We use a Latent moderated structural equation approach to assess the relationships between our variables within our model. Structural equation models consist of two main components, a structural component, where the relationships between variables can be examined. Like in our case the relationship between crowd support, referee bias and team performance. The measurement part exists to measure latent structures in the data, these latent structures are able to capture complex variable and measures, which would be hard to define to a single variable or metric. The measurement component is usually conducted through a confirmatory or exploratory factor analysis.

Game outcomes are not the only measure of team performance we look at. Soccer is a low scoring game in a sense that a single goal or single action can decide the entire outcome of a match. Luck with a team their only shot deflected in the goal and the opponent missing big chances can mean an outcome not congruent with how the match developed. To delve deeper into we use several metrics to quantify team performance on a deeper level than purely outcome based. Examples could be Expected Goals, a number that symbolizes the amount of goals normally would have been scored based on the amount and the quality of chances created by the team, thus removing luck and misfortune out of the analysis.

To analyze the role of referee bias on team performance we first need to examine whether crowd support is a significant predictor of referee bias. Our construct of referee bias is a factor score for multiple metrics of referee decision making, such as number of fouls and number of yellow and red cards. We also deem the spread(difference) between number of fouls, yellow cards and red cards between the home and away team to be relevant. Such that this spread would expected to be 0 when controlling for factors such as team strength but if a referee has a certain bias towards teams this number could be positive or negative, where a positive(negative) number implies a referee on average gives more(less) fouls;yellow or red cards to the home team. Indicating a bias against(in favour) of the home team. We believe that the factor analysis of several measures of referee decisions would be a valid strategy to integrate a latent construct of referee bias into our model.

Following our factor analysis, the measurement part of our model is structural we proceed with our regression model to examine the relationship between our dependent and independent variables. Since our model comprises of both moderating and mediating variables, a moderated mediation model is required to correctly analyze the effects of the various independent variables on our dependent variable team performance. Multiple models to measure moderated mediation exist currently. Examples include Path Analysis, Product Indicator analysis and Latent moderated Structural Equations. (Feng, Song, Zhangh, Zheng and Pan, 2020) find that Latent moderated Structural Equations outperforms Path Analysis and Product Indicator analysis in all settings studied. Therefore, we will proceed with a Latent moderated Structural Equations model to estimate our model.

**SPI: S**trength metric of teams on a scale measured before the beginning of each match. The metric incorporates recent results and the strength of the teams played previously and therefore is also a measure of recent form.

**Age difference:**  this variable captures the differences in age between the home and away team, where a positive value implies the home team have an older squad on average than the away team, and a negative value implies the away team have older squad players.

**Importance of match:** This variable is calculated by … and …. and represents the importance of the match outcome on the league developments for both teams their perspective. There are 2 values, one value for the away team and one value for the home team

**Covid:** a dummy variable that indicates whether or not the game was played before or after the start of the covid-19 pandemic, with a 1 representing post-covid matches.

**Crowd occupancy:** The ratio of the number of attendants to the number of seats in the stadium. Due to data collection and computation limits, the crowd occupancy is calculated as the total number of attendants within a league divided by the average capacity for all teams in the league. For example a league where the clubs on average have a stadium capacity of 50.000 with 4.000.000 fans in total visiting 200 matches means an average attendance of ((4000000/200))/50.000 = 0.4.

**Expected goals:** Expected goals are measured as the expected goals that would have been made from a teams chances within a match, based on average player and average situation. It is a sum of the quality of chances created by a team. A shot for open goal with no opponent player in tackling distance has a very high xg chance since it will be converted into a goal almost all the time. A 40 yard shot that gets deflected by a defender on the way and flies in with luck has a low xg value, as such shots have a low probability of producing a goal. Expected goals metric thus takes out luck out of the performance of a team and can be seen as a good metric of match performance and match outcomes.

**Spread in fouls:** The number of fouls committed by the home team - the number of fouls committed by the away team. A positive number means the home team has committed more fouls than the away team in the specific match, and a negative number implies the away team committed more fouls. 0 implies an equal number of fouls for both teams.

**Spread in yellow cards :** The number of yellow cards for the home team - the number of yellow cards for the away team. A negative number can be interpreted as the away team receiving more yellow cards than the home team. A positive number means the number of yellow cards received by the home team exceeds those of the away team. 0 implies an equal number of yellow cards for home and away team. This could be 0 for both but also 1 or 2 or more for each side.

**Goals:** This variable denotes the number of goals scored by teams int he particular match.

Some teams always play well away from home while having difficulties at home. An example is Ajax in European competitions over the last few seasons, being unbeaten away from home for nearly 2 years while playing formidable sides, but barely beating any side in their own Johan-Cruijff Arena. Sources of these difference between could stern from tactics, playing style and playing mentality.

The manifestation of the Bosman rule in 1995, guaranteeing a free transfer for players who are out of contract had an immense impact on players relationships with clubs. Players more than ever are able to easily switch between clubs, dramatically decreasing the average tenure for football players at their clubs.

We therefore hypothesize that teams with high shares of foreign players will have less difference in performance between home and away games, both to a lesser innate bond with the club and the decreased support of fans for the clubs they don’t recognize anymore.

Table 3 variable operationalization

|  |  |  |
| --- | --- | --- |
| **Variable** | **Type** | **Explanation** |
| Covid | Nominal | 1 = during covid pandemic, 0 = before |
| Yellow card spread | Interval | 0 = equal number of yellow cars for home and away |
| Foul Spread | Interval | 0 = equal number of Fouls committed both teams |
| Percentage of points | Ratio | 0 = no points 1 = only home wins |
| Expected goals | Ratio | 0 = no chances created |
| Goals | Interval | 0 = no goals 1 = 1 goal etc. |
| Proportion Foreigners | Ratio | 0 = 0 percent 1 = 100 percent |
| Age difference | Ratio | 0 = no age difference |
| Occupancy rate | Ratio | 1 = sold out stadium, 0 = no spectators |
| SPI | Ratio | 0 = no strength , 100 = strongest possible |
| Importance | Ratio | 0 = non important 100 = extremely important |
| Probability | Ratio | 0 = no chance of win 1 = 100 percent change |

The findings on the significance of travelling for home advantage are more ambiguous. (Pollard, da Silva & Nísio, 2008 ) find that travel distance seem to play a role albeit a minor one in conceiving of home advantage. Especially since their research is focusing on Brazil, where distances between teams can be very large. Distances that would impossibly exist in European countries. This is similar to what (Nevill&Holder) found in their research, where they found that distance only mattered when distance traveled was of such magnitude that different time zones would be crossed. (Goddard, 2006; Clarke & Norman, 1995) however, do seem to find a significant effect for travel distance on away team performance even with reasonably small distances between 0-200 miles.

This website provides a multitude of interesting statistics on numerous aspects of football, they provide data on player level, club level and league level. We manually collect the following variables: Firstly we take the average age of the players used by the team within the season to serve as a proxy for the age of players. Secondly, we use the share of minutes played by foreigners as a measurement for the extent of foreign influence within squads. Crowd occupancy data is gathered by using the average occupancy of each team over the course of the season. Crowd size is measured by the average attendance over the entire season per team.

n the form of the number of goals expected to score against an average opponent on neutral terrain. This offensive rating is the combined with a

Percentage points at Home

Additionally, by first performing a factor analysis, we try to use as much data as possible and combine different variables of team performance and referee bias to obtain more meaningful data on these measures. Concepts such as referee bias and team performance are difficult to capture in a single number. We use various aspects of referee decisions such as fouls, yellow cards and red cards. Team performance can be measured in numerous ways. Looking at outcome related variables such as points collected our goals scored can give insight into home advantage, however, secondary performance indicators such as shots, corners and expected goals(measure to compute quality of chances created in a match) can be valuable as well. We explore whether these multiple variables can be captured into overarching measures of team performance. Removing unnecessary indicators from the analysis in the process.

After data collection and construction of the final datasets we first perform an exploratory factor analysis to find the relevant variables for our analysis and to construct latent measures of concepts that are difficult to capture in a single metric, such as referee bias and team performance. We then apply a regression model to the data in order to draw valid conclusions on the existence of home advantage.