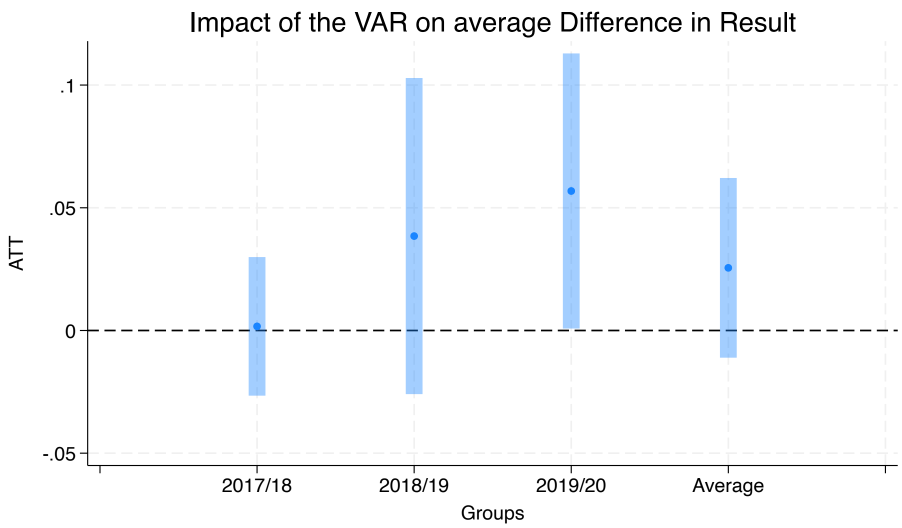
Staggered Callaway and Sant'Anna Difference-in-Differences Effects

	(1) avg_DiffResult
VAR_use	0.026 (0.019)
Observations	768

Note. All values have been rounded to three decimal places. VAR_use is the aggregated effect of the use of the VAR in all different treatment timings and groups on the dependent variable. Standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

In line with this result, figure 8 shows positive treatment effects for each treatment group but only significant at the 5% significance level for the treatment group that got treated from the start of 2019/20.

Figure 8



In summary, VAR implementation has not led to a significant decrease in the alignment of the predicted and real match outcomes.

5.2 Checks on the Assumptions

To check whether the control units are valid counterfactuals for the treatment units, the parallel trends assumption is tested by plotting the treatment trends and running statistical tests. The SUTVA is validated by argumentation.

5.2.1 Conditional Parallel Trends Assumption

For all staggered DiD models, Figures 9 to 16 visually display the trends of each treatment group. By examining the pre-treatment trends, it can be argued whether the parallel trends assumption is likely to hold or be violated. For example, in Figures 9 and 14, which show yellow cards and goals respectively, the trends appear to be parallel, indicating that the parallel trends assumption is likely to hold for these outcomes. In the other figures, there is also no strong evidence of non-parallel trends, except for fouls in Figure 11, where two treatment groups show positive trends while the others show negative trends. This suggests that for most outcomes, the parallel trends assumption is reasonably valid.

Figure 9

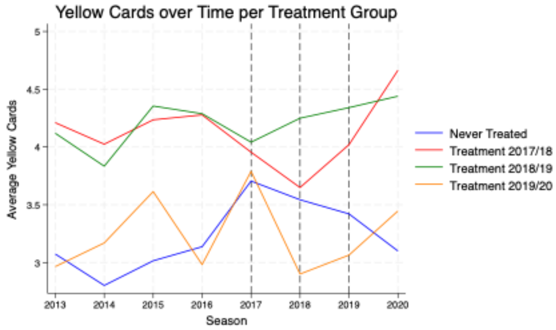


Figure 10

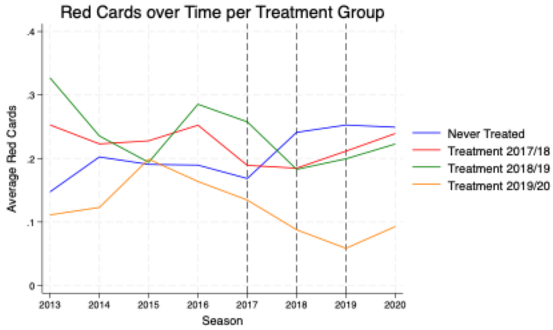


Figure 11

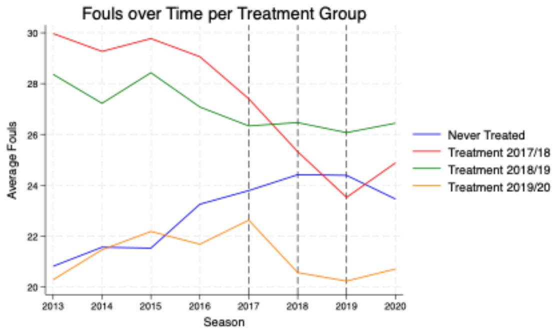


Figure 12

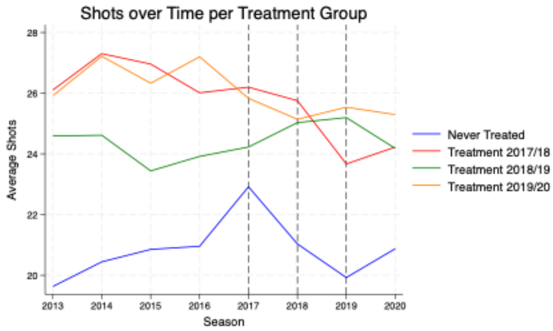


Figure 13

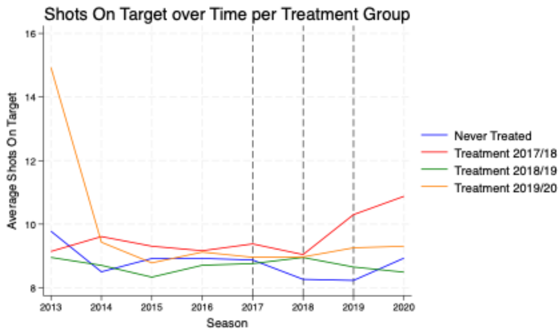


Figure 14

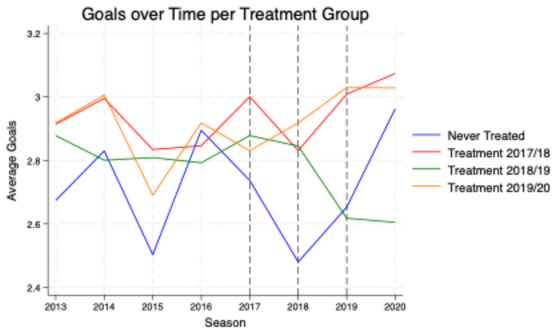


Figure 15

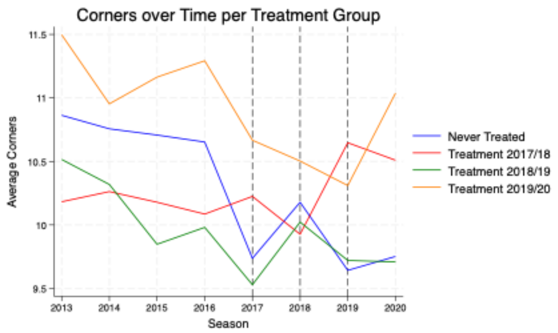
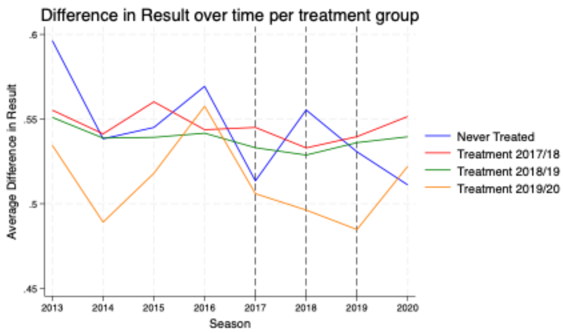


Figure 16



Next, the pre-treatment trend tests are conducted. Contrary to the visual inspection, the statistical tests indicate non-parallel trends. As shown in Table 5, all p-values are significant at the 1% level, rejecting the null hypothesis that there is no trend in the outcome variable across the pre-treatment periods. This suggests that the conditional parallel trends assumption does not hold for any of the models.

Table 5

Pre-treatment Trends Tests for Conditional Parallel Trends Assumption

	(1) avg_yellows	(2) avg_reds	(3) avg_fouls	(4) avg_shots	(5) avg_SOT	(6) avg_goals	(7) avg_corners	(8) avg_Diff Result
Chi2(15)	135.688	83.797	111.924	150.632	402.713	42.410	91.493	57.241
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Obs.	448	448	448	448	448	768	448	768

Note. All values have been rounded to three decimal places. Chi2(15) is the chi-square statistic with 15 degrees of freedom, a test to specifically test the null hypothesis (H_0) that there is no trend in the outcome variable across the pre-treatment periods. The p-value is the probability of observing a chi-square statistic more extreme than the chi-square statistic, assuming the null hypothesis is true.

However, the visual inspection of the trends (Figures 9 to 16) indicates that the trends appear to be more parallel than suggested by the chi-squared tests. Particularly when considering the scale of the plots and the relatively stable nature of most variables in the long run.

Despite the visual indications of parallel trends in many cases, the chi-squared tests reject the null hypothesis for all outcomes, indicating non-parallel trends. These controversial results highlight the importance of considering both visual and statistical evidence. The visual plots suggest that the parallel trends assumption might be less of an issue than the chi-squared test indicates, especially for certain outcomes. However, from these diverging results it is crucial to acknowledge that it is not certain that all trends are parallel, which could introduce bias in the estimated treatment effects. This discrepancy could come from substantial differences between the treatment and control groups or leagues that evolve over time, such as variations in playing styles, team strategies, formations, referee styles, or league policies.

5.2.2 SUTVA

Regarding the SUTVA, it is reasonable to argue that there is no interference between units and consistent treatment application across units and time. Firstly, European football leagues operate independently, meaning VAR implementation in one league is unlikely to directly influence match characteristics (e.g., goals, fouls, cards) in another league. For instance, the introduction of VAR in the Bundesliga should not affect matches in the Premier League. Additionally, teams of course cannot change from leagues in which they want to play based on VAR implementation, preventing any anticipation of treatment by joining a league with or without VAR.

Moreover, there is uniform implementation within leagues. Besides some testing periods in practice games or lower leagues, once VAR is introduced, it is used in all matches within that league, ensuring any observed changes in match characteristics are attributable to VAR. Despite strong reasons to believe the SUTVA holds, there might be inconsistencies in VAR application within a league. Garretsen et al. (2024) for instance found that younger, less experienced, and lower-ranked VARs significantly let the on-field referee review their decisions more often. Consequently, they see the initial decision significantly more often getting confirmed. So, it can be concluded that although the introduction of the VAR was intended to be an objective add-on for refereeing, the VAR application matters for the decision-making process, and it can differ between referees or leagues. Ultimately, the final decision on the field still lies with the referees, who might interpret the same situation differently, leading to inconsistent rule enforcement. This issue persists even though decisions have become more consistent post-VAR introduction. The non-parallel trends indicate potential league-specific differences, suggesting that other league- or application-specific factors can also impact the fulfillment of the SUTVA.

Inconsistencies can also arise from factors like malfunctioning VAR equipment, varying availability of camera angles between stadiums, and league-specific guidelines or protocols. Evolving guidelines for VAR usage and external pressure from media and fans can influence how referees and VAR officials apply the technology. Given these potential league-specific differences in guidelines and protocols, it is plausible that these variations contribute to the difficulties in meeting this assumption.

5.2.3 Two-Way Fixed Effects Model

Lastly, to further validate the findings and illustrate the advantages of the staggered DiD approach, a Two-Way Fixed Effects (TWFE) model was estimated for each outcome. As shown in Table 6, the TWFE model results differ a lot from those of the staggered DiD models. For average yellow cards, red cards, and shots on target, the signs of the coefficients are opposite to those found in the staggered DiD models. This indicates that the direction of the effect changes depending on the model used. On the other hand, fouls, shots, goals, corners, and the difference in result show the same sign as in the staggered DiD models, suggesting some consistency across methodologies.

Moreover, another notable difference lies in the statistical significance of the results. The TWFE model indicates that only the average number of fouls has a significant effect, whereas the staggered DiD models showed significant effects for average yellow cards and goals. This controversy in results underscores the added value of the staggered DiD approach in capturing the different treatment effects for each different group at a different treatment timing of VAR implementation that the TWFE model overlooks as it only estimates the difference between treated and not treated.

Table 6

Two-Way Fixed Effects models

	(1) avg_yellows	(2) avg_reds	(3) avg_fouls	(4) avg_shots	(5) avg_SOT	(6) avg_goals	(7) avg_corners	(8) avg_Diff Result
VAR_use	-0.160	-0.011	-2.679***	-0.367	0.268	-0.032	0.327	0.013
	(0.117)	(0.021)	(0.523)	(0.325)	(0.219)	(0.068)	(0.219)	(0.010)
Obs.	448	448	448	448	448	768	448	768

Note. All values have been rounded to three decimal places. VAR_use is the aggregated effect of the use of the VAR on the dependent variables. Standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

6. Discussion

The implementation of Video Assistant Referee (VAR) in men's European football leagues between 2017 and 2020 has resulted in significant changes to some match characteristics. However, the effect of VAR on the predictability of match outcomes remains uncertain.

6.1 Implications for Match Characteristics

The findings reveal that the introduction of VAR has had a mixed impact on match characteristics. Specifically, the analysis indicates a significant increase in the average number of yellow cards issued. This could be due to stricter rule enforcement, potentially influenced by the increased supervision provided by the VAR. While the VAR itself cannot directly issue yellow cards, its presence might influence referees to be more vigilant in identifying and penalizing offenses.

Moreover, the observed decrease in the average number of goals scored per match under VAR indicates that the system might be influencing offensive strategies or referees' decisions regarding goal-scoring opportunities. This could be due to the increased likelihood of disallowed goals upon VAR review, leading teams to adopt more conservative attacking approaches. This outcome is critical for teams and coaches as it underscores the need for strategic adjustments in response to VAR's presence.

Despite the observed significant changes, this study found no substantial impact on the number of fouls, red cards, shots, shots on target, or corners. Indicating that while VAR affects certain aspects of match play, other elements remain relatively stable. The unchanged frequency of red cards might imply that the severity of offenses leading to red cards is less sensitive to VAR intervention compared to yellow cards, even since the VAR does not interfere in yellow card situations. In summary, the indirect behavioral changes outweigh the direct effect of the VAR intervention. Similarly, the stable number of shots and corners indicates that teams' overall offensive efforts and set-piece opportunities are not drastically influenced by the VAR.

6.2 Predictability of Match Outcomes

Contrary to expectations, the predictability of match outcomes, as measured by alignment with pre-match bookmaker odds, did not show significant changes post-VAR implementation. This finding suggests that while VAR improves decision accuracy and diminishes referee biases, according to recent literature, it does not necessarily translate into less predictable match outcomes. This could be due to potential new inconsistencies arising from the implementation of the VAR, canceling out the improved change in refereeing accuracy and biases, leading to no significant change in predictability.

Existing literature supports the notion that fans value the uncertainty of match outcomes as a core element of football's appeal. Therefore, the stable predictability of results post-VAR might contribute positively to maintaining spectator interest and engagement. Unless not decreasing the predictability and improving spectator engagement, stable predictability highlights a critical balance between improving decision accuracy and preserving the engagement and unpredictability that define the sport.

6.3 Robustness and Methodological Considerations

The robustness checks conducted, including the tests for the conditional parallel trends' assumption and the SUTVA, raised some concerns about the validity of these findings. While the plots did not show concerns about the trends being non-parallel, the chi-squared test results did. This makes it extra difficult to conclude these conflicting findings, but as the plots indicate, there is no need to worry about the non-parallel trends. However, it cannot be denied that the results should be interpreted with some caution until the statistical results also indicate that the trends are parallel.

To get to statistically parallel trends, it could help to account for league- or team-specific characteristics when generalizing the impact of VAR. Different teams and leagues may have unique playing styles, player behavior, or even the physicality of the game, which are not evenly distributed across all teams and leagues and do not remain the same over the sample period. Ignoring these changing factors could lead to non-parallel trends, biased estimates, and incorrect conclusions about the overall effect of VAR. But only if the factors are significantly changing the effect of the VAR.

On the other hand, referee practices, and regulatory environments that differ per league can affect the SUTVA assumption. For instance, the way referees in the Bundesliga interpret and apply VAR might differ from their counterparts in the Premier League or La Liga. Such variations could lead to differential impacts of VAR across leagues, complicating the comparison and generalization of results. However, VAR implementation is quite the same across all leagues and as VAR implementation in one league is not influencing match characteristics in another league, this provides enough evidence for the SUTVA to hold. Nevertheless, due to the complex context of the VAR's effects, a tailored approach is essential in evaluating VAR's effects.

6.4 Economic Contribution

This research adds economic value in highlighting the importance of using advanced econometric techniques, such as the staggered DiD approach, to account for variations in treatment effects over time and across different groups. Because, while the TWFE model could provide useful insights into the effects of certain treatments, the staggered DiD method offers more detailed and reliable estimates of the impact of treatments that are not implemented at the same time for different groups. This research shows that using a staggered DiD approach estimates different results than using TWFE for the effect of the VAR on match characteristics and the predictability of match outcomes in European football leagues.

Furthermore, this paper showed that the VAR system increased the average number of yellow cards and decreased the average number of goals scored. As the VAR cannot be used for yellow cards, this means the decision-making of the on-field referees or players is likely to be impacted by the oversight of the VAR. This means that referees and/or players have less bias in their decision-making and/or have less moral hazard in refereeing or playing. This reduction in biases demonstrates that VAR or system oversight affects behaviors, providing essential insights for designing and optimizing oversight mechanisms in other fields. This parallels how stricter oversight by higher courts impacts district judges' sentencing decisions, highlighting the broader implications of oversight systems on decision-making processes.

Lastly, this study also contributes to finding that the VAR is not significantly affecting many aspects of the game such as red cards, fouls, shots, shots on target, corners, and the predictability of the game. This finding suggests that the beautiful game remains to be the beautiful game with the VAR as a tool to improve referee accuracy as Spitz et al. (2020) found that VAR increased correct decisions from 92.1% to 98.3%.

6.5 Future Research Directions

Future research should explore the differences across leagues and competitions that affect match characteristics and outcomes. Comparing and statistically verifying the differences across leagues in playing styles, player behavior, physical abilities, and qualities and incorporating those differences as covariates in the regressions, will help isolate the true effect of VAR from other confounding factors. This approach will lead to more accurate and reliable conclusions.

Furthermore, new research should explore the impact of the VAR on player behavior and referees' decision-making. This will allow us to not only understand how technological innovations such as the VAR impact the game, but also how oversight reviewing (in general) impacts decision-making in other economic situations. However, such research may also provide answers to the question of where the impact of VAR on the number of yellow cards and/or goals comes from.

7. Conclusion

In conclusion, the implementation of VAR has significantly impacted certain match characteristics in men's European football leagues. Notably, it has led to a significant increase in yellow cards, alongside a decrease in goals scored per match. However, its influence on the predictability of match outcomes appears limited.

It's important to acknowledge that these findings may be influenced by limitations in this study. While there are more reasons to accept both assumptions, it can also be argued that neither is true. Therefore, these results should be interpreted with caution. Nevertheless, these findings highlight the complex nature of VAR's influence on football, indicating the need for ongoing research and context-specific evaluations to fully understand and optimize the role of the Video Assistant Referee in football.

8. References

Albanese, A., Baert, S., & Verstraeten, O. (2020). Twelve eyes see more than eight. Referee bias and the introduction of additional assistant referees in soccer. PloS One, 15(2), e0227758. https://doi.org/10.1371/journal.pone.0227758

Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to home advantage in English Premiership football. Journal Of Sports Sciences, 25(11), 1185–1194. https://doi.org/10.1080/02640410601038576

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. Journal Of Econometrics, 225(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Cox, A. (2015). Spectator demand, uncertainty of results, and public interest. Journal Of Sports Economics, 19(1), 3–30. https://doi.org/10.1177/1527002515619655

Da Silva, M. L., Filho, V. C. B., De Lima E Silva, L., De Alkmim Moreira Nunes, R., De Lourdes Preciado, M., Barreira, D., & Campaniço, J. (2023). Video assistant referee in soccer: a scoping review. Retos: Nuevas Tendencias en Educación Física, Deportes Y Recreación, 50, 1163–1171. https://doi.org/10.47197/retos.v50.95409

Dohmen, T., & Sauermann, J. (2015). Referee bias. Journal Of Economic Surveys, 30(4), 679–695. https://doi.org/10.1111/joes.12106

Dufner, A., Schütz, L., & Hill, Y. (2023). The introduction of the video assistant referee supports the fairness of the game – An analysis of the home advantage in the

German Bundesliga. Psychology Of Sport And Exercise, 66, 102386. https://doi.org/10.1016/j.psychsport.2023.102386

Erikstad, M. K., & Johansen, B. T. (2020). Referee bias in professional football: favoritism toward successful teams in potential penalty situations. Frontiers in Sports And Active Living, 2. https://doi.org/10.3389/fspor.2020.00019

Fontenla, M., & Izón, G. M. (2018). The effects of referees on the final score in football. https://www.redalyc.org/journal/5723/572366594004/html/

Fischman, J. B., & Schanzenbach, M. M. (2011). Do standards of review matter? The case of federal criminal sentencing. The Journal Of Legal Studies, 40(2), 405–437. https://doi.org/10.1086/659262

Forrest, D., Simmons, R., & Buraimo, B. (2005). Outcome uncertainty and couch potato audience. Scottish Journal Of Political Economy, 52(4), 641–661. https://doi.org/10.1111/j.1467-9485.2005.00360.x

Garretsen, H., Stoker, J., Alessie, R., & Talsma, B. (2024). Who rules in terms of the video assistant referee? Decision making in dutch football. EUR. https://www.eur.nl/en/ese/media/2024-03-garretsen

Han, B., Chen, Q., Lago-Peñas, C., Wang, C., & Liu, T. (2020). The influence of the video assistant referee on the Chinese Super League. International Journal Of Sports Science & Coaching, 15(5–6), 662–668. https://doi.org/10.1177/1747954120938984

Holder, U., Ehrmann, T., & König, A. (2021). Monitoring experts: insights from the introduction of video assistant referee (VAR) in elite football. Journal Of Business Economics/Zeitschrift Für Betriebswirtschaft, 92(2), 285–308. https://doi.org/10.1007/s11573-021-01058-5

Inside FIFA. (2023). Video assistant referee (VAR). https://inside.fifa.com/technical/football-technology/football-technologies-and-innovations-at-the-fifa-world-cup-2022/video-assistant-referee-var

Johnson, S., & Fort, R. (2022). Match outcome uncertainty and sports fan demand: an agnostic review and the standard economic theory of sports leagues. International Journal Of Empirical Economics, 01(02). https://doi.org/10.1142/s281094302250007x

Kubayi, A., Larkin, P., & Toriola, A. (2021). The impact of video assistant referee (VAR) on match performance variables at men's FIFA World Cup tournaments. Proceedings Of The Institution Of Mechanical Engineers, Part P: Journal Of Sports Engineering And Technology, 236(3), 187–191. https://doi.org/10.1177/1754337121997581

Lago-Peñas, C., Rey, E., & Kalén, A. (2019). How does video assistant referee (VAR) modify the game in elite soccer? International Journal Of Performance Analysis in Sport, 19(4), 646–653. https://doi.org/10.1080/24748668.2019.1646521

McCormick, R. E., & Tollison, R. D. (1984). Crime on the court. Journal Of Political Economy, 92(2), 223–235. https://doi.org/10.1086/261221

Nevill, A., Balmer, N., & Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. Psychology Of Sport And Exercise, 3(4), 261–272. https://doi.org/10.1016/s1469-0292(01)00033-4

Sors, F., Grassi, M., Agostini, T., & Murgia, M. (2020). The sound of silence in association football: Home advantage and referee bias decrease in matches played without spectators. European Journal Of Sport Science, 21(12), 1597–1605. https://doi.org/10.1080/17461391.2020.1845814

Spitz, J., Wagemans, J., Memmert, D., Williams, A. M., & Helsen, W. (2020). Video assistant referees (VAR): The impact of technology on decision making in association football referees. Journal Of Sports Sciences, 39(2), 147–153. https://doi.org/10.1080/02640414.2020.1809163

Sport&Strategie. (2019). KNVB neemt het voortouw met de VAR, maar waarom? Sport&Strategie. Pieter van der Meer. https://www.sportenstrategie.nl/sportinnovatie/knvb-neemt-het-voortouw-met-de-var-maar-waarom/

Winand, M., Schneiders, C., Merten, S., & Marlier, M. (2021). Sports fans and innovation: An analysis of football fans’ satisfaction with video assistant refereeing through social identity and argumentative theories. Journal Of Business Research, 136, 99–109. https://doi.org/10.1016/j.jbusres.2021.07.029