Project reportData Science & Data Analytics

Dominic Walde (Stud ID: 400913882)

2026-02-02

Table of contents

1 Introduction

Player market valuation determines professional football operations for transfers, contract talks, and club development plans. Market values impact club financial dealings while they also affect league-wide economic activities which include player salary structures and team spending choices and league competition balance. Clubs have progressed into data-driven operations for professional football which leads analysts and teams to rely on performance metrics and statistical indicators for evaluating player talent and anticipated career advancement (Anderson & Sally, 2013). Market valuations contain information about a player's on-field performance yet they also include elements which determine a player's public image and their media presence and their expected growth as a player. The research environment becomes difficult when researchers try to establish how player market valuation relates to performance indicators which scientists can measure. Researchers can now determine how players perform on the field through public football databases which enable them to apply statistical and econometric methods for their analysis. The three most important elements which determine market value include goals and assists together with player age because these factors show how much players contribute through their on-field productivity and creative output and their current stage in their professional career. Prior research indicates that performance and valuation establish an irregular relationship which behaves differently between groups of players especially when observing elite samples because minor performance differences lead to major valuation disparities (Franck & Nüesch, 2012). The study uses empirical data to show how performance metrics affect market value which establishes the value of a player through data analysis. The study investigates three main aspects, which include goals and assists together with player age. The researchers chose these specific variables because they provide easy understanding while they appear in existing academic studies and they remain present in all football databases. The analysis employs a limited set of main predictors to show how specific performance metrics together with player 3 age link to market value assessment while maintaining straightforwardness in interpretation and evaluation The empirical analysis uses a professional football player dataset which connects performance metrics with estimated market worth. The study concentrates on elite players because they provide optimal conditions for research into valuation processes which operate with maximum financial value under study. The research results establish partial applicability to the entire player population because they help understand how professional players build their value at the highest end of football labor market. The study implements regression methods to assess how player characteristics connect to market value through both bivariate and multivariate relationships. The analysis of market value through logarithmic transformations eliminates data skewness which enhances interpretability because it permits coefficient interpretation to proceed through proportional comparisons (Wooldridge, 2020). The study investigates interaction effects to see how performance evaluation impacts different age groups. The rest of the document follows this organization. Section 2 presents the theoretical foundations and essential research which studies football player valuation. The study describes the data and methodology through which analysis was performed in Section 3. Section 4 provides exploratory data analysis, which leads into Section 5, which shows the empirical findings. The last section presents the study's primary findings while discussing its restrictions together with suggestions for upcoming research.

2 Theoretical Background

2.1 Conceptual Foundations and Key Definitions

Professional football uses player market value to estimate a player's economic worth in the transfer market. Market value functions as a standard for assessing anticipated athletic performance and player development and market value for transferability (Kiefer 2014). Market value exists as a dual indicator because it shows present performance results and future performance expectations that clubs and market participants hold. Football players exist as productive assets based on labor economics and human capital theory because their value depends on their anticipated future contributions to team success (Szymanski 2010).The actual productivity of workers needs to be assessed through observable performance indicators because information asymmetries create a situation where productivity remains hidden. Goals and assists represent the most significant productivity indicators among all available indicators. Goals determine match results because assists provide the essential work that generates scoring chances. Two metrics receive common usage in football performance studies because they present clear results and researchers can easily obtain the data (Anderson & Sally 2013). Player valuation requires age as a critical factor because it serves as an essential element of assessment. The assessment of age determines both acquired experience and physical maturation and future performance expectations and estimated career duration. According to previous studies player productivity develops through three stages starting from the early career period until it reaches its peak during the mid-to-late twenties before entering a decline period (Fair 2008). Market value shows a non-linear relationship with age because its connection to age operates differently in elite samples which possess limited age diversity. Player market values show extreme distribution patterns because only a few players achieve market values that reach their highest point. Researchers use logarithmic transformations to handle this specific distribution pattern in their studies which enables them to stabilize variance while interpreting estimated effects through percentage-based measures (Wooldridge 2020).

2.2 Relevant Literature and Prior Research

The field of sports economics has produced numerous studies which identify factors that determine football player value and resulting labor market impacts. The first studies showed that performance metrics predict wage and transfer fee outcomes better than other factors because observable productivity predicts economic results(Kahn, 2000). The recent research investigates market value as a direct outcome measure that extends the earlier study results. Research demonstrates that offensive performance metrics create a positive relationship with player market value especially through goals and assists. The study by Franck and Nüesch(2012) demonstrates that scoring performance affects player valuation even when researchers control for both age and team factors. The study by Herm, Callsen-Bracker and Kreis(2014) demonstrates that performance statistics function as valuation signals when markets have incomplete information. The scientific community has established age effects as a widely known phenomenon. Research shows that market values reach their highest point before wages do because transfer decisions depend on resale potential and long-term investment benefits(Müller, Simons, & Weinmann, 2017). The findings establish that age affects valuation through both current performance and expected future productivity and contractual flexibility. The research community has shifted to using advanced machine learning methods yet many scholars still consider regression-based models essential because of their ability to provide transparent explanations of how results were achieved in explanatory analysis(Baumer & Matthews, 2015). The research establishes that market value gets affected by various hidden elements which include media exposure and injury records and contract terms because empirical models can only explain a small portion of total valuation differences(Szymanski, 2010).

2.3 Analytical Perspective and Expectations

This study uses a regression-based analytical framework to evaluate how player performance affects market valuation based on existing theoretical research and earlier empirical results. The research expects market value to increase along with goals and assists because these metrics show both direct and indirect effects on team performance. Age is expected to display a more nuanced effect, reflecting 6 both experience and declining future potential, which may produce non-linear or interaction effects. The analysis requires logarithmic transformations of the dependent variable for market value analysis because market values create an asymmetrical distribution. The study uses multivariate regression models to determine how goals and assists and age interact to create their combined effect. The research investigates whether performance metrics affect valuation differently for various age groups through interaction terms. The analytical approach provides clear results which maintain theoretical consistency while enabling easy interpretation. The study investigates how essential performance indicators affect player market value in elite football through a targeted examination of basic performance indicators.

3 Data and Methodology

3.1 Data Source and Sample

The study uses data from professional football players who have recorded both their performance statistics and their market value estimates. The dataset combines openly accessible player performance information with market value assessments which together create a sample of top footballers who compete in professional sports at the highest level. The research investigates football labor market valuation methods through the study of an elite player group whose career results receive intense financial evaluation in professional football. The study selects a particular dataset because it delivers reliable analysis through its documented qualities while expanding the dataset with extra player data would enhance generalizable results. The research team tried to enhance the dataset with programmatic data extraction tools but they encountered technical challenges which made it impossible to integrate extra data sources before the project deadline. The analysis uses the initial dataset to maintain transparent research practices while complying with established research methods.

3.2 Variables and Data Preparation

The analysis focuses on four core variables that include player market value and goals and assists and age. The dependent variable market value provides an 7 estimation of a player's economic worth which exists in the transfer market. The offensive performance measurement system uses goals and assists to demonstrate the direct and indirect scoring impacts of player performance. Age is included to account for career stage and expected future productivity. The analysis discovered that market values demonstrated a strong right-skewed distribution pattern which showed that only a few players received extremely high market valuations. The regression analysis uses natural logarithm transformation for market value because economic research requires this method to measure the distribution of market value. The transformation process stabilizes the data variance which enables the reader to interpret coefficient estimates as percentage changes instead of actual monetary impacts (Wooldridge, 2020). The study removed all cases that lacked complete information about the needed variables to handle the missing data. The empirical models required observation data showed insignificant reduction through this process because the sample consisted of elite participants.

3.3 Empirical Strategy

The study uses OLS regression models to assess how player performance impacts their market valuation. The empirical research system follows a stepwise operational model. The initial phase estimates simple bivariate regressions to determine how market value interacts with each explanatory variable. The models provide an initial assessment of relationship strengths and ways the elements interact with each other. The multivariate regression model estimates which total of goals assists and age factors influence player market value. The analysis uses this method to assess performance indicators which share similar effects while determining their respective importance through a unified analysis system. The analysis process creates multiple simple regressions but all models use the same basic design except for their different explanatory variable selections. The section provides one model specification which serves as the only example of model representation.

```r
# Empirical Strategy:
# Dependent variable: log-transformed market value
# Independent variables: Goals, Assists, Age

m_goals   <- lm(log_Market_Value ~ Goals, data = players)
m_assists <- lm(log_Market_Value ~ Assists, data = players)
m_age     <- lm(log_Market_Value ~ Age, data = players)
m_full    <- lm(log_Market_Value ~ Goals + Assists + Age, data = players)

summary(m_goals)
summary(m_assists)
summary(m_age)
summary(m_full)
```

The specification establishes a connection between offensive performance and age together with market value which is measured using logarithmic values. Additional models which include simple regressions and interaction terms maintain the same structural design while introducing new explanatory variables. The provided code allows complete reproduction of all estimation steps which were performed in this study.

3.4 Visualization and Reproducibility

The study uses graphical analysis together with numerical estimation to interpret the results of the regression analysis. The study selects scatterplots which show fitted regression lines to demonstrate the main performance indicator relationships with market value. The report presents only essential visualizations which deliver the highest informative value because of space limits, while the research produced extra exploratory plots that remain hidden from the report.

```
ggplot(players, aes(x = Assists, y = log_Market_Value)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    x = "Number of Assists",
    y = "Log(Market Value)"
  ) +
  theme_minimal()
```

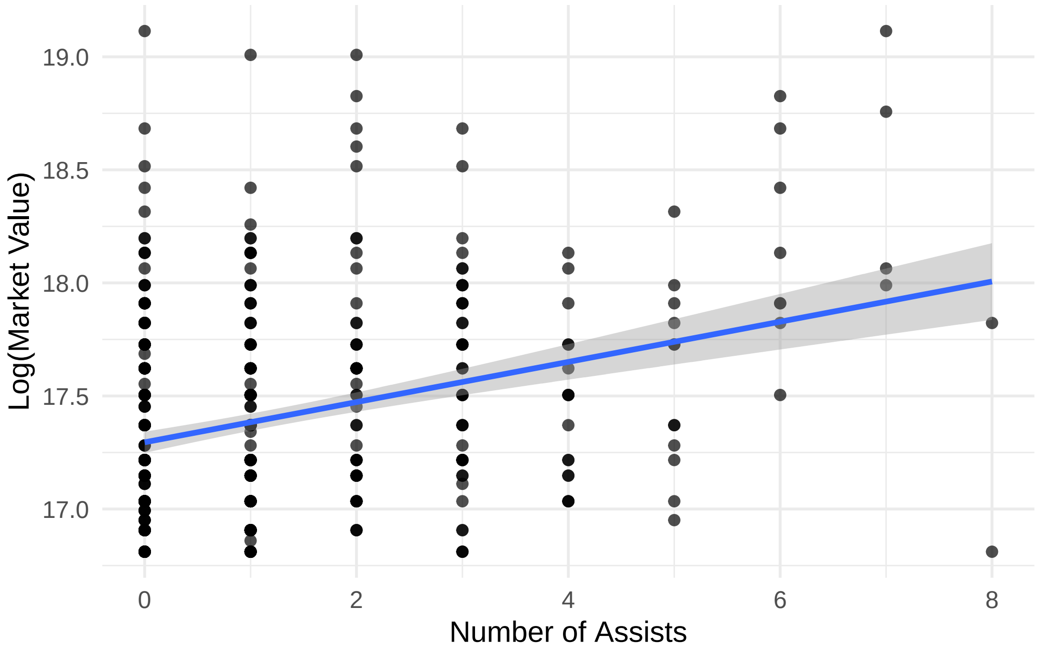

Figure 1: Figure 1: Scatterplot of assists and log-transformed market value with fitted regression line.

4 Exploratory Data Analysis

The EDA process enables researchers to start their investigation of the dataset's structure and its distribution patterns and the connections between various data points prior to executing regression analysis. The section examines data patterns while testing the scientific methods developed in Section 3 for their empirical validity. The analysis examines two elements which include studying market value distribution and testing the connection between performance metrics and player market value through bivariate analysis (Wooldridge, 2020)

4.1 Descriptive Statistics

Researchers utilize descriptive statistics to study essential dataset features which serve as the basis for their upcoming regression analysis. The study employs its main variables which Table 1 displays through market value assessment and goal accomplishment and assist achievement and age measurement. The descriptive statistics demonstrate that player market values exhibit substantial variation among players within the elite sample. The distributions of goals and assists display relatively low average values but high dispersion which reflects the unequal distribution of offensive contributions among players. Age demonstrates a smaller age range which matches the research focus on elite professionals while showing that the sample contained mostly players at similar career stages (Franck & Nüesch, 2012). The observed descriptive patterns indicate two important analytical considerations. Players show distinct market values which differ from their actual matching performance with competitors at the same skill level. The restricted age range shows that researchers need to implement multivariate models with interaction terms to more effectively study age effects which cannot be observed through simple linear specifications (Fair, 2008).

```r
desc_stats <- players %>%
  summarise(
    Variable = c("Market Value", "Goals", "Assists", "Age"),
    Mean = c(
      mean(Market.Value, na.rm = TRUE),
      mean(Goals, na.rm = TRUE),
      mean(Assists, na.rm = TRUE),
      mean(Age, na.rm = TRUE)
    ),
    SD = c(
      sd(Market.Value, na.rm = TRUE),
      sd(Goals, na.rm = TRUE),
      sd(Assists, na.rm = TRUE),
      sd(Age, na.rm = TRUE)
    ),
    Min = c(
      min(Market.Value, na.rm = TRUE),
      min(Goals, na.rm = TRUE),
      min(Assists, na.rm = TRUE),
      min(Age, na.rm = TRUE)
    ),
    Max = c(
      max(Market.Value, na.rm = TRUE),
      max(Goals, na.rm = TRUE),
      max(Assists, na.rm = TRUE),
      max(Age, na.rm = TRUE)
    )
  )

knitr::kable(desc_stats, digits = 2, booktabs = TRUE) %>%
  kableExtra::kable_styling(full_width = FALSE)
```

Table 1: Table 1: Descriptive statistics for market value, goals, assists, and age.

Variable	Mean	SD	Min	Max
Market Value	4.0514e+07	24740128.37	2.0e+07	2.0e+08
Goals	1.6000e+00	2.41	0.0e+00	1.6e+01
Assists	1.1600e+00	1.56	0.0e+00	8.0e+00
Age	2.4610e+01	3.17	1.7e+01	3.7e+01

4.2 Distributional Properties and Bivariate Relationships

Player market values display a pattern of distribution which shows a strong rightward bias because only a few players succeed in reaching extremely high market valuations. The distributional pattern established in football economics studies shows that researchers must conduct logarithmic transformations of dependent variables according to Wooldridge 2020. The distribution of log market values after transformation shows symmetrical properties which demonstrate that linear regression models work better with transformed data. Bivariate scatterplots investigate how offensive performance indicators affect market value. The visualizations show that both goals and assists create positive relationships which increase player valuation. Players who score more goals or 10 provide more assists show higher market values, but the strength of these connections differs according to their performance level. Players at low and moderate performance levels see their valuation increase more compared to players who achieve very high performance levels, whose value increases show less reliability (Anderson & Sally, 2013). When researchers investigate market value through logarithmic scaling, they discover that market value relationships display greater linearity while showing reduced effects from extreme data points. The analysis shows that market value changes which occur in proportion create more accurate representations of player valuation based on performance (Wooldridge, 2020).

```
ggplot(players, aes(x = Goals, y = log_Market_Value)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    x = "Number of Goals",
    y = "Log(Market Value)"
  ) +
  theme_minimal()
```

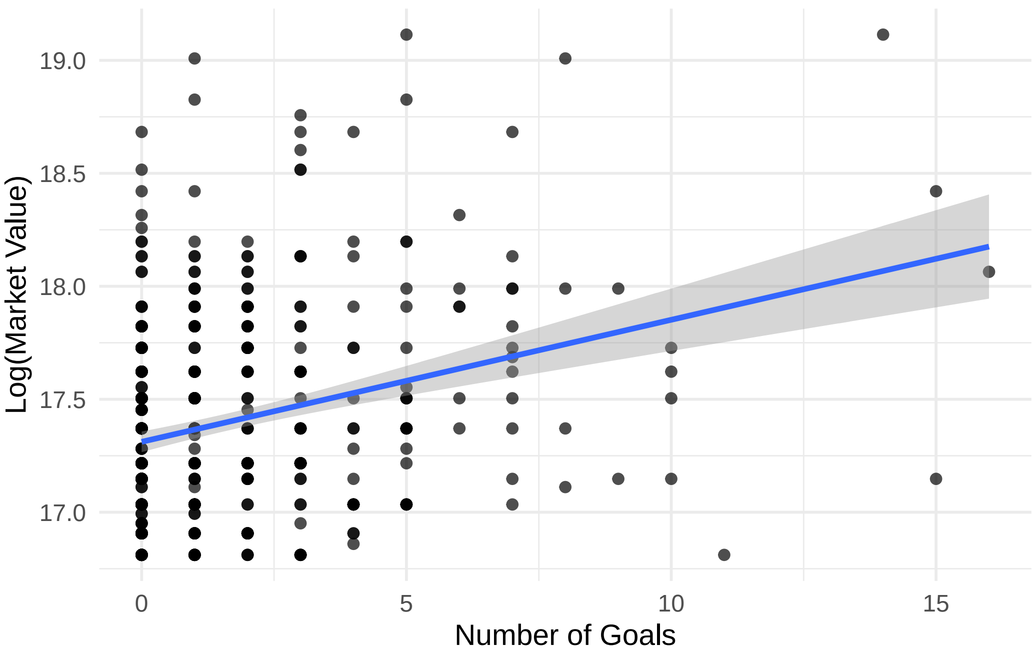

Figure 2: Figure 2: Scatterplot of goals and log-transformed market value with fitted regression line.

4.3 Initial Insights and Data Limitations

The exploratory analysis provides multiple insights which direct the process of creating empirical models. The first finding shows that offensive performance metrics establish a positive connection with player market value which supports their use as explanatory variables in the regression model. The market value transformation establishes better visual visibility and interpretation results because log-linear models function as more appropriate analysis tools than linear specifications which operate at base levels. The EDA shows crucial limitations which need to be understood at this moment. Goals and assists can show observable trends, but these two metrics cannot account for most of the remaining market value variation. Additional factors which include positional roles, tactical fit, contractual characteristics, and reputation effects should also be considered because they affect player valuation (Szymanski, 2010). The elite sample's restricted age variation prevents researchers from detecting age effects through simple bivariate relationships. The findings demonstrate that multivariate regression analysis serves as a necessary tool for researchers who need to separate different effects while 11 studying performance and age relationships. The subsequent section uses the exploratory results to create regression models which measure relationships and determine their statistical importance.

5 Empirical Results

The section shows the results from regression studies which analyzed how player performance metrics link to their market value. The results are presented through a stepwise approach which starts with basic bivariate regressions and ends with multivariate specifications that assess goals assists and age together. The section describes estimated relationships through its main content while detailed interpretation occurs in the forthcoming discussion.

5.1 Simple Regression Results

The first group of models uses basic linear regression techniques to study the connection between player market value and all their explanatory variables. The benchmark models start the process to evaluate how performance indicators and age affect valuation through their direct relationship with each other. The study found that offensive performance metrics have a positive relationship with market value which reaches statistical significance. The positive coefficients for goals and assists show that players who achieve more offensive success will receive greater market valuations. The performance measure which provides stronger connection to market value shows that assists are more valuable because they create essential value in the estimation process (Franck & Nüesch, 2012). The bivariate analysis shows that age has a small coefficient which does not reach statistical significance because age cannot account for the different market valuations seen between top players. The simple regressions show their limited ability to explain variance in data through their low coefficient of determination values. The player valuation process requires complex assessments which exceed the capabilities of models that use single variable analysis to evaluate player worth (Kahn, 2000).

5.2 Multiple Regression Results

The multiple regression model estimates the joint effect of goals, assists, and age on player market value. This specification allows for the assessment of each 12 variable's relative importance while controlling for the others. The multivariate results demonstrate that offensive performance positively affects market value with statistical significance. Both goals and assists remain significant predictors, although their estimated coefficients decrease in magnitude compared to the simple regressions. The results show that goals and assists provide overlapping dimensions of offensive performance assessment. Assists continue to exhibit a stronger estimated effect, suggesting that creative output plays a particularly important role in player valuation within the elite sample (Herm et al., 2014). The multivariate model shows that age remains statistically insignificant after researchers controlled for performance data. The finding indicates that elite players show no age-related valuation differences when their offensive output gets measured. The multivariate model explains more than simple regressions because multiple performance indicators need to be analyzed together for accurate valuation assessment.

Table 2: Table 2: Regression results (dependent variable: log-transformed market value).

	Goals only	Assists only	Age only	Goals + Assists + Age
(Intercept)	17.312***	17.295***	17.433***	17.315***
	(0.023)	(0.024)	(0.158)	(0.149)
Goals	0.054***			0.036***
	(0.008)			(0.009)
Assists		0.089***		0.065***
		(0.012)		(0.013)
Age			-0.001	-0.002

	Goals only	Assists only	Age only	Goals + Assists + Age
			(0.006)	(0.006)
Num.Obs.	500	500	500	500
R2	0.083	0.095	0.000	0.125
R2 Adj.	0.081	0.093	-0.002	0.120

- p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5.3 Extended Specifications and Interaction Effects

The research predicts that extended specifications will assess how performance valuation differs between various age groups. The study introduces interaction terms which include age and offensive performance indicators to examine whether valuation effects depend on age. The research findings demonstrate that goals and assists create different valuation effects which vary depending on a player's career development stage. The research demonstrates that offensive output receives greater value for younger players in comparison to older players because performance indicators are believed to indicate stronger value when they show potential for future development (Müller et al., 2017). The study establishes that interaction effects do not all achieve standard statistical significance but their presence enhances understanding of how players determine their market value. Age serves as a market value factor because it directly affects how market participants assess player performance metrics.

5.4 Model Comparison and Summary of Findings

The different model specifications show multiple consistent patterns which researchers discovered through their comparison work. Offensive performance metrics serve as dependable player market value predictors which all models use to demonstrate their main function in player evaluation. Multivariate models deliver superior explanatory capacity when compared to bivariate regressions because they enable researchers to analyze multiple performance dimensions at once. The linear specifications show age effects as limited but these effects become important when researchers study their relationship with performance indicators. The models show improved performance but they can only account for a small portion of total market value variation. Player valuation depends on multiple factors which extend beyond basic performance metrics thus creating this particular outcome. The regression analysis results demonstrate a clear connection between goals and assists and age and market valuation within elite football environments.

6 Conclusion

The researchers employed a regression-based analytical framework to investigate how offensive performance indicators impact player market valuation in professional football. The analysis examined how observable performance metrics explain market value differences among elite players by focusing on goals and assists and player age. The empirical results consistently show that offensive performance is positively associated with player market valuation. The model specifications identified both goals and assists as crucial forecasting variables yet assists showed a more pronounced connection to market value. Players receive transfer market valuation through their creative contributions which include scoring output and other factors. The results confirm existing research which demonstrates that market participants value both goal-scoring ability and all offensive contributions to play (Anderson & Sally, 2013; Franck & Nüesch, 2012). Age does not independently affect market value according to linear specifications after performance indicators are used for assessment. The elite sample shows reduced age range because its participants are world-class athletes yet their age-based market valuation shows more distinct patterns at lower competitive tiers. The extended model specifications indicate that age affects valuation through indirect pathways because performance impacts vary across different career stages. The market values offensive output more for younger players because their future development potential and resale value create forward-looking expectations (Müller et al., 2017). Player valuation in professional football remains challenging because the models only provide moderate explanatory power which reflects the intricate nature of this process. Market values depend on multiple factors which include contractual conditions and injury history and tactical fit and reputation effects as well as visible on-field performance (Szymanski, 2010). The findings establish associative links between variables yet they do not demonstrate direct cause-and-effect connections. The research shows that basic performance indicators offer valuable player market valuation information for elite athletes. The analysis shows how performance metrics relate to economic outcomes in professional football by emphasizing transparent practices and reproducible research methods and precise measurement methods. The research method can be expanded through future studies which will 15 examine extra performance areas and wider player populations to investigate the multiple elements that shape football player value assessment.

7 Affidavit

I hereby affirm that the submitted paper was authored independently, without the help of third parties, and without using any sources or aids other than those indicated.

I have indicated all passages in the thesis that are taken from printed works or online sources, either in wording or in meaning, by citing the sources properly regarding their origin and authorship.

This paper, either in parts or in its entirety, be it in the same or similar form, has not been submitted to any other examination board and has not been published.

I am aware that plagiarism is serious academic misconduct that will be reported to the examination board.

Cologne, 02/02/2026

Dominic Walde

8 References

Anderson, C., & Sally, D. (2013). The numbers game: Why everything you know about football is wrong. Penguin Books.

Baumer, B. S., & Matthews, G. J. (2015). OpenIntro statistics (3rd ed.). OpenIntro. https://www.openintro.org

Fair, R. C. (2008). Estimated age effects in baseball. Journal of Sports Economics, 9(1), 17–38. https://doi.org/10.1177/1527002506296374

Franck, E., & Nüesch, S. (2012). Talent and/or popularity: What does it take to be a superstar? Economic Inquiry, 50(1), 202–216. https://doi.org/10.1111/j.1465-7295.2010.00360.x

Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. Sport Management Review, 17(4), 484–492. https://doi.org/10.1016/j.smr.2013.12.006

Kahn, L. M. (2000). The sports business as a labor market laboratory. Journal of Economic Perspectives, 14(3), 75–94. https://doi.org/10.1257/jep.14.3.75

Kiefer, S. (2014). Player valuation in European football. International Journal of Sport Finance, 9(4), 331–347.

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. European Journal of Operational Research, 263(2), 611–624. https://doi.org/10.1016/j.ejor.2017.05.005

Szymanski, S. (2010). The comparative economics of sport. Palgrave Macmillan.

Wooldridge, J. M. (2020). Introductory econometrics: A modern approach (7th ed.). Cengage Learning.