**CCMVV4037E**
**The Economics of Sports**
Battista Severgnini
Spring 2024

# The Relationship between Player Market Value and Team Performance in Club and National Team Football

Rasmus Clausen, Cand.merc.(mat.) - Copenhagen Business School

Number of characters: 34,026    Number of Pages: 15

**Abstract**

This paper investigates the relationship between player market values and team performances in football, comparing national teams to club teams. Using a proposed mathematical model, data from both national and club teams are analyzed, focusing on factors such as player market values, the variance of these values, and team cohesion. Structural differences between national and club football are explored to understand the variations in factors influencing team performances. Previous findings suggest that economic factors, as reflected in market values, explain 70% of club team performances, while football experts qualitatively discuss different dynamics in national team football. Further analysis will refine these conclusions and provide insights into the key factors, including case studies on countries like Norway and Croatia. **Key words:** *Football, Sports Economics, Team performance, Player market values, Econometrics, National team, Club team*

## 1 Introduction

In November 2023, the Danish national team in football secured their place for UEFA Euro 2024. Despite topping qualification Group H in front of Slovenia and Finland, the Danish media heavily criticized the performances and the tactics of coach Kasper Hjulmand, suggesting the squad underperformed given their quality. While Denmark successfully qualified for the Euros and World Cups for the fourth consecutive time, their neighboring rivals Sweden and Norway failed to qualify for the finals in Germany. Interestingly, both failed Nordic nations featured some of the world's most expensive players, such as Erling Haaland, Martin Ødegaard, and Alexander Isak, in their squads.

Especially Norway stands out as one of the most peculiar examples of underperformance despite possessing a considerable amount of talent and market value. According to Transfermarkt, the average player market value of the Norwegian team is 19.8 million euros, ranking them as the 11th most valuable national team globally. Given the assumption that player market values of national teams highly influence performances, the case of Norway may suggest a need for a new coach, as they have not qualified for a major tournament since 2000.

However, some argue that success in national team football is not solely determined by player values. Take Croatia, for example, a relatively small ex-Yugoslavian country currently ranked as the 20th most valuable national team. Despite their moderate market value, Croatia impressively finished as runners-up and third place in the last two FIFA World Cups. Football experts often attribute their success to a cohesive and experienced squad that has accumulated extensive playing time together over the past decade.

In club football, it is widely recognized that higher budgets tend to lead to better performances over extended periods. According to findings by Davidsen & Hammer, club finances contribute to 70 percent of performance outcomes over a five-year span.[1] However, as observed in the cases of the Nordic and Croatian national teams, the dynamics of national team football may differ slightly. This paper examines the relationship and degree of influence between the market value of national teams' players and their performances, drawing comparisons with club football.

First, the paper sets the scene by introducing the intended models and hypotheses in Section 2: Model & Hypothesis where the proposed models and hypotheses are introduced, setting the chronological framework for the paper. In Section 3: Data & Methodology, Transfermarkt player values are introduced as the primary data source for market values, accompanied by the Elo rating system, which tracks performances in both national team and club team football. To fit the models outlined in Section 2 by simple and multivariable linear regression, the paper uses the method of Ordinary Least Squares, which is also explained here in section 3.

In Section 4: Results, a summary of the results from the regression of the presented models are provided. This enables the decomposition of different factors influencing national team performances, facilitating the analyses in Section 5: Analysis & Discussion. This section consists of two analyses. First, the hypotheses will be approached and tested in Analysis A, followed by Analysis B, where national team performances will be decomposed into the different factors found earlier in the paper. Ultimately, the aim is to address the research question: Does the average player market value less significantly influence the performance of national teams compared to club teams in football? If so, what are the primary factors influencing national team performances?

[1] Davidsen & Hammer 2021

## 2 Model & Hypothesis

As the paper examines the relationship between the player market values of a squad and the team's performances, a proposed mathematical model can be expressed in its simplest form as:

$$P = \alpha_1 V + \beta + \epsilon \tag{1}$$

Here, the dependent variable $P$ represents the team's performance over a medium to long term period measured by Elo rating, while the explanatory variable $V$ signifies the average player market value of a team, making the parameter $\alpha_1$ the relevant coefficient to estimate through linear regression. An intercept $\beta$ is included for now to fit the model, but its significance will be analyzed later in the paper. $\epsilon$ denotes the error term, representing the residual of the linear regression.

First, model (1) will be fitted to national team data. Subsequently, the same model (1) will be applied to club team data, allowing for the examination of similarities and differences between the two different datasets. By comparing the results of the two regressions side by side, several comparisons will be of interest. One such comparison involves the fitted slopes $\alpha_1$ and intercepts $\beta$, which may reveal expected differences in performance per additional million euros of player market value between the two football formats. While the slope may provide valuable insights, the significance of the parameter $\alpha_1$ and the residual of the model (1) regression are even more compelling. As motivated in Section 1, it is now hypothesized that the fit of national team data may result in a less significant slope parameter $\alpha_1$ and larger residuals $\epsilon$ compared to the club team data, potentially leading to a lower R-squared value in the national team data regression. This hypothesis is constructed based on a series of structural differences between national team football and club football.

**Box 1: Hypothesis 1**
When applying model (1), it is hypothesized that the club team regression may yield a significantly higher $R^2$-value compared to the national team regression.

Firstly, players are owned by their respective club teams, where daily training takes place. In 2023, European national teams were only assembled as a unit five times throughout the year, with each period referred to as a national team term, lasting approximately seven days. However, during these seven days, European national teams play two matches, meaning that a national team aggregates around 15 to 20 training days together over the course of an entire year. To compare the number of matches and training sessions between national teams and club teams, one can consider Danish superstar Christian Eriksen. For Manchester United, a fit Eriksen could play up to 68 matches in 2023 across all comptetions. In contrast, the Danish national team only played 10 matches in the same year. This disparity likely contributes to the Danish national team being less tactically aligned and agile compared to Manchester United.

Furthermore, even minor player injuries can have a more significant impact on a national team's season compared to a club team's season. Missing a national team term could result in missing 20%

of all matches in a year. Consequently, the starting XI for a national team may differ significantly between matches. For example, the Danish team changed five of eleven players from matchday 2 to matchday 3 in the Euro 2024 qualification due to injuries to regular starters. Such extensive changes are rare in club football, where five substitutions between matches typically occur due to significant losses or resting key players for important games.

Due to the mentioned differences between national team football and club football, the paper will include an extended model as a supplement to (1). The purpose of this extended model is to enhance the explanatory power of national team performances.

$$P = \alpha_1 V + \alpha_2 C + \alpha_3 S + \beta + \epsilon \tag{2}$$

Here, the fundamental model (1) is extended with two additional variables: $C$ and $S$, each multiplied by its respective $\alpha$-coefficient. $C$ quantifies the average number of national team appearances by the squad, while $S$ measures the variance of the player market values within the given national team squad. As the extended model (2) aims to yield comparable explanation in national team football and club football, it is constructed to fit the dynamics of national team football by multiple linear regression. Consequently, model (2) will exclusively be applied to national team data. It is now hypothesized that this fit may result in all estimated coefficients ($\alpha_1$, $\alpha_2$, $\alpha_3$) being statistically significant with positive or negative values as outlined in Box 2.

**Box 2: Hypothesis 2**
When applying model (2) to national team football, it is hypothesized that $\alpha_1$ and $\alpha_2$ are positive, while $\alpha_3$ is negative. All three coefficients are expected to be statistically significant.

This hypothesis is based on experimental reasoning regarding the extended determinants influencing national team performances, which will be further analyzed and illustrated later in the paper if validated. Explained briefly, a greater number of national team caps by the players in the squad, $C$, are expected to result in better performances, implying that higher values will have a positive influence. Conversely, a wider gap between the most and least valuable players, $S$, is expected to yield poorer performances compared to national teams with more evenly skilled players, suggesting that higher values will have a negative influence.

# 3 Data & Methodology

To explore the relationship between a team's market value and its performance over a medium to long term period, data for the two specified parameters are required in model (1). When fitting the extended model (2) for the national teams, data for international caps and variance of the team's market value are needed too. In the following subsections, the data choice of average player market value and average international caps from Transfermarkt, as well as Elo ratings from eloratings.net and clubelo.com, will be justified. Additionally, considerations regarding the number and selection of players and clubs in the models will be presented. All data was collected as of April 18, 2024, and primarily scraped using Python.

### 3.1 Market value data from Transfermarkt

Transfermarkt is a primary data source for player market values in football analysis. Despite the opacity of many transfer deals, it has built trust and credibility within the football community. The website covers a wide range of leagues and had 39 million unique monthly visitors in 2021, highlighting its significance in assessing player valuations and understanding market trends.[2] Dutch studies have even shown that top clubs themselves cite Transfermarkt values in their official financial documents, further underscoring its influence within the industry.[3]

Given the significant reputation of their player valuations, it might come as a surprise that Transfermarkt doesn't use an algorithm for their regular and biannual market value updates. The platform utilizes various pricing models to determine the important numbers, with a major reliance on input from the its community. This active community contributes to the valuation process alongside Transfermarkt's employees and voluntary data scouts of every region or nation globally. These contributors engage in written discussions regarding whether specific players have increased or decreased their values over the preceding six months, leading to the decision of the updated values. Critics may argue that Transfermarkt's player market values are not solely algorithm-driven, but these valuations rely on a comprehensive list of factors divided into three categories: 'most important factors', 'individual transfer modalities', and 'situational conditions'. This approach ensures a nuanced assessment of player market values.[4]

The absence of a Transfermarkt algorithm stems from the complex nature of real-world player valuations. Market structures, squad registration regulations, and player preferences vary by league and country, complicating valuations. For instance, player priorities, like living in London versus Newcastle or choosing a smaller club in southern Italy over a Europa League team in northern Norway, add to this complexity. Furthermore, studies have shown that Machine Learning models often predict transfer prices significantly inaccurately, highlighting the challenges of algorithmic approaches in this domain and underscoring the value of community input at Transfermarkt.[5] While other data sources such as KPMG or CIES exist, Transfermarkt is chosen for this paper due to its widespread usage and reputation. The variance of player market values within the national teams for model (2) is calculated from the Transfermarkt values too, which also serve as the data source for the players' international caps. The data manipulation to address squad size variation is done using Python, calculating measures based on the top 18 most valuable players in each team.

[2] The New York Times 2024
[3] Follow The Money 2020
[4] Transfermarkt 2021
[5] Aydemir et al. 2022

## 3.2 Team performance data based on Elo ratings

When measuring team performance in modern football, several options exist. For national teams, the traditional benchmark has been the official FIFA World Ranking. However, this system faced criticism in the 2000s and 2010s for its points calculation from friendly matches and perceived bias towards European and South American teams. In response, FIFA revised its methodology in 2018 to calculate points based on relative strength, inspired by the Elo system used in chess.[6]

Due to the FIFA World Cup occurring every four years, top national teams can compete against nations from other confederations, showcasing the relative strengths of teams from different regions. In contrast, the FIFA Club World Cup, until recently, was not comparable. In the 2022 FIFA World Cup in Qatar, 32 teams from every region participated. Meanwhile, the 2022 FIFA Club World Cup featured only one club team from each of the six confederations, each being the respective Champions League winner. This discrepancy results in limited opportunities to compare the levels of clubs from different confederations. With European clubs winning 17 out of the latest 18 FIFA Club World Cup tournaments, it is evident that UEFA clubs are significantly ahead of other regions in the world.[7] This dominance is further underscored by the composition of the latest Brazil squad called up for the Copa America 2024, where 21 of the 23 players have contracts with European clubs, despite the Brazilian football league having a good reputation globally.[8] With these factors in mind, this paper will solely utilize club team performance data from UEFA clubs.

UEFA distributes an official coefficient ranking for European club teams based on performance in the Champions League (CL), Europa League (EL), and Conference League (ECL). However, this ranking excludes national competitions like the English Premier League and can not be directly compared to the Elo-inspired FIFA World Ranking for national teams, complicating performance assessments across club and national football.[9]

**Box 3: The eloratings.net formula**[a]

$R_n = R_o + K \times (W - W_e)$

Here, $R_n$ is the new rating, $R_o$ is the old (pre-match) rating, and $K$ is the weight constant for the tournament played, e.g. 60 for World Cup finals and 20 for friendly matches. $K$ is then adjusted for the goal difference in the game. It is increased by half if a game is won by two goals, by $\frac{3}{4}$ if a game is won by three goals, and by $\frac{3}{4} + \frac{(N-3)}{8}$ if the game is won by four or more goals, where $N$ is the goal difference. $W$ is the result of the game (1 for a win, 0.5 for a draw, and 0 for a loss), while $W_e$ is the expected result (win expectancy), either from the chart or the formula: $W_e = \frac{1}{10^{(-dr/400)}+1}$. Here, $dr$ equals the difference in ratings plus 100 points for a team playing at home.

[a] World Football Elo Ratings 2024

[6] FIFA 2018

[7] Transfermarkt 2024

[8] BBC 2024

[9] UEFA 2024

An alternative and arguably better approach is to use the unofficial Elo ratings calculated at European club level by the website clubelo.com. This provides a more comprehensive view of club team performances relative to one another, considering a wider range of competitions and matches every week.[10] As a positive outcome of using this method, comparable data for national teams' performances is available at eloratings.net, because both websites use the exact same formula, ensuring consistency and comparability between club and national team performance ratings.[11] Despite minor differences between the Elo formula used by clubelo.com and FIFA's World Ranking, using comparable data sources enhances the robustness of this paper's analysis. Thus, eloratings.net and clubelo.com are chosen for measuring medium to long-term performances.

### 3.3 Selection of data

The Transfermarkt player market values are accessible for 205 of the 210 national teams registered on the official FIFA World Ranking. Additionally, the performances of all these national teams are measured using Elo ratings at eloratings.net. However, it's important to note that the methodology documentation of World Football Elo Ratings explains that "Ratings tend to converge on a team's true strength relative to its competitors after about 30 matches," indicating that "Ratings for teams with fewer than 30 matches should be considered provisional." While most national teams have historically played 30 matches, many nations rarely participate in the final FIFA World Cup, while others rarely play matches at all. Considering that club team performance data is only available for European clubs, including every national team in the models (1) and (2) would yield unclear estimation results and challenge the assumption of linearity. Furthermore, fitting numerous national teams into the model could be methodologically incorrect, particularly as smaller nations may not have the majority of their players competing in Europe. A reduced dataset also reduces the number of players with limited reputations, thereby making the estimation results more robust, given that Transfermarkt values are community-based to some degree.

The Club Elo Ratings provided by clubelo.com cover teams in the first division of 27 countries and a few second-tier leagues, totaling 526 European club teams. However, most of these clubs primarily compete in national competitions, which makes it challenging to compare teams internationally as the level of competition varies significantly. In the 2023/2024 season, 96 clubs compete in the final stages of the international UEFA CL, EL, and ECL, providing a basis for objectively assessing the relative strengths of clubs from different countries.

Considering these factors and a statistical preference for datasets of equal length, the models in this paper will incorporate the top 75 national teams from the World Football Elo Ratings, alongside the top 75 club teams from the Club Elo Ratings.

[10] Football Club Elo Ratings 2015
[11] World Football Elo Ratings 2024

## 3.4 The OLS method

Ordinary Least Squares (OLS) is a method used to estimate the parameters of a linear regression model. It finds the line - or hyperplane in higher dimensions, as in the case of model (2) - that minimizes the sum of the squared differences between the observed and predicted values of the dependent variable. In other words, OLS seeks to find the best-fitting line through the data points by adjusting the slope and intercept of the line to minimize the sum of the squared residuals. However, OLS relies on several key assumptions, including that the relationship between the variables is linear, the errors are normally distributed with a mean of zero, and there is no multicollinearity or heteroscedasticity in the data.[12] These assumptions will be further analyzed later in the paper, if relevant. This estimation method will be implemented in R, and the results will be presented in the following section.

# 4 Results

When later fitting the extended model (2) with three explanatory variables using multiple linear regression, it is statistically advantageous to take the logarithm of each factor before running the regression. This transformation helps address potential issues such as skewness or heteroscedasticity in the data, which could violate the assumptions of linear regression. By taking the logarithm of the independent variables, the impact of extreme numerical values diminishes, and a more symmetrical distribution of the data will be achieved. This addresses, for example, the issue where the average number of international caps for a given squad is 25, while the average player market value for the same squad is 25 million euros, as the latter numerical value is a million times larger. To ensure comparability between the estimations of model (1) and (2), the logarithm of team market values is taken before fitting the baseline (1) model below.

## 4.1 Fitting the simple model (1)

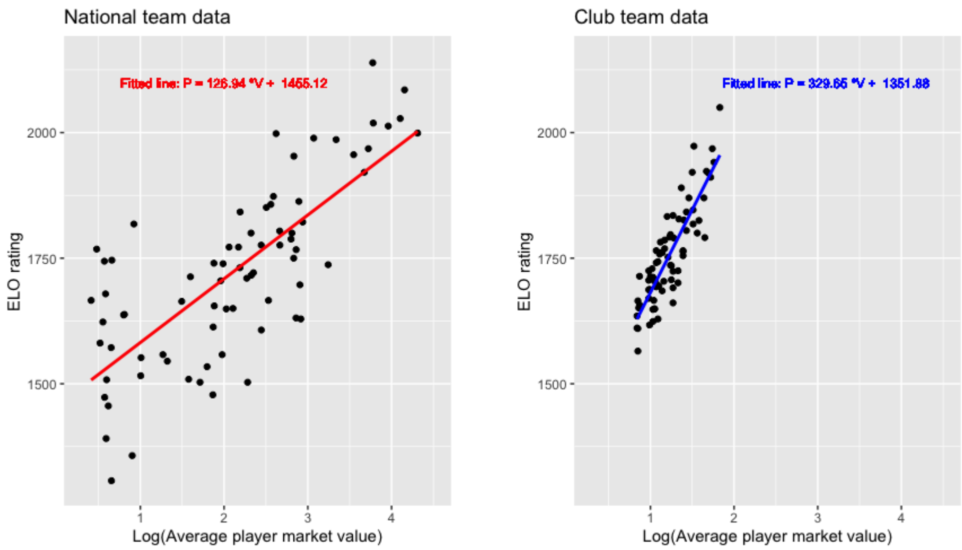


Figure 1: Model (1) estimations

[12]Wackerly et al. 2008

The two plots for the top 75 national teams and European club teams, respectively, reveal both similarities and differences between the datasets. Firstly, both the OLS-fitted lines have statistically significant positive slopes at a 0.1% significance level, which is also illustrated in Table 1. These $\alpha_1$-coefficients are estimated to be 126.94 for national team football and 329.65 for club teams, meaning a 1% increase in the squad's average player market value will increase the given team's Elo rating by 126.94 points for a national team and 329.65 points for a club team. Another observation is that the data points of the club teams are generally closer to the fitted line compared to national team data. This conclusion is also reflected in the adjusted $R^2$-values outlined in Table 1, with a value of 0.74 for club teams but only 0.56 for national teams.

| **Coefficients** | **National teams** | **Club teams** |
|---|---|---|
| Intercept, $\beta$ | 1455.12*** | 1351.88*** |
| | (30.44) | (28.52) |
| Value_log_mean, $\alpha_1$ | 126.94*** | 329.65*** |
| | (12.95) | (22.82) |
| Adjusted $R^2$ | 0.5625 | 0.7374 |

Signif. codes: *** $p < 0.001$

Table 1: Summary of Model (1) Linear Regression Results

In this paper, the adjusted $R^2$-values are preferred because they are more comparable between different dimensions of data, as adding more variables to a regression model will always increase the simple $R^2$-value.[13] For now, it is concluded that the simple model (1) has higher explanatory power for club team data, indicating that other factors are needed to make the national team model as precise as the club team model. Finally, the estimated intercepts are also presented in Table 1, but although significant at the 0.1% level, they are not very informative given the hypotheses of this paper. Before delving into the regression model (2), the estimated equations for the simple model (1) are outlined:

$$\hat{P}_{NT1} = 126.94V + 1455.12 + \epsilon \tag{3}$$

$$\hat{P}_{CT1} = 329.65V + 1351.88 + \epsilon \tag{4}$$

## 4.2 Fitting the extended national team model (2)

Seeking to enhance the explanatory power of the national team model, two additional independent variables were introduced into the regression model in (2). These variables supplement the average player market value, $V$, with the average number of international caps, $C$, and the variance of player market values within the squad, $S$. As previously mentioned, all explanatory variables undergo log-transformation. Table 2 presents the estimates and comparison between the simple model (1) and the extended model (2) for national teams exclusively.

[13] Chihara & Hesterberg 2011

| **Coefficients** | **Simple NT Model** (1) | **Extended NT Model** (2) |
|---|---|---|
| Intercept, $\beta$ | 1455.12*** | 1108.86*** |
| | (30.44) | (107.32) |
| Value_log_mean, $\alpha_1$ | 126.94*** | 201.47*** |
| | (12.95) | (26.10) |
| Caps_log, $\alpha_2$ | - | 126.97*** |
| | - | (34.71) |
| Value_log_var, $\alpha_3$ | - | -51.14*** |
| | - | (14.31) |
| Adjusted $R^2$ | 0.5625 | 0.6696 |

Signif. codes: *** $p < 0.001$

Table 2: Comparison of Model (1) and (2) Linear Regression Results for National teams

The extension of the simple national team model with additional factors behind team performances aimed to enhance explanatory power. The most notable outcome from the multiple linear regression results is therefore the adjusted $R^2$, which increased from 0.56 in model (1) to 0.67 for model (2), bringing it closer to 0.74 for the club team model (1). This improvement is attributed to the fitted hyperplane with all coefficients being statistically significant at the 0.1% level, resulting in the following estimation:

$$\hat{P}_{NT2} = 201.47V + 126.97C - 51.14S + 1108.86 + \epsilon \tag{5}$$

In this more refined model applied to national team data, the coefficient $\alpha_1$ increased from 126.94 to 201.47 although the larger standard deviation should be considered. This suggests that a 1% increase in the average player market value leads to a rise of 201.47 points in the team's Elo rating, thereby improving its performance. Similarly, the positive coefficient $\alpha_2$ indicates that a higher average number of international caps in the team results in better performance, with the Elo rating rising by 126.97 points on average for every 1% increase in caps. Conversely, the negative estimated coefficient $\alpha_3$ signifies that a larger variance in the market values of players for a given national team adversely affects its performance, resulting in an average reduction of 51.14 Elo points for every 1% increase in variance.

# 5 Analysis & discussion

Having presented the estimation results in the preceding Section 4, it becomes possible to delve deeper into the focal points of this paper. In Section 2, two hypotheses regarding the drivers of national team and club team performances were formulated. To either validate or refute these hypotheses, they will now be examined and discussed in the initial segment of this analysis, denoted as Analysis A. Furthermore, the extended estimation model (2) applied to national team data yielded Equation (5), which will serve as a tool to decompose the performance factors for specific countries, including those highlighted in Section 1. The decomposition will follow the analysis of hypotheses and will be named as Analysis B.

### 5.1 Analysis A: Approaching and testing the hypotheses

The first of the paper's two hypotheses is as follows: *When applying model* (1)*, it is hypothesized that the club team regression may yield a significantly higher* $R^2$*-value compared to the national team regression.* Upon initial examination of the data presented in Section 4, this assertion may seem plausible, given that the adjusted $R^2$-value for club teams is 0.74, whereas the national team's explanatory power only amounts to 0.56. However, statistical analysis necessitates a formal hypothesis test to draw any meaningful conclusions from estimated values. Therefore, an appropriate approach is to conduct an F-test to assess differences in variance between the two datasets for club teams and national teams respectively, under a null hypothesis that the variance in the two estimations is equivalent.[14] This exercise is conducted using R. Given the extremely small p-value ($5.20 * 10^{-14}$, see Appendix) of the F-test comparing the club team and national team datasets, the statistical null hypothesis is rejected. This result validates Hypothesis 1 of the paper, indicating that the average player market value provides greater explanatory power when describing club team performances compared to national team performances. While a more in-depth analysis could involve examining normally distributed residuals in both datasets, this exercise is left out due to space constraints.

Continuing the investigation of the paper's hypotheses, the second hypothesis posited: *When applying model* (2) *to national team football, it is hypothesized that* $\alpha_1$ *and* $\alpha_2$ *are positive, while* $\alpha_3$ *is negative. All three coefficients are expected to be statistically significant.* Upon estimating the coefficients resulting in Equation 5, all parameters were found to be significant, even at a 0.1% level, as indicated in Table 2. This high level of significance finds further tests unnecessary to validate Hypothesis 2. Instead of further testing, one would delve into another aspect of constructing this hypothesis. The primary aim was to increase the adjusted $R^2$-value. To examine whether the models' explanatory power truly improved from estimation model (1) to model (2) using national team data, a partial F-test is conducted in R. Once again, the p-value ($5.35 * 10^{-5}$, see Appendix) is very small, leading to the rejection of the null hypothesis, indicating that there indeed is an increase in the adjusted R-values. Adding more variables to a statistical model is not always the best approach. As noted in the methodology section, the implementation of linear regression depends on various assumptions.

Firstly, it is imperative to analyze and criticize the potential multicollinearity among the three explanatory variables by calculating their correlations. The correlations including international caps are computed to be 0.207 and 0.221, which are deemed non-problematic. More notably, the correlation between the average player market value and the variance in player market values within a given team is found to be 0.902. While this violates the assumptions of the model, it stems from the inherent relationship where teams with higher-valued players tend to have greater variance. Despite the importance of addressing potential multicollinearity, the inclusion of both variables in the models is justified by the corresponding coefficients being respectively positive and

[14] Chihara & Hesterberg 2011

negative to a very small level of significance. Therefore, both factors are retained in the model.

Furthermore, one could criticize that the Transfermarkt player market values themselves, to some extent, are influenced by team performances as outlined in Section 3. This inherent relationship may potentially bias the conclusions and prompt a discussion on causality. However, it is worth noting the conclusion drawn in Section 3 that Transfermarkt's data remains the most reliable data source in the market, which complicates further addressing of the issue. As a final note on the statistical considerations, no problematic patterns are found in this residual analysis, so this aspect will not be further analyzed in the paper.

## 5.2 Analysis B: Decomposing national team performances

Upon reviewing the plot in Figure 1, which illustrates the fitting of the simple model (1) to national team data, it became evident that an extended model was necessary to address the larger residuals. By examining the upper right corner of the plot in Figure 2, it is observed that countries such as Croatia, Uruguay, and Argentina outperform their expected level of Elo points considering their average player market values, while the Nordic countries Sweden and Norway achieve poorer results than expected.

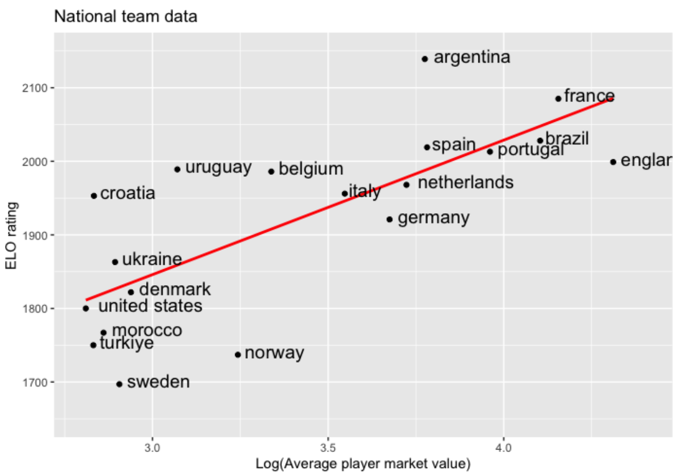


Figure 2: National team model (1) estimation

With Equation 5 estimated using multiple linear regression and all coefficients found to be statistically significant, it is now possible to decompose the performances of national teams into the model's three explanatory variables: average player market value, average international caps, and variance of the team's player market values. To conduct this analysis, each estimated coefficient will be multiplied by the corresponding data for a given team, while holding all other inputs constant at zero. By repeating this process for all three coefficients, each multiplied by their respective variables, it becomes possible to predict the expected Elo rating of a national team. This prediction is derived by summing together the three modeled factors and the constant intercept, $\beta$. Under the assumption that no other relevant factors are omitted from the model, the residual can then be interpreted as the extent to which countries either overperform or underperform their expected performance level.

Figure 3 illustrates the decomposition of selected elite national teams. The blue, yellow, and green bars indicate the extent to which each modeled factor contributes to their performances. The purple dot represents the true Elo rating of the teams after subtracting the less relevant $\beta$-coefficient, while the red bar illustrates factors omitted in the model. It is these omitted factors that can be interpreted as either overperformance or underperformance, which makes the red bar in each country particularly interesting to examine.

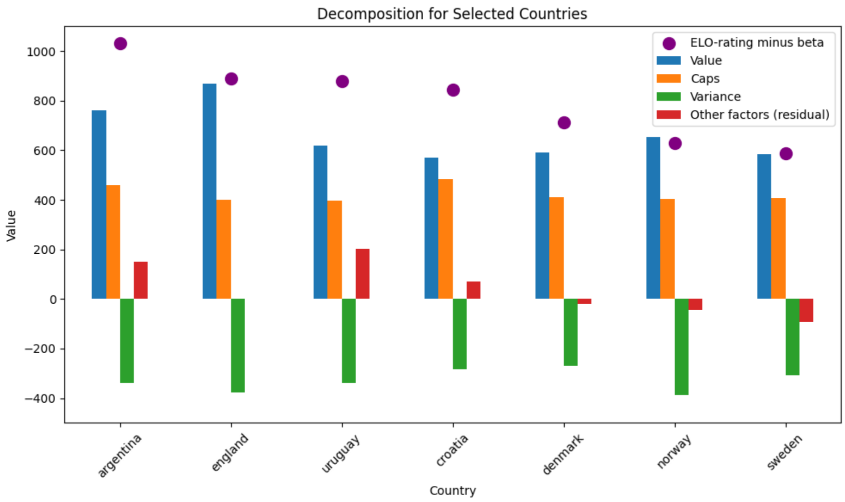


Figure 3: Decomposing the national team model (2) estimation

Comparing the illustrations of the simple and extended model in Figure 2 and Figure 3 respectively, the Nordic countries still face challenges, although to a lesser extent than in the simple model. From the decomposition, it is evident that Norway's performances are notably hindered by the variance in their player values relative to their large average market value. Conversely, Croatia benefits from a wealth of experience within the team, reflected in the high average number of international matches played, indicating strong cohesion among the players. Examining the smaller group of national teams included in 3, it appears that South American teams outperform their expected levels. Conversely, when considering the worst-performing teams relative to their expected levels according to the model in Figure 4, several African teams are included. This observation could indicate regional disparities in the factors of team performance. Alternatively, it could suggest that players of equivalent quality have varying market values based on nationality, such as English players being priced higher than their South American counterparts.[15]

[15]Bell et al. 2024

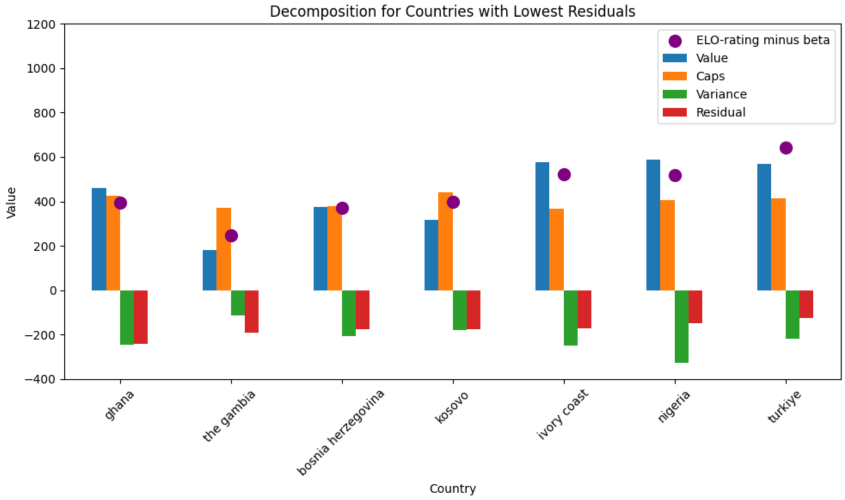


Figure 4: Decomposing the national team model (2) estimation

Lastly, when examining the top national teams globally based on the calculated expected Elo rating, Portugal emerges as a potential frontrunner with the highest number of Elo points, despite ranking fifth on the true list. This observation is highlighted in Table 3, which compares the top 10 national teams. Notably, France maintains its second position in both rankings, showcasing a good fit from the model. On the other hand, England experiences a notable discrepancy, ranking third in the true Elo ranking but dropping to sixth based on the expected Elo rating. Argentina's performance stands out in this illustration as in Figure 3, securing the top spot in the true ranking despite being predicted to hold the fourth position. Arguments like these underscore the nuances in performance evaluation methods and the complexities of assessing team rankings.

| **Team** | **Expected Ranking** | **True Ranking** |
|---|---|---|
| Portugal | 1 | 5 |
| France | 2 | 2 |
| England | 3 | 6 |
| Argentina | 4 | 1 |
| Netherlands | 5 | 10 |
| Brazil | 6 | 3 |
| Germany | 7 | 13 |
| Italy | 8 | 11 |
| Spain | 9 | 4 |
| United States | 10 | 22 |

Table 3: Comparison of Expected and True Elo Rankings for Top 10 National Teams

# 6 Conclusion

This paper investigates whether the average market player value significantly impacts national team performances compared to club teams in football, with a primary focus on identifying the key factors influencing national team success. Consistent with previous research, the findings reveal that economic strength explains approximately 70% of club football performance over the medium to long term. However, the average player market value of national teams does not explain their performances to the same extent as it does for club teams. The explanatory power for national teams increases when the variables of average number of international caps and variance in player market values are added to the model, bringing the adjusted $R^2$ closer to that of the simple club team model.

It can be concluded that a higher average number of international caps significantly enhances a national team's performance, as exemplified by Croatia's otherwise surprising successes. Conversely, a greater variance in player values within the squad leads to significantly poorer performances. This is the primary factor hindering national success for Norway's superstars Haaland and Ødegaard.

Finally, it is important to note that, based on a decomposition of national team performances, there may be regional factors that the estimated models do not account for. Additionally, a critical view should be, and has been, taken towards the paper's statistical and econometric considerations.

# 7 References

## 7.1 Books and Research Papers


- Aydemir, A. E., Temizel, T. T., and Temizel, A. (2022). A Machine Learning Ensembling Approach to Predicting Transfer Values
- Bell, A. R., Brooks, C., and Brooks, R. (2024). Are English football players overvalued?
- Chihara, L. M., and Hesterberg, T. C. (2011). Mathematical Statistics with Resampling and R
- Davidsen, M., and Hammer, D. (2021). How Hard Can It Be?
- Wackerly, D., Mendenhall, W., and Scheaffer, R. L. (2008). Mathematical Statistics with Applications


## 7.2 Articles


- BBC (2024). Casemiro left out of Brazil squad for Copa America. [Online - accessed 17/5-2024] https://www.bbc.com/sport/football/articles/cekl7r9pvryo
- FIFA (2018). Revision of the FIFA / Coca-Cola World Ranking. [Online - accessed 17/5-2024] https://digitalhub.fifa.com/m/f99da4f73212220/original/edbm045h0udbwkqew35a-pdf.pdf
- Follow The Money (2020). How the volunteers of data website Transfermarkt became influential players at European top football clubs. [Online - accessed 14/5-2024] https://www.ftm.eu/articles/transfermarkt-volunteers-european-football
- Football Club Elo Ratings (2015). The system or how this works. [Online - accessed 18/5-2024] http://clubelo.com/System
- The New York Times (2021). The Wisdom of the Crowd. [Online - accessed 14/5-2024] https://www.nytimes.com/2021/08/12/sports/soccer/soccer-football-transfermarkt.html
- Transfermarkt (2021). Transfermarkt Market Value explained - How is it determined? [Online - accessed 14/5-2024] https://www.transfermarkt.co.in/ transfermarkt-market-value-explained-how-is-it-determined-/view/news/385100
- Transfermarkt (2024). FIFA Club World Cup. [Online - accessed 17/5-2024] https://www.transfermarkt.com/fifa-club-world-cup/erfolge/pokalwettbewerb/KLUB
- UEFA (2024). UEFA Rankings. [Online - accessed 18/5-2024] https://www.uefa.com/nationalassociations/uefarankings/club/about/
- World Football Elo Ratings (2024). World Football Elo Ratings. [Online - accessed 18/5-2024] https://eloratings.net/about

# 8 Appendix

## 8.1 F-test for differences in variance between the club- and national team datasets

```
> # Take R2-values from the two models
> R2_modelclub <- summary(logmodelclub)$adj.r.squared
> R2_modelnational <- summary(logmodelnational)$adj.r.squared
> # Calculate F-statistic
> F_statistic <- ((R2_modelclub - R2_modelnational) / (1 - R2_modelclub)) / (R2_modelnational / (nrow(national75) - 2))
> # Number of observations (75)
> n <- nrow(national75)
> # Degrees of freedom
> df_modelclub <- 1
> df_modelnational <- n - 2
> # Calculate p-value
> p_value <- 1 - pf(F_statistic, df_modelclub, df_modelnational)
> p_value
[1] 5.195844e-14
```

## 8.2 Partial F-test for improved estimation model

```
> # Take R2-values from the two models
> R2_logmodel <- summary(logmodelnational)$adj.r.squared
> R2_exmodel <- summary(exmodelnational)$adj.r.squared
> # Calculate extra variables
> extra_variables <- length(coef(exmodelnational)) - length(coef(logmodelnational))
> # Calculate F-statistic
> F_statistic <- ((R2_exmodel - R2_logmodel) / extra_variables) / ((1 - R2_exmodel) / (n - length(coef(exmodelnational)) - 1))
> # Calculate p-value
> p_value <- pf(F_statistic, extra_variables, n - length(coef(exmodelnational)) - 1, lower.tail = FALSE)
> print(p_value)
[1] 5.354522e-05
```