The Impact of VAR: Balancing Technology and Tradition in Soccer

Keegan Burr

2024-12-11

Introduction

Soccer, known to many as football, is one of the most popular sports in the world. Clubs, fans, leagues, sponsors, and broadcasters invest billions of dollars into the sport each year. Fans are passionate, and the stakes are high. The games mean so much to so many, and because of this, the game is ever-evolving. One of the most significant changes in the sport's history was the implementation of Video Assistant Referees, also known as VAR. According to the English Premier League, VAR is "a qualified referee who watches the match via a number of screens and can view slow-motion replays, enabling them to advise the on-field referee." It was initially introduced to be used only in the event of a 'clear and obvious error,' but since VAR's introduction, the system has sparked constant debate. Many question the need for technology in a beloved sport that has been played for hundreds of years. In a study done by Jack Kenyon Brown, Brown examines how VAR has influenced what is considered the best league in the world, the English Premier League. His analysis utilizes generalized linear models (GLMs) with negative binomial random components to capture non-normal distributions of count data, specifically looking at variables such as goals, penalties, offsides, etc. In this paper, I will critically examine and evaluate Brown's methodology, to understand the validity of his conclusions about VAR's impact on the Premier League.

Analysis of Methods

The methodology section in Brown's study is both in-depth and multifaceted. Brown uses a combination of traditional statistical tests, alongside visual aids, and finally, a generalized linear model (GLM) to attempt to understand the overall impact VAR has had on the Premier League. In this section, in an attempt to validate Brown's methods and results, and to deepen my understanding, I will explore the rationale behind key decisions Brown made. Additionally, I will identify gaps within the methodology that could raise questions, and provide potential improvements and additional insights that strengthen the analysis.

Starting from the base level, we begin with the data collection for the analysis. Brown's study utilized publicly available soccer data from FBref.com, leveraging packages found in RStudio like XLM and RCurl for web scraping the data. The dataset covered all 3,800 Premier League games across the last ten completed seasons, 2014-2024. The decision to use this particular dataset is commendable given its comprehensiveness, as it included dozens of data points for each match. However, my first question regarding Brown's methodology is why would he choose to retain only ten variables.

For the analysis, Brown chose variables such as goals, fouls, and penalty kicks awarded. His reasoning behind selecting these variables was, "we selected 10 variables of interest that we felt may be most impacted by the implementation of VAR." All variables he selected and the reason of interest can be found below in table one. However, I believe by using this approach he excluded other potentially relevant variables like possession percentage, dribbles attempted, a team's position in the league, etc. By excluding these kinds of variables, Brown potentially limited the scope of his analysis. Instead, I believe Brown should've retained all variables that were available within the dataset and made an assessment of their influence. This could be done by simply testing variables' significance levels with a hypothesis test. Then determine if the p-value associated with each variable contributes significantly to the outcome, and if not, remove any variables with a p-value exceeding the predetermined threshold (likely significance level of 0.05). This would have provided a data-driven approach to identifying significant predictors, ensuring that no potentially important variables were overlooked.

Variables Table 10 Variables of Interest

Variable	*Reason of Interest*
Goals	Goals are the most important part of the game
Fouls	Has player discipline improved because of players being scared to get fouls checked by VAR?
Penalty Kicks Awarded	Has VAR lead to an increase in awarded penalties now that every possible instance is double-checked using VAR slow-motion replay technology?

Variable	*Reason of Interest*
Penalty Kicks Scored	Has the waiting time during VAR checks affect the number of penalties scored?
Shots	Has VAR affected the attacking tendencies of the game?
Offsides	Has VAR technology led to an increase of catching people offsides?
Red Cards	Have red cards gone down because of players scared to get fouls checked by VAR? Or, have red cards gone up because of VAR slow-motion replay technology?
Yellow Cards	Have yellow cards gone down because of players scared to get fouls checked by VAR?
Tackles	Have tackles gone down because of players scared to get fouls checked by VAR?
Interceptions	Have interceptions gone down because of players scared to get fouls checked by VAR?

In regards to the assumption testing, Brown conducted normality tests using the Shapiro-Wilk, Anderson-Darling, and Kolmogorov-Smirnov tests (These test formulas can be viewed below). These tests are designed to evaluate whether a dataset follows a normal distribution. The Shapiro-Wilk test, in particular, is sensitive to deviations from normality in smaller sample sizes, while the Anderson-Darling test emphasizes discrepancies in the tails of the distribution. The Kolmogorov-Smirnov test compares the sample distribution to a theoretical normal distribution. In Brown's analysis, these tests would confirm that all ten variables of interest did not strictly follow a normal distribution. Although the Central Limit Theorem would suggest normality should emerge in large samples, these results showed deviations that would merit consideration. So, Brown used histograms and the visualizations supported these findings.

Test Name	*Formula*
Shapiro-Wilk	$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
Anderson-Darling	$A^2 = -N - S$ where $S = \sum_{i=1}^{N} \frac{(2i-1)}{N}[\ln F(Y_i) + \ln(1 - F(Y_{N+1-i}))]$

Test Name	*Formula*
Kolmogorov-Smirnov	$D = \max_{1\leq i\leq N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i)\right)$

Following this, Brown performed Breusch-Pagan and Levene's tests for equal variance. The Breusch-Pagan test's purpose is to detect heteroscedasticity or non-constant variance, in regression models, while a Levene's test evaluates the equality of variances across groups. These two tests would reveal that of the ten variables, two variables, offsides, and interceptions, violated the equal variance assumption. This in turn would lead to the decision to use Welch's T-test ($t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$). The reasoning behind using this test was because a Welch's T-test does not assume equal variance, and thus would help mitigate these equal variance violations. To further strengthen his analysis by addressing the non-normal distribution of the data, Brown included a Mann-Whitney U-Test. The Mann-Whitney U-test is particularly effective in comparing two independent groups when the data does not meet normality assumptions. In this analysis, it allowed Brown to assess differences in variables like penalty kicks or fouls awarded between pre-VAR years and post-VAR years without the need for parametric approximations. Overall this ensures that the insights drawn were not skewed by deviations from normality.

After, Brown chose to use a generalized linear model (GLM) with a negative binomial distribution that was particularly well-suited for this study. The reason is that this approach is extremely effective as the negative binomial GLM is specifically designed for overdispersed count data, where the variance exceeds the mean. This is something that was evident in part of the match statistics, making this method a natural fit. Other alternative methods like Poisson regression would have not been as useful, as Poisson regression assumes equal mean and variance. Additionally, with logistic regression, we would only output a binary outcome and although the simplicity could be valuable in some contexts, in the context of Brown's study, binary outcomes would reduce the richness of the data.

After utilizing the aforementioned methods, for each response variable, Brown decided to perform each of the methods on three different subsets of EPL match data. This was in an attempt to understand the immediate, intermediate, and long-term impact VAR had on the Premier League. To do this, Brown created a one-year comparison (comparing the season before VAR to the season after), a three-year comparison, and a five-year comparison. Although this focus on pre- and post-comparisons provided flexibility in understanding the effects, its usage raises questions about the influence of confounding factors during these periods. Changes in coaching staff, rule changes, and external events like COVID-19 likely had some kind of impact on the data. However, it would be difficult to address this as of course no control group of an equivalent English Premier League did not use VAR during the same years that we can compare to, so an alternative method to account for the lack of a control group would've been to focus less on larger time intervals, and instead use more granular intervals. Intervals such as half-season or quarterly intervals would likely have helped reveal short-term trends that were more difficult to contextualize in larger periods. Recognizing these confounding variables in this study is essential, which is why, despite identifying some statistically significant differences, Brown ultimately concluded that it would be statistically irresponsible to attribute a drastic effect on the Premier League solely to VAR.

Analysis of Normative Consideration

The introduction of VAR technology in soccer has sparked significant debate regarding its impact on the fairness and integrity of the sport. Applying John Rawls' theory of "Justice as Fairness," VAR's uneven distribution across different leagues raises concerns about equity. Leagues with fewer resources do not have the same advantage, creating a disparity that undermines fairness in the sport. In this section, I will argue that while VAR theoretically enhances fairness, from a Rawlsian and Utilitarian perspective, such inequalities can only be justified if steps are taken to ensure that all have access to the technology and preserve the elements that make soccer meaningful. Without such measures, while VAR theoretically enhances fairness, its real-world application raises significant ethical questions concerning equity, tradition, and the dehumanization of the game.

Although VAR was designed with the intent of making refereeing 'fairer,' we must ask whether its implementation has actually created new inequalities. This brings us to a critical distinction between fairness as equity and fairness as equality. Fairness as equality would suggest that all teams and leagues should have access to the same resources and technology to ensure a level playing field. However, fairness as equity considers that different contexts require different levels of support, meaning that certain teams or leagues may need more resources to achieve the same standard of fairness. From the perspective of John Rawls' concept of 'Justice as Fairness,' these disparities in resource allocation would be ethically problematic. In particular, Rawls argues that inequalities are morally justifiable only if they benefit the least advantaged members of society. In the case of VAR, wealthier teams and leagues such as the Premier League can afford to implement the technology, leading to more accurate and fair refereeing decisions for those particular teams, but not for all.

In addition to Justice as Fairness, Utilitarianism offers another critical lens to assess the ethical implications of VAR. Utilitarianism seeks to maximize overall happiness and minimize harm for the greatest number of people. On one hand, VAR has the potential to improve fairness by reducing errors in critical moments, leading to more justified outcomes. However, alternatively, excessive reliance on VAR may unintentionally harm the game's emotional appeal and the enjoyment of fans. Soccer is not only about accuracy in decision-making but also about unpredictability and the emotional connection fostered by the people. Over-reliance on technology could diminish the role of referees. This shift toward a more mechanical, predictable game might reduce the excitement that comes from the unpredictability, ultimately undermining the overall enjoyment of the sport.

Furthermore, the over-reliance on VAR can have long-term consequences for the sport's integrity. Since the beginning, soccer has embraced human error as part of the game. By continually deferring to technology, the authority of referees could begin to be undermined. This may lead to a situation where the sport becomes even more dependent on technology, eliminating the role of human agency in the decision-making process. This ties directly to Utilitarianism, as it shows the potential harm caused by removing human judgment. So from a utilitarian standpoint, the goal would be to find a balance between fairness with the preservation of the qualities that make soccer so beloved. If VAR's overuse leads to a less enjoyable experience, an Utilitarian would argue implementation outweighs the benefits of fairness.

While the introduction of VAR has brought about significant differences in refereeing, its implementation has also raised an important normative question. Rawls' Justice as Fairness reminds us of the importance of addressing inequalities of VAR. Additionally, Utilitarianism reminds us to consider the broader consequences of technology's impact on the enjoyment of the sport. To align with both ethical frameworks, the implementation of VAR should be approached in a way that ensures equitable access across leagues and carefully considers the balance between fairness and the preservation of the human elements that make soccer meaningful.

Conclusion

Brown's analysis of VAR in the Premier League offers a rigorous statistical framework for examining its impact on various match statistics. Using generalized linear models with negative binomial random components, he effectively captures the intricacies of count data, addressing the non-normal distributions that are typical in this context. While Brown's research reveals some intriguing trends regarding VAR's influence, he acknowledges that it would be statistically irresponsible to conclude that VAR has had a drastic effect on the Premier League due to the presence of confounding variables. This approach emphasizes the need for caution in interpreting the results and highlights opportunities for future research, refining the analysis.

While VAR was introduced to enhance transparency and trust in officiating, it has often led to confusion and controversy, with even replayed decisions leaving players and fans uncertain. This ambiguity challenges the fairness and integrity of the game, raising the question of whether the benefits of more accurate decision-making justify the disruption to the sport's traditions. After critically examining and evaluating Brown's methodology, Brown's findings highlight the need for careful consideration of VAR's broader implications, both statistically and normatively, to determine whether it truly serves the best interests of football.

References

Cal Poly Digital Commons. The Impact of Video Assistant Referee (VAR) on the English Premier League. Accessed 11 Dec. 2024, https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?params=/context/theses/article/4475/&path_info=The_Impact_of_Video_Assistant_Referee___VAR___on_the_English_Premier_League.pdf.

Premier League. “Video Assistant Referee: How It Will Work in 2019/20.” Premier League, 31 July 2019. Accessed 11 Dec. 2024, https://www.premierleague.com/news/1293321?os=fuzzscan0xxtr&ref=app#!.

National Center for Biotechnology Information. “PMC11047540.” PubMed Central. Accessed 11 Dec. 2024, https://pmc.ncbi.nlm.nih.gov/articles/PMC11047540/.

Statology. “Breusch-Pagan Test: Definition and Examples.” Statology. Accessed 11 Dec. 2024, https://www.statology.org/breusch-pagan-test/.

Datatab. “Test of Normality – How to Conduct and Interpret It.” Datatab Tutorial. Accessed 11 Dec. 2024, https://datatab.net/tutorial/test-of-normality.

U.S. National Institute of Standards and Technology. “Normal Probability Plot.” Engineering Statistics Handbook, https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm. Accessed 11 Dec. 2024.

U.S. National Institute of Standards and Technology. “Histogram.” Engineering Statistics Handbook, https://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm. Accessed 11 Dec. 2024.

U.S. National Institute of Standards and Technology. “Goodness-of-Fit Tests.” Engineering Statistics Handbook, https://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm. Accessed 11 Dec. 2024.

University of Virginia Library. “Getting Started with Negative Binomial Regression Modeling.” UVA Library Research Data Services + Sciences. Accessed 11 Dec. 2024, https://library.virginia.edu/data/articles/getting-started-with-negative-binomial-regression-modeling.