CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Consequences of Implementation of Video Assistant Referee in Fortuna Liga

Bachelor's thesis

Author: Ondřej Habáň

Study program: Economics and Finance

Supervisor: prof. PhDr. Ladislav Krištoufek, Ph.D.

Year of defense: 2022

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 3, 2022

Ondrej Haban

Abstract

The thesis deals with the issue of the Video Assistant Referee in football. It evaluates the consequences of its implementation in Czech Fortuna Liga on the sample of 678 matches held during two and half seasons. The results from the models designed to treat count data were compared with relevant literature. In the form of both simple and multiple regression with additional control variables was investigated the relationship between VAR and the set of match-changing incidents, including yellow cards, red cards and penalty kicks, and the relationship between VAR and errors of on-pitch referees. The terms presence of VAR, VAR interventions and VAR as the whole were differentiated. Whereas a significant statistical association of VAR as the whole was not revealed for yellow and red cards, a 56% increase in the number of penalties associated with VAR as the whole significantly performed. Furthermore, the negative and highly significant 118% association of the presence of VAR was reckoned in the case of errors of on-pitch referees. Subsequently, the percentage decreased due to VAR interventions, however, not sufficiently to reveal a negative and significant association in errors of on-pitch referees for VAR as the whole. The exception created errors based on factual decisions.

JEL Classification	Z21, F21, Z29, O33
Keywords	football, VAR, match-changing incidents, errors of on-pitch referees, R Studio, count data models

Title	Consequences of Implementation of Video Assistant Referee in Fortuna Liga
Author's e-mail	`81767649@fsv.cuni.cz`
Supervisor's e-mail	`ladislav.kristoufek@fsv.cuni.cz`

Abstrakt

Tato bakalářská práce se zabývá problematikou videorozhodčího ve fotbale. Hodnotí následky jeho implementace v české Fortuna ligy na vzorku 678 zápasů, konaných v rámci období dvou a půl sezony. Výsledky modelů, připravených pro závislé proměnné nabývající pouze přirozená čísla, byly porovnány s relevantní literaturou. Ve formě jednoduchého a složeného modelu s přidanými proměnnými byl vyšetřován vztah mezi videorozhodčím a množinou důležitých zápasových incidentů zahrnujících žluté karty, červené karty a pokutové kopy a vztah mezi videorozhodčím a chybami rozhodčích na hřišti. Termíny přítomnost videorozhodčího, intervence videorozhodčího a VAR jako celek byly rozlišeny. Zatímco signifikantní statistická asociace videorozhodčího jako celku nebyla odhalena pro žluté ani červené karty, u penalt se 56% nárůst spojený s videorozhodčím signifikantně prokázal. Kromě toho byla naměřena negativní a vysoce signifikantní asociace spojená s přítomností videorozhodčího ve výši 118% v případě chyb rozhodčích na hřišti. Poté toto procento pokleslo kvůli intervencím videorozhodčího, avšak ne dostatečně na to, aby byla objevena negativní a signifikantní asociace v chybách rozhodčích na hřišti pro VAR jako celek. Výjimku tvořily chyby na základě faktuálních rozhodnutích.

Klasifikace JEL	Z21, F21, Z29, O33
Klíčová slova	fotbal, VAR, důležité zápasové incidenty, chyby rozhodčích na hřišti, R Studio, modelování počtu událostí

Název práce	Konsekvence implementace videorozhodčího ve Fortuna lize

E-mail autora	81767649@fsv.cuni.cz
E-mail vedoucího práce	ladislav.kristoufek@fsv.cuni.cz

Acknowledgments

I would like to thank to prof. PhDr. Ladislav Krištoufek, Ph.D., the supervisor of my thesis, who has supplied me a valuable feedback every time I was seeking for a helping hand. His remarks and recommendations meant a lot for me and I was always trying to take from them a piece of advice for a future work. It was my great pleasure to cooperate with him. Furthermore, big thanks belong also to my dear friend and colleague Bc. Vilém Krejcar, who provided me crucial tips regarding the final form of the thesis. Lastly, I would like to appreciate my colleagues from study and leisure group Sedmipočetníci, who motivated me to grow up during the whole study program.

Typeset in FSV LaTeX template with many thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Haban, Ondrej: *Consequences of Implementation of Video Assistant Referee in Fortuna Liga.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2022, pages 115. Advisor: prof. PhDr. Ladislav Krištoufek, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

NFL National Football League

GLT Goal-line Technology

VAR Video Assistant Referee

IFAB International Football Association Board

FIFA Féderation Internationale de Football Associations

KU Leuven Katholieke Universiteit Leuven

EPL English Premier League

GZD grey-zone decision

LOG Laws of the Game

OFR on-field review

IAAP Implementation Assistance and Approval Programme

VOR Video Operation Room

RRA Referee Review Area

AVAR Assistant Video Assistant Referee

RO replay operator

DOGSO denying an obvious goal-scoring opportunity

KNVB Royal Netherlands Football Association

FAČR Football Association of Czech Republic

F:L Fortuna Liga

LFA League Football Association of Czech Republic

NB Negative Binomial

GLM generalized linear model

TSL Turkish Super League

KR FAČR Committee of Referees of Football Association of Czech Republic

LDV limited dependent variables

OLS Ordinary Least Squares

QP Quasipoisson

MLE maximum likelihood estimation

CT test Cameron and Triveldi's test

QMLE quasi-maximum likelihood estimation

ZIP zero inflated Poisson

ZINB zero inflated Negative Binomial

AIC Akakike Information Criteria

QAIC quasi-Akakike Information Criteria

R controls control variables related to the figure of on-pitch referee

T controls control variables related to characteristics of teams and the match before the kick-off

M controls control variables regarding how the match proceeded

BS backward selection

VIF Variance Inflation Factor

Adj R^2 Adjusted R squared

UEFA Union of European Football Associations

UCL UEFA Champions League

UEL UEFA Europa League

UECL UEFA Europa Conference League

BC before consultation

AC after consultation

AC2 after consultation 2

Chapter 1

Introduction

Sport has been connected with technology for more than 100 years. Already at the end of the 19^{th} Century, the photo finish was used for the evaluation of a horse race on the east coast of the United States as the one of the first examples in the sport of using technology that should have helped officiates make more precise decision[1]. Plenty of sports had been joining themselves to the sport-technology carousel over the 20^{th} Century and the beginning of 21^{st} Century. We could adduce examples of the instant replay system using by National Football League (NFL) that was introduced in 1985 or the tennis Hawk-Eye system that has been utilized from 2002[2,3]. In 2012 came officiating technology also into football by introducing Goal-line Technology (GLT) i.e., a system that determines whether the whole of the ball crossed the goal line[4]. A few years later, in 2016, football leagues over the world started to exploit another officiating technology-Video Assistant Referee (VAR), which became a subject of interest in our thesis[5].

At this moment, we provide a brief definition of the subject of our interest from the first principle of VAR protocol created by International Football Association Board (IFAB) of Féderation Internationale de Football Associations (FIFA). It says that VAR is a match official with independent access to match footage, who may assist a referee only in the event of a clear and obvious error or a serious missed incident[6]. In other words, VAR was set to fix unam-

[1] Source:`https://bit.ly/3y3RRCl`.
[2] Source:`https://bit.ly/3LAWYOu`.
[3] Source:`https://bit.ly/3ybSkCm`.
[4] Source:`https://fifa.fans/3vUEf9R`.
[5] Source:`https://sites.duke.edu/wcwp/2019/04/01/a-brief-history-and-defense-of-var/`.
[6] Source:`https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

biguous mistakes of pitch-based referees. Afterward, we dedicate more space also for the remaining principles and other parts of VAR protocol. At present, we move along to outline several views from which it might be beneficial to study the issue of VAR and to put forward the fractions of society to which we, through the thesis, aim.

We could divide the society, which may benefit from VAR-regarding studies into two units-people whose job is football-related and people who consume football in their leisure. Both groups have a common demand for football data. We can demonstrate it on an example of each group. Even though many football clubs use data in their decision-making process, two particular clubs have recently received media attention because of the widespread usage of data-English Brentford FC and Danish FC Midtjylland[7]. As the example of football fans demanding the data, we can adduce Czech company Livesport s.r.o.-provider of sport results and statistics. The company already overdid 100 million users over the world[8].

Furthermore, we divided possible benefits from exploring the issue of VAR into several categories due to their purpose. We might benefit from studying topics of the efficiency of VAR interventions, the efficiency of the process of communication between VAR and the pitch, and the reaction on VAR implementation from on-pitch referees. Moreover, we might reach an additional value from studying the impact of VAR on the game itself, i.e., on incidents that happen directly during the match.

Each category of possible benefits will be detailly discussed with related literature in the course of the thesis. Furthermore, a part of these possible benefits inspired us in our research. We will concretely investigate what impact does VAR have in Czech Fortuna Liga (F:L) on various indicators that are somehow related to the game itself and the performance of referees. As the indicators that are related to the game, we decided to select based on the related literature the number of yellow cards awarded in a match, the number of red cards awarded in a match, and the number of penalty kicks given in a match[9]. (Holder *et al.* (2022); Carlos *et al.* (2019); Lago-Peñas *et al.* (2020);

[7]Source:`https://www.scisports.com/state-of-the-football-analytics-industry-in-2021/`.
[8]Source:`https://bit.ly/3EZXy5T`.
[9]Source:`https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

Gürler & Polat (2021)). The second group of indicators, which we will work with, are indicators that are related to the performance of on-pitch referees. This set of variables was introduced to measure the number of errors of on-pitch referees as individuals in situations when a reviewable match-changing incident, i.e., a goal, a penalty, or a direct red card, has to be evaluated[10].

The following chapters of the thesis are structured as follows. In Chapter 2, we discuss the importance of VAR and also its criticism. Chapter 3 regards the methodology of VAR, i.e., the issue of VAR protocol will be reopened. Chapter 4 provides information about the history of VAR. In Chapter 5, we mutually compare results from related studies, which dealt with both the issue of VAR and match-changing incidents and the issue of VAR and errors of on-pitch referees. Starting with Chapter 6 we get to our research, where we firstly further develop potential benefits that motivated us for the research. Then, in Chapter 7, Chapter 8 and Chapter 9, we become acquainted with data and with models that we will be exploiting-both in a general way and on concrete cases. And in Chapter 10, we finally present results from the research on the sample of F:L matches. Chapter 11 serves for a final conclusion.

[10]Source:https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.

Chapter 2

Importance and criticism of VAR

2.1 Importance of VAR

To perceive the importance of officiating technologies in football is relatively straightforward, and it is based on the previously mentioned VAR protocol and the protocol related to GLT. In our opinion, the first principle of VAR protocol, which, as we know, says that VAR was set to fix unambiguous mistakes of pitch-based referees, does not provide a brief definition of this technology only. It also covers a crucial part of the importance of VAR because unambiguous mistakes of pitch-based referees can affect the ultimate outcome of the game[1] (Leveaux (2010)). We could introduce several well-known examples, which happened in the first decade of 21st Century, i.e., in the period when the officiating technologies were not implemented in football, but in several earlier-mentioned sports were. The first of them happened in 2009. France played against Ireland in the second leg of the play-off match that could have one of the teams brought to the World Cup. In extra-time, Thierry Henry, a French attacker, clearly handled the ball using his hand in the opponent's box and passed the ball to his teammate, who scored the game-winning goal[2]. A situation, which might have (or should have) been canceled by VAR, sent the French team to World Cup. The second example took place one year later during the above-mentioned World Cup. In the quarter-final match between England and Germany, while Germany led 2-1, Frank Lampard, an English midfielder, wiped his shot over the crossbar and scored a clear goal that on-pitch referees did not recognize because the ball immediately flew out from the goal[3]. Using GLT, it might not

[1] Source:https://sites.duke.edu/wcwp/2019/04/01/a-brief-history-and-defense-of-var/.
[2] Source:https://bit.ly/3vuV5ND.
[3] Source:https://bit.ly/3vuV5ND.

have (or again, it should not have) happened. From the examples, we can see that the momentum of the game can be changed by a single decision of on-pitch referees (Leveaux (2010)). It is not even uncommon for a pitch-based referee to be identified by a team or its fans as the reason for losing and to be blamed for influencing the final result of a game by either not enforcing the rules or being biased (Leveaux (2010)).

Another incentive of the importance of officiating technologies in football, especially VAR, could be deducted from data. In later paragraphs, we decided to devote a separate space to the research of Katholieke Universiteit Leuven (KU Leuven), whose research from the 2016-2018 period directly preceded the full adoption of VAR to the practice[4]. For this paragraph, we mention the research of Spitz *et al.* (2020), who supported KU Leuven study by several outcomes, especially how correct were researched referees with the usage of VAR and without the usage of VAR. From 9094 situations that might have affected the game, on-pitch referees were correct in 92.1% on them (Spitz *et al.* (2020)). After interventions of VAR, the accuracy of on-pitch referees increased to 98.3% (Spitz *et al.* (2020)). In logistic regression model, a final decision, i.e., after possible consultation with VAR, was significantly better than an initial decision, i.e., without possible consultation with VAR (Spitz *et al.* (2020)). We will come back to KU Leuven research including the study of Spitz *et al.* (2020) in Chapter 4. Firstly, we properly define VAR itself, explain how it should work and for which purposes, i.e., which situations it should solely investigate.

2.2 Criticism of VAR

Finally, as we discussed the importance of VAR, we decided to devote several lines also to its criticism. Negative responses to VAR are generally related to interruption of the natural flow of the game, lack of excitement, and debatable decisions (Van den Berg & Surujlal (2020)). Especially the first and the last response from the list might be related to the first principle of VAR protocol, which indirectly encourages VAR officiates to intervene only if it is necessary. In other words, there is no need to be involved in grey-zone decision (GZD), which the Technical Director of IFAB characterizes as situations for which there is no conclusive reference decision. Thus more than one decision could be supported

[4]Source:`https://bit.ly/3s1NzaR`.

(Spitz *et al.* (2020)). In the research chapters of the thesis, we will discuss GZD's as they might create some issues.

A survey of how people perceive VAR in English Premier League (EPL) conducted organization YouGov in January and August 2020. In January, 60% of 1419 respondents declared that VAR performed so far either very badly (26%) or fairly badly (34%). In August, when answered 914 respondents, the ratio improved to 50% (21%-very badly and 29%-fairly badly). A graphical representation of another question from the survey (from January 2020) could be seen in Figure 2.1. Researchers also asked about the enjoyability of matches, which is related to the above-mentioned excitement. Slightly more than two-thirds of researched people thought that VAR made matches either a lot less enjoyable (29%) or a little less enjoyable (38%), which corresponds with our initial reasons for negative responses to VAR. Again, in August 2020, numbers slightly improved to 62%. In both surveys, more than 70% of respondents agreed that the association should keep VAR being used. However, they should introduce changes. Researchers decided to include in the surveys the first question about the performance but about tennis and cricket, and the results were distinctly better[5][6].

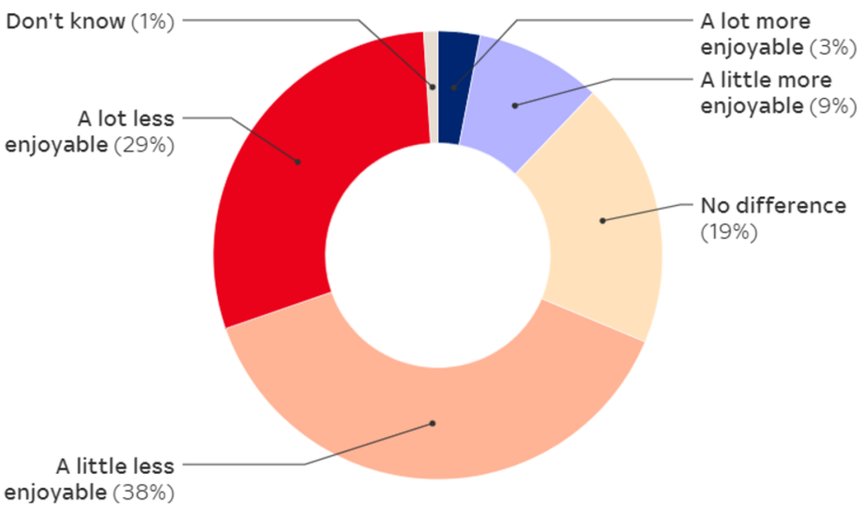

Figure 2.1: How much more or less enjoyable has VAR made watching EPL matches?

[5]Source:`https://bit.ly/3OPk9X6`.
[6]Source:`https://bit.ly/3F27sDM`.

Chapter 3

Methodology of VAR

3.1 Principles of VAR

The third chapter of the thesis is devoted mainly to VAR protocol, which inclusion to Laws of the Game (LOG) in 2018 could be seen as a historical action in the world of football (Samuel *et al.* (2020)). LOG are the universal football document. It includes all football rules, protocols, and approaches, from the act of kicking corners to the regulation of how referees or players should be dressed. IFAB issues LOG every year, and they are the same for all football throughout the world[1].

VAR protocol itself starts with the principles of VAR. Previously we already mentioned the first principle, which could also work as a definition of VAR. We also discussed several thoughts that from this principle spring. For importance, we just remind its essence, i.e., VAR as a match officiate with independent access to match footage assists the referee only in the case of clear and obvious error or serious missed incident, which means only in the event of goal or no goal, penalty or no penalty, a direct red card and mistaken identity, i.e., situation, when referee cautions or sends off the wrong player. VAR does not involve in any other situation than the above-mentioned, which also contains a single yellow card, the second yellow card, or a foul that was committed anywhere else than in the box[2].

Furthermore, we can find other 11 principles stated in the protocol. For the scope of the thesis, we decided to mention only several of them. Firstly, the

[1]Source: `https://bit.ly/3F2aDvc`.
[2]Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

on-pitch referee has always to make a decision, i.e., it is not permitted to give no decision and wait for VAR. Such a decision can also be to continue playing, which works in practice, for example, in the case of not giving a penalty. The original decision of the pitch-based referee cannot be then overturned unless VAR clearly shows the decision was in contraction with the first principle, i.e., a clear and obvious error or serious missed incident[3]. This principle uncovers more about GZD's. Together with the first principle, it gives us a piece of evidence of not overturning such events.

Secondly, there are principles that define the relationship between a pitch-based referee and VAR, especially regarding who eventually decides. They tell us that only the pitch-based referee can initiate a review. VAR and other officiates (linesmen and the fourth official) could only recommend a review. Moreover, the final decision has to be taken by the pitch-based referee only. The final decision could be preceded by either just information from VAR or on-field review (OFR). OFR's are situations when the pitch-based referee comes to the monitor to see retakes from the incident[4]. Implications from these principles are that pitch-based referees are still responsible for the decision-making outcome. They do not even have to accept the recommendation that an incident may be reviewed and overturned. These principles give VAR an advisory role.

Thirdly, review processes are not time-limited. While a review process lasts, transparency has to be maintained, i.e., players and team officiates cannot surround referees, and referees must be visible all the time the review lasts. And lastly, if a play was stopped and then was restarted, an on-pitch referee cannot undertake the review except for mistaken identity or red card cases[5]. That is why referees nowadays rather wait right after a suspicious incident related to a possible penalty or a possible goal.

3.2 VAR components

The general principles of VAR were stated, and we move along to the components from which VAR consists and how these components should look to fulfill requirements. The requirements are given by Implementation Assistance and

[3]Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.
[4]Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.
[5]Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.

Approval Programme (IAAP) that was approved by IFAB in 2018[6]. Before VAR system can be used in a live competitive match, the competition organizer must successfully perform technology tests in all competition stadiums where it will be used. All technology tests must be recorded and made available to FIFA upon request[7].

VAR technology generally consists of three components: Video Operation Room (VOR), Referee Review Area (RRA) and Referee communication system. All of these components have to fulfill minimal requirements. VOR is a space located near the stadium or at the stadium, where VAR referees examine the match footage. There have to be at least two VAR referees, i.e., VAR and Assistant Video Assistant Referee (AVAR) and also replay operator (RO). For examining purposes, there have to be available at least four exactly defined types of cameras to VAR and AVAR (two in the central position and two for checking the offside line). A layout of how VOR should like can be seen in Figure 3.1. Each person in VOR has a specific list of tasks. We mention that for AVAR it is, for example controlling the live tape, while VAR checks an incident and RO is responsible for the technical setup. VOR has to be constantly monitored by VOR camera for the sake of transparency, and only authorized persons are allowed to enter the room[8].

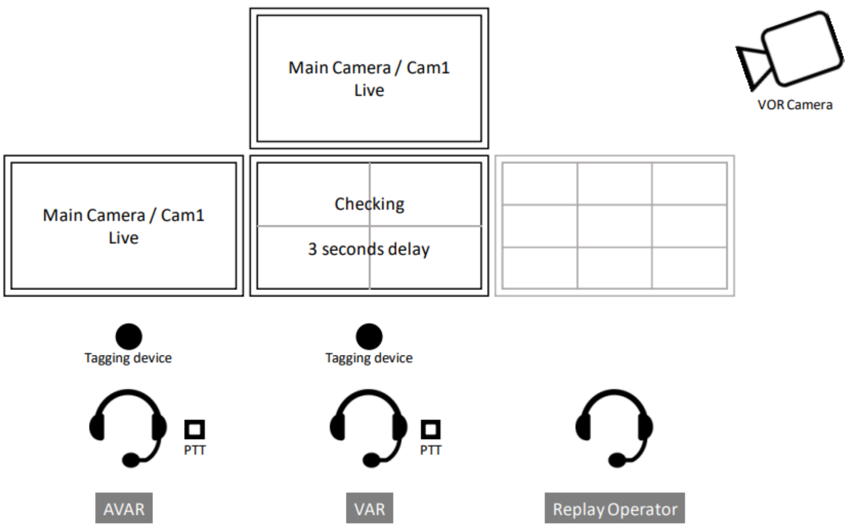

Figure 3.1: Video Operation Room layout

[6]Source: https://fifa.fans/3KyOAPU.
[7]Source: https://fifa.fans/3MYHHaz.
[8]Source: https://fifa.fans/3MYHHaz.

RRA should be an outdoor cabled monitor located by the side of the field, where a pitch-based referee does the OFR's. A pitch-based referee should come to RRA in the case of previously discussed incidents to see the match footage. The footage is controlled by VAR, a pitch-based referee has to communicate with VOR if a different tape is necessary[9]. The referee communication system serves for the communication between the field and VOR. An on-pitch referee and linesmen have to be open-mic to themselves and VOR. On the other hand, VOR team works on the push-to-talk principle for the communication with the pitch. All communication has to be for the transparency reasons recorded[10]. FIFA emits a list of organizations that have a certificate to develop or manufacture products used in VAR process i.e., FIFA Quality Programme[11].

3.3 Reviewable match-changing incidents

The following paragraph will be devoted to the list of events that could be reviewed by VAR called reviewable match-changing decisions or incidents. As it was stated in Section 3.1 there are four categories of such events. Firstly, it is a goal or no goal decision, which means that VAR can call a pitch-based referee to change the goal to the no goal and otherwise. Reasons why a goal should or should not be allowed, are based on committing a specific type of event, which is against the rules, while the goal is scored or while the goal is built-up. A limitation might be that it is not exactly stated when a goal is started to be built up. Specific above-mentioned events are handball, offside, foul, and situations when the ball is out of play before scoring a goal. The second category is related to penalties. VAR could intervene in the decision, whether a penalty kick should be changed to no penalty kick and otherwise. Again, there is a list of events that is worked within those situations. We could find there: handball, offside, foul, and ball out of the play of attacking team, before the penalty was awarded such as in the case of goal decision. Also, a location can be discussed by VAR i.e., whether the challenge happened in the penalty box (resulting in the penalty) or outside the penalty box (resulting in the free kick). Thirdly, we have direct red card decisions. VAR could intervene in the match to change no card or yellow card into the direct red card and otherwise, i.e., to change the direct red card into a milder punishment. For such situations

[9]Source: `https://fifa.fans/3MYHHaz`.
[10]Source: `https://fifa.fans/3MYHHaz`.
[11]Source: `https://fifa.fans/3Ky0APU`.

VAR exploits denying an obvious goal-scoring opportunity (DOGSO) parameter, which means it evaluates the position of attacking and defending players to see whether the red card would be appropriate. It also evaluates whether the challenge was serious, violent (for example biting), or insulting (for example abusing)[12]. The last category-mistaken identity was already explained.

3.4 VAR procedure

The last topic of Chapter 3 will regard the procedure of decision-making co-operation between the pitch and VOR. The exact procedure that a reviewable match-changing incident may go through consists of four steps: original decision, check, review, and final decision. Some incidents could go through just the first two steps, which holds in cases when it is decided that the original decision was correct. Furthermore, some of them could go through the whole process. The first step, i.e., the original decision, is the initial decision made by a group of on-pitch referees. As it was already said, a referee has to make this decision. No decision is not permitted because it leads due to IFAB to the weak officiating with many reviews and significant problems if there is a technology failure. The decision could be delayed for an evident attacking situation when a player is about to score a goal or has a clear run into or towards the opponent's box. However, at the end of such actions, there still has to be an original decision[13].

The second step is the check. VAR automatically checks footage of every potential or actual goal, penalty, direct red card, or mistaken identity with the use of several camera angles and slow motions. If the check does not indicate a clear and obvious error or a serious missed incident, VAR either do a silent check, i.e., does not communicate at all, or just confirm to a pitch-based referee that there was not an error and the incident is solved. However, if the check indicates a clear and obvious error or a serious missed incident, VAR informs a referee, who can decide whether the situation will be reviewed or will not[14].

The third step is thus the review. For this step, the game has to be necessarily stopped. If it has not been stopped yet, a referee has to do so when the ball is

[12] Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.
[13] Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.
[14] Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.

in a neutral zone or a neutral situation. VAR describes to the referee what can be seen on TV replays, and the referee can then either go to RRA to view the replay footage, i.e., OFR before making a final decision or make a final decision based on the referee's own perception (if necessary, including the perception of other on-pitch referees) and information from VAR. For some situations, OFR is more appropriate than a direct final decision, and otherwise. OFR is usually appropriate when the need for a subjective decision is eligible, e.g., the intensity of a foul or handball considerations. In such decisions, there might be some nuances between how VAR sees the situation and how does the referee. VAR-only review (and the direct final decision) is usually appropriate for factual decisions, e.g., the position of an offense player, point of contact, or ball out of play. In the case of such decisions, there are more apparent borders between yes and no. These borders are either literally painted, i.e., the borders of the pitch, or they can be virtual, i.e., the virtual offside line. OFR might not bring any added value in these cases. Nevertheless, OFR's can be used for factual decisions either, if it will help to manage players and sell the decision. The review process should be completed as efficiently as possible, but the accuracy of the final decision is more important than the speed. The review is followed by the last step, which is the final decision. After the final decision, the play could be restarted[15].

[15]Source: https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.

Chapter 4

History of VAR

4.1 Worldwide history

In Chapter 4, we introduce the history of VAR. The first part will be devoted to the global history, the second part the debated research of KU Leuven, and finally, we also briefly describe how VAR has been developed in Czech F:L. Although VAR has started to be utilized in 2016, its origin in football could be found already in the early 2010s, when the Dutch project Refereeing 2.0 started to exist (Murray & Howitt (2019))[1]. Royal Netherlands Football Association (KNVB) informed about the project and its purpose in 2013/14 season, while its pilot phase has already operated for several years. The project's purpose is to generally improve and maintain the quality of refereeing by supporting the role of technology in football. And thus, VAR and GLT were also included into the project[2]. In the top Dutch football league-Eredivisie were initiated the first mock trials with VAR in the 2012/13 season. Those trials also included off-line testing i.e., the usage of VAR without affecting matches[34]. First mock trials with GLT took place even several years sooner because the technology was firstly adopted already on the 2012 World Cup (Murray & Howitt (2019))[5]. In 2014, KNVB began informally petitioning IFAB to introduce video-assistance in football matches. Their proposals were heard one year later, in 2015. The new president of FIFA, Gianni Infantino, held a meeting in Zürich to consider the Dutch proposal for VAR[6]. The idea was well-received.

[1] Source:https://sites.duke.edu/wcwp/2019/04/01/a-brief-history-and-defense-of-var/.
[2] Source: https://www.knvb.com/themes/new-laws-of-the-game/refereeing-2.0.
[3] Source:https://sites.duke.edu/wcwp/2019/04/01/a-brief-history-and-defense-of-var/.
[4] Source: https://www.wired.co.uk/article/var-football-world-cup.
[5] Source: https://www.wired.co.uk/article/var-football-world-cup.
[6] Source: https://www.wired.co.uk/article/var-football-world-cup.

Therefore, in March 2016, at 130th FIFA annual meeting, IFAB gave a green light for an experimental phase, in which VAR was trialed with a view to its possible permanent introduction into the game. It was clear that there was a need for tight control to evaluate the results effectively. Therefore, the control remained in the hands of IFAB and all subjects interested in taking part in the trials were subject to one protocol used by all(Gallardo *et al.* (2019))[7]. The first test matches (under the head of IFAB), i.e., friendly international matches between Italy and Spain and Italy and Germany, took place in the same month as the annual meeting. They were followed by two years period, which IFAB set for the trials to be held(Gallardo *et al.* (2019))[8]. During this period VAR appeared for the sake of experiment in several countries over the world, including the US, Australia, or South Korea. Also, the well-known European leagues and cups such as German Bundesliga, Italian Serie A, or English FA Cup were not omitted, such as worldwide events, e.g., 2017 FIFA Confederations Cup[9][10][11]. From the 2017/18 season, several leagues even implemented VAR technology to all matches all season long. The period of trials was finished in 2018 on 132nd FIFA Annual Business Meeting by submitting a report regarding the information on VAR experiment[12]. The report, which was created in cooperation with KU Leuven and which provided IFAB crucial data, will be the topic of the next section.

4.2 Leuven research

The report bears the full name Information on the Video Assistant Referee (VAR) Experiment including provisional results, and consists of the summary of research results collected by KU Leuven since the beginning of VAR experiment in March 2016 and the list of questions and answers related to VAR[13]. We will focus only on the first-mentioned part because the majority of those questions were already answered in Chapter 3. Firstly, we describe how wide the experiment was in terms of participants. More than 20 national associations and competitions took part in the research. Some of them were already

[7] Source: `https://www.wired.co.uk/article/var-football-world-cup`.
[8] Source: `https://www.wired.co.uk/article/var-football-world-cup`.
[9] Source: `https://www.wired.co.uk/article/var-football-world-cup`.
[10] Source:`https://sites.duke.edu/wcwp/2019/04/01/a-brief-history-and-defense-of-var/`.
[11] Source: `https://bit.ly/3s1NzaR`.
[12] Source: `https://bit.ly/3s1NzaR`.
[13] Source: `https://bit.ly/3kveHuw`.

mentioned. We add that Football Association of Czech Republic (FAČR) also participated. We can count 804 of them in terms of competitive matches, which were included in the research. The trials were also made on further 700 training, exhibition, or friendly matches. However, those matches were not included in the study due to compatibility reasons[14]. In 804 matches, 3947 checks for possible reviewable incidents were made. Nevertheless, the majority of them were background checks, i.e., the checks that do not interfere with the game[15]. After these checks, a review is usually not necessary. The average number of checks per match was lower than five. The median check time of VAR was 20 seconds. Moreover, most of the checks have, due to the report, no impact on the flow of the game[16]. From checks, we leap to the following step of VAR procedure, i.e., reviews. As it was said that the majority of checks were background, it is, according to VAR protocol, not surprising that in 68.8% matches, there was no review. Furthermore, only 5.2% matches had more than one review[17]. Now, we get to the impact of VAR from the results of the experiment. Before and during the experiment, researchers showed that a clear and obvious error occurs on average in one match out of three. The accuracy of initial decisions of pitch-based referees related to reviewable incidents only was 93%. This accuracy increased to 98.9% when VAR could have intervened in these decisions, i.e., VAR brought additional 5.9% accuracy into the decision-making process of referees[18]. The similarly-minded results that we outlined in Chapter 2 were slightly different because we demonstrated the impact of VAR on the research of Spitz *et al.* (2020), who enlarged KU Leuven research for other matches (their dataset consisted of 2195 matches across 13 associations). The additional accuracy, which in the case of the study of Spitz *et al.* (2020) was equal to 6.2%, supported the results of KU Leuven report. The report also suggests that 100% accuracy is impossible due to human perception and subjectivity in decision-making. Indeed, a clear and obvious error was not corrected in one out of twenty matches[19]. We could also perceive the impact of VAR from another angle of view. The researchers declared that VAR had a decisive impact on the outcome of 8% matches and 24% matches were positively affected by VAR i.e., in 24% matches VAR changed an initial incorrect decision[20]. The median

[14]Source: https://bit.ly/3kveHuw.
[15]Source:https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/.
[16]Source: https://bit.ly/3kveHuw.
[17]Source: https://bit.ly/3kveHuw.
[18]Source: https://bit.ly/3kveHuw.
[19]Source: https://bit.ly/3kveHuw.
[20]Source: https://bit.ly/3kveHuw.

duration of a review was 60 seconds (70 seconds in OFR cases only). The average time lost due to VAR represented less than 1% of playtime, which had, according to the researchers, only a small impact on the overall time of play. The small impact was demonstrated on free kick and throw-in interruptions, which cost overall playtime 9.5%, respectively 8% on average[21].

4.3 History in Czech Republic

Our historical excursion ends where the research part of the thesis takes place, i.e., in the first Czech football league called Fortuna Liga. F:L started to utilize VAR in the end of 2017[22]. From the previous paragraphs, we know that before VAR could be put into practice, several technology tests have to precede. F:L was not an exception. In the beginning of 2017 League Football Association of Czech Republic (LFA) i.e., the interest grouping of all Czech professional football clubs, which controls and organizes professional competitions such as F:L, signed a contract with FIFA and IFAB and got involved into VAR testing program[23][24]. Firstly, VAR was used in the off-line mode. This phase finished in May 2017, when LFA succeeded at the official inspection of FIFA and IFAB and received the very high-quality grade. Therefore, VAR project could have switched itself to the online mode and could have been actively used during matches. On 3$^{\text{rd}}$ December 2017, VAR was for the first deployed to F:L match in the online mode. The match between AC Sparta Praha and FK Mladá Boleslav, broadcasted by O2 TV, was a historical milestone for VAR technology in Czech football[25]. According to the official webpage of F:L, in the course of the match, no OFR's happened. However, VAR was used for the sake of one VAR-only review and two silent checks. In the spring phase of the 2017/18 season, VAR was on average implemented to one match each round. In later seasons, the number of matches per round to which VAR was deployed had been increasing. From the spring phase of the 2020/21 season, VAR has not missed any single match of F:L[26].

[21] Source: https://bit.ly/3kveHuw.
[22] Source: https://www.lfafotbal.cz/videorozhodci.
[23] Source:https://www.lfafotbal.cz/o-nas.
[24] Source: https://www.lfafotbal.cz/videorozhodci.
[25] Source: https://www.lfafotbal.cz/videorozhodci.
[26] Source: https://www.lfafotbal.cz/videorozhodci.

Chapter 5

Relevant literature

In Chapter 5, we aim to present several studies and other relevant literature that motivated us to create the research part of the thesis that will follow from Chapter 6. The chapter is divided into two subfields of study: VAR and match-changing incidents, VAR and errors in refereeing. When we described the research of KU Leuven and the part of the research of Spitz *et al.* (2020), we already touched both subfields. In the following paragraphs, further studies will be added for both issues.

5.1 VAR and match-changing incidents

Firstly, we introduce relevant literature regarding the relationship between the presence and interventions of VAR and match-changing incidents. We start with for us well-known research of Spitz *et al.* (2020). We have not mentioned yet a part of their study, which regards the impact of VAR on the change in particular categories of reviewable match-changing incidents, i.e., goals, penalties, and red cards. In their dataset, the most significant proportion of all checks generated red card checks (39.3%), and the most significant proportion of all reviews made penalty reviews (43.9%). Nevertheless, all our three categories of match-changing incidents reached at least a 20% proportion in both check and review cases. On the other hand, mistaken identity cases formed only 0.2% of all checks and 1.1% of all reviews. Therefore we decided not to deal with them further (Spitz *et al.* (2020)). Furthermore, later we add to our consideration also yellow cards. As we know from the definition of review, a referee has the opportunity to change the initial decision. The researchers called situations when a review changes no to yes an extra incident, e.g., an extra penalty. This way, it was in all 2195 matches of the dataset of Spitz *et al.* (2020) created 76 extra penalties, 126 extra red cards, and 114 fewer goals. The results bring us the idea of positive biases in the number of penalties and red cards and negative bias in the number of goals, which might exist in matches where VAR is present.

However, this idea is not supported by the research of Holder *et al.* (2022). The researchers constructed a dataset from German Bundesliga and Italian Serie A matches focused on penalty kicks and red cards. The dataset consisted of five seasons without VAR and two seasons with VAR (in both competitions VAR was introduced at the beginning of the 2017/18 season, and from this moment, it participated in all matches all seasons long)(Holder *et al.* (2022)). The results from their study could be seen in Figure 5.1 (for penalties) and in Figure 5.2 (for red cards). The figures show us the total number of penalties and red cards in a particular season (from 2012/2013 to 2018/2019) of both competitions. Numbers were more or less constant over time (Holder *et al.* (2022)). From the 2017/2018 season, there are divisions in the graphs. Lighter grey shows us the number of penalties and red cards before VAR possibly intervened in the situation a darker grey after. After the introduction of VAR, the total number of initially given penalty kicks decreased by more than 25% and the number of initially given red cards by more than 30% (Holder *et al.* (2022)). After the initial decision, VAR generally created extra incidents, and trends rather started to look like in the previous seasons (Holder *et al.* (2022)). This finding did not support the idea of positive biases in these incidents due to VAR presence. However, it might support a different assumption, i.e., VAR influences the decision-making behavior of referees (Holder *et al.* (2022)). This assumption is in the dataset of Holder *et al.* (2022) based on the idea that the presence of VAR vice versa creates negative biases in the number of penalties and red cards, and VAR interventions shift the biases to normal. The assumption also fits the second subfield. Therefore it will be reopened in Section 5.2.

In this paragraph, we rather present the research of Carlos *et al.* (2019). Their sample represents a part of the sample, which Holder *et al.* (2022) used as it consists of 1024 matches played in Serie A and Bundesliga in the 2016/17 and 2017/18 seasons. Thus, in half of the matches of their sample VAR was present, and in the second half, it was not (Carlos *et al.* (2019)). The researchers using Poisson regression model or, in the case of overdispersion Negative Bino-

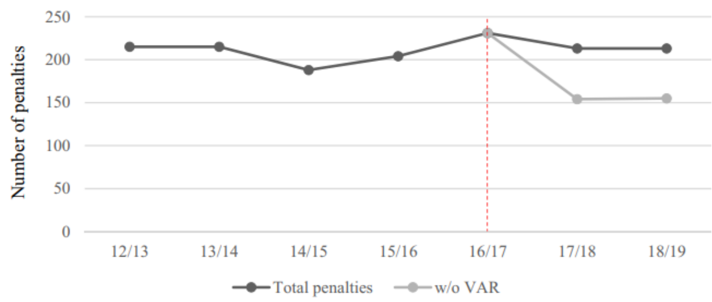

Figure 5.1: Seasonal development in number of penalties in Bundesliga and Serie A in context with VAR introduction

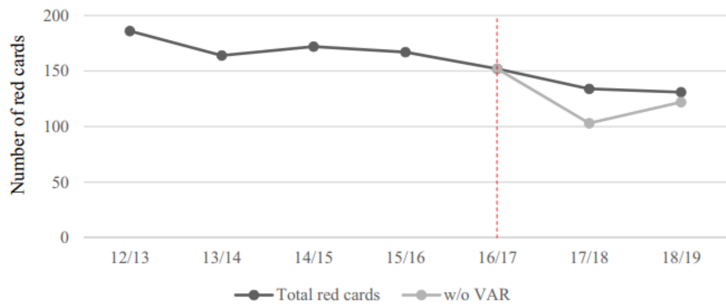

Figure 5.2: Seasonal development in number of red cards in Bundesliga and Serie A in context with VAR introduction

mial (NB) model (will be explained in the research chapters) aimed to capture the sign, the magnitude, and the significance of the relationship between the presence and interventions of VAR and a bunch of incidents (including our match-changing incidents) (Carlos *et al.* (2019)). From Figure 5.3 we can see that the results regarding red cards and penalties confirm, what we saw from Figure 5.1 and Figure 5.2 between 2016/2017 and 2017/2018 seasons. Neither in the case of penalties nor the case of red cards was the relationship significant (both magnitudes were negative, but p-values were very high) (Carlos *et al.* (2019)). A study of Lago-Peñas *et al.* (2020) based on Spanish La Liga during the 2017/2018 and 2018/2019 seasons supported the findings of penalties and red cards from Bundesliga and Serie A. In the dataset equally divided into matches with VAR and without VAR, the researchers using models from generalized linear model (GLM) family again investigated the sign, the magnitude, and the significance of VAR presence and interventions in relation to same incidents as Carlos *et al.* (2019) (Lago-Peñas *et al.* (2020)). By neither penalties nor red cards were found a significant relationship. This time all magnitudes were positive. However, p-values were again very high (Lago-Peñas *et al.* (2020)). The topic of VAR and match-changing incidents also studied Gürler & Polat (2021). Their research consisted of 34 match weeks with VAR and 60 match weeks without VAR played in Turkish Super League (TSL) (Gürler & Polat (2021)). In the case of red cards, the statistical effect of VAR was not significant such as in other mentioned studies (Gürler & Polat (2021)). Penalties were not part of Gürler & Polat (2021) study.

		95% CI			
Variable	Estimate	Lower	Upper	Z-ratio	P-value
Fouls	−1.24	−2.02	−0.46	−3.104	0.002
Goals	−0.21	−0.41	0.00	−1.954	0.051
Offsides	−0.41	−0.71	−0.11	−2.711	0.007
Penalties	−0.03	−0.10	0.04	−0.781	0.435
Playing Time 1st half	0.20	0.06	0.34	2.819	0.005
Playing Time 2nd half	0.12	−0.12	0.36	0.985	0.325
Red Cards	−0.02	−0.07	0.04	−0.634	0.526
Total Playing Time	0.32	0.06	0.59	2.365	0.018
Yellow Cards	−0.47	−0.72	−0.22	−3.644	0.000

CI = Confidence interval

Figure 5.3: Statistical reletionship between the presence of VAR and chosen incidents in Bundesliga and Serie A

However, neither the research of Lago-Peñas *et al.* (2020) nor the study of Gürler & Polat (2021) provide information how many penalties or red cards were awarded after the consultation with VAR such as the study of Holder *et al.* (2022) does. Moreover, neither La Liga nor TSL was part of the dataset of Spitz *et al.* (2020). Therefore, we know that in 13 countries, including Germany and Italy, there were altogether created extra red cards and extra penalties due to VAR. In Germany and Italy, those extra incidents returned the number of penalties and red cards to the trends of last seasons, and thus they were not significant. They were not significant in Turkey and Spain either, but we can only assume that the process was similar to other countries.

While we found out from three, respectively, four studies that there are not significantly more or fewer red cards and penalties in matches with VAR, we may not say the same about goals and yellow cards. From the research of Carlos *et al.* (2019), in matches where VAR was present, spectators could have seen a significantly (at 90% level) lower number of goals (Figure 5.3). The negative statistical relationship between goals and VAR was also found by Gürler & Polat (2021), even at 95% level. However, the study of Lago-Peñas *et al.* (2020) did not support this trend. In Spanish La Liga, there was detected even a positive statistical connection between goals and VAR. However, the resulted p-value was not close to being significant (Lago-Peñas *et al.* (2020)). Nevertheless, from the study of Spitz *et al.* (2020), we know that the number of extra goals attributed by VAR was also negative. Therefore, without data, which we had in the case of red cards and penalty kicks from the study of Holder *et al.* (2022), we cannot say which part of a possible negative relationship between VAR as the whole and goals might refer to VAR presence and which to VAR interventions.

Carlos *et al.* (2019) also found that in matches with VAR, there was a significantly (at 99% level) lower number of yellow cards. However, their data were not supported by Gürler & Polat (2021) and Lago-Peñas *et al.* (2020). In these studies, a significant relationship between VAR and yellow cards was not revealed. In the case of Gürler & Polat (2021), the sign of the relationship was negative, in the case of Lago-Peñas *et al.* (2020), even positive. From other incidents, we could see that a significant negative relationship can also be found for offsides (Carlos *et al.* (2019); Lago-Peñas *et al.* (2020)) and for fouls (Carlos *et al.* (2019)).

5.2 VAR and errors in refereeing

The next paragraphs will be devoted to the topic of VAR and errors in refereeing. This topic is related to what suggested Holder *et al.* (2022): VAR influences the decision-making behavior of referees. Firstly, we discuss results from a case study of Samuel *et al.* (2020). The researchers focused on the issue of VAR in Israeli Ligat Ha'Al, where they studied besides issues of conscious decisions of referees and their coping also perceptions of referees regarding the implementation of VAR (Samuel *et al.* (2020)). For the purpose of this paragraph, we focus on the first-mentioned topic. The dataset included 212 matches officiated with VAR until March 2020 and their comparison to previous matches, to which VAR had not been implemented yet. In Israel VAR went through the pilot phase in the play-off stage of the 2018/2019 season, and it appeared in all played matches of the 2019/2020 season (Samuel *et al.* (2020)). The number of critical errors of on-field referees has significantly (at 95% level) increased just in the 2019/2020 season (81 errors) in comparison to the 2018/2019 season (69 errors) and seasons 2017/2018 (49 errors) and 2016/2017 (52 errors) (Samuel *et al.* (2020)). After VAR interventions 77.5% of those errors were corrected (Samuel *et al.* (2020)). Nevertheless, corrections are not the purpose of this section. In KU Leuven study and the research of Spitz *et al.* (2020), we already saw that there is a positive and significant effect of VAR in terms of how it is capable of rectifying incorrect decisions that on-pitch referees made. In this section, we investigate why VAR might have to correct more errors than how many errors occur in matches where VAR is not present. The results from the studies of Samuel *et al.* (2020) and Holder *et al.* (2022) from three competitions, to which VAR was implemented, i.e., Israel, Italy, and Germany support themselves in this assumption.

We could study football referees from many levels. For this chapter, we take the view that defines a referee as a human athlete (such as a football player) who performs an essential role in ensuring that sports competitions come off smoothly (Phillips & Fairley (2014); Samuel *et al.* (2017); Slack *et al.* (2013)). A referee is identified with a specific activity, i.e., officiating matches, for which has to be well-prepared to perform the activity in the best way (MacMahon (2015); Philippe *et al.* (2009)). The environment of elite refereeing is highly competitive (Samuel *et al.* (2017)). Moreover, referees are, during officiating, usually judged by everyone who has something in common with a particular match[1]. In the following paragraphs, we will be discussing a concept of stress that could generally change the likelihood of making a mistake (Atsan (2016)). We will also investigate whether the stress might be connected with the presence of VAR and thus the presence of VAR with errors in refereeing. Besides the term stress, which we generally approach as a label of some worries[2]., we will also work with concepts of pressure, anxiety, competence, self-confidence, and authority.

Such an occupation as a football referee brings stress (Soriano Gillué *et al.* (2018)). The stress may harm the performance by influencing all stages of the decision-making process and, therefore, also the decision itself (Soriano Gillué *et al.* (2018); Atsan (2016)). Soriano Gillué *et al.* (2018) researched possible sources of stress for a football referee in the sample of 127 referees from Catalan regional leagues. The referees fulfilled a questionnaire, where they evaluated their main stressors during matches by assigning them the number from 1 to 7 (1 was the lowest and 7 the highest). Among 20 possible reasons, we were able to find "Committing a technical error" reason in the third place with a mean value of 5.07 and "Contradicting the decision of a colleague" in the sixth place with a mean value of 4.42 (Soriano Gillué *et al.* (2018)). Although Catalan

[1] Source: `https://bit.ly/3ML7jrk`.

[2] Source: `https://dictionary.cambridge.org/dictionary/english/stress`.

regional referees probably have not ever had an experience with VAR, there might be, in our opinion, a possibility that these reasons, i.e., a technical error or a disagreement with a colleague, would transfer to matches with VAR and VAR might these stressors even deepen, because with VAR both an amount of technology and the number of referees increase. A part of the study of Samuel *et al.* (2020) was related to perceptions of referees regarding VAR implementation. Eleven elite referees from Ligat Ha'Al were given Likert-type subscales that should have measured both a perceived control over the event and a perceived significance of the event on the range from 1 to 5, i.e., 1-not at all/very negative and 5-very much/very positive (Samuel *et al.* (2020)). As the events were also chosen, "pressure increase" and "pressure decrease". The mean of "pressure increase" was equal to 1.55 and the mean of "pressure decrease" to 2.45 (Samuel *et al.* (2020)). Therefore, the referees felt that the pressure that they experienced during matches, where VAR was present, decreased rather than increased (Samuel *et al.* (2020)). And because the increase of pressure generally results in a state of stress, the study of Samuel *et al.* (2020) on the contrary supports the idea that the stress does not increase in VAR matches (Stephenson *et al.* (2022)). Nevertheless, possible limitations of this assumption might be a small sample of Israel-only referees and also the fact that the pressure is not a sole stressor, as will be shown in the following paragraphs(Samuel *et al.* (2020)).

Stephenson *et al.* (2022) also claimed that another state that generally increases the stress is anxiety. For this purpose, we mention the study of Johansen & Haugen (2013). Their study aimed to examine the level of anxiety among 83 Norwegian top-class football referees and its impact on their officiating. Moreover, they aimed to predict the level of anxiety according to several factors, among which we could also find perceived competence of refereeing (Johansen & Haugen (2013)). Data showed that referees who perceived that their competence was weaker or average (compared with their colleagues) scored significantly higher anxiety than did other referees (Johansen & Haugen (2013)). In our opinion, the feeling of competence of on-pitch referees might differ in matches where VAR is present. Although, according to VAR protocol, both the initial decision and the final decision are made by on-pitch referees, VAR is another agent that plays a role in the decision-making process and, therefore, the felt competence of referees might decrease.

Other states that might be connected with stress are authority and self-confidence. Regarding the authority, it was suggested that leadership positions are associated with a lower level of stress, and therefore more authority means less stress[3]. Kolbinger & Lames (2017) proposed a potential threat of technology officiating ads (to which we include VAR) for the authority of referees. Furthermore, self-confidence is, directly and indirectly, connected with stress (Galanakis *et al.* (2016)). However, in the case of VAR Samuel *et al.* (2020) and their Likert-type subscales did not show that VAR might decrease the self-confidence of on-pitch referees (mean value equaled 1.91 on the scale). Nevertheless, we already mentioned the limitations of this study.

The purpose of the section regarding refereeing mistakes was mainly about discussing, proposing, and assuming. We suggested several factors that are generally directly or indirectly related to stress, i.e., pressure, anxiety, self-confidence, authority, and competence. The stress may have a negative impact on the performance, including the issues of decision-making and errors. However, there is no unambiguous proof that the presence of VAR on match increases in the case of on-pitch referees any of these states of mind or stress itself. Nevertheless, two studies showed us that more on-field mistakes in matches with VAR happen (Samuel *et al.* (2020); Holder *et al.* (2022)). Therefore, the above-mentioned factors might have their role in this bias. Moreover, there might exist other circumstances that could enter the issue. In our opinion, the most explainable arguments from those we mentioned are a lack of competence and authority of on-pitch referees.

[3] Source: `https://stanford.io/376zdPg`.

Chapter 6

Motivation for the research

The rest of the thesis will be devoted to the own research conducted on the sample of F:L matches. The research aims to establish on the results from the related studies discussed in Chapter 5. Therefore, we will be investigating two main topics: the relationship between VAR and match-changing incidents and the relationship between VAR and errors of on-pitch referees. We will assume possible benefits from studying these topics in the following paragraphs.

Regarding the issue of VAR and match-changing incidents, we previously mentioned that we focus on yellow cars, red cards, and penalties. We selected such incidents based on the studies of Holder *et al.* (2022), Carlos *et al.* (2019), Lago-Peñas *et al.* (2020), Spitz *et al.* (2020) and Gürler & Polat (2021). Benefits from investigating this issue might be broad because as we suggested in Chapter 1, the demand for football data is widespread among fans and clubs. Moreover, the benefits might not remain only for clubs and fans. We could also mention sports-betting industry, for which data are a valuable tool for making odds[1]. Furthermore, football and refereeing authorities may benefit from our results, too, because as Holder *et al.* (2022) suggested, VAR biases in match-changing incidents (red cards and penalties concretely) can open the topic of the decision-making behavior of referees. Moreover, as Carlos *et al.* (2019) suggested: bias in the number of yellow cards awarded in matches with VAR might tell us how players are aware of VAR being present in the match. If the bias was negative, it might have been a signal of players being less aggressive because they are aware of VAR and a potential threat of a red card. Therefore, our motivation to choose exactly these statistics could be explained by the extra

[1] Source: `https://bit.ly/3KxC1Tb`.

value that they may bring. And that is why we decided to include also yellow cards, even though they are not defined as match-changing incidents due to VAR protocol[2]. On the other hand, we decided not to include the number of goals scored in variables' selection because we did not find this extra value from studying it in relation to VAR from the related studies.

Furthermore, we assume that refereeing authorities (Committee of Referees of Football Association of Czech Republic (KR FAČR) in the case of the Czech Republic) might benefit from studying how the implementation of VAR affects on-pitch referees in terms of making errors. This way, the authorities might improve established approaches in the training of referees or their approach to VAR protocol. Based on the studies of Holder *et al.* (2022) and Samuel *et al.* (2020) we will follow up this issue. All dependent variables will be discussed more in-depth in Chapter 9.

[2]Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

Chapter 7

Data

7.1 Observations and time division

Our research was conducted on 678 matches of F:L, which gradually took place during the 2018-2021 period. This time range includes the entire 2018/19 and 2019/20 seasons (both primary and secondary superstructure parts) and the fall half of the 2020/21 season. Games that were supposed to be held at the end of 2020 but were postponed for various reasons to the beginning of 2021 were also included in the research. All researched games were divided into five phases based on the part of the particular season in which they were played. Each F:L season has two phases, i.e., the fall phase and the spring phase. This layout can be seen in Table 7.1. The first two phases record 2018/19 games, the second two 2019/20 games, and the last one 2020/21 games. We decided not to segregate matches postponed to 2021 to a separate phase because they are only six. Therefore, they might not fully represent a sample of various matches that usually takes place during a normal phase. We approached phase partition to limit the possible time effect, which will be discussed more detailly in Chapter 8. Phases were chosen instead of seasons because of the general nature of registration periods, i.e., transfer windows in European football. European football clubs, including the clubs in our research, have two opportunities during a year to acquire and sell players, i.e., registration periods. The duration of these periods differs across European countries. However always, one of them passes off in summer and another in winter, when some competitions, including F:L, walk through a break. In the case of F:L both the summer break and the winter break are relatively well merged by the registration periods[1].

[1]Source: `https://fifa.fans/3ORAudO`.

Table 7.1: Matches played due to phase

Phase	**Matches played**
1	152
2	129
3	160
4	111
5	126

7.2 Data collection and number of variables

Data were collected from several webpages that focus on football statistics and information. In this section, we outline just a broader origin of the data, which will be extended for specific variables in Chapter 9 and Appendix A. A part of the dataset that can be considered as game-related statistics such as goals, cards or fouls was processed from *Livesport*[2] and *TotalCorner*[3] webpages and the official webpage of *Fortuna Liga*[4]. Variables regarding individually-orientated player and referee statistics were also gathered from the official site of *Fortuna Liga* and *Transfermarkt*[5] webpage. And lastly, information that offer the evaluation of referees' work in F:L games were collected from Communiques of KR FAČR that are available on the official webpage of *FAČR*[6]. The total number of exploited variables exceeded 30.

7.3 VAR - variable of our interest

In Chapter 7, we decided to devote a unique space also to the specific variable, which we will treat as an independent variable of our interest in all models. *VAR* is a binary variable that takes only two possible levels: 1 if the Video Assistant Referee was present on the particular match of F:L and 0 if it was not. In Table 7.2, we could see how all 678 games from our dataset can be divided based on VAR presence. The total ratio between matches with VAR and others is relatively even. Slightly more than 53% of matches disposed VAR system.

[2]Source: `https://www.livesport.cz/`.
[3]Source: `https://www.totalcorner.com/`.
[4]Source: `https://www.fortunaliga.cz/`.
[5]Source: `https://www.transfermarkt.com/`.
[6]Source: `https://www.fotbal.cz/`.

Table 7.2: Matches played due to *VAR*

VAR	**Matches played**
0	318
1	360

From Section 4.3, we know that VAR technology has been deployed to F:L matches in online mode from the 2017/18 season. Nevertheless, its coverage was initially sparse. LFA has been gradually adding more matches with VAR to each round, and from the spring phase of the 2020/21 season (which is not part of our dataset), VAR coverage has been 100% in F:L[7]. The increasing usage of VAR in F:L matches, which we should be aware of, could be seen from Table 7.3. The figure gets together the percentual usage of VAR based on the *phase*. Initially, the ratio was equal to 36%. However, it gradually raised to 83%. Information whether there was VAR available during the game or was not were collected from the official webpage of *Fortuna Liga*.

Table 7.3: VAR usage rate

Phase	**With VAR**	**Without VAR**	**VAR usage rate**
1	55	97	36%
2	48	81	37%
3	79	81	49%
4	74	37	67%
5	104	22	83%

[7]Source: `https://www.lfafotbal.cz/videorozhodci`.

Chapter 8

Methodology

Chapter 8 serves to deliver information about the models' construction in a general way. Firstly, we present an universal approach that we will exploit for selecting a typology of benchmark and subsequent models. We aim to explain the logic behind choosing such an approach based on the characteristics of our dependent variables. Then, we provide information about control variables, including their division and motivation for adding them to multiple models. Finally, we describe a procedure through which we will be extracting the best candidate models from a selected group of control variables.

8.1 Nature of our dependent variables

We suppose that our four dependent variables (number of yellow cards awarded in the match, number of red cards awarded in the match, number of penalties given in the match, and number of errors of on-pitch referees as individuals in the case of reviewable match-changing incidents) have a common characteristic. None of them can gain a different value than positive and integer. In other words, all of them are theoretically defined on $N+$ including zero. However, this holds only on the theoretical level. As we know, a football match is usually constrained by slightly more than 90 minutes of playtime, when a number of possible events such as our response variables can occur[1]. This boundary practically excludes our dependent variables from gaining extreme values, e.g., the highest number of yellow cards across all F:L matches played during the 1994-2021 period was 12[2]. The conception of an extreme value might differ in

[1]Source: `https://bit.ly/3F27sDM`.
[2]Source: `https://www.fortunaliga.cz/`.

the case of each response variable. In some cases, such as the variables regarding awarded cards, the conception is more straightforward than for penalties or errors because of other constraints such as a limited number of players nominated to the match or the fact that a team cannot continue playing with less than seven players[3]. Nevertheless, for the sake of our study, we suggest that our dependent variables belong to the group of limited dependent variables (LDV) because they are restricted in their range by gaining only positive integer values or zero(Wooldridge (2009)). We also perceive several constraints limiting them from gaining extreme values.

Such LDV such as we described at the end of the last paragraph can also be incorporated into a narrower kind of dependent variables called count variables (Wooldridge (2009)). Count variables can take only non-negative integer values: $(0, 1, 2, \dots)$ and could be restricted to cases, when a variable takes only few values including zero, which corresponds to the description of our response variables (Wooldridge (2009); Kasyoki (2016))[4]. Among other examples of count variables, we can quote the number of kids ever born to a woman or the number of times someone is arrested during a year (Wooldridge (2009)). These examples also work with nonnegative integer values that are somehow restricted by time and empirics, which strengthened our beliefs that our regressands are counts. Thus we can state families of models that are typically applied to such as variables. Moreover, several studies including the researches of van der Wurp *et al.* (2019), Klemp *et al.* (2021) or Barbiero (2019) treated football statistics as count data.

8.2 General idea of models

For LDV and thus also for count variables, a linear model estimated by Ordinary Least Squares (OLS) might not provide the best possible fit of explanatory variables because of several restrictive assumptions, which our response variables might not fulfill (Wooldridge (2009)). Nevertheless, the linear model is basically taken as an informative econometric benchmark. Thus we decided to follow this approach and use it as the first benchmark model (Wooldridge (2009))[5]. However, there are several models which might work better for count

[3]Source: `https://bit.ly/3F27sDM`.
[4]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[5]Source: `https://bit.ly/3vVW6ND`.

data. The starting point for count data analysis is Poisson regression model (Cameron & Trivedi (2013)). Nevertheless, for this model we also register some restrictive assumptions that should be met (Kasyoki (2016); Roback & Legler (2021))[6]. Due to the assumptions, researchers introduced several further models as an extension to Poisson regression model to treat count data such as Quasipoisson (QP) regression model, NB models, hurdle models, zero-inflated models or Weibull-count model (Wooldridge (2009); Kasyoki (2016))[7][8]. In the next paragraphs, we will discuss these assumptions in relation to our dataset. We also get through several above-mentioned dilative models. Nevertheless, we decided to take Poisson regression model as the second benchmark (and the workhorse) model as it did in their football-related studies Carlos *et al.* (2019) or Klemp *et al.* (2021) (Winkelmann (2015)). We will approach other models only if the assumptions are not met.

8.3 Poisson regression model

In this section of Chapter 8, we present the basics of our workhorse model, i.e., Poisson regression model. The model is based on the idea of Poisson distribution, which should take a dependent variable instead of normal distribution (Wooldridge (2009)). A random variable Y following Poisson distribution should count a number of events per unit of time or space. The number of events should depend only on the length or on the size of the interval between two events λ (Roback & Legler (2021)). Poisson distribution can be seen in Equation 8.1. The expected value of random variable Y that went through the Poisson process should equal to λ and its standard deviation to the square root of λ, which gives us equality in the relationship between the expected value of Poisson random variable and its variance (Roback & Legler (2021))[9]. This relationship, which is depicted in Equation 8.2 will matter when we will be making decisions, whether to stick with Poisson regression models or instead use different models.

[6]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[7]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[8]Source: `https://bit.ly/3vzFs7K`.
[9]Source: `https://bit.ly/3LBPcnh`.

$$P(Y = y) = \frac{\exp(-\lambda)\lambda^y}{y!}; y = 0, 1, \ldots, \infty \tag{8.1}$$

$$E(Y) = Var(Y) = \lambda \tag{8.2}$$

To be able to use maximum likelihood estimation (MLE) framework, i.e., the way we estimate β's of our further regressions, we need to specify a distribution of Y_i, given all explanatory variables X_i[10]. A profitable approach that we can look through in Equation 8.3 is to model expected value of Y_i given X_i as an exponential function (Wooldridge (2009)). From Equation 8.2 we also know that both sides of Equation 8.3 are equal to λ_i[11]. Coefficients can be easily interpreted by taking logarithm as we can see in Equation 8.4 (Wooldridge (2009)).

$$E(y|x_1, x_2, \ldots, x_k) = \exp(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k) \tag{8.3}$$

$$\log(E(y|x_1, x_2, \ldots, x_k)) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \tag{8.4}$$

Once we have a notion how Poisson regression model looks like, we need to find a mean how to estimate β's of Equation 8.4. We use the same approach, which we would work with, if we were dealing with binary variables, i.e., MLE, which purpose is to find a maximum of the function depicted in Equation 8.5 generally called $\hat{\beta}_{MLE}$. Its interpretation is not the same as in OLS cases and will be explained in the results chapter of the thesis (Wooldridge (2009); Cameron & Trivedi (2013)).

$$\zeta(\beta) = \sum_{i=1}^{n}(y_i \mathbf{x_i}\beta - \exp(\mathbf{x_i}\beta) - \log(y_i!)) \tag{8.5}$$

8.4 Poisson regression assumptions and alternative models

Section 8.4 will be devoted to the list of assumptions that Poisson regression should satisfy. The first of them was already mentioned: a dependent variable

[10] Source: `https://bit.ly/3LBPcnh`.
[11] Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.

should be a count per unit of time or space, i.e., it should be a Poisson response, which our regressands satisfy (Roback & Legler (2021))[12]. The second assumption is related to the independence of observations (Roback & Legler (2021))[13]. One observation basically cannot provide information on another[14]. As an observation, we mean, in our case, a final match outcome in terms of a dependent variable. The dependence of observations was discussed in several football-related studies utilizing Poisson regression. However, their studies were rather modeling football results, and thus their conception of dependence regarded home goals versus away goals in a single match, where the dependence might be more straightforward (Barbiero (2019); Lidén (2016); Dixon & Coles (1997)).

Even though we do not divide the regressands into a home and away (we count them as total), this assumption might create a discussion also in our case - it depends on how we perceive the nature of the dataset. It is apparent that we deal with a substantially smaller number of football clubs than the number of matches in our research (678 matches of 19 teams in total). Each team plays a game every week against a different opponent. Each pair of teams usually meets twice a year on different playgrounds (from the 2018/19 season, the superstructure part of F:L has been imposed, i.e., some teams do not meet twice but three times a season)[15]. Since our dataset consists of two and half seasons and every year, just one to three teams swap themselves in the league table (it means that one to three teams relegate to the second tier and the same number promotes to F:L), we can find a non-negligible number of matches of same pairs of teams playing on the same pitch in our dataset[16]. However, we realized that these mutual matches did not occur immediately after each other. Moreover, in most cases, these matches were separated at least by one registration period, during which plenty of changes might run through a football club. We can assume changes in the composition of a team, in injuries, in staff, in bans from playing, in the atmosphere within a squad, in short-term goals, or in position in the table, e.g., it was discovered that on average, 39.7% of squads of F:L teams represented new singings (signed within

[12]Source: `https://bit.ly/3OI7c1g`.
[13]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[14]Source: `https://bit.ly/3OI7c1g`.
[15]Source: `https://www.livesport.cz/`.
[16]Source: `https://www.livesport.cz/`.

one year) during 2009-2017 period[17]. As Scarf (2017) suggested, even if some teams met previously, their next match and its story might not be the same.

We are aware of the fact that league matches generally might have some characteristics resulting from previous games, e.g., based on affection between teams or position in the table that in some way predetermine them to be more likely that a specific event such as our dependent variables in their games occur, e.g., clashes between rivals were ranked highly in terms of yellow cards in our dataset. Nevertheless, due to the dynamics of the team sports environment and the separation of mutual matches in our dataset, we decided to assume all observations to be independent (Kleinert *et al.* (2012)).

Another problem that we investigated was related to possible time effects. In our dataset, we dispose of 678 observations over many points of time. We can divide them due to an hour of kick-off, a day of kick-off, a round of kick-off, a phase of kick-off, or a season of kick-off. In Chapter 7, we outlined several reasons why in our opinion, the most appropriate way to divide them may be due to phases. Nevertheless, in none of these division periods, we would not be able to find exactly the same individuals, especially if we considered what suggested Scarf (2017). Therefore, from the definition, our data are not panel (Wooldridge (2009)). We decided to understand our data as pooled cross-sections and control for possible fixed effects made by winter or summer breaks by the previously discussed *phase* variable. (Wooldridge (2009); Raffalovich & Chung (2015)).

The third assumption already mentioned in non-complete form in Equation 8.2 put mean and variance of Poisson model into equal relationship-both conditionally on x_i and unconditionally (Wooldridge (2009); Roback & Legler (2021))[18]. The state when the third assumption holds, is called equi-dispersion[19]. Its violations could be either underdispersion or overdispersion (Wooldridge (2009)) The situation, when variance is greater than mean, i.e., overdispersion is for count data more common and it can occur in the case of heterogeneity (Wooldridge (2009); Winkelmann (2015)). Overdispersion is illustrated in Equation 8.6. Researchers suggested several means, how to deal

[17]Source: `https://www.football-observatory.com/IMG/sites/mr/mr34/en/`.
[18]Source: `https://bit.ly/3LBPcnh`.
[19]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.

with overdispersion e.g., QP regression model or NB models, i.e., NB1 model or NB2 model (Wooldridge (2009); Roback & Legler (2021); Kasyoki (2016); Winkelmann (2015); Cahoy *et al.* (2020)). The same models work also for underdispersion, which is less common for count data (Wooldridge (2009); Harris *et al.* (2012); Cahoy *et al.* (2020))[20]. For both underdispersion and overdispersion that might occur in our data, we decided to combine QP regression model and NB2 model for particular cases that will be explained in the following paragraphs. The possible usage of QP regression model when equi-dispersion assumption is not maintained, can be found in studies of Wooldridge (2009) and Roback & Legler (2021). For NB2 model, we can trace these pieces of information to researches of Kasyoki (2016) and Cahoy *et al.* (2020). We preferred NB2 model to NB1 model based on the study of Kasyoki (2016), nevertheless, several studies suggested, both NB models should work, when equi-dispersion assumption does not hold (Cahoy *et al.* (2020))[21]. Both QP regression model and NB2 model often give similar results, however, they vary in the way the variance of the model is computed (Ver Hoef & Boveng (2007)). Whereas, the variance of QP regression model is a linear function of the mean, the variance of NB2 model is a quadratic function of the mean (Ver Hoef & Boveng (2007))[22].

$$Var[y_i|\mathbf{x_i}] > E[y_i|\mathbf{x_i}] \tag{8.6}$$

This difference changes the appearance of the test, through which we will control the equi-dispersion assumption[23]. The test for equi-dispersion assumption is called Cameron and Triveldi's test (CT test) (Cameron & Trivedi (1990)). It tests the null hypothesis, i.e., equi-dispersion against the alternative hypothesis with variance in the form that can be seen in Equation 8.7. For cases when we will be tending to use QP regression model to treat possible overdispersion or underdispersion, we use for QP regression model recommended form of the test and put $g(.)$ equal to $g(E[y_i])$. In situations, when NB2 model will be preferred to deal with possible overdispersion or underdispersion, we use for NB2 model recommended form of the test and set $g(.)$ equal to $g(E[y_i]^2)$ (Cameron & Trivedi (1990))[24]. If the model does not pass a specific form of the test, i.e., α will be significantly (at 95%) different from zero, we will work

[20]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[21]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[22]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[23]Source: `https://rdrr.io/cran/AER/man/dispersiontest.html`.
[24]Source: `https://rdrr.io/cran/AER/man/dispersiontest.html`.

with either QP regression model or NB2 model instead of Poisson regression model. Once we have a form of the model selected, the remaining aim does not change, i.e., we attempt to estimate β's. For NB2 model, we will work with MLE framework, such as in the case of Poisson regression model[25]. For QP regression model we will exploit quasi-maximum likelihood estimation (QMLE) framework (Wooldridge (2009)).

$$H_0 : Var[y_i] = E[y_i], H_A : Var[y_i] = E[y_i] + \alpha g(E[y_i]) \tag{8.7}$$

The last possible issue regarding Poisson regression is the excess of zeros, i.e., a situation when the proportion of zeros is too high in a (dependent) random variable (Kasyoki (2016)). Researches again suggested several ways to deal with the excess: imposing hurdle models or imposing zero-inflated models (Kasyoki (2016); Winkelmann (2015); Hu *et al.* (2011))[26]. The difference between these two models can be found in the origin of zeros, to which they are set (Hu *et al.* (2011)). Hurdle models assume all zeros to be structural, i.e., they do not happen by chance (Hu *et al.* (2011)). For clarification, we exhibit an example of structural zeros. In the situation, when we count the number of items that customers buy in a store, structural zero would be a decision of an individual that cannot be overturned not to buy anything or not to go to the store at all[27]. Zero is the only possible outcome for this customer[28]. On the other hand, zero-inflated models assume zeros can be either structural or sampling (Hu *et al.* (2011)). The sampling (or random) zeros are those which happen by chance, e.g., the customer from the previous example goes to the store with the will to buy something but changes the mind and ends up with nothing, i.e., the outcome of this customer can be either zero or count (Hu *et al.* (2011))[29]. In our case, we suppose zeros that occur in the dependent variables to be solely sampling. In our opinion, football teams basically cannot make a decision before the kick-off that they will not kick a penalty or will not get a card. We believe that the same holds for referees and mistakes. Therefore, if necessary, we will treat the possible excess of zeros by zero-inflated models, which might bring better analysis, even though we assumed our zeros to be exclusively random. In zero-inflated models, a dependent variable is modeled

[25]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[26]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[27]Source: `https://bit.ly/3yduOVA`.
[28]Source: `https://bit.ly/3s4P3Rv`.
[29]Source: `https://bit.ly/3yduOVA`.

as a mixture of Bernoulli distribution and count distribution, e.g., Poisson or NB[30]. From the type of the count distribution comes the exact name of the model â€“ zero inflated Poisson (ZIP) model or zero inflated Negative Binomial (ZINB) model (Hu *et al.* (2011)). As NB models are one of the solutions for overdispersion or underdispersion, ZINB fits better than ZIP for such cases (Hu *et al.* (2011); Roback & Legler (2021); Kasyoki (2016)). For demarcating a threshold, whether to consider ZIP model or ZINB model in the very first decision, we exploit the study of Kasyoki (2016), whose research recommends using hurdle models if the proportion of zeros in a random variable is higher than 0.3. For zero-inflated models, we did not find such a threshold, rather recommendations to follow criteria in models' selection (Mohri & Roark (2022)). Therefore, we decided for the following approach. If the excess of zeros in the dependent variable is higher than 0.3, we, based on CT test compare Poisson regression model or NB2 model with ZIP model or ZINB model due to Akakike Information Criteria (AIC). This way, we get to the above-mentioned division between situations, when we will prefer NB2 model and when QP regression model. QP model is generally preferred to NB2 model in this thesis, primarily due to the book of Wooldridge (2009). Nevertheless, it will be considered only when the proportion of zeros in a dependent variable fits into the 0.3 threshold, and thus zero-inflated models will not be taken into account. On the other hand, when the excess of zeros in a dependent variable exceeds the initially stated threshold and zero-inflated models will be taken into account, NB2 model will be considered. We decided for this partition due to the characteristics of QP regression model and NB2 model and their comparability with ZINB model. We mentioned above that we will be comparing the former models with zero-inflated models through AIC. AIC was selected instead of log-likelihood because zero-inflated models do not consist of the same number of parameters as the former models (a dependent variable is modeled by two distributions instead of one)[31]. However, QP regression model works with different criteria than NB2 model or ZINB model called quasi-Akakike Information Criteria (QAIC)[32]. Therefore, we decided to compare models closer to each other in terms of criteria.

To conclude Section 8.4 we summarize the general approach that we decided to stick to within the determination of models. Firstly, we run the first bench-

[30]Source: `https://bit.ly/3yduOVA`.
[31]Source: `https://bit.ly/3yduOVA`.
[32]Source: `https://cran.r-project.org/web/packages/bbmle/vignettes/quasi.pdf`.

mark linear model estimated by OLS. Secondly, we proceed with the second benchmark model, i.e., Poisson regression model. Thirdly, we state whether the excess of zeros might be an issue in the model from the 0.3 threshold. Then we check whether the model is equi-dispersed using CT test under the above-described conditions. If both rules are satisfied, i.e., the excess of zeros in the dependent variable is lower than 0.3 and equi-dispersion holds, we will stick with Poisson regression model. If the excess of zeros is under the threshold and equi-dispersion assumption does not hold, we will be exploiting QP regression model. And in other situations, i.e., equi-dispersed and excess of zeros or not equi-dispersed and excess of zeros, we compare either Poisson regression model or NB2 model (based on CT test) with the corresponding zero-inflated model due to AIC. Commented and illustrated will be only results of the ultimately selected form of a model. The approach will be repeated twice because for each of our four topics, we will work with simple and multiple model.

8.5 Information about control variables

The following part of Chapter 8 consists of a general reflection about independent control variables that we will be exploiting in multiple models. In the case of each response variable, we divided control variables into three universal fields i.e., control variables related to the figure of on-pitch referee (R controls), control variables related to characteristics of teams and the match before the kick-off (T controls), and control variables regarding how the match proceeded (M controls). Behind each field of controls, we can suppose a logic, why it might be beneficial to add it to our regressions.

8.5.1 R controls

Exploiting R controls may be related to the phenomenon of potential inconsistency across referee squad. Dobson *et al.* (2007), who studied variations of on-pitch referees in amounts of yellow and red cards awarded in EPL during the 1996-2003 period, found significant differences in standards of referees in terms of these disciplinary actions. Therefore, we decided to use several means to control for possible inconsistency in awarding cards and penalties or committing errors across F:L referee squad. For this purpose, we will be using variables counting for the average number of cards and penalties per referee from the previous season. We also assumed that a proxy variable designed to capture a strain of the referee in terms of managing matches right after each other might affect the consistency. And finally, we constructed variables related to the experience of a referee in terms of whistled matches in F:L and European competitions. This concept was inspired by studies of Picazo-Tadeo *et al.* (2017), and Holder *et al.* (2022). Even though the experience was not a significant predictor in any of these football-related studies, both research teams decided to include it in their considerations.

8.5.2 T controls

Several studies suggested that various match outcomes, such as the number of cards and penalties, might be influenced by which teams play (Erikstad & Johansen (2020); Soares & Shamir (2016); Jones *et al.* (2003)). The researches of Erikstad & Johansen (2020) and Soares & Shamir (2016) focused on relationships among indices of success of a team and the number of yellow cards and penalty kicks awarded in a match. Soares & Shamir (2016) found a negative and significant (at least at 90%) relationship between the size of the team's budget and the number of yellow cards per foul awarded in the match by that team in three out of four investigated competitions. Using the similar approach, a positive and significant relationship (again at least at 90%) was detected between the number of penalties awarded in favor of the team (normalized by the number of shots inside the box) and the mentioned budget of the team in two out of four competitions (Soares & Shamir (2016)). Moreover, in one out of four competitions, a positive and significant relationship was revealed between yellow cards per foul awarded by the team and the rank of the team in the table (Soares & Shamir (2016)). The same ratio was detected for normalized penalties and the rank (Soares & Shamir (2016)). Erikstad & Johansen (2020) found out in the sample of Norwegian Eliteserien matches that the two most successful local teams were given in their matches 110% penalties that should have been (due to the expert panel), and their opponents in games against the top teams only 12.5%. Jones *et al.* (2003) examined whether the awareness of referees that a team is considered to be aggressive could increase the number of cards given to that team. Using the experimental and the control group of referees, they found a significant increase (Jones *et al.* (2003)).

We mentioned several examples of why we decided to control for which teams play a particular match and which characteristics we could attribute to the match before kick-off. The list of variables that we decided to use for this purpose is broader than in the case of R controls. The common characteristic of all T controls is that we can measure them before a specific match bridge itself to the first minute (unlike M controls that will follow). We will be again using several variables, counting for the average number of the specific incident (cards, penalties, defensive battles, and passes to box) per team from the previous season. Moreover, for the sake of T controls list, we constructed several variables interconnected with above-mentioned studies of Jones *et al.* (2003), Erikstad & Johansen (2020) and Soares & Shamir (2016). Firstly, a variable capturing the rival relationship between teams, where we assumed the idea of aggressivity of Jones *et al.* (2003). Secondly, a variable counting the distance between teams in the league table before the kick-off (Soares & Shamir (2016)). And lastly, a control variable regarding the success rate of the team based on studies of Erikstad & Johansen (2020) and Soares & Shamir (2016). We also assumed that season, i.e., spring, summer, fall, or winter, the presence of spectators in stands (distinguishable due to Covid-19 restrictions), and an hour of kick-off might be taken into account, and thus we created appropriate controls for these concepts.

8.5.3 M controls

Finally, the last group of control variables bears various statistics that happened in a particular match. Unlike the T controls, we are not able to collect them before the particular match finishes. Under M controls group, we assumed that each of our regressands, i.e., cards, penalties, and errors, is naturally connected with several incidents. We can introduce the example from the mentioned study of Soares & Shamir (2016), which standardized yellow cards by fouls and penalties by shots from the box. Therefore, yellow cards, red cards, penalties, fouls, defensive battles, goals, shots on goal, dribbling moves, passes to chance, attacks, corners, and various modifications of these statistics were included in the list of M controls. Since our models will be constructed for descriptive purposes (not predicting), and as we said, we approach matches as wholes, i.e., we do not distinguish in what minute a particular incident happened, we supposed that this specific group of controls can be used for explaining the relationship between VAR and the dependent variables.

8.5.4 Motivation for exploiting control variables

In the case of each dependent variable, controls will be divided into the above-mentioned three fields. However, for each response variable, exact controls assumed for the baseline model will be particularly different. The list of all used control variables (including their division and information about their purposes and their origins) can be found in Appendix A.

In our opinion, we could divide the general motivation to work with R controls, T controls and M controls in the relationship between VAR and the set of dependent variables into two segments. Firstly, as it was said in Section 4.3, VAR had been exploited in F:L only at several matches per round for an approximal length of three seasons (including two and half seasons of our research)[33]. Those matches might not have been selected randomly, as F:L used for VAR project only broadcast facilities of O2 TV channel, which (unlike nowadays) did not dispose of the right to broadcast all matches all season long (Klír (2019)). The selection of F:L matches, to which VAR was deployed in its pilot phases could be seen in the example from our dataset: VAR occurred in 157 out of 160 games, where at least one of the participants were either AC Sparta Praha or SK Slavia Praha. This selection is specific in comparison to other competitions that we previously mentioned in relation to the match-changing incidents, i.e., Bundesliga, Serie A, La Liga, and TSL. For these competitions, VAR was implemented suddenly, i.e., after a particular summer break VAR has taken part all season long. Therefore, the samples of related studies consisted of a season without VAR and a season with VAR (Carlos *et al.* (2019); Lago-Peñas *et al.* (2020)) or a particular number of match weeks without VAR and a particular number of match weeks with VAR (Gürler & Polat (2021)). We are aware of potential biases that the distribution of VAR to F:L matches may have made. We assumed that these biases might be the case for all four dependent variables we aimed to investigate. Moreover, we suppose that the distribution of VAR to F:L matches might have participated in creating possible inconsistencies among the referee squad and possible biases among M controls. Secondly, we assume that sources of potential biases among match statistics and potential inconsistencies among the referee squad of F:L may be broader than the first motivational segment suggests. We suppose that a part of these phenomena might have been created just randomly. Nevertheless, both assumed factors, i.e., the selection

[33] Source: https://www.lfafotbal.cz/videorozhodci.

of F:L matches, to which VAR was implemented in the case of our dataset in comparison to the related studies, and random factors encouraged us to divide the controls into three categories and include them to final models in the most efficient ways to control for possible biases.

8.6 Construction of final models

Finals models will be constructed based on several steps. Firstly, we run a simple Poisson regression for all discrete and continuous selected controls in relation to the dependent variable. We will investigate the relationship between residuals and fitted values to detect whether possible non-linear transformations of a predictor might be included[34]. If the plots suggest to us that they might, we compare variants by Pseudo R squared (Cameron & Windmeijer (1996)[35]. We will also explore the plot of residuals and leverages to identify and remove possible outliers[36].Secondly, we create the correlation matrix with a linear form of all considered controls and our dependent variable. We, meanwhile, focus just on the correlation between the controls and the dependent variable. Thirdly, we state a baseline model. For the sake of the best candidate model selection, we decided for stepwise backward selection (BS) i.e., the baseline model will consist of all considered control variables (full model)[37][38]. For the baseline model, we perform Variance Inflation Factor (VIF) to reveal possible multicollinearity issue[39]. If the issue is reported, i.e., VIF for a control variable is greater than five, we will eliminate the variable with a lower correlation with the regressand from the process of BS (James *et al.* (2013)). As the next step, we gradually subtract from the baseline model control variables based on the correlation coefficient with the response variable, i.e., firstly, we subtract the control variable with the lowest correlation with the response variable, then the second lowest, etc. We selected the correlation coefficient as it is a statistical measure of the strength of the relationship between two variables[40]. During the process we keep variables *VAR* and *phase* in the model. *VAR* is the independent variable of our interest, and *phase* variable was set to control

[34]Source: `https://online.stat.psu.edu/stat462/node/117/`.
[35]Source: `https://bit.ly/3y4JwhE`.
[36]Source: `https://bit.ly/3y5Vg3u`.
[37]Source: `https://bit.ly/3kvnzAE`.
[38]Source: `https://quantifyinghealth.com/stepwise-selection/`.
[39]Source: `https://www.investopedia.com/terms/v/variance-inflation-factor.asp`.
[40]Source: `https://www.investopedia.com/terms/c/correlationcoefficient.asp`.

for possible fixed effects. For only nominal variable *temp*, we decided to check its interconnectedness with the dependent variable through a simple Poisson regression. If the majority of the estimates of its levels are significant, at least at 90%, we keep the variable through the whole BS and then decide whether to keep it in the best candidate model or not. For variables where non-linear transformations were detected, we eliminate all their forms at once based on the correlation coefficient of the linear version of variables. All models in BS will be run both in the form of our workhorse model, i.e., Poisson regression model, and in the form of the econometric benchmark, i.e., the linear model estimated by OLS. Particular stepwise models will be compared through AIC (for Poisson regression model) and Adjusted R squared (Adj R^2) (for linear model estimated by OLS) as these measures penalizes for additional variables[41][42].We decided to stop BS by the first variable, which correlation with the dependent variable is higher than for *VAR* i.e., that control variable creates our stopping rule[43]. If this model is not better off than the previous (in terms of both AIC and Adj R^2), we will not continue with BS. In possible cases, when AIC and Adj R^2 will suggest different conclusions, we will prefer AIC as linear models estimated by OLS are in our thesis just for the sake of an econometric benchmark. When we find the best candidate model, we again firstly check VIF. If the multicollinearity issue is still not a case, we will possibly transform its form to either QP regression model, NB2 model, or zero-inflated models based on the approach that we determined in Section 8.4.

[41]Source: `https://www.scribbr.com/statistics/akaike-information-criterion/`.
[42]Source: `https://bit.ly/37UMiLK`.
[43]Source: `https://quantifyinghealth.com/stepwise-selection/`.

Chapter 9

Concrete models

We have just reached the part of the thesis where we introduce concrete models that might help us to support the studies from Chapter 5. The general approach in all four topics will be: firstly, we become more familiar with a dependent variable. Secondly, we present a selected set of independent variables, which we will use as control variables in multiple models. And finally, in context with previous studies, we state hypotheses. Results and discussions will be included in later parts of the thesis. The independent variable of our interest- *VAR* was presented in Section 7.3.

9.1 Yellow cards

9.1.1 Descriptive statistics

We decided to start with the relationship between *VAR* and match-changing incidents, where we begin with yellow cards models. Referees use yellow cards as caution messages to players for fouls or inappropriate behavior on the pitch. Two yellow cards mean a red card and sending-off from the game[1]. Information about yellow cards were collected from *Livesport*[2] webpage. Only yellow cards given to players were included in the research, i.e., yellow cards for staff members were not included. A variable that counts the total number of yellow cards awarded in a match by both teams was given *ycards* name.

Firstly, we look at the distribution of *ycards* variable. For this purpose, we plot in Figure 9.1 a histogram of our response variable. From the figure, it

[1] Source: `https://bit.ly/3F2aDvc`.
[2] Source: `https://www.livesport.cz/`.

can be seen that the distribution of yellow cards is right-skewed. The mean of the distribution is approximately equal to 4.4 and the median to 4. Moreover, because we are aware that we work with count data, we noticed that *ycards* variable took only 13 possible values - $(0, 1, \ldots, 12)$ with the mode equal to 4 that summarizes Table 9.1. Zero value is represented only in 2.5% of cases. Therefore, possible excess of zeros will not be treated (Kasyoki (2016)), and the final model will be either Poisson or QP.

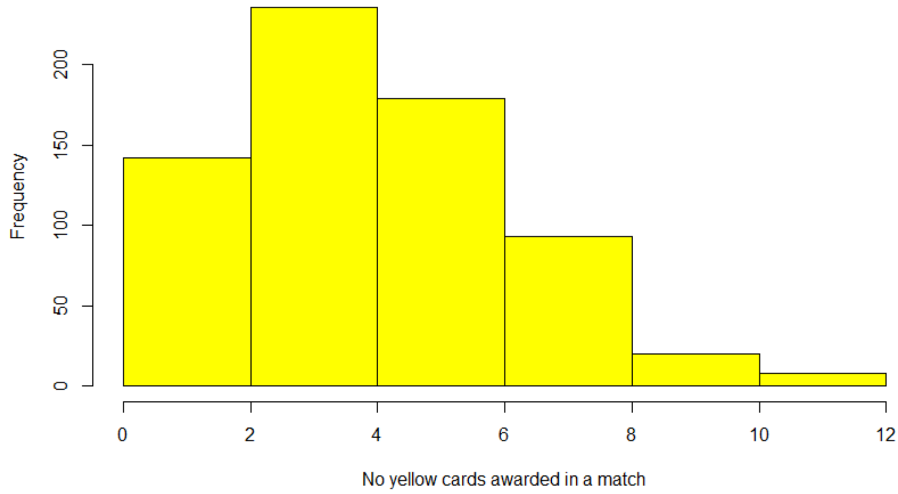

Figure 9.1: Histogram of *ycards*

Table 9.1: Frequencies of values of *ycards*

ycards	0	1	2	3	4	5	6	7	8	9	10	11	12
frequency	17	45	80	115	121	91	88	53	40	16	4	6	2

9.1.2 Control variables

In the next paragraph, we would like to introduce R controls, T controls, and M controls that might help us in relationship between *VAR* and *ycards*. The selected set of control variables can be seen on the correlation matrix depicted in Table 9.2. The matrix includes all variables that we were considering (except *phase* and *temp*, to which we approach specially, as it is mentioned in Section 8.6). However, to select the most suitable model, some of them were eliminated in process of BS, as it was discussed in the same section. The variables that we kept in the model are bolded. Additional information about all control variables, including their division into R controls, T controls and M controls, can be found in Appendix A. All variables that count for average value per match, i.e., *refyc*, *refrc*, *teambattsea*, *teamrcsea* and *teamycsea*, were computed for the previous season, e.g., all values of these variables for 2019/20 season were measured in 2018/19 season. Moreover, for variables *refyc* and *refrc*, we stated that a referee has to manage at least three games in the previous season (or at least in the season before) to count an observation. The base group of variable *group* was created by matches of primary parts of seasons and the base group of nominal control *temp* by winter matches. Variable *farintab* was standardized by number of finished rounds before the match. As *primetime* games were evaluated evening matches, which were played separately, at 5 p.m. or later (in F:L, we can usually find from two to four *primetime* matches per round)[3]. As matches, where a greater *rivalry* can be found, were assumed Prague derby (AC Sparta Praha versus SK Slavia Praha), Little Prague derbies, including Vrsovice derby (AC Sparta Praha or SK Slavia Praha versus Bohemians 1905), Podjestedske derby (FC Slovan Liberec versus FK Jablonec), Silesian derby (FC Banik Ostrava versus SFC Opava), and one of Moravian derbies (1. FC Slovacko versus FC Fastav Zlin). Moreover, several interregional rivalries, i.e., games between Sparta or Slavia versus FC Viktoria Plzen, and games between Sparta or Slavia versus FC Banik Ostrava, were also included[4]. As a *successful team* was considered, due to the historical table of F:L, either SK Slavia Praha or AC Sparta Praha[5]. In cases of *farintab*,*goalsdiff* and *teamycsea* variables, we detected from graphs of residuals and fitted values possible non-linear transformations. The transformations were supported by pseudo R squared of concerning models. Therefore, we included variables *ftsq* ($farintab^2$), *gdsq* ($goalsdiff^2$), and *teamycseasq* ($teamycsea^2$) to the baseline model. On the other hand, variables *farintab* and *teamycsea* were eliminated because *ftsq* and *teamycseasq* showed better fit in terms of Pseudo R squared as sole variables. Moreover, several observations were subtracted due to outliers in *refrc* variable.

The baseline multiple model included all variables from Table 9.2, variables *phase*, *temp* and the predictor of our interest *VAR*, because VIF values were lower than five for all separate independent variables, and all levels of nominal variable *temp* were significantly different from the base group in terms of *ycards*

[3]Source: `https://www.livesport.cz/`.
[4]Source: `https://bit.ly/3w370Bs`.
[5]Source: `https://www.fortunaliga.cz/`.

Table 9.2: Correlation matrix for *ycards*

Predictor	**Correlation**
fouls	0.32
rcards	0.28
goalsdiff (Q+L)	-0.21
rivalry	0.16
teamycsea (Q)	0.13
farintab (Q)	-0.11
primetime	0.11
sucteam	0.10
group	-0.08
teamrcsea	0.08
refyc	0.07
corona	-0.06
zerohalf	0.04
battles	0.03
refrc	-0.02
teambattsea	0.01
ispen	0.01

(at 90%). The correlation coefficient between *VAR* and *ycards* was equal to 0.07, therefore, *BS* should have consisted of eight steps. Nevertheless, none of the stepwise models without variable *teamrcsea*, which correlation coefficient with *ycards* was just higher than the coefficient of *VAR*, performed better than the previous one, therefore, the selection eventually contained eleven steps, i.e., up to variable *primetime*, which elimination made both stepwise models worse. From the selection, we obtained twelve linear models estimated by OLS and the same number of Poisson regression models. During the process, the following Poisson regression model was never better than the previous one in terms of AIC, except the model without variable *refyc*, and certainly the last model without *primetime* variable. Furthermore, several additional stepwise models estimated by OLS performed better than their neighbors in the selection. This held for variables *battles*, *zerohalf*, *group*, and *sucteam*. Therefore, we decided to add variables *refyc* and *primetime*, whose elimination made both AIC and Adj R^2 worse back to the stepwise regression. Furthermore, we performed forward selection with variables, whose elimination made worse just Adj R^2. We gradually added these variables back due to the correlation matrix, i.e., firstly variable *sucteam*, which correlation with *ycards* was highest among these four controls, etc. This way, we decided to construct the best candidate model from all control variables, whose correlation coefficient with *ycards* is better than for *primetime* and variables *primetime*, *refyc*, *phase*, *temp* and *VAR*. When we compared the best candidate with and without variable *temp*, both measures were better, when *temp* was included to the model, therefore, we decided for keeping this control. The model constructed in such a way performed the best in terms of AIC. VIF of all predcitors of the best candidate model stayed under the threshold.

9.1.3 Hypotheses

Firstly, we just briefly remind why did we choose *ycards* as one of our dependent variables. The first reason is a broad demand for football statistics by both people whose job is football-related and people who consume football in their leisure. The second reason is an extra value that might result from studying the relationship between *VAR* and *ycards*, i.e., value in awareness of players of VAR (Carlos *et al.* (2019)). And the last reason that we have not mentioned yet is the interconnectedness of *ycards* with match-changing incidents through red cards, e.g., it was revealed in the sample of UEFA Europa League (UEL) matches that prior to red card 65% of yellow cards had been awarded by the team, who got subsequently the red card[6].

The suggestion of Carlos *et al.* (2019) will be a workhorse idea for our hypotheses. The intensity of a foul or consideration of a handball belongs to the category of subjective decisions. Unlike in cases of factual decisions, e.g., offsides, the border between yes and no is not strict[7]. Therefore, players might not know whether they get a yellow card or a red card for specific misconduct. However, as Carlos *et al.* (2019) suggested, they might be aware of VAR. Although in TSL and La Liga was not revealed a negative and significant relationship between *VAR* and *ycards*, in our dataset we will assume, what we saw in Bundesliga and Serie A (Carlos *et al.* (2019); Lago-Peñas *et al.* (2020); Gürler & Polat (2021)). We suppose a negative significant (at 90%) relationship between *VAR* and *ycards* in the multiple model with R controls, T controls and M controls (**H1**). Nevertheless, in the case of the simple model, we do not assume the relationship to be negative and significant because we suppose that

[6]Source: `https://bit.ly/3vV3bhl`.
[7]Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

possible bias from the omission of the control variables that we described in Subsection 8.5.4 will be positive (**H2**).

9.2 Red cards

9.2.1 Descriptive statistics

The second match-changing incident that we present in our study is red cards. Red cards, such as yellow cards, belong to the category of disciplinary actions that an on-pitch referee awards for various misconducts. A player or a staff member awarded by red card is sent off from the game, i.e., it is a stronger action than a yellow card. Red cards could be given directly or as a sum of two yellow cards[8]. Information about red cards were collected from *Livesport*[9] webpage. The variable that counts for the number of red cards in the match was given *rcards* name. Red cards awarded by staff members were not counted.

Table 9.3 shows us the distribution of *rcards* variable. We could see that the variable took over 678 observations only three values, i.e., $(0, 1, 2)$. As the proportion of zeros in the random variable is equal to 83%, i.e., it is higher than the 0.3 threshold, we will be detecting a possibility of zero-inflated models' usage (Kasyoki (2016)). The mean of *rcards* is equal to 0.19, which is comparatively smaller than the mean of red cards across 87 competitions over the world from the 2015-2020 period, which was equal to 0.25[10].

Table 9.3: Frequencies of values of *rcards*

rcards	0	1	2
frequency	561	105	12

9.2.2 Control variables

In Table 9.4 is depicted the correlation matrix with control variables, which we decided to use for *rcards* regression (as previously: except predictors *phase* and *temp*). The control variables that we decided to keep in the final model

[8]Source: `https://bit.ly/3F2aDvc`.
[9]Source: `https://www.livesport.cz/`.
[10]Source: `https://football-observatory.com/IMG/sites/mr/mr57/en/`.

(after the process of BS) are bolded. Additional information about all control variables, including their division into R controls, T controls and M controls, can be found in Appendix A. The list of variables is similar to the set that we presented for yellow cards, because as we already mentioned, these punishments are close. Again, all variables that count for average value per match, i.e., *refyc*, *refrc*, *teambattsea*, *teamrcsea* and *teamycsea* were computed for the previous season such as in Subsection 9.1.2. The variables *temp*, *rivalry*, *primetime*, *sucteam*, *group*, *refyc*, *refrc* and *farintab* were explained in the same subsection. Variables *teamycsea*, *battles* and *timeoffirstyc* were suspicious from non-linear transformations due to the plots of residuals and fitted values. Non-linearity was, in the case of these controls, supported by pseudo R squared of concerning models. Therefore, we included variables *teamycseasq* (*teamycsea*^{2}), *battlessq* (*battles*^{2}), and *tofysq* (*timeoffirstyc*^{2}) to the baseline model. On the other side, control *timeoffirstyc* was eliminated from the selection, because its squared version fitted better as the sole variable. Moreover, several observations were subtracted due to outliers in *refrc* variable.

Table 9.4: Correlation matrix for *rcards*

Predictor	**Correlation**
ycards	0.28
battles (Q+L)	-0.15
timeoffirstyc (Q)	-0.12
group	-0.09
refyc	0.09
teamycsea (Q+L)	0.06
primetime	0.05
teambattsea	-0.05
teamrcsea	0.04
farintab	-0.04
ispen	0.04
sucteam	0.04
zerohalf	-0.04
goalsdiff	0.03
fouls	0.03
refrc	-0.01
corona	-0.01
rivality	0.01

As none of the independent variables were suspicious from reporting multicollinearity (from our VIF threshold) and the majority of the levels of the only nominal variable *temp* were, unlike in the previous case, not significant, the baseline model included all variables from Table 9.4, variable *phase* and response variable of our interest *VAR*. Variable *temp* will be added at the end to the best candidate model for evaluation, whether keep it or not. The correlation coefficient between *VAR* and *rcards* was equal to 0.08, therefore, BS should have contained fourteen steps. Nevertheless, Poisson regression model without variable *refyc*, which correlation coefficient with *rcards* was just higher than the coefficient of *VAR*, performed worse than the previous model. Therefore the selection eventually contained an additional step, i.e., eliminating on control variable *group*, which made both measures worse. From the selection, we reached sixteen linear models estimated by OLS and sixteen Poisson regression models. During the selection, none of Poisson regression models was better than the previous one in terms of AIC. The only Poisson regression model, whose AIC increased, was the last model, from which variable *group* was eliminated. However, we cannot say the same about the stepwise linear models estimated by OLS. The linear models, from which were subsequently eliminated variables *refyc*, *teamycsea*, *teamycseasq*, *teambattsea*, *farintab*, and *goalsdiff*, performed worse in terms of Adj R^2 than the previous model. Therefore, we approached to the same way as we did it in the case of *ycards*. Firstly, we added variable *group*, which elimination made worse both measures back to the stepwise regression. Secondly, we performed the forward selection with the above-mentioned controls, whose subtraction made worse just Adj R^2. During the forward selection, we held the same approach as for *ycards*, i.e., we added the controls back in turn based on Table 9.4. This way, we constructed the best candidate model from all control variables, whose correlation coefficient with *rcards* was better than for *group* and variables *group*, *goalsdiff*, *phase*, and *VAR*. When we compared the best candidate with and without variable *temp*, both measures were better when *temp* was not included in the model. The model constructed in such a way performed the best in terms of AIC. VIF of all predictors of the best candidate model stayed under the threshold.

9.2.3 Hypotheses

We just briefly remind the logic behind selecting *rcards* variable as one of our regressands. Firstly, as we discussed: the demand for football statistics is widespread and red cards belong to the relevant category of match-changing incidents[11]. Secondly, such as in the case of yellow cards, researching the relationship between *rcards* and *VAR* might bring us extra value in studying the decision-making behavior of on-pitch referees (Holder *et al.* (2022)). From the study of Spitz *et al.* (2020) we know that VAR created a positive extra number of red cards in 2195 matches across 13 competitions. However, as we saw in the research of Holder *et al.* (2022) conducted on Bundesliga and Serie A, those extra cards did not increase the total number of red cards in comparison to previous seasons. Therefore, Holder *et al.* (2022) suggested that VAR influences the decision-making behavior of on-pitch referees in the form that the presence of VAR creates a negative bias in the number of red cards and VAR interventions subsequently shift the number to normal. In F:L, during our 678 researched games, we registered 21 extra red cards awarded due to reviews, which is more than 16% of all red cards in the dataset. Therefore, we could aim hypotheses on supporting Holder *et al.* (2022). The variable counting for the total number of red cards in the match minus the extra red cards from VAR reviews was marked as *rcards2*.

From the studies of Carlos *et al.* (2019), Lago-Peñas *et al.* (2020), Gürler & Polat (2021) and Holder *et al.* (2022) we assume *VAR* not to be significant in relation to *rcards* in the multiple model with R controls, T controls and M controls (**H3**). Moreover, we suppose (keeping the controls fixed) a negative and significant (at 90% level) relationship between *VAR* and *rcards2*, i.e., if we subtract red cards awarded after VAR reviews (**H4**). **H3** and **H4** together can support the assumption of Holder *et al.* (2022) about VAR influencing the decision-making behavior. On the other hand, we suppose the relationship between *VAR* and *rcards* in the simple model to be positive and significant (at 90%), because we assume that possible bias due to the omission of the control variables that was described in Subsection 8.5.4 will be positive (**H5**). This way we also assume the relationship between *VAR* and *rcards2* not to be significant in the simple model (**H6**).

[11]Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

9.3 Penalties

9.3.1 Descriptive statistics

The third match-changing incident we aim to explore in relation to *VAR* is penalty kicks. A penalty is awarded if a player commits a direct free kick offense inside the penalty area, i.e., inside the box, of the own team[12]. The research of *InStat* reported that from almost 100 000 penalty kicks across the world since 2009, more than 75% of them resulted in the goal (74% in our dataset)[13]. Therefore, we can take a penalty as a big opportunity to score a goal and change the match. For our research, we count all penalty kicks (not even scored ones). Data was collected from *Livesport*[14] webpage and the variable that counts for the number of penalties in the match bears *pen* name.

The distribution of *pen* variable in 678 F:L matches is similar to the distribution of *rcards* in terms of excess of zeros. Although the excess is lower than 83%, it is still high enough to be greater than our threshold (Kasyoki (2016)). Values and their frequencies of *pen* variable summarizes Table 9.5. Since the proportion of zeros in *pen* is equal to 70.5%, we will be considering also zero inflated models based on AIC. The average number of penalty kicks-0.33 is slightly lower in comparison to European top five competitions (due to Union of European Football Associations (UEFA) ranking) from 2019/20 and 2020/21 seasons, where the mean was equal to 0.36[15]. Variable *pen* takes again a limited number of values across our dataset, i.e., $(0, 1, 2, 3)$.

Table 9.5: Frequencies of values of *pen*

pen	0	1	2	3
frequency	478	179	17	4

9.3.2 Control variables

The correlation matrix with the control variables, which we decided to work with in *pen* multiple regression, is illustrated in Table 9.6 (except controls *phase*

[12]Source: `https://bit.ly/3F2aDvc`.
[13]Source: `https://instatsport.com/football/article/penalty_research`.
[14]Source: `https://www.livesport.cz/`.
[15]Source: `https://bit.ly/3F4HjV9`.

and *temp*). The control variables that we decided to keep in the final model (after the process of BS) are again bolded. Additional information about all control variables, including their division into R controls, T controls and M controls, can be found in Appendix A. A bunch of variables that we selected for *pen* regression is more different than the sets chosen for cards' regressions. Both cards and penalty kicks are awarded for a particular misconduct as a punishment. Nevertheless, penalties contain another component, which is an opportunity to score the goal through the shot from 12 yards (10.97 meters) distance, i.e., from shorter distance than a regular free kick[16]. All variables that count for average value per match, i.e., *refpen*, *teamptbsea*, and *teampensea* were computed for the previous season, as we did it in before. Variables *rivalry*, *primetime*, *sucteam*, *group*, *temp*, and *farintab* were already discussed in Subsection 9.1.2. For variable *refpen* holds the same as for *refyc* and *refrc* predictors. Variables *dribbles* and *passtogoal* were computed as the sum of all dribbling moves or passes to goal, i.e., passes to a scoring chance of all forwards and midfielders in the match. Dribbling moves and passes to goal of defenders were estimated and added to this sum in the form of average from the previous season. The composition of these variables in such a form was enforced by the way, these statistics are gathered on *Fortuna Liga*[17] webpage. Variables *refpen*, *teamptbsea*, *farintab*, *rcards*, *attacks*, *fouls*, *dribbles*, *passtogoal*, and *timeoffirstyc* were suspicious from non-linear transformations from the plots of residuals and fitted values. In all of these cases, pseudo R squared measures encouraged us to transform these variables from the linear way to the quadratic. Therefore, we included variables *refpensq* ($refpen^2$), *teamptbseasq* ($teamptbsea^2$), *ftsq* ($farintab^2$), *rcsq* ($rcards^2$), *attsq* ($attacks^2$), *fsq* ($fouls^2$), *dribblessq* ($dribbles^2$), *ptgsq* ($passtogoal^2$). and *tofysq* ($timeoffirstyc^2$) to the baseline model. On the other hand, controls *farintab*, *rcards*, *fouls*, *passtogaol*, and *timeoffirstyc* were subtracted from the model, because their exclusive squared versions fitted better as sole predictors. Several observations were eliminated due to outliers in *dribbles*, *passtogoal*, *attacks*, and *shotsongoal* predictors.

The majority of the levels of only nominal predictor *temp* did not perform as significant in Poisson regression model, therefore we eliminated them from all steps of BS but the last. Furthermore, none of the independent variables exceeded our threshold for multicollinearity determined by VIF. Therefore, our

[16]Source: `https://bit.ly/3F2aDvc`.
[17]Source: `https://www.fortunaliga.cz/`.

Table 9.6: Correlation matrix for *pen*

Predictor	Correlation
shotsongoal	0.13
refpen (Q+L)	-0.12
rcards (Q)	0.09
sucteam	0.09
attacks (Q+L)	-0.08
battles	-0.08
primetime	0.07
timeoffirstyc (Q)	-0.07
teamptbsea (Q+L)	0.06
dribbles (Q+L)	-0.05
rivalry	0.04
ptg (Q)	0.04
corona	0.04
farintab (Q)	0.03
fouls (Q)	0.02
corners	0.02
group	-0.02
teampensea	$<\|0.01\|$

baseline model contained all variables from Table 9.6, and certainly variables *VAR* and *phase*. The correlation coefficient between *VAR* and *pen* was equal to 0.12 (just two controls reported higher correlation with *pen*), thus BS process included seventeen steps. Firstly, we gradually subtracted sixteen control variables, i.e., from *teampensea* variable to *rcards* (respectively *rcsq*) variable, which had the highest correlation coefficient with *pen* among all controls, whose coefficient was lower than 0.12. From these steps of BS, we obtained seventeen linear and seventeen Poisson regression models. Secondly, we eliminated also variables *refpen* and *refpensq*, whose correlation with *pen* was just higher than for *VAR*. Both AIC and Adj R^2 of the models without *refpen* and *refpensq* worsened, therefore, we decided not to continue with the selection. In the previous seventeen steps of BS, in two cases both measures worsened (variables *tofysq* and *rcsq*). In none of the cases happened that just one of the measures got worse, therefore, we added variables *tofysq* and *rcsq* back to the stepwise regression and the model built from variables *shotsongoal*, *refpen*, *refpensq*, *rcsq*, *tofysq*, *VAR*, and *phase* was taken as the best candidate model. Indeed, the model performed the best among all stepwise models in terms of AIC. The addition of nominal variable *temp* did not make none of our measures better-off. VIF of all predcitors of the best candidate model stayed under the threshold.

9.3.3 Hypotheses

The reasons why did we include *pen* variable to our set of dependent variables are similar to the reasons for red cards. Again firstly, we investigate penalty kicks because they belong to the bunch of match-changing statistics, and we are aware of a broad demand for football data[18]. Secondly, penalties may bring us the same extra value as red cards, i.e., in studying the decision-making behavior of on-pitch referees (Holder *et al.* (2022)). In the dataset of Spitz *et al.* (2020) VAR all over created a positive number of extra penalties. However, in Bundesliga and Serie A, those extra penalties just equalized the negative bias in context with previous seasons (Holder *et al.* (2022)). Extra penalties were also detected in our dataset, i.e., in 678 F:L matches, we found in total 43 extra penalty kicks due to VAR interventions. Those 43 penalties mean slightly more than 19% of all penalties kicked across all matches. We could again aim hypotheses on the research of Holder *et al.* (2022), i.e., that VAR influences the decision-making behavior of on-pitch referees (in a possible form that VAR presence creates a negative bias in the number of penalties and VAR interventions shift this number to the normal). The variable that takes the total number of penalties in the match minus or plus awarded or canceled penalties after VAR reviews was given *pen2* name. Unlike in the case of red cards, VAR created across 678 F:L matches also negative extra value in the number of penalty kicks. The above-mentioned number 43 is the total count if we add $+1$ for every penalty awarded due to VAR review and subtract -1 for every penalty canceled due to VAR review.

Due to the studies of Carlos *et al.* (2019), Lago-Peñas *et al.* (2020), Holder *et al.* (2022), we suppose that the relationship between *VAR* and *pen* will not be statistically significant in the multiple model (**H7**). Moreover, we assume in the multiple model the negative and significant (at 90% level) relationship between *VAR* and *pen2*, i.e., if we count for awarded penalties before VAR interventions only (**H8**). **H7** and **H8** together can support the suggestion of Holder *et al.* (2022) that VAR presence influences the decision-making behavior of on-pitch referees. Nevertheless, in the simple model we suppose that the

[18]Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

relationship between *VAR* and *pen* will be positive and statistically significant (at 90% level), because we assume that possible bias due to the omission of the control variables that was described in Subsection 8.5.4 will be positive (**H9**). This way we suppose the relationship between *VAR* and *pen2* not to be significant in the simple model (**H10**).

9.4 Errors of on-pitch referees

9.4.1 Descriptive statistics

Finally, we present models that investigate relationships between the presence and interventions of VAR and the number of errors made by on-pitch referees in game-changing situations. As we know from the theoretical chapters, VAR can intervene only in the case of a reviewable match-changing incident, i.e., goal, penalty, direct red card, or mistaken identity[19]. This study deals with all of them except mistaken identity. As we previously mentioned, we decided to omit mistaken identity based on the research Spitz *et al.* (2020), which found only a small proportion of these incidents both for checks and for reviews. Moreover, the source that will be noted in the following paragraphs, which collects information about mistakes of referees in F:L, does not present this type of error at all.

To be able to count for errors of on-pitch referees in F:L, we created a set of variables. This set could be divided into two parts based on at which phase of VAR procedure mistakes were made. The first part of the set is not related to actions of VAR itself but rather to actions of referees that are directly on the pitch. It consists of four variables connected with three categories of errors in match-changing incidents, i.e., errors regarding goal decisions, penalty decisions, and red card decisions. The last variable is the sum of these three. We marked these four variables as *referbcg*, *referbcp*, *referbcrc* and *referbc*. Each of them is a count that can take zero or positive integer values. However, as we discussed in Section 8.1, the dependent variables, which we work with, are constrained by the duration of a football match and the empirics. Variable *referbcg* counts for errors of the pitch-based referee in the particular match in goal situations, variable *referbcp* in penalty situations and variable *referbcrc* in red card situations. The last one from the list-*referbc* counts for all errors of

[19]Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

the pitch-based referee in a particular match. In the case of these variables, it does not matter whether there was or it was not VAR present on the match because they take into account errors of on-pitch referees before a possible consultation with VAR could happen (that is why all of them includes before consultation (BC) sign).

The second part of the set consists of eight variables, which count for errors of referees in the same situations, but after possible consultation with VAR. Names of the variables were chosen with the same logic as in the first part of the set. We just switched BC sign for after consultation (AC) sign. Therefore, we have variables *referacg*, *referacp*, *referacrc* and *referac*. In matches to which VAR was not deployed, before consultation and after consultation means the same, and thus the variables will be equal in all observations. The second part of the set also includes modifications of four AC variables marked by after consultation 2 (AC2) sign. In Chapter 3, we discussed situations, to which VAR cannot intervene, e.g., the second yellow card situations[20]. That is why we designed variables *referac2*, *referacg2*, *referacp2* and *referacrc2*, which count for all errors of on-pitch referees that remained after possible consultation with VAR, without particular errors, which VAR might have corrected, if it had a permission to do so.

To conclude the essence of our two sets of variables, the first BC set focuses on the work of on-pitch referees and evaluates their decisions. On the other hand, results from the second set might tell us something about the work of VAR itself and its collaboration with on-pitch referees, i.e., OFR 's and VAR-only reviews. Moreover, in the second set, we dispose of AC2 variables that suggest how the work of VAR and the collaboration with on-pitch referees could have looked like if the competence of VAR had been modified.

The data was collected from Communiques of KR FAČR that can be found in the official webpage of *FAČR*[21]. Committee provides reports that evaluate how referees fulfilled their expectations every week of every season in our research. The reports are open to the public, and we used them to transform qualitative information into a measurable way. We were aware of the ambiguity difficulties that this step might evoke. Therefore, we were extremely cautious

[20] Source: `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.
[21] Source: `https://www.fotbal.cz/`.

in evaluating what was an error and which incident belonged among GZD's. Due to the communiques, in several situations, on-pitch referees or VAR stood by not clear problems. We have already divided football decisions into factual and subjective. This division tells us that borders in football might not always be unambiguous. Furthermore, F:L officiates might not always dispose of all camera angles or technologies they wish they had, e.g., the virtual offside line that operates, for instance, in EPL, has not been implemented yet to F:L[22].

In this paragraph, we present descriptive statistics of all twelve mentioned variables counting for errors of on-pitch referees. In Table 9.7, we can see which values at which frequencies took all the responses that we described. The most before consultation referees' mistakes that were committed in one match were five. The median of all variables is equal to zero, which implies that in more than half games, a mistake was not made at all. It also tells us which family of models will also be considered. Because the excess of zeros is greater than the 0.3 threshold in all twelve cases, we will be comparing zero-inflated models with Poisson regression model or NB2 model based on AIC (Kasyoki (2016)). Most mistakes were made due to penalty reasons, followed by goal and red card reasons at a similar rate. If we counted the mean of total BC mistakes, it would equal 0.35. If we did the same for total AC mistakes, the average would improve to 0.19 (for the category AC2 it would be even 0.17). The difference between *referac* and *referac2* in terms of total errors is equal to 15. In 13 cases out of 15, the difference was made due to the second yellow card reasons. Furthermore, the averages of all particular types of errors are also lower for AC variables. Therefore, because we know that BC and AC variables are equal in matches, to which VAR was not implemented, we can state that the cleaned impact of VAR was positive in all categories of mistakes in F:L matches, i.e., the total number of errors was generally lower after possible interventions of VAR such as it was found in the study of Spitz *et al.* (2020) and KU Leuven research. Nevertheless, we would like to establish our hypotheses rather on studies of Holder *et al.* (2022) and Samuel *et al.* (2020) and investigate impacts after the consideration of possible positive bias in the number of errors in matches, where VAR was present. In other words, we would like to investigate VAR as the whole (not only its interventions).

[22]Source: `https://www.premierleague.com/news/1488423`.

Table 9.7: Frequencies of values of errors responds

referbc	0	1	2	3	4	5
frequency	488	150	35	4	0	1
referbcg	0	1	2			
frequency	619	54	5			
referbcp	0	1	2			
frequency	572	99	7			
referbcrc	0	1	2	3		
frequency	622	53	2	1		
referac	0	1	2	3		
frequency	569	94	11	4		
referac2	0	1	2	3		
frequency	581	84	10	3		
referacg	0	1	2			
frequency	648	28	2			
referacg2	0	1	2			
frequency	650	26	2			
referacp	0	1	2			
frequency	624	52	2			
referacp2	0	1	2			
frequency	624	52	2			
referacrc	0	1	2			
frequency	640	36	2			
referacrc2	0	1	2			
frequency	652	25	1			

9.4.2 Control variables

This subsection will be devoted to the bunch of control variables we will be working on within the multiple model. Unlike in the cases of other dependent variables, we deal with not one but twelve regressands. For several reasons, we construct the multiple model just for the case of total BC errors, i.e., *referbc* variable. One of the reasons is based on the essence of our dependent variables that we presented in the previous subsection. We basically do not suppose that AC and AC2 errors can be qualitatively explained by the same set of controls as it was designed for BC errors when $VAR = 1$. In our opinion, not only R controls, T controls and M controls would enter this relationship, but also several characteristics of VAR and AVAR. Other reasons for dealing with *referbc* only in the multiple model can be explained by the limited scope of the thesis and our interest in establishing hypotheses on the studies of Holder *et al.* (2022), and Samuel *et al.* (2020). Therefore, we left more profound regressions of after consultation variables and before consultation variables for the particular match-changing incidents for further research.

In Table 9.8, we can observe the correlation matrix with our designed controls for *referbc* regression without variables *temp* and *phase*. The controls that we decided to include to the final model are again bolded. Additional information about all control variables, including their division into R controls, T controls and M controls, can be found in Appendix A. The set of control variables that we decided to work with contains several variables that were not used in the regressions of match-changing incidents. We decided to include more R controls, as errors of referees are the topic of this section. For variable *eurref*, appearances of referees in European competitions were counted for the previous season in two European cups-UEFA Champions League (UCL) and UEL, because UEFA Europa Conference League (UECL) had not existed yet. As a *byeweek* was considered a spare week, during which a referee was not nominated to F:L match, e.g., if a referee was nominated to match in 17$^{\text{th}}$ round and 19$^{\text{th}}$ round, *byeweek* would equal to one in 19$^{\text{th}}$ round as 18$^{\text{th}}$ week was spare for that referee. Variables *rivalry*, *primetime*, *group*, *temp*, and *sucteam* were discussed in the previous sections. Variable *expref* counts for an actual number of previously managed games in F:L of a referee before each particular match. *Corners* variable was suspicious from non-linear transformations based on the plots of residuals and fitted values. Pseudo R squared encouraged us to include this variable to the selection as a quadratic one, i.e., *cornerssq*. On the other hand, variable *corners* was no longer used as *cornerssq* fitted better as the sole predictor.

The baseline multiple model contained all variables from Table 9.8 and variables *VAR* and *phase*. The majority of the levels of nominal variable *temp* was not significantly different from the base group in terms of *referbc* and all separate predictors performed sufficiently in terms of VIF. The correlation coefficient between *VAR* and *referbc* was equal to 0.21 (only *pen* control reported higher one). BS process included twelve steps. We decided to stop the selection by variable *cards*, which is the first variable with lower correlation coefficient with *referbc* than *VAR*, because if we stopped by the next variable, as it was stated in the general approach, we would have just *VAR* and *phase* variables

Table 9.8: Correlation matrix for *referbc*

Predictor	**Correlation**
pen	0.27
cards	0.19
sucteam	0.17
primetime	0.16
rivalry	0.09
corners (Q)	-0.09
adjgoals	0.05
corona	0.05
zerohalf	-0.03
byeweek	0.02
eurref	-0.02
group	0.01
expref	-0.01

left in the model. During the selection, we firstly gradually eliminated eleven control variables, i.e., from *expref* variable to *sucteam* variable, which was the first variable with lower correlation with *referbc* than *cards*. From these steps of BS, we reached twelve linear regression models estimated by OLS and twelve Poisson regression models. Secondly, we eliminated also variable *cards*. Both AIC and Adj R^2 of the last models got worse, therefore, we decided not to continue with the selection. In the case of previous steps of BS, elimination of two particular control variables worsened both measures (variables *adjgoals* and *cornerssq*). Moreover, after the subtraction of *sucteam*, *corona* and *expref* controls, just AIC improved. Therefore, we added the variables, whose elimination made both measures worse to 12^{th} step of BS and compared it through forward selection with the regressions, where were except these variables also variables *sucteam*, *corona* and *expref*. This way we detected that the best candidate model consists of variables *pen*, *cards*, *cornerssq*, *adjgoals*, *phase* and *VAR*. The addition of variables *temp*, *sucteam* and *corona* did not make AIC better. VIF of all predcitors of the best candidate model stayed under the threshold.

9.4.3 Hypotheses

Reasons why the counts of errors of on-pitch referees were involved in our research were discussed previously. We just remind that we suppose possible benefits from studying this issue for football and refereeing authorities. If we found enough evidence to support suggestions and results of Samuel *et al.* (2020) and Holder *et al.* (2022) regarding more mistakes in matches, where VAR was present, Czech football and refereeing authorities might reconsider some approaches.

In the following paragraphs we specify several assumptions regarding our issue for both simple and multiple models. In the case of all BC variables we suppose, what studies of Samuel *et al.* (2020) and Holder *et al.* (2022) suggested i.e., the presence of VAR influences the decision-making behavior of on-pitch referees as we can find more mistakes in matches with VAR. We assume the relationships (between *VAR* and the variables from BC set) to be positive and significant in all models: at 95% level in the case of simple models: *referbc* (**H11**), *referbcg* (**H12**), *referbcp* (**H13**), and *referbcrc* (**H14**) and at 90% level in the case of multiple model (**H15**), where we deal with *referbc* variable only. We decided to lower the significance level for the multiple model, because we assume that possible bias in the number of errors of on-pitch referees due to the omission of the control variables that was described in Subsection 8.5.4 will be positive.

Regarding all AC variables except *referacg*, we suppose that the impact of VAR in terms of correcting mistakes, which is generally positive as we can see from Table 9.7 or from studies of Spitz *et al.* (2020) and KU Leuven, will not be high enough to see significantly (at 95% level) lower number of after consultation errors of on-pitch referees in matches with VAR in simple models: *referac* (**H16**), *referacp* (**H17**), and *referacrc* (**H18**). On the other hand, for the variable that counts for goal errors, we suppose that a decrease in AC mistakes in matches, where VAR was present will be significant (at 95%), since the most common reason for goal interventions in our dataset was an offside (**H19**)[23]. And we know that offsides belong to category of factual decisions. For variables bearing AC2 mark, we assume the same hypotheses, because the total difference among AC and AC2 variables was only 15 (also **H16**, **H17**, **H18** and **H19**).

[23]Source: `https://www.fotbal.cz/`.

Chapter 10

Results

Chapter 10 connects to what was established in Chapter 8 and Chapter 9. The following paragraphs will present results from the regressions of the best candidate models in the form suitable to the assumptions. The chapter is divided into four sections based on the demonstrated relationship. In all of these sections, we firstly reveal the results from the simple model, then from the multiple model, and finally, we discuss the results and compare them with related studies. For both simple and multiple models, we will be interested also in equi-dispersion assumption (made on Poisson regression model)[1]. Previously, we were on the concrete cases talking over just the issue of excess of zeros, which was due to our threshold apparent from descriptive statistics (Kasyoki (2016)) For dependent variables, by which the excess was found i.e., *rcards*, *pen* and all variables counting for errors, we will be considering ZIP, ZINB, Poisson regression model and QP regression model based on CT test and AIC (Cameron & Trivedi (1990); Hu *et al.* (2011); Roback & Legler (2021); Kasyoki (2016))[2]. For *ycards* response variable, where the excess of zeros was not found, we will make a decision whether to stick with Poisson regression model or rather select QP regression model based on CT test (Cameron & Trivedi (1990))[3]. Nevertheless, for all models that we will possibly interpret in this chapter, the process of extracting information from β coefficients is not as straightforward as for linear models estimated by OLS. The sign of a relationship and its significance can be interpreted from specific pieces of information that we provide through tables at the very first moment. However, the magnitude of a relationship is unlike for OLS counted from $\exp(\beta)$. We demonstrate it on an example of a binary

[1]Source: `https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf`.
[2]Source: `https://rdrr.io/cran/AER/man/dispersiontest.html`.
[3]Source: `https://rdrr.io/cran/AER/man/dispersiontest.html`.

predictor x: e.g., if $\exp(\beta_x)$ is equal to 1.26, we can associate 26% increase of a dependent variable when $x = 1$. Continuous predictors work similarly but with unitary increase[4]. For zero-inflated models, in some cases, when several predictors acquire high values, we might be forced to reschedule these predictors due to the singularity issue[5]. We decided for standardization as this form of adjustment because it does not change the performance of the model, which will help us in the comparison of former models and zero-inflated models[6].

10.1 Yellow cards

10.1.1 Simple model

We begin with *ycards*, where we supposed negative and significant relationship with *VAR* in the multiple model (**H1**), but not in the simple model (**H2**). Already from the summary of distributions of *ycards*, conditionally on *VAR*, which can be observed in Table 10.1 and also from knowledge about positive correlation coefficient between *VAR* and *ycards* from Subsection 9.1.2, it could be deducted that **H2** probably will not be rejected. Distributions of *ycards* when $VAR = 1$ and when $VAR = 0$ seem to be relatively similar in all points except the mean, which is higher by 0.3, when VAR was present on the match.

Table 10.1: Distribution of *ycards* conditionally on *VAR*

	Min	1Q	Med	Avg	3Q	Max
ycards | VAR = 0	0	3	4	4.24	6	11
ycards | VAR = 1	0	3	4	4.54	6	12

For Poisson regression model, the detected dispersion was significantly higher than CT test affords for equi-dispersion assumption, therefore, we were encouraged to use simple QP regression model as the excess of zeros is not the case for *ycards* (Cameron & Trivedi (1990))[7]. The model supported our appearance from Table 10.1. The sign of the relationship is positive and moreover is significant at 90% level (Table 10.2), therefore, we cannot reject **H2**. The exponentiated coefficient could be interpreted that *ycards* increases by 7% if

[4]Source: `https://bit.ly/37SKfrR`.
[5]Source: `https://bit.ly/3s1VULt`.
[6]Source: `https://bit.ly/3vZzIDa`.
[7]Source: `https://rdrr.io/cran/AER/man/dispersiontest.html`.

Table 10.2: Results from *ycards* simple QP model

regression model	*ycards* QP
β	0.07
$\exp\beta$	1.07
VAR total assoc.	7%
SE	0.04
t statistic	1.71
p value	0.09
sign.	>90%

$VAR = 1$. From the simple model, we could have supposed that players in F:L are during the play not aware of VAR being present as they get more yellow cards (Carlos *et al.* (2019)). However, it might be, in our opinion, relatively early to say so because we have not run the multiple model with controls from Table 9.2 yet. Moreover, if we ran this relationship through the linear model estimated by OLS, the Adj R^2 would be just 0.03. Therefore, we let conclusions for the latter part of this chapter.

10.1.2 Multiple model

The control variables that we decided to include to the multiple model regarding *ycards* and *VAR* are bolded in Table 9.2. We considered the model, which is constructed this way, as the best candidate because it ranked highest among other candidates in terms of AIC. After the examination of CT test (with appropriate form $g(.)$), the dispersion parameter α was significantly (at 95% level) underdispersed (α was equal to -0.1 with corresponding z statistic equal to 2.1). Therefore, we transformed the model to QP regression model as it is outlined in Section 8.4. Results from QP regression model are depicted in Table 10.3. When we keep the controls in the model, the estimate of variable *VAR* is still positive (*VAR* was associated with a 2% increase of *ycards*). However, it is no longer significant. The corresponding p value is equal to 0.74, therefore, we can reject hypothesis **H1**. Regarding the estimates of the controls, the strongest connection with *ycards* showed *fouls* variable, which is, in our opinion, justifiable as fouls belong to the misconduct category, for which yellow cards are awarded. Unlike for T controls and M controls, the only R con-

Table 10.3: Results from *ycards* multiple QP model

predictor	β	$\exp\beta$	**SE**	**t statistic**	**p value**	**sign.**
VAR	0.02	1.02	0.05	0.34	0.74	<90%
refyc	0.04	1.04	0.03	1.57	0.12	<90%
teamycseasq	0.01	1.01	0.01	2.72	0.01	>99%
ftsq	-0.05	0.95	0.03	-2.00	0.05	>95%
primetime	0.10	1.10	0.05	2.10	0.04	>95%
rivalry	0.20	1.22	0.06	3.20	<|0.01|	>99%
rcards	0.24	1.28	0.04	6.56	<|0.01|	~100%
goalsdiff	-0.10	0.91	0.03	-2.82	<|0.01|	>99%
gdsq	0.01	1.01	0.01	1.00	0.32	<90%
fouls	0.02	1.02	<|0.01|	6.72	<|0.01|	~100%
temp_aut	-0.13	0.88	0.05	-2.36	0.02	>95%
temp_spr	-0.18	0.83	0.06	-2.87	<|0.01|	>99%
temp_sum	-0.15	0.86	0.06	-2.65	0.01	>99%
phase	-0.02	0.98	0.02	-1.54	0.13	<90%

trol that went through to the best candidate model-*refyc*, was not significant. On the other hand, the positive and significant statistical relationship with *ycards* was except for *fouls* variable revealed for the number of red cards in the match. The binary variables controlling for the rival relationship between the teams and hour of kick-off also showed a positive and significant connection with *ycards*. The non-linear transformations of variables *goalsdiff*, *teamycsea* and *farintab* also performed as significant. Furthermore, we could observe a significant decrease in yellow cards in all seasons towards winter, which was in the base group. The coefficient of variable *phase* that was set to limit possible fixed effect was negative but not significant. The intercept of the model was equal to 1.8. Adj R^2 of the linear version of the model stopped on value 0.24 and AIC of Poisson regression model on number 2645. In the next subsection, the main discussion will come back to the variable of our interest. We compare the result from the simple and multiple models with the outcomes from previously mentioned studies.

10.1.3 Discussion

We just remind that the expectations from adding controls to *ycards* model were discussed earlier in Subsection 9.1.3. Although the control variables changed the statistical relationship between *VAR* and *ycards* the direction we assumed, the relationship remained positive in the multiple model. Therefore, we could have rejected hypothesis **H1**. **H1** was set to support the assumption of Carlos *et al.* (2019) from Bundesliga and Serie A about players being aware of VAR during the game and thus behaving more cautious because of the fear from sending-off. On the other hand, we could mention the study of Aksum *et al.* (2020), which investigated the visual fixations of players during the match in Norwegian Eliteserien. The researchers suggested that players focus mainly on the ball, a teammate, or an opposing player during the match. Although the study does not tell us that players do not have the refereeing squad in their minds, it suggests that there are plenty of incentives in different locations that players perceive during the match, and we could assume that the fear from sending-off retreats to the background.

The negative and significant relationship between *VAR* and *ycards* in the simple model was not supported also by other studies from Chapter 5 conducted on La Liga and TSL, where the relationship was not significant such as in our multiple model (Lago-Peñas *et al.* (2020); Gürler & Polat (2021)). Nevertheless, our simple model brought a result that is different from all three previously mentioned studies (that aimed at the same relationship without control variables) because β coefficient from the simple QP regression was even significant on the positive side. Therefore, we could not have rejected hypothesis **H2**. We were not able to explain why the presence of VAR itself on the match should have generated a greater amount of yellow cards, therefore, we rather stuck with the assumption from Subsection 9.1.3 that this result might have been biased due to the omission of related controls. In Subsection 8.5.4 we discussed more detailly why the omission of controls might cause biases in our dependent variables.

Despite the multiple model did not support the assumption of Carlos *et al.* (2019), it supported the assumption from the last paragraph that the simple model was biased due to the omission of control variables. β coefficient of *VAR* was no longer significant in relation to *ycards* as its p value was approaching one. Therefore, we obtained a piece of evidence that VAR does not increase yellow cards as it suggested the simple model.

10.2 Red cards

10.2.1 Simple models

We continue with the regressions regarding *rcards* and *rcards2* variables. For the total number of red cards in the match, i.e., *rcards* variable we assumed in the simple model positive and significant (at 90% level) relationship with *VAR* (**H5**). However, for the adjusted variable *rcards2*, which does not count with red cards awarded after VAR interventions, we, on the other hand, supposed the relationship with *VAR* not be significant in the simple model (**H6**). In Table 10.4 we present the results from Poisson regression models, which were selected for both *rcards* and *rcards2* regressions. We decided for Poisson regression models, because for both dependent variables the dispersion parameter α was not significantly (at 95% level) different from zero after the examination of CT test and Poisson regression model performed better than ZIP model in terms of AIC.

Table 10.4: Results from *rcards* and *rcards2* simple Poisson models

regression model	*rcards* Poisson	*rcards2* Poisson
β	0.37	0.06
$\exp\beta$	1.44	1.06
VAR pres. assoc.	-	6%
VAR total assoc.	44%	-
SE	0.18	0.19
z statistic	2.02	0.32
p value	0.04	0.75
sign.	>95%	<90%

From Table 10.4 we can see that both **H5** and **H6** will not be rejected. The relationship between *VAR* and *rcards* was indeed positive. Moreover, its significance exceeded our expectations as it fit even to 95% level. On the other side, variable *VAR* was not significant in relation to *rcards2*. The estimate remained positive. However, p value from the regression was not close to any threshold of significance. Therefore, in the simple models, the presence and interventions of VAR were associated with 44% increase of *rcards* and the presence of VAR exclusively with 6% increase in *rcards2*. Despite the results from the simple models were different from what we saw in the related researches of Holder *et al.* (2022), Carlos *et al.* (2019), Lago-Peñas *et al.* (2020) and Gürler & Polat (2021), we leave conclusions for the discussion subsection. As we mentioned in Subsection 9.2.3, we suppose that the omission of control variables will create a positive bias in terms of red cards. Without results from the multiple model, we meanwhile do not assume that VAR may generate a greater number of red cards awarded during the match.

10.2.2 Multiple models

In the multiple model, we supposed the relationship between the presence of VAR on the match (including its interventions for *rcards*) and the number of red cards awarded in the match or its adjustment to be different than in the simple model due to the additional control variables. Firstly, we assumed the coefficient of *VAR* will not be significant in relation to *rcards* (**H3**). Secondly, we supposed that the same coefficient will be negative and significant (at 90%) in relation to *rcards2* (**H4**). The control variables that we decided to work within the best candidate model can be found in Table 9.4. Unlike in the case of *ycards*, for both *rcards* and *rcards2* the proportion of zeros in the variable was higher than our initially stated threshold. Moreover, after the examination of CT test, both multiple Poisson regression models performed as significantly underdispersed (z statistics of α's was close to -3). Therefore, the decision was between NB2 model and ZINB model in both cases. When we compared these variants by AIC, both in the case of *rcards* regression and *rcards2* regression, ZINB model showed lower AIC, thus, we selected this model for both regressions. Nevertheless, before examining the regressions, we were forced to standardize several controls due to the issue discussed at the beginning of the chapter. Concretely, we adjusted variables *tofysq*, *battlessq* and *battles* (the last control just for *rcards2* regression). The new variables took *stofysq*, *sbattlessq* and *sbattles* names. After the standardization, we applied ZINB model on both *rcards* and *rcards2* regression.

The results from the regressions can be seen in Table 10.5 and Table 10.6. As it was said, when we defined zero-inflated models, the distribution of the models is a mixture of Bernoulli distribution (values marked by *BIN* in tables) and NB2 distribution (values marked by *CNT* in tables). For the sake of the results, we will focus mainly on the count distribution part because, as it was said, we generally assume zeros in the dependent variables to be exclusively

Table 10.5: Results from *rcards* multiple ZINB model

predictor	**dist.**	β	$\exp\beta$	**SE**	**z statistic**	**p value**	**sign.**
VAR	CNT	0.03	1.03	0.21	0.16	0.87	<90%
group	CNT	-1.05	0.35	0.51	-2.06	0.04	>95%
ycards	CNT	0.27	1.31	0.05	5.91	<|0.01|	∼100%
stofysq	CNT	0.27	1.31	0.16	1.62	0.10	<90%
battles	CNT	-0.07	0.93	0.02	-3.07	<|0.01|	>99%
sbattsq	CNT	1.10	3.01	0.47	2.27	0.02	>95%
goalsdiff	CNT	0.07	1.07	0.07	0.99	0.32	<90%
phase	CNT	<|0.01|	1.00	0.07	<|0.01|	1.00	<90%
VAR	BIN	-69.7	-	97.9	-0.71	0.48	<90%
group	BIN	39.6	-	65.7	0.60	0.55	<90%
ycards	BIN	12.1	-	15.1	0.80	0.42	<90%
stofysq	BIN	77.0	-	102	0.76	0.45	<90%
battles	BIN	-2.83	-	2.96	-0.96	0.34	<90%
sbattsq	BIN	36.2	-	39.9	0.91	0.36	<90%
goalsdiff	BIN	-19.4	-	24.7	-0.79	0.43	<90%
phase	BIN	25.8	-	137	0.46	0.64	<90%

Table 10.6: Results from *rcards2* multiple ZINB model

predictor	**dist.**	β	$\exp\beta$	**SE**	**z statistic**	**p value**	**sign.**
VAR	CNT	-0.13	0.88	0.22	-0.56	0.58	<90%
group	CNT	-1.24	0.29	0.59	-2.10	0.04	>95%
ycards	CNT	0.26	1.30	0.05	5.27	<|0.01|	∼100%
stofysq	CNT	-0.21	0.81	0.16	-1.27	0.21	<90%
battles	CNT	-2.21	0.11	0.55	-4.04	<|0.01|	∼100%
sbattsq	CNT	2.29	9.86	0.60	3.83	<|0.01|	∼100%
goalsdiff	CNT	0.23	1.26	0.08	2.90	<|0.01|	>99%
phase	CNT	<|0.01|	1.00	0.08	0.06	0.95	<90%
VAR	BIN	10.6	-	13.3	0.80	0.42	<90%
group	BIN	3.48	-	>1000	<|0.01|	1.00	<90%
ycards	BIN	-60.7	-	60.2	-1.00	0.31	<90%
stofysq	BIN	-114	-	112	-1.02	0.31	<90%
sbattles	BIN	248	-	243	-1.02	0.31	<90%
sbattsq	BIN	403	-	392	1.03	0.31	<90%
goalsdiff	BIN	70.3	-	68.7	1.02	0.31	<90%
phase	BIN	-205	-	200	-1.03	0.31	<90%

sampling. This assumption was supported by the fact that none of the β coefficients of the predictors from the binomial parts were significant. However, significant were not β coefficients of variable *VAR* in the count parts of the models either. If we compare the sign and the significance of the statistical relationship between *VAR* and *rcards* from the simple and the multiple model, we can see that *VAR* was in the multiple model no longer a significant predictor of the dependent variable. Keeping other predictors fixed, the presence and interventions of VAR were associated with just a 3% increase of *rcards*. Therefore, we cannot reject hypothesis **H3**. Nevertheless, keeping the control variables in the model, *VAR* was not a significant predictor of *rcards2* either. The coefficient was unlike in the simple model negative (the presence of VAR was associated with a 12% decrease in *rcards2*). However, its p value was way too high to label the predictor as significant. Thus, we can reject hypothesis **H4**. Regarding the controls: variables *group*, *ycards*, *battles* (*sbattles* respectively), and *sbattlessq* were significant both in relation to *rcards* and *rcards2*. The signs of the coefficients of these variables also were the same in both models, i.e., the statistical connection with red cards regressands was positive for the number of yellow cards in the match, negative for binary variable *group* and non-linear for variables counting for the number of defensive battles in the match. The only control variable which was significant just in one model was variable *goalsdiff*. Adj R^2 of the linear versions of the models was equal to 0.11 for both regressands. AIC of ZINB models was equal to 592 (for *rcards* regression) and 520 (for *rcards2* regression). In the discussion subsection, we will come back to the relationship between *VAR* and the dependent variables counting for red cards in the match.

10.2.3 Discussion

Before we begin with the discussion, we just briefly remind that *VAR* performed as a significant predictor for *rcards* in the simple model, but not in the multiple model (thus, we could not have rejected both **H3** and **H5**). However, for the same predictor, the significant statistical connection was revealed neither in the simple model nor in the multiple model in relation to the adjusted number of red cards, i.e., variable *rcards2* (therefore, we could have rejected **H4**, but could not have **H6**).

Generally, the results from red cards regressions remind the results from *ycards* regressions, if we compare them to the related studies of Carlos *et al.* (2019), Gürler & Polat (2021), Holder *et al.* (2022), Lago-Peñas *et al.* (2020). Whereas the researchers did not reveal either positive or negative statistical connection between *VAR* and *rcards* in the simple model in four European competitions (Bundesliga, Serie A, La Liga, and TSL), in F:L the coefficient of *VAR* was positive and significant in relation to *rcards* in the simple model. This way, we revealed that the presence and following interventions of VAR were statistically associated with a 44% increase in *rcards*. The only study from the above-mentioned that worked with the adjusted number of red cards, i.e., counting only incidents, about which was decided without VAR, was the research of Holder *et al.* (2022). They found out that *VAR* decreased *rcards2* by 30% totally in Bundesliga and Serie A. Our simple model revealed even an 6% increase, which was not significant, though. The reason why we were investigating such relationships was to establish on the suggestion of Holder *et al.* (2022) that the presence of VAR influences the decision-making behavior of on-pitch referees. A result, which would support this suggestion, might consist in the significantly different results between *rcards* and *rcards2* regressions. Ideally, it would be under the circumstances, which described Holder *et al.* (2022), i.e., negative and significant coefficient of *VAR* for *rcards2* shifted not to be significant for *rcards*. Nevertheless, we already discussed the specifics of our dataset, and thus, even though we control for various other factors in multiple models, there still might be some positive biases in the number of red cars in matches, where VAR was present. Given that we suppose that KR FAČR provides qualitative evaluations, we decided to consider significantly different results as situations when either one coefficient is significant and the second is not, or the statistical associations of *VAR* projected to *rcards* and *rcards2* differ by more than 30% due to the study of Holder *et al.* (2022). And such results were revealed in F:L in the simple model.

Nevertheless, we would have liked to rather establish conclusions on multiple models due to Subsection 8.5.4 and Subsection 9.2.3, where we discussed the control variables and assumed positive bias due to their omission in the number of red cards in the match. This assumption also supported the fact that the simple models' results were positively biased compared to the results of related studies. More importantly, these possible biases might not have been the same both in the case of *rcards* regression and *rcards2* regression, and thus, the difference between the results could have tightened and even stopped being significant in the multiple model. Therefore, despite the fact that a significant difference between the results from *rcards* regression and *rcards2* regression was found in the simple model, the multiple models may have shown us different results.

The additional control variables in fact tightened the difference between the results from *rcards* regression and *rcards2* regression. *VAR* was no longer significant in relation to *rcards*, as we assumed. However, *VAR* remained not significant in relation to *rcards2* either. Even though the presence of VAR statistically decreased initially given red cards by 12% and following VAR interventions changed this percentage to positive 3%, none of the coefficients ended up significant keeping the controls fixed. Therefore, although we were able to find a 15% difference that VAR interventions created (38% in the simple models), the difference was not due to the above-mentioned consideration significant, and thus, we cannot support a part of the assumption of Holder *et al.* (2022), i.e., VAR influence decision-making behavior of on-pitch referees in terms of red cards.

10.3 Penalties

10.3.1 Simple models

The next variables on the list are *pen* and *pen2*. For these variables, we constructed the hypotheses similarly as for *rcards* and *rcards2*, because through these variables we together aim on study of Holder *et al.* (2022). Therefore, for the total number of penalties in the match, i.e., *pen* variable, we assumed in the simple model positive and significant (at 90% level) relationship with *VAR* (**H9**) and for the adjusted variable *pen2*, which adds to *pen* variable penalty kicks that was withdrawn by VAR and subtracts from *pen* variable penalties that was awarded due to VAR, we supposed the relationship with *VAR* not be significant at all in the simple model (**H10**). The proportion of zeros exceeded 0.3 in both dependent variables. For *pen2* response, the dispersion parameter α was not significantly different from zero during CT test, therefore, we were selecting from Poisson regression model and ZIP model. Poisson regression model performed better in terms of AIC, thus, we decided to use it for *pen2* regression. Nevertheless, for not adjusted *pen* variable, the dispersion param-

Table 10.7: Results from *pen* and *pen2* simple models

regression model	*pen* NB2	*pen2* Poisson
β	0.45	0.12
$\exp\beta$	1.57	1.13
VAR pres. assoc.	-	13%
VAR total assoc.	57%	-
SE	0.14	0.15
z statistic	3.25	0.80
p value	$<\|0.01\|$	0.42
sign.	>99%	<90%

eter ended up significantly underdispersed after performing CT test. Thus, we narrowed the options for *pen* variable between NB2 model and ZINB model. As the first mentioned model performed better in terms of AIC, we selected it for *pen* regression.

The results from the regression can be observed in Table 10.7. We could not have rejected both **H9** and **H10**. The predictor of our interest had a positive statistical relationship with *pen*, which significance was even above 99%, i.e., it exceeded our expectations. The presence and interventions of VAR were associated with a 57% increase in penalty kicks. On the other hand, *VAR* has not remained significant in relation to *pen2*. Although the estimate stayed positive, its p value exuded significance. The increase due to the presence of VAR declined to 13% for *pen2*. Therefore, the results from the regressions of penalty kicks were somehow similar to the results that we obtained in Subsection 10.2.1 (also in the way they were different from the results of related studies of Holder *et al.* (2022) Carlos *et al.* (2019), Lago-Peñas *et al.* (2020)). However, we leave conclusions for the discussion subsection, as we did in other cases because, as we mentioned previously, we assume the omission of control variables makes these results positively biased. Meanwhile, we do not suppose that VAR might create a greater number of penalties during the match.

10.3.2 Multiple models

The hypotheses that we supposed for response variables *pen* and *pen2* in the multiple model again shade the hypotheses for red cards regressands. Therefore, we have hypothesis **H7**, i.e., the coefficient of *VAR* will not be significant in relation to *pen* keeping the controls fixed and hypothesis **H8**, i.e., the coefficient of *VAR* will be negative and significant (at 90% level) in relation to *pen2* keeping the controls fixed. The control variables that we selected for these regressions are depicted in Table 9.6. However, in both regressions, we dealt with the possible issue of excess of zeros (determined by the 0.3 threshold) and the issue of underdispersion, which was revealed after the examination of CT test (both α's were negative and significantly different from zero). Therefore, final models were selected due to AIC from NB2 model and ZINB model. Unlike in the case of red cards regressions, ZINB model did not perform better than NB2 model in both cases, but just for *pen* regression. For *pen2* regression, NB2 model showed lower AIC, and therefore, was chosen instead ZINB model. As we were working with ZINB model, we were forced to standardize several control variables. It regarded variables *tofysq* and *shotsongoal*. The new variables took *stofysq* and *sshotsongoal* names.

The results from ZINB model with *pen* in the position of the dependent variable can be seen in Table 10.8. In Table 10.9 are illustrated the results from NB2 model for *pen2*. In the case of ZINB model, we will focus mainly on the results from the count distribution part as it is described in Subsection 10.2.2 (*CNT* values in tables). Moreover, this part is comparable with NB2 model as they came from the same distribution (Hu *et al.* (2011)). The results in terms of coefficients of *VAR* are different than for red cards regressions because they barely changed from the results of the simple models. The relationship between *VAR* and *pen* in the count distribution part of multiple ZINB model was positive and significant (such as in the simple model). Keeping the control variables fixed, the significance remained higher than 99%, and the interpretation of the coefficient decreased just by 1%, i.e., the presence and interventions of VAR were associated with a 56% rise in the number of penalty kicks. On the other side, the relationship was still not significant when we switched *pen* for *pen2*. Despite the coefficient was not significant, it was still positive and its interpretation changed just by 2% (from 13% increase to 11% increase in *pen2*, if $VAR = 1$). Therefore, we can reject both **H7** and **H8**. Regarding the control variables: *shotsongoal* (*sshotsongoal* respectively) and at least one from the pair of controls counting for average penalties per match for a particular referee from the previous season remained significant in both models. Furthermore, variable *rcsq* was significant just in the first ZINB model. Other controls: *tofysq* (*stofysq* respectively) and *phase* did not show a significant connection in none of the models. Adj R^2 of the linear version of the model was equal just to 0.05 for *pen* regression and it was even lower, i.e., 0.03 for *pen2* regression. AIC of ZINB model (*pen* regressand) was equal to 848 and the same measure for NB2 model (*pen2* response) 765. In the discussion subsection, we will come back to the relationship between *VAR* and the dependent variables counting for penalties in the match.

Table 10.8: Results from *pen* multiple ZINB model

predictor	**dist.**	β	$\exp\beta$	**SE**	**z statistic**	**p value**	**sign.**
VAR	CNT	0.44	1.56	0.16	2.80	0.01	>99%
refpen	CNT	-3.23	0.04	0.94	-3.43	<|0.01|	~100%
refpensq	CNT	3.79	44.1	1.21	3.13	<|0.01|	>99%
sshotsongoal	CNT	0.17	1.18	0.07	2.38	0.02	>95%
rcsq	CNT	0.16	1.18	0.09	1.82	0.07	>90%
stofycsq	CNT	-0.06	0.94	0.08	-0.74	0.46	<90%
phase	CNT	0.02	1.02	0.05	0.35	0.73	<90%
VAR	BIN	-15.2	-	46.8	-0.33	0.75	<90%
refpen	BIN	161	-	146	1.10	0.27	<90%
refpensq	BIN	175	-	167	1.05	0.30	<90%
sshotsongoal	BIN	-65.3	-	58.4	-1.12	0.26	<90%
rcsq	BIN	-4.22	-	>1000	<|0.01|	1.00	<90%
stofycsq	BIN	52.5	-	46.9	1.12	0.26	<90%
phase	BIN	18.0	-	19.6	-0.91	0.36	<90%

Table 10.9: Results from *pen2* multiple NB2 model

predictor	β	$\exp\beta$	**SE**	**z statistic**	**p value**	**sign.**
VAR	0.10	1.11	0.17	0.61	0.54	<90%
refpen	-2.42	0.09	1.03	-2.35	0.02	>95%
refpensq	1.41	4.08	1.24	1.14	0.26	<90%
shotsongoal	0.06	1.06	0.02	2.38	0.02	>95%
rcsq	0.16	1.17	0.10	1.59	0.11	<90%
tofycsq	<|0.01|	1.00	<|0.01|	-1.41	0.16	<90%
phase	0.04	1.04	0.06	0.73	0.47	<90%

10.3.3 Discussion

Despite the hypotheses stated for penalty regressions reminded us the hypotheses for red card regressions, we cannot say the same about the results that we obtained. The coefficient of the variable of our interest was positive and significant both in the simple and in the multiple model in relation to *pen* (therefore, we could have rejected **H7**, but could not have **H9**). On the other hand, neither in the simple nor in the multiple model *VAR* remained a significant predictor for the adjusted number of penalties counted by *pen2* variable (thus, we could have rejected **H8**, but not **H10**). The difference consists in that *VAR* withstood significant in the multiple model in relation to *pen*.

Nevertheless, the results from the simple models also, in this case, remind the results from yellow and red cards simple regressions and their comparison to the related studies. Whereas, the researchers did not reveal in their studies a significant statistical relationship between *VAR* and *pen* in the simple model in three European competitions (Bundesliga, Serie A, La Liga), in F:L the coefficient of *VAR* performed as positive and highly significant in relation to *pen* in the simple model (Carlos *et al.* (2019); Lago-Peñas *et al.* (2020); Holder *et al.* (2022)). This way, we detected a 57% increase (even higher than for *rcards*) in the number of penalty kicks in the match, which was associated with the presence and following interventions of VAR. The study of Holder *et al.* (2022) worked also with the adjusted penalties, i.e., *pen2*. The researchers found that the presence of VAR was associated with a 25% decrease in *pen2* in the sample of Bundesliga and Serie A. In F:L we revealed even a 13% increase, but the coefficient was not significant.

As the suggestion of Holder *et al.* (2022) that the presence of VAR influences the decision-making behavior of on-pitch referees is a bearing idea also for this issue, our approach in supporting this suggestion was the same as for red cards. We just changed the minimal difference to 25%, due to the study of Holder *et al.* (2022). The rest remained the same as in Subsection 10.2.3, we just applied the approach on penalty regressions. Although VAR interventions solely were associated with a 44% increase in penalty kicks in the simple model and thus we could have marked this difference as significant, we left broader conclusions for the multiple model from several times discussed reasons.

However, unlike in the case of red cards, the differences in the interpretation of the coefficients between the simple and the multiple models were small for penalties issue (just 1% for *pen* and 2% for *pen2*). Therefore, the difference between the results from the multiple regressions remained almost the same towards the results from the simple regression, i.e., the results from the simple models were not biased due to the omission of chosen controls. In the multiple model the presence of VAR increased initially awarded penalties by 11%, and the following VAR interventions magnified this percentage to 56%, i.e., we counted 45% difference due to VAR interventions, which exceeded 25% stated threshold. Moreover, the coefficient of *VAR* remained significant in relation to *pen* and not significant in relation to *pen2*. Therefore, we can particularly support the assumption of Holder *et al.* (2022) i.e., VAR influences the decision-making behavior of on-pitch referees in penalty situations.

We were wondering whether small differences between the simple and the multiple model could not have been made by an inappropriate choice of control variables. Even though Poisson regression model with such controls reported the lowest AIC, we also investigated, whether the percentual increase in *pen* associated with *VAR* did not decline rapidly in the baseline model. Nevertheless, keeping all controls fixed, *VAR* was still associated with a 54% increase in the number of penalties.

10.4 Errors of on-pitch referees

10.4.1 Simple models

The last topic, for which we have not presented results yet, is errors of on-pitch referees. We just remind that errors were covered by a set of twelve variables: four of them counting for mistakes before possible consultation with VAR and remaining eight after possible consultation with VAR. In the form of a simple model, we aimed to go through all twelve variables. The hypotheses that we stated for them can be found in Subsection 9.4.3. Nevertheless, we firstly mention which model did we decide for in the case of each response in relation to *VAR*. As the proportion of zeros in all dependent variables is greater than the initially stated threshold, we considered except Poisson regression model and NB2 model also ZIP model and ZINB model. The majority of simple Poisson models passed CT test as a significant overdispersion or underdispersion was not detected for them. Therefore, for these variables, the considered models narrowed to Poisson regression model and ZIP model only. However, this did not hold for variables *referac* and *referac2*, i.e., variables counting for the total number of errors of an on-pitch referee in the match after possible consultation, for which parameter α was higher than CT test allowed, and thus, these models were overdispersed. Therefore, for these two particular cases, the choice was between NB2 model and ZINB model. For all of the mentioned options, we counted AIC. As this measure was not better for any option from zero-inflated models family, we decided for NB2 model in the case of variables *referac* and *referac2* and for Poisson regression model in the case of other ten variables.

The results from the regressions can be seen in Table 10.10, Table 10.11 and Table 10.12. We firstly provide a report on BC variables. As we can see, the presence of VAR reported a positive and significant relationship with all of them. However, a significance varied between errors of on-pitch referees from goal reasons and others. The predictor of our interest showed a statistical connection, which significance approached 100% in relation to variables *referbc*, *referbcp* and *referbcrc*. The magnitude of the significance even exceeded 95% level stated in hypotheses **H11**, **H13** and **H14**. Therefore, we cannot reject **H11**, **H13** and **H14**. The presence of VAR was associated with 124% increase in total mistakes, 134% in penalty mistakes and 190% in red card mistakes. Furthermore, we cannot reject hypothesis **H12**, either. Despite the significance of variable *VAR* decreased to 95% level in relation to *referbcg*, this level was assumed by **H12**. In this case, the presence of VAR was associated with 69% increase in *referbcg*. Hypotheses **H11**, **H12**, **H13** and **H14** supported the suggestions from studies of Holder *et al.* (2022) and Samuel *et al.* (2020), i.e., the presence of VAR statistically influence decision-making behavior of on-pitch referees as they make more mistakes in matches, to which VAR was implemented.

Secondly, we present the results of after consultation variables. We just remind that this set of variables was established to reveal whether VAR interventions are able to fix a possible larger amount of mistakes in matches where VAR was present. In matches, to which VAR was not implemented, AC and BC variables are equal. Moreover, this set was divided on AC and AC2 variables. The second part of the set (AC2 variables) was created by adjusting AC variables by subtracting from them errors, which VAR could have fixed if it had permission to do so. The total difference between AC and AC2 variables was

Table 10.10: Results from total errors responds simple models

regression model	*referbc* Poisson	*referac* NB2	*referac2* NB2
β	0.81	-0.22	-0.50
$\exp\beta$	2.24	0.80	0.61
VAR pres. assoc.	124%	-	-
VAR total assoc.	-	-20%	-39%
SE	0.14	0.19	0.20
z statistic	5.60	-1.14	-2.44
p value	$<\|0.01\|$	0.26	0.01
sign.	~100%	<90%	>95%

Table 10.11: Results from goal errors responds simple Poisson models

regression model	*referbcg* Poisson	*referacg* Poisson	*referacg2* Poisson
β	0.52	-0.91	-1.14
$\exp\beta$	1.69	0.40	0.32
VAR pres. assoc.	69%	-	-
VAR total assoc.	-	-60%	-68%
SE	0.26	0.38	0.41
z statistic	1.99	-2.39	-2.75
p value	0.05	0.02	0.01
sign.	>95%	>95%	>99%

Table 10.12: Results from penalty and red card errors responds simple Poisson models

regression model	*referbcp* Poisson	*referbcrc* Poisson	*referacp* Poisson	*referacrc* Poisson	*referacrc2* Poisson
β	0.85	1.07	-0.34	0.50	-0.20
$\exp\beta$	2.34	2.90	0.71	1.64	0.82
VAR pres. assoc.	134%	190%	-	-	-
VAR total assoc.	-	-	-29%	64%	-18%
SE	0.21	0.31	0.27	0.33	0.39
z statistic	4.03	3.49	-1.26	1.50	-0.51
p value	$<\|0.01\|$	$<\|0.01\|$	0.21	0.14	0.61
sign.	~100%	~100%	<90%	<90%	<90%

equal to 15. On the majority of the difference participated red cards reasons. Penalty reasons did not participated at all, i.e., variables *referacp* and *referacp2* are equal in all observations. Thus, we included only one of these responses in the tables. The results from the regressions are illustrated on the same list of tables as BC errors. The cleaned impact of VAR, i.e., the impact of VAR interventions, is apparent, if we compare the coefficients of *VAR* predictor in relation to AC and BC errors. Those for AC errors are generally lower. However, the stated hypotheses also regarded the significance of the coefficients. Variable *VAR* performed as significant for both AC and AC2 version of a variable only in the case of goal errors. The relationship between *VAR* and *referbcg* was negative and significant at 95% level (for *referacg2* even at 99%). The 95% level was assumed in hypothesis **H19** for both variables. Therefore, we cannot reject it. The presence and following interventions of VAR were thus associated with a 60% decrease in goal errors (68% for the adjusted version). Nevertheless, the coefficients for other particular types of errors were not significant though. For red card errors, the coefficient of *VAR* was even positive in relation to the non-adjusted version of the regressand. However, 64% increase due to VAR presence and interventions was not significant. If VAR had permission to fix the second yellow card reasons, β coefficient of the variable of our interest would have improved to be negative (with an associated 18% decrease in *referacrc2*), but not sufficiently to be significant. Therefore, we cannot reject hypothesis **H18**. Hypothesis **H17** cannot be rejected either, because despite the presence and interventions of VAR were associated with a 29% decrease in penalty errors, the coefficient was not significant. We end this paragraph with the hypothesis **H16**, which we can reject just partly. *VAR* was indeed not significant in relationship with *referac*, as 20% decrease in *referac* associated with *VAR* brought p value larger than 0.1. However, it became significant when we subtracted from the dependent variable 15 errors that VAR could have fixed if it had permission to do so, i.e., β coefficient of *VAR* was negative and significant at 95% level in relation to variable *referac2*, and the associated decrease shifted to 39%.

10.4.2 Multiple models

From the reasons mentioned in Subsection 9.4.2, we present a multiple model just for the dependent variable *referbc*. The selection of controls was described in the same subsection. In Subsection 9.4.3, we assumed that β coefficient of *VAR* will be positive and significant at 90% level in relation to *referbc* keeping the controls fixed (**H15**). Nevertheless, firstly, we dealt with possible issues of excess of zeros (the proportion of zeros in *referbc* was higher than 0.3) and overdispersion or underdispersion. After the examination of CT test, parameter α was not significantly different from zero, i.e., equi-dispersion assumption held. Therefore, the choice tightened between Poisson regression model and ZIP model. Based on AIC of these models, we decided for ZIP model.

Table 10.13: Results from *referbc* multiple ZIP model

predictor	dist.	β	$\exp\beta$	SE	z statistic	p value	sign.
VAR	CNT	0.78	2.18	0.16	4.97	<|0.01|	~100%
pen	CNT	0.48	1.62	0.09	5.16	<|0.01|	~100%
adjgoals	CNT	0.06	1.06	0.04	1.44	0.15	<90%
scornerssq	CNT	-0.03	0.97	0.08	-0.40	0.69	<90%
cards	CNT	0.09	1.10	0.03	3.61	<|0.01|	~100%
phase	CNT	-0.08	0.92	0.05	-1.61	0.11	<90%
VAR	BIN	136	-	113	1.20	0.23	<90%
pen	BIN	-226	-	305	-0.74	0.46	<90%
adjgoals	BIN	-26.2	-	20.6	-1.28	0.20	<90%
scornerssq	BIN	126	-	97.7	1.29	0.20	<90%
cards	BIN	-18.7	-	14.8	-1.27	0.21	<90%
phase	BIN	-115	-	89.3	-1.29	0.20	<90%

The results from the regression can be seen in Table 10.13. As we were working with ZIP model, we were forced to standardize the control variable *cornerssq*. The new variable took *scornerssq* name. We will again focus mainly on the results from the count distribution part (*CNT* in table). The coefficient of variable *VAR* remained in the multiple positive and highly significant. The increase associated with our predictor of interest in the total number of errors before possible consultation with VAR declined from 124% to 118%, keeping the selected set of control variables fixed. The significance even exceeded the expectations in hypothesis **H15**, which could not have been rejected. Regarding other predictors, as significant ones proved themselves variables *pen* and *cards*. Both coefficients were positive and highly significant. Adj R^2 of the linear version of the model was equal to 0.13 and AIC of Poisson regression model to 951. In the discussion subsection, we will come back to the relationship between *VAR* and the responses counting for errors of on-pitch referees. We compare the results from the simple models, and the multiple model with the related studies of Samuel *et al.* (2020) and Holder *et al.* (2022)

10.4.3 Discussion

Before we conclude the thesis in the next chapter, we will discuss the results from errors regressions. As the coefficients of VAR were positive and statistically significant in all five regressions regarding BC errors, we can support the studies of Samuel *et al.* (2020) and Holder *et al.* (2022), which suggested that the presence of VAR influences decision-making behavior of on-pitch referees the way they make more mistakes in match-changing incidents in matches, where VAR is present. From the simple regressions, the presence of VAR was associated with a 69% increase in goal errors, 134% increase in penalty errors, and 190% increase in red card errors. Furthermore, we reckoned 124% growth in total errors in the simple model and 118% in the multiple model. The significance of the coefficient from the multiple model was, for us, quite unexpected (we stated the hypothesis on 90% level). That is why we investigated the percentual association and the significance again for the baseline Poisson regression model. Nevertheless, as the percentual increase was still equal to 82% and the coefficient was significant at 99% level, we can conclude that keeping the selected set of controls fixed, we can still reckon a significantly higher number of on-pitch referees' errors in matches with VAR.

We were also interested in potential causes of the issue of errors of pitch-based referees. Several paragraphs regarding possible causes were included to Section 5.2. We discussed several topics related to the state of stress and worries, including anxiety, competence, authority, pressure, and self-confidence. We also put forward several studies and assumed how these concepts might be interconnected with the presence of VAR. Nevertheless, as it was said, we could not certainly conclude which of these factors can contribute because most of the studies regard these issues in football generally. Just Samuel *et al.* (2020) studied pressure and self-confidence in relation to VAR in the form of a questionnaire. However, we discussed the limitations of this study (Samuel *et al.* (2020)). Moreover, we admit that other concepts might take part in the issue. Since we do not dispose of the capacity to measure the blood pressure of on-pitch referees during matches, we can only assume causes of the phenomenon (Gasperin *et al.* (2009)). Therefore, eventually, we aimed mainly to reveal, whether there is another piece of evidence of the positive relationship between the presence of VAR and errors of on-pitch referees in F:L and leave possible causes and their interconnection with VAR for further research.

The aim of the second part of the set of errors variables was to express how VAR presence and its interventions are associated with numbers of errors and how they might be if the competence of VAR widened. Moreover, from these associations and the associations of BC responses, we can extract the percentages for VAR interventions solely. Previously in the thesis, we called this phenomenon the cleaned impact of VAR. The cleaned impact of VAR took part in a slightly different form also in the studies of and KU Leuven, and Spitz *et al.* (2020). The researchers proved the cleaned impact of VAR to be positive, i.e., a lower number of errors remains after VAR interventions. Also, in F:L matches, we were able to detect that the cleaned impact of VAR resulted in a lower number of mistakes in all categories of errors.

Nevertheless, in our opinion, more important numbers consist in the coefficients of *VAR* in relation to AC and AC2 errors, which interpretation gives us how VAR presence and VAR interventions together are associated with errors variables, i.e., which associations creates VAR have as the whole. Although, all the coefficients of *VAR* except *referacrc* regression were negative, just the coefficients of *referacg*, *referacg2* and *referac2* were significant. The most goal reviews in our dataset were due to offside reasons, i.e., factual decisions. On the other hand, the majority of penalty and red card reviews belonged to the category of subjective decisions. Therefore, F:L data suggested that VAR as the whole improved the outcome regarding factual decisions, which projected to goal decisions. Even for BC variables regarding goals, the associated increase was not as high as for other errors. On the other hand, our data also suggested that VAR as the whole had not significantly improved the outcome regarding subjective decisions, which projected to penalty, red card, and eventually total decisions. In the case of total errors, there was a 20% decrease associated with VAR as the whole, for penalty errors 29% decrease and for red card errors even 64% increase, which would have changed to 18% decrease, if VAR had permission to intervene the second yellow card decisions and if had evaluated all of them correctly. Nevertheless, none of these percentages were significant.

Data suggested that one of the possible participating reasons can be found in a higher associated increase in BC errors because the percentages of subsequent VAR interventions were reaching similar values as for goal errors. Except for the possible causes of this phenomenon, which we discussed earlier, we can also consider a particular inconsistency of KR FAČR in their reports. This idea is based on the possibility that the approach of the committee towards VAR as a new project might have been overtried. In other words, matches where VAR was present might have been devoted greater attention than other matches. And the greater attention might have resulted in positive biases in BC variables if $VAR = 1$ (especially in the case of subjective decisions). Nevertheless, as we do not dispose of the capacity to evaluate the reports of the committee in that way, we supposed that the committee approached all the matches the same as we did in Subsection 10.2.3.

Chapter 11

Conclusion

In the last chapter of the thesis, we summarize the outcomes of our research, their consequences related to the studies mentioned in Chapter 5 and their possible limitations. Firstly, we were investigating the relationship between the presence and interventions of VAR and several match-changing incidents, i.e., yellow cards, red cards, and penalties. In the case of yellow cards, β coefficient of the variable of our interest, VAR, did not perform as significant in the multiple model, where we kept the designed set of control variables fixed. Therefore, we did not support the assumption of Carlos *et al.* (2019), who suggested that a negative relationship between ycards and VAR might be due to players being aware of VAR during the game and thus behaving more cautious because of the fear from sending-off. On the other hand, we mentioned the study of Aksum *et al.* (2020), which revealed a number of various not-referee incentives that a football player focuses on during the match.

For both red cards and penalty kicks, we created two predictors. Both were counting for a number of red cards (respectively penalties) in the match. However, one of them was adjusted by the red cards (respectively penalties) that were awarded due to or canceled by VAR. Thus, we could have expressed statistical associations of VAR presence, VAR interventions, and these two together. In red cards and penalties regressions, we aimed on the suggestion of Holder *et al.* (2022) that VAR influences the decision-making behavior of on-pitch referees. We decided to possibly support this suggestion through a significant difference in the interpretation between VAR coefficient of not-adjusted regressand and adjusted regressand, e.g., pen and pen2. Eventually, we could have supported the suggestion of Holder *et al.* (2022) partly. In the multiple models of rcards and rcards2, none of β coefficients of VAR was significant, and the difference between the interpretations of the coefficients was lower than 30%. Therefore, we could not have supported that VAR influences the decision-making behavior of on-pitch referees in red card decisions. On the other hand, we could have supported the same assumption, but for penalty kicks, because the difference between the interpretations of coefficients of VAR was equal to 45% for this issue, i.e., VAR interventions were associated with a 45% increase in the number of penalties, which exceeded 25% threshold from the study of Holder *et al.* (2022).

Nevertheless, we are aware that even though we controlled for several factors, especially the results for pen and pen2 were still shifted in comparison to the study of Holder *et al.* (2022), where the coefficient of VAR went from negative to close to zero between the adjusted response and non-adjusted response (ours went from close to zero to positive). In our opinion, there still might be unexplained positive bias in the number of penalties due to the specifics of our dataset, which we went through in Subsection 8.5.4 This assumption was also supported by low Adj R^2 of linear versions of penalties models. We suppose that if we were able to mild the bias, the results would have shifted closer to the study of Holder *et al.* (2022). We also concede that the bias might not be the same for both adjusted and non-adjusted versions of the dependent variable, and therefore, eventually, the percentage may decrease. We conclude that the results from penalties regression supported the related study, however further studies might develop this issue hereafter.

Secondly, we were investigating the presence and interventions of VAR in relation to errors of on-pitch referees. For this sake, we created the set of twelve dependent variables divided based on until which part of VAR procedure an error survived and which type of game-changing situation an error was regarded. The presence of VAR showed a positive and significant relationship with all types of errors, i.e., on-pitch referees made in the case of our dataset significantly more mistakes in matches where VAR was present. As this held even in the multiple model (for counts of total errors), we could have supported the studies of Samuel *et al.* (2020) and Holder *et al.* (2022). In the multiple model, the presence of VAR was associated with a 118% increase in referbc. However, subsequently, the statistical impact of VAR interventions, i.e., the cleaned impact of VAR, in part proved itself as it was negative in all types of mistakes. Nevertheless, we were also interested in the statistical impact of VAR as the whole and how this impact would change if we widened the competence of VAR. The presence and interventions of VAR together showed a negative and significant association with errors variables only in the case of goal errors and total errors with the broader competence of VAR. We concluded that VAR as the whole improved the outcome only for factual decisions, but not for subjective decisions. As one of the possible reasons, we mentioned higher percentual associations of VAR presence for errors based on subjective decisions. We discussed several possible causes, including several psychical states of the referee and their connection to the presence of VAR on the match or the inconsistency of KR FAČR. Nevertheless, eventually, we left a study of possible causes for further research.

One of the general limitations of the research is, in our opinion, the structure of the dataset in comparison to the related studies. Since the structure of our dataset was partly enforced by in which year the thesis was written, we suppose that further studies on the topic of VAR in F:L can compare whole seasons with and whole seasons without VAR. Especially for penalties and errors of on-pitch referees, where the difference between simple and multiple models was lower, such studies might bring partly different results. Our study might serve as a starting point for researchers who aim to extend this topic.

Bibliography

AKSUM, K. M., L. MAGNAGUAGNO, C. T. BJORNDAL, & G. JORDET (2020): "What do football players look at? an eye-tracking analysis of the visual fixations of players in 11 v 11 elite football match play." *Frontiers in Psychology* **11**.

ATSAN, N. (2016): "Decision-making under stress and its implications for managerial decision-making: A review of literature." *International Journal of Business and Social Research* **6**: p. 38.

BARBIERO, A. (2019): "Alternative count regression models for modeling football outcomes." In "MathSport International," pp. 16–24. Athens University of Economics and Business.

VAN DEN BERG, L. & J. SURUJLAL (2020): "Video assistant referee: Spectator and fan perceptions and experiences." **12**: pp. 1309–8063.

CAHOY, D., E. DI NARDO, & F. POLITO (2020): "Flexible models for overdispersed and underdispersed count data." .

CAMERON, A. & P. K. TRIVEDI (1990): "Regression-based tests for overdispersion in the poisson model." *Journal of Econometrics* **46(3)**: pp. 347–364.

CAMERON, A. & F. WINDMEIJER (1996): "R-squared measures for count data regression models with applications to healthcare utilization." *Journal of Business Economic Statistics - J BUS ECON STAT* **14**: pp. 209–220.

CAMERON, A. C. & P. K. TRIVEDI (2013): *Regression Analysis of Count Data*. Econometric Society Monographs. Cambridge University Press, 2 edition.

CARLOS, L.-P., R. EZEQUIEL, & K. ANTON (2019): "How does video assistant referee (var) modify the game in elite soccer?" *International Journal of Performance Analysis in Sport* **19(4)**: pp. 646–653.

DIXON, M. J. & S. G. COLES (1997): "Modelling association football scores and inefficiencies in the football betting market." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **46(2)**: pp. 265–280.

DOBSON, S., P. DAWSON, J. GODDARD, & J. WILSON (2007): "Are football referees really biased and inconsistent?: Evidence on the incidence of disciplinary sanction in the english premier league." *Journal of the Royal Statistical Society Series A* **170**: pp. 231–250.

ERIKSTAD, M. K. & B. T. JOHANSEN (2020): "Referee bias in professional football: Favoritism toward successful teams in potential penalty situations." *Frontiers in Sports and Active Living* **2**.

GALANAKIS, M., A. PALAIOLOGOU, G. PATSI, I.-M. VELEGRAKI, & C. DARVIRI (2016): "A literature review on the connection between stress and self-esteem." *Psychology* **07**: pp. 687–694.

GALLARDO, M., M. BETANCOR, & A. BENÍTEZ (2019): *The Use of Video Technologies in Refereeing Football and Other Sports.*

GASPERIN, D., G. NETUVELI, J. DIAS DA COSTA, & M. PATTUSSI (2009): "Effect of psychological stress on blood pressure increase: A meta-analysis of cohort studies." *Cad. Sade Publica* **25**.

GÜRLER, C. & V. POLAT (2021): "Video assistant referee's effect on football: Turkish super league case." *Revista Brasileira de Educacao* **13**: pp. 118–124.

HARRIS, T., Z. YANG, & J. W. HARDIN (2012): "Modeling underdispersed count data with generalized poisson regression." *The Stata Journal* **12(4)**: pp. 736–747.

HOLDER, U., T. EHRMANN, & A. KÖNIG (2022): "Monitoring experts: insights from the introduction of video assistant referee (var) in elite football." *Journal of Business Economics* **92**.

HU, M.-C., M. PAVLICOVA, & E. NUNES (2011): "Zero-inflated and hurdle models of count data with extra zeros: Examples from an hiv-risk reduction intervention trial." *The American journal of drug and alcohol abuse* **37**: pp. 367–75.

JAMES, G., D. WITTEN, T. HASTIE, & R. TIBSHIRANI (2013): *An Introduction to Statistical Learning: with Applications in R.* Springer.

Johansen, B. & T. Haugen (2013): "Anxiety level and decision-making among norwegian top-class soccer referees." *International Journal of Sport and Exercise Psychology* **11**: pp. 215–226.

Jones, M., G. Paull, & J. Erskine (2003): "The impact of a team's aggressive reputation on the decisions of association football referees." *Journal of sports sciences* **20**: pp. 991–1000.

Kasyoki, A. (2016): "Statistical models for count data." *Science Journal of Applied Mathematics and Statistics* **4**: p. 256.

Kleinert, J., J. Ohlert, B. Carron, M. Eys, D. Feltz, C. Harwood, L. Linz, R. Seiler, & M. Sulprizio (2012): "Group dynamics in sports: An overview and recommendations on diagnostic and intervention." *Sport Psychologist* **26**: pp. 412–434.

Klemp, M., F. Wunderlich, & D. Memmert (2021): "In-play forecasting in football using event and positional data." *Scientific Reports* **11**.

Klír, J. (2019): "Moderní technologie ve fotbale a jejich využití z pozice rozhodĂŤích [online]." SUPERVISOR : Oldřich Racek.

Kolbinger, O. & M. Lames (2017): "Scientific approaches to technological officiating aids in game sports." *Current Issues in Sport Science (CISS)* **2**.

Lago-Peñas, C., M. Gómez, & R. Pollard (2020): "The effect of the video assistant referee on referee's decisions in the spanish laliga." *International Journal of Sports Science &amp Coaching* **16(3)**: pp. 824–829.

Leveaux, R. (2010): "Facilitating referee's decision making in sport via the application of technology." *Communications of the IBIMA* .

Lidén, J. (2016): "Bivariate models to predict football results." .

MacMahon, C. (2015): *Sports officials and officiating: Science and practice.* Routledge.

Mohri, M. & B. Roark (2022): "Structural zeros versus sampling zeros." pp. 1–7.

Murray, M. & R. Howitt (2019): *Sport science.* Hodder Education.

Philippe, F. L., R. J. Vallerand, J. Andrianarisoa, & P. Brunel (2009): "Passion in referees: Examining their affective and cognitive experiences in sport situations." *Journal of Sport and Exercise Psychology* **31(1)**: pp. 77 – 96.

Phillips, P. & S. Fairley (2014): "Umpiring." *Journal of Leisure Research* **46(2)**: pp. 184–202.

Picazo-Tadeo, A. J., F. González-Gómez, & J. Guardiola (2017): "Does the crowd matter in refereeing decisions? evidence from spanish soccer." *International Journal of Sport and Exercise Psychology* **15(5)**: pp. 447–459.

Raffalovich, L. & R. Chung (2015): "Models for pooled time-series cross-section data." *International Journal of Conflict and Violence* **8**: pp. 209–221.

Roback, P. & J. Legler (2021): *Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R.*

Samuel, R. D., Y. Galily, E. Filho, & G. Tenenbaum (2020): "Implementation of the video assistant referee (var) as a career change-event: The israeli premier league case study." *Frontiers in Psychology* **11**.

Samuel, R. D., Y. Galily, & G. Tenenbaum (2017): "Who are you, ref? defining the soccer referee's career using a change-based perspective." *International Journal of Sport and Exercise Psychology* **15(2)**: pp. 118–130.

Scarf, P. (2017): "Football scores, the poisson distribution and 30 years of final year projects in mathematics, statistics and operational research." *MSOR Connections* **15**: p. 61.

Slack, L. A., I. W. Maynard, J. Butt, & P. Olusoga (2013): "Factors underpinning football officiating excellence: Perceptions of english premier league referees." *Journal of Applied Sport Psychology* **25(3)**: pp. 298–315.

Soares, J. & L. Shamir (2016): "Quantitative analysis of penalty kicks and yellow card referee decisions in soccer." *American Journal of Sports Science* **4**: p. 84.

Soriano Gillué, G., Y. Ramis, M. Torregrossa, & J. Cruz (2018): "Sources of stress inside and outside the match in football referees." *Apunts. Educacion Fisica y Deportes* pp. 22–31.

Spitz, J., J. Wagemans, D. Memmert, A. Williams, & W. Helsen (2020): "Video assistant referees (var): The impact of technology on decision making in association football referees." *Journal of Sports Sciences* **39**: pp. 1–7.

Stephenson, M. D., B. Schram, E. F. D. Canetti, & R. Orr (2022): "Effects of acute stress on psychophysiology in armed tactical occupations: A narrative review." *International Journal of Environmental Research and Public Health* **19(3)**.

Ver Hoef, J. & P. Boveng (2007): "Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data?" *Ecology* **88**: pp. 2766–72.

Winkelmann, R. (2015): "Counting on count data models." *IZA World of Labor* p. 148.

Wooldridge, J. M. (2009): *Introductory Econometrics: A Modern Approach.* ISE - International Student Edition. South-Western.

van der Wurp, H., A. Groll, T. Kneib, G. Marra, & R. Radice (2019): "Generalised joint regression for count data with a focus on modelling football matches." .

Appendix A

Information about used control variables

A.1 Used R Controls

refyc - average number of yellow cards per referee in the last season; taken from Fortuna Liga
refrc - average number of red cards per referee in the last season; taken from Fortuna Liga
refpen - average number of penalties per referee in the last season; taken from Fortuna Liga
expref - number of guided matches in F:L in the career of referee; taken from Fortuna Liga
eurref - appearence of referee in European competitions (yes or no); taken from Transfermarkt
byeweek - presence of bye week for the referee (yes or no); taken from Fortuna Liga

A.2 Used T Controls

teambattsea - average number of battles per teams in the last season; taken from Fortuna Liga
teamycsea - average number of yellow cards per teams in the last season; taken from Fortuna Liga
teamrcsea - average number of red cards per teams in the last season; taken

from Fortuna Liga
teampensea - average number of penalties per teams in the last season; taken from Fortuna Liga
teamptbsea - average number of passes to box per teams in the last season; taken from Fortuna Liga
farintab - standardized distance between teams in the table before the match; taken from Fortuna Liga
group - part of the season (primary or superstructure); taken from Livesport
primetime - hour of kick-off (primetime or not); taken from Livesport
rivalry - presence of rival relationship between teams (yes or no); taken from Wikipedia
sucteam - presence of successful club (yes or no); taken from Fortuna Liga
corona - presence of capacity restrictions (yes or no); taken from Fortuna Liga
temp - season (spring, summer, fall or winter); taken from Livesport

A.3 Used M Controls

rcards - number of red cards in the match; taken from Livesport
ycards - number of yellow cards in the match; taken from Livesport
cards - number of cards in the match; taken from Livesport
pen - number of penalties in the match; taken from Livesport
adjgoals - number of goals without scored penalties in the match; taken from Livesport
fouls - number of fouls in the match; taken from Fortuna Liga
battles - number of defensive battles in the match; taken from Fortuna Liga
shotsongoal - number of shots on goal in the mach; taken from Livesport
attacks - number of attacks in the mach; taken from Total Corner
corners - number of corners in the mach; taken from Livesport
dribbles - number of dribbling moves in the mach; taken from Fortuna Liga
ptg - number of passes to goal in the mach; taken from Fortuna Liga
goalsdiff - difference between home and away goals in the match; taken from Livesport
zerohalf - 0:0 result in the first half (yes or no); taken from Livesport
ispen - presence of penalty in the match (yes or no); taken from Livesport
timeoffirstyc - minute of the first yellow card awarded; taken from Livesport

Appendix B

Additional sources

In this part of the Appendix, alternative sources to academic research used in text will be listed.

1) *A brief history of technology in sport.* Shira Springer, 2012.Available from `https://bit.ly/3y3RRCl`.

2) *History of instant replay.* NFL. Available from `https://bit.ly/3LAWYOu`.

3) *Hawk-Eye in Tennis.* Hawk-Eye Innovations, 2015. Available from `https://bit.ly/3ybSkCm`.

4) *Goal-line Technology.* FIFA. Available from `https://fifa.fans/3vUEf9R`.

5) *A Brief History (And Defense) of VAR.* Matthew Farrell, 2019, Duke University. Available from Source link.

6) *Video Assistant Referee (VAR) Protocol.* IFAB. Available from `https://www.theifab.com/laws/latest/video-assistant-referee-var-protocol/`.

7) *State of the football analytics industry in 2021.* Sci Sports, 2021. Available from Source link.

8) *Livesport slaví jubileum. Udržel tržby, má sto milionů uživatelů a mění tvář.* Robert Sattler, 2021, Forbes. Available from `https://bit.ly/3EZXy5T`.

9) *Top 22 Worst Refereeing Decisions in World Football History.* Vijay Murali, 2011, Bleacher Report. Available from `https://bit.ly/3vuV5ND`.

10) *Experiments with Video Assistant Referees (VARs) enter next stage.* IFAB, 2017. Available from `https://bit.ly/3s1NzaR`.

11) *YouGov - VAR.* YouGov, 2020. Available from `https://bit.ly/3OPk9X6`.

12) *YouGov Results - VAR August 2020.* YouGov, 2020. Available from `https://bit.ly/3F27sDM`.

13) *Laws of the Game 21/22.* IFAB, 2021. Available from `https://bit.ly/3F2aDvc`.

12) *Implementation Assistance and Approval Programme for VAR technology (IAAP).* FIFA. Available from `https://fifa.fans/3Ky0APU`.

13) *VAR IAAP-Technology.* FIFA. Available from `https://fifa.fans/3MYHHaz`.

14) *Refereeing 2.0.* KNVB. Available from `https://www.knvb.com/themes/new-laws-of-the-game/refereeing-2.0`.

15) *The inside story of how FIFA's controversial VAR system was born.* Joao Medeiros, 2018, Wired. Available from `https://www.wired.co.uk/article/var-football-world-cup`.

16) *Information on the Video Assistant Referee (VAR) experiment, incl. provisional results.* IFAB, 132th annual business meeting, January 2018. Available from`https://bit.ly/3kveHuw`.

17) *Videorozhodčí.* LFA. Available from Source link.

18) *O nás.* LFA. Available from `https://www.lfafotbal.cz/o-nas`.

19) *The pressure and coping of football referees: A sports psychology angle.* Victor Thompson, 2018, Sports Psychologist. Available from https://bit.ly/3ML7jrk.

20) *Stress.* Cambridge Dictionary. Available from https://dictionary.cambridge.org/dictionary/english/stress.

21) *More authority means less stress, say Stanford and Harvard psychologists.* Max McClure, 2012, Stanford University. Available from https://stanford.io/376zdPg.

22) *Live and Clicking: The Big Data Relationship Between Sport and Betting.* Stuart Barraclough, 2020, IE Business School, Madrid. Available from https://bit.ly/3KxC1Tb.

23) *Worldwide registration periods calendar.* FIFA. Available from https://fifa.fans/3ORAudO.

24) *Count Data Models.* Bauer College of Business, University of Houston. Available from https://www.bauer.uh.edu/rsusmel/phd/ec1-22.pdf.

25) *Linear Regression.* Econometrics Academy. Available from https://bit.ly/3vVW6ND.

26) *Football data analysis:An example with the Countr package.* Tarak Kharrat et al, 2019, Cran R Project. Available from https://bit.ly/3vzFs7K.

27) *The Poisson Regression Model.* Faculteit Economie en Bedrijfswetenschappen, KU Leuven. Available from https://bit.ly/3LBPcnh.

28) *Poisson Regression Analysis using SPSS Statistics.* Laerd Statistics. Available from https://bit.ly/3OI7c1g.

29) *The importance of squad stability:evidence from European football.* Raffaele Poli et al, 2018, CIES Football Observatory Monthly Report. Available from https://www.football-observatory.com/IMG/sites/mr/mr34/en/.

30) *dispersiontest: Dispersion Test.* R Package Documentation. Available from `https://rdrr.io/cran/AER/man/dispersiontest.html`.

31) *What is the difference between zero-inflated and hurdle models?.* Stack Exchange. Available from `https://bit.ly/3yduOVA`.

32) *Models for excess zeros using pscl package (Hurdle and zero-inflated regression models) and their interpretations.* R Studio Pubs Static. Available from `https://bit.ly/3s4P3Rv`.

33) *Dealing with quasi- models in R.* Ben Bolker, 2021, Cran R Project. Available from `https://cran.r-project.org/web/packages/bbmle/vignettes/quasi.pdf`.

34) *Residuals vs. Fits Plot.* STAT 462, Penn State. Available from `https://online.stat.psu.edu/stat462/node/117/`.

35) *R-squared, Adjusted R-squared and Pseudo-R-squared.* Time Series Analysis, Regression and Forecasting. Available from `https://bit.ly/3y4JwhE`.

36) *Identifying Outliers in Linear Regression-Cookâ€™s Distance.* Christian Thieme, 2021, Towards Data Science. Available from `https://bit.ly/3y5Vg3u`.

37) *Backward Selection-a way to final model.* Changxia Shao, 2019, PharmaSUG. Available from `https://bit.ly/3kvnzAE`.

38) *Understand Forward and Backward Stepwise Regression.* Quantifying Health. Available from `https://quantifyinghealth.com/stepwise-selection/`.

39) *Variance Inflation Factor (VIF).* Investopedia. Available from `https://www.investopedia.com/terms/v/variance-inflation-factor.asp`.

40) *Correlation Coefficient.* Investopedia. Available from `https://www.investopedia.com/terms/c/correlationcoefficient.asp`.

41) *Akaike Information Criterion | When How to Use It*. Rebecca Bevans, 2020, Scribbr. Available from https://www.scribbr.com/statistics/akaike-information-criterion/.

42) *R-Squared vs. Adjusted R-Squared: What's the Difference?*. Investopedia. Available from https://bit.ly/37UMiLK.

43) *List of association football club rivalries in Europe*. Wikipedia. Available from https://bit.ly/3w370Bs.

44) *The Effects of the Red Cards in the Knockout Competitions: An Analysis of the UEFA Europa League*. Fernando Lera Lopez et al, 2019. Available from https://bit.ly/3vV3bhl.

45) *Backward Selection-a way to final model*. Changxia Shao, 2019, PharmaSUG. Available from https://www.lexjansen.com/pharmasug-cn/2019/PO/Pharmasug-China-2019-PO77.pdf.

46) *Global study of penalty cards in professional football*. Raffaele Poli et al, 2020, CIES Football Observatory Monthly Report. Available from https://football-observatory.com/IMG/sites/mr/mr57/en/.

47) *Massive Research of Penalties by InStat*. InStat. Available from https://instatsport.com/football/article/penalty_research.

48) *How significant are penalties in European football?*. Antionio Duran, 2020, Driblab. Available from https://bit.ly/3F4HjV9.

49) *How offsides are determined by VAR*. Premier League, 2020. Available from https://www.premierleague.com/news/1488423.

50) *Interpret Poisson Regression Coefficients*. Quantifying Health. Available from https://bit.ly/37SKfrR.

51) *zeroinfl "system is computationally singular" whereas no correlation in predictors*. Stack Overflow. Available from https://bit.ly/3s1VULt.

52) *When and why to standardize a variable.* Deepanshu Bhalla, Listen Data. Available from `https://bit.ly/3vZzIDa`.