European Sport Management Quarterly

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/resm20

Predicting transfer fees in professional European football before and during COVID-19 using machine learning

Yanxiang Yang, Joerg Koenigstorfer & Tim Pawlowski

To cite this article: Yanxiang Yang, Joerg Koenigstorfer & Tim Pawlowski (2024) Predicting transfer fees in professional European football before and during COVID-19 using machine learning, European Sport Management Quarterly, 24:3, 603-623, DOI: 10.1080/16184742.2022.2153898

To link to this article: https://doi.org/10.1080/16184742.2022.2153898

EUROPEAN SPORT MANAGEMENT QUARTERLY
2024, VOL. 24, NO. 3, 603–623
https://doi.org/10.1080/16184742.2022.2153898

Routledge
Taylor & Francis Group

Predicting transfer fees in professional European football before and during COVID-19 using machine learning

Yanxiang Yang [a], Joerg Koenigstorfer [a] and Tim Pawlowski [b]

[a]Technical University of Munich, Munich, Germany; [b]University of Tuebingen, Tuebingen, Germany

ABSTRACT

Research question: Our study aims to extend findings from previous efforts exploring the factors associated with transfer fees to and from all big five league clubs in European football (men) by building upon advances in machine learning, which allow to depart from linear functional forms. Furthermore, we provide a simple test of whether the transfer market has changed since the beginning of the COVID-19 pandemic.

Research methods: A fully flexible random forest estimator as well as generalized and quantile additive models are used to analyze smooth (non-linear) effects across different quantiles of scraped data (including remaining contract duration) from transfermarkt.de ($n = 3{,}512$). While we train our models with a randomly drawn subsample of *before*-COVID-19 transfers, we compare the prediction accuracy for two subsets of test data, that is, *before* and *during* COVID-19.

Results and findings: Since our findings suggest several non-linear predictors of transfer fees, moving beyond linearity is insightful and relevant. Moreover, our models trained with *before*-COVID-19 data significantly *underestimate* the actual transfer fees paid *during* COVID-19 particularly for high- and medium-priced players, thus questioning any cooling-off effect of the transfer market.

Implications: In the discussion of our findings, we showcase how moving beyond linearity and modeling quantiles can be revealing for both research and practice. We discuss limitations such as sample selection issues and provide directions for future research.

ARTICLE HISTORY
Received 23 July 2021
Accepted 25 November 2022

KEYWORDS
Transfer market; soccer; transfer fee; COVID-19; machine learning

Introduction

In professional football, talented players are the clubs' most valuable resources. Player registrations give clubs the exclusive rights to a player's services and these registrations can be exchanged (purchased or loaned) on the international market. In 2021, more than 18,000 international permanent transfers were made with revenues of almost US$ 5 billion (FIFA, 2022). Permanent transfers are those when a player is permanently engaged by the buying club. The transfer fee for players reflects the financial compensation for this movement of under-contract players, in addition to player wages. But what are the key determinants of transfer fees? How can we make comparably accurate

CONTACT Joerg Koenigstorfer joerg.koenigstorfer@tum.de Technical University of Munich, Campus D Uptown Munich, Georg-Brauchle-Ring 60/62, Munich 80992, Germany

Supplemental data for this article can be accessed here https://doi.org/10.1080/16184742.2022.2153898

predictions for such fees? Moreover, did the COVID-19 pandemic affect the relevance of common predictors and the accuracy of predictions based on pre-COVID-19 evidence?

In spite of the existing empirical work on the determinants of transfer fees, which began with the seminal study conducted by Carmichael and Thomas (1993), previous studies commonly rely on simple linear regressions and explore a rather limited set of variables for comparably small samples (see Appendix Table 1 for an overview, supplementary material). One recent exception is McHale and Holmes' (2022) analysis of football transfer fees. They used both advanced performance metrics (based on data from *Instat*) and player ratings (based on data from *sofifa*) and employed machine learning methods. Yet, the authors do not describe the nature of non-linear relationships between variables in detail and do not consider COVID-19-specific effects on the transfer market. To date, no study has yet explored empirically whether and to which extent the transfer market has changed since early 2020. In this regard, Parnell et al. (2021) argue that, in times of the COVID-19 pandemic, 'clubs and leagues must consider more critically the role of employment contracts, transfer/labour markets, […] and connective potential within their network' (p. 25). Hence, one might argue that teams have simply reduced their overall spending on transfer fees due to COVID-19 related financial constraints (*cost argument*; Quansah et al., 2021). At the same time, however, teams might have reduced their set of players to be recruited, thus somewhat decreasing their bargaining power. In this case, they might have spent more money per player to strategically strengthen their team (*focus argument*). Overall, it remains largely unknown whether and to what extent the transfer market has changed since the beginning of the pandemic.

Our study aims to explore these issues. More precisely, we investigate the relevance of a large set of determinants of players' transfer fees in the big five European football leagues (men), that is, the English Premier League, the French League 1, the German Bundesliga, the Italian Serie A, and the Spanish La Liga, before and during the pandemic. The determinants include variables of player background characteristics, player performance, selling- and buying-club characteristics, as well as time. By building upon some advances in machine learning and considering different quantiles of transfers, our analysis departs from the assumption of linear functional forms, thus potentially improving prediction accuracy. While we train our models with a randomly drawn subsample of *before*-COVID-19 transfers, we compare the prediction accuracy for two subsets of test data, that is, *before* as well as *during* COVID-19. In this regard, we aim to partly close a research gap identified by Parnell et al. (2021), who provide conceptual arguments for how COVID-19 might have changed the football ecosystem.

Methods

Data

We scraped a longitudinal transfer dataset of the big five leagues over 14 seasons (from 2008/09 to 2021/22) from transfermarkt.de. The website provider actively searches for valid and reliable sources of transfer fees and relevant characteristics (including data when a player was transferred as well as selling- and buying-club characteristics and player-related data), involving about 80 staff members and crowdsourced knowledge.

Such an approach is useful for the present study, because transfer fees are often not made public.[1] Previous studies suggest that the website is a highly reliable database and one of the best sources for football transfers (Depken & Globan, 2021). It has been often used in the sport management, sport economics, and labor market literature (e.g. Coates & Parshakov, 2021; Feuillet et al., 2020; Gyimesi & Kehl, 2021; Herberger & Wedlich, 2017; Müller et al., 2017; Ramos-Filho & Ferreira, 2021).

All transfers that fulfill the following inclusion criteria are considered eligible for our study: (a) transfers which involved a fee (excluding free and loan transfers similar to McHale and Holmes (2022)[2], (b) at least either the selling club or the buying club was from one of the big five leagues, and (c) the transfer took place between the seasons 2008/09 and 2021/22 (until 2 February 2022).[3] Overall, 7,918 transfer records from seasons 2008/09 to 2021/22 are included in the descriptive analysis. Thereof, 3,512 records for transfers that occurred between 2015/16 and 2021/22 which contain information on the remaining contract duration of players are used in our analysis.[4]

Measures

The dependent variable is the natural logarithm of the transfer fees (in real EUR; reference year: 2021). The real transfer fee is converted from the nominal fee, inflated or deflated by the monthly consumer price index (CPI) for recreation and culture (including sports) in 2021 February prices. Given that at least either the selling club or the buying club is from one of the European big five leagues, the harmonized CPI (oecd.org) is used for the real fee inflation or deflation, enabling international comparisons of inflation and deflation rates. Given that transfer fees are highly skewed to the right (as evidenced by histogram plots), we use the natural logarithm (Dobson & Gerrard, 1999; Gyimesi & Kehl, 2021).

As possible predictors of transfer fees we consider several variables that are partly used in earlier studies. The variables cover five domains, that is, (1) player background characteristics, (2) player performance, (3) selling-club characteristics, (4) buying-club characteristics, as well as (5) time specific variables. Within these domains, we select variables that are conceptually linked to transfer fees, are accessible at the time of the study for a large dataset of transfers, and for which no multicollinearity concerns exist.

In line with previous studies (see Appendix Table 1 for a summary of previous findings, supplementary material), we use the following five measures of (1) player background characteristics: age, height, nationality, position, and remaining contract duration. As measures of (2) player performance, we use available player performance statistics as well as player injury history.[5] Both (3) selling- and (4) buying-club characteristics provide measures for assessing the bargaining power of the respective clubs (Carmichael & Thomas, 1993; Dobson et al., 2000; Dobson & Gerrard, 1999). The variables used in our analysis are both club performance-related (such as sporting and financial performance) and size-related (such as the spectator numbers). Lastly, football transfer markets are constantly developing, and the transfer fees may vary over (5) time due to reasons such as newly signed broadcasting contracts (Depken & Globan, 2021). Table 1 reports all definitions and descriptions of the variables that are considered as predictors in our study including their arguments for inclusion.

Table 1. Definitions of variables and reason for their inclusion.

Variable	Description and reason for inclusion	References
Transfer fee	Natural logarithm of real transfer fee (EUR)	
Player characteristics		
Age	Age at time of transfer (years); indicator of experience	[1–14]
Height	Height in cm; indicator of scoring and heading abilities	[1–3,11,12]
Nationality	Continent of birth (dummies: Europe, Asia, Africa, South America, North America) [a]	[1,5,7]
Position	Dummies: attacker, defender, midfielder, goalkeeper [b]	[1–4,6,7,8–14]
Remaining contract duration	Remaining contract duration at time of transfer (in days); indicator of negotiating power	[4,7,8,11]
Player performance during the previous season		
UEFA Champions League	Dummy: player played in the UEFA Champions League [a]	[6,7,9,10], partly [4]
Appearances	Number of appearances for the club [a]	[2,3,6,9,10,13,14]
Substitution on	Number of substitutions on [a]	[12]
Substitution off	Number of substitutions off [a]	[12]
Minutes played	Total minutes played [a]	[1,3,9,11,12]
Points (/1,000MP)[d]	Average points per match multiplied with number of appearances[a]	
Goals (/1,000MP)[d]	Number of goals [a]	[2,3,4,9–14]
Assists (/1,000MP)[d]	Number of assists for goals [a]	[1,4]
Yellow cards (/1,000MP) [d]	Number of yellow cards; indicator of aggressive play	[1,12]
Injury proneness	Age-relative frequency of recorded injuries before transfer divided by number of days in injury before transfer, i.e. (injury days/frequency)/age; indicator of proneness to injuries	[11]
Selling-club characteristics during the previous season		
Arrivals	Number of players that were transferred into the club [c]	Partly [3]
Departures	Number of players that were transferred out the club [c]	Partly [3]
Transfer income	Transfer income [c]	Partly [3,6]
Transfer expenditure	Transfer expenditure [c]	Partly [3,6,11]
Spectators	Number of spectators; indicator of club's size	
UEFA club coefficient	Ranking of UEFA club coefficients; indicator of club's international performance	
League ranking	Club's league position; indicator of club's national performance	[3,6,13,14]
League	Dummies: Premier League, other English leagues, Ligue 1, other French leagues, Bundesliga, other German leagues, Serie A, other Italian leagues, La Liga, other Spanish leagues, other European leagues, South American leagues, other non-European leagues	[1,4,5]
Buying-club characteristics during the previous season		
Arrivals	Number of players that were transferred into the club [c]	Partly [3]
Departures	Number of players that were transferred out the club [c]	Partly [3]

(*Continued*)

Table 1. Continued.

Variable	Description and reason for inclusion	References
Transfer income	Transfer income [c]	Partly [3,6]
Transfer expenditure	Transfer expenditure [c]	Partly [3,6,11]
Spectators	Number of spectators; indicator of club's size	
UEFA club coefficient	Ranking of UEFA club coefficients; indicator of club's international performance	
League ranking	Club's league position; indicator of club's national performance	[3,6,13,14]
League	Dummies: Premier League, other English leagues, Ligue 1, other French leagues, Bundesliga, other German leagues, Serie A, other Italian leagues, La Liga, other Spanish leagues, other European leagues, South American leagues, other non-European leagues	[1,4,5]
Time effects		
Transfer window	Dummy: summer or winter	[6,8]
Transfer seasons	Coded from 1 (season 2015/16) to 7 (season 2021/22)	[4–6,8,9,13]

Notes. MP: minutes of playing time. UEFA: Union of European Football Associations.
[a]Indicators of ability and performance.
[b]Player positions are classified as: attacker (centre-forward, left winger, right winger, second striker); defender (centre-back, left-back, right-back); midfielder (attacking midfield, central midfield, left midfield, right midfield); and goalkeeper.
[c]Indicator of financial status of the team.
[d]Performance metrics are calculated per 1,000 min of playing time (Coates & Parshakov, 2021). References: [1] Ante (2019); [2] Carmichael et al. (1999); [3] Carmichael and Thomas (1993); [4] Coates and Parshakov (2021); [5] Depken and Globan (2021); [6] Dobson and Gerrard (1999); [7] Feess et al. (2004); [8] Garcia-del-Barrio and Pujol (2020); [9] Gerrard and Dobson (2000); [10] Reilly and Witt (1995); [11] McHale and Holmes (2022); [12] Ruijg and van Ophem (2015); [13] Speight and Thomas (1997a); [14] Speight and Thomas (1997b).

The predictors except dummies are scaled before the analysis with a mean of 0 and a standard deviation of 1. We removed highly correlated predictors ($r > 0.8$; Hair, 2009), such as the number of spectators per match (the total sum is included instead), the number of games missed due to players' injuries (injury days and frequencies are included as a combined measure), player squad, and players' goal rate (the total number of goals is included). We assessed the multicollinearity between predictors via the variance inflation factors (VIF, with values below 10 being an indication of no severe multicollinearity concerns) (O'brien, 2007). The multicollinearity criteria are met for all remaining variables except for the dummy variables.

Modeling

We randomly split the *before*-COVID-19 data into a training set (70%, $N = 1{,}903$) and a test set (30%, $N = 816$). The training data are used to train our models, while the test data are used to assess our models.[6] The *during*-COVID-19 data ($N = 793$) are then used for further predictions in order to test whether the transfer market has changed since the beginning of the pandemic.

Model development

We start our analysis with training a simple linear model performed with the 'caret' R package (Kuhn, 2008). Linear models have a high degree of interpretability, whereas they are sensitive to variance and as such may have low predictive accuracy (James et al., 2013). We also train generalized and quantile additive models (GAMs and QAMs), which move beyond linearity while keeping additivity (Hastie & Tibshirani, 2017). As such, these models are much more flexible while maintaining to a large extent the interpretability of simple linear regression models. We use the 'mgcv' R package (Wood & Wood, 2015) to perform the automatic smoothing parameter estimation and select the optimal model with the residual maximum likelihood (REML) score. The QAM further considers the conditional distribution of the highly skewed transfer fees, that is, at the 10th, 25th, 50th, 75th, and 90th percentiles.[7] The estimation is performed with 'qgam' (Fasiolo, Wood, et al., 2020) and visualized by 'mgcViz' R packages (Fasiolo, Nedellec, et al., 2020). Finally, we train a fully flexible model employing the random forest (RF) estimator. The RF is a regression tree-based ensemble model. It is generated by bootstrapped training samples with a random subset of the predictors (James et al., 2013). For training the model, we use the 'randomForest' R package (Liaw & Wiener, 2002) with 500 trees, 2/3 of the training data, and a minimum of five nodes to each tree. Variable importance is assessed via the mean decrease accuracy (MDA) and mean decrease Gini (MDG) (Friedman et al., 2001). MDA measures how much accuracy the model loses by excluding each variable. MDG is based on calculating the loss function per splits of trees. The higher MDA and MDG, the higher the importance of the variable to the model. Ten-fold cross-validations are performed for the training set in all models (James et al., 2013).

Prediction task

Beside the assessment of the model performance using the test data of our *before*-COVID-19 sample, we explore whether the transfer market has changed since the

beginning of the pandemic by applying the fitted models to predict the transfer fees *during* COVID-19. We also test whether the model-predicted results differ for each of the big five leagues (according to the league affiliation of the corresponding buying club). Lastly, we conducted analyses to assess whether and how the transfer market changed for the most expensive vs. mid-level and least expensive players. In this regard, we focus on the groups of players that belong to the <33rd percentile, between the 33rd and 66th percentile, and above the 66th percentile of the transfer fees.

We evaluate the model performance with the R^2 statistic and the root mean squared error (RMSE) in both the test set and the during-COVID-19 set. A higher R^2 indicates a greater explanatory power of the model. The RMSE assesses the square root of the mean of squared differences between the model-predicted values and actual values; a smaller RMSE indicates a better predictive performance (Hyndman & Koehler, 2006).

Results

Descriptive statistics

Table 2 provides an overview of the descriptive statistics of real transfer fees and all predictors for the full sample. It further presents these statistics for transfers that occurred before (n = 2,719) and during the pandemic (n = 793). The average real transfer fee before (during) COVID-19 is €7.72 (€8.40) million. The overall average real transfer fee is €7.88 million (ranging from €0.001 to €217 million with a median of €3.7 million).[8]

When mean summer (winter) real transfer fees before COVID-19 are compared with mean summer (winter) real transfer fees during COVID-19, non-significant differences emerged (summer: M_{Before} = €7.85 million, SD = 11.99 vs. M_{During} = €8.49 million, SD = 12.34; $t(1{,}029) = -1.17$, $P = 0.24$; winter: M_{Before} = €7.10 million, SD = 10.41 vs. M_{During} = €7.95 million, SD = 10.58; $t(233) = -0.85$, $P = 0.40$). There are, however, fewer transfers during both summer and winter during COVID-19 compared to before (Appendix Table 2, see supplementary material).

In what follows, we first assess the relationships between the various predictors and real transfer fees, making use of models that allow us to depart from linear functional forms. Second, we test whether the transfer market has changed since the beginning of the pandemic by assessing the predictive power of our models trained with *before*-COVID-19 data.

Predictors of transfer fees

Table 3 reports the findings for the simple linear regression model as well as the GAM and QAM. Several estimates of our simple linear regression model are fully in line with previous findings and hardly differ compared to the GAM and QAM. For instance, (i) compared to the position of attacker, all other positions are negatively associated with transfer fees. Also, (ii) UEFA Champions League appearance as well as (iii) the involvement of a Premier League club as the buying or selling club are associated with higher transfer fees.

For many predictors, however, moving beyond linearity (with both GAM and QAM) and modeling quantiles instead of the mean (with QAM) reveals several important

Table 2. Descriptive statistics.

Variable	Full sample (n = 3,512)	Before COVID-19 (n = 2,719)	During COVID-19 (n = 793)
Real transfer fee (EUR million)	7.88 ± 11.81	7.72 ± 11.74	8.40 ± 12.04
Player characteristics			
Age (years)	24.49 ± 3.61	24.51 ± 3.59	24.42 ± 3.66
Height (cm)	182.64 ± 6.57	182.49 ± 6.64	183.14 ± 6.31
Nationality			
Europe	2,544 (72.4%)	1,962 (72.2%)	582 (73.4%)
Asia	67 (1.9%)	50 (1.8%)	17 (2.1%)
Africa	451 (12.8%)	370 (13.6%)	81 (10.2%)
North America	23 (0.7%)	11 (0.4%)	12 (1.5%)
South America	427 (12.2%)	326 (12.0%)	101 (12.7%)
Position			
Attacker	1,243 (35.4%)	955 (35.1%)	288 (36.3%)
Defender	1,074 (30.6%)	820 (30.2%)	254 (32%)
Midfielder	992 (28.2%)	792 (29.1%)	200 (25.2%)
Goalkeeper	203 (5.8%)	152 (5.6%)	51 (6.4%)
Remaining contract duration (days)	680.05 ± 361.39	698.14 ± 365.72	618.05 ± 339.11
Player performance			
UEFA Champions League (yes)	566 (16.1%)	371 (13.6%)	195 (24.6%)
Appearances	29.20 ± 11.27	29.52 ± 11.23	28.12 ± 11.34
Substitution on	5.19 ± 5.28	5.06 ± 5.24	5.63 ± 5.38
Substitution off	6.67 ± 5.84	6.38 ± 5.63	7.67 ± 6.45
Minutes played	2,147 ± 992	2,186 ± 998	2,012 ± 961
Points[a]	21.12 ± 13.14	20.27 ± 11.86	24.06 ± 16.49
Goals[a]	2.01 ± 2.50	1.99 ± 2.56	2.03 ± 2.29
Assists[a]	1.32 ± 1.39	1.31 ± 1.37	1.35 ± 1.45
Yellow cards[a]	1.99 ± 1.52	1.99 ± 1.47	1.99 ± 1.68
Injury frequency	2.93 ± 3.58	2.89 ± 3.64	3.06 ± 3.33
Days in injury	98.57 ± 133.26	96.35 ± 133.65	106.18 ± 131.72
Injury proneness	1.13 ± 1.62	1.07 ± 1.53	1.32 ± 1.86
Selling club characteristics			
Arrivals	23.92 ± 11.88	24.39 ± 12.34	22.3 ± 10.01
Departures	24.48 ± 12.71	24.96 ± 13.19	22.83 ± 10.77
Transfer income (EUR million)	41.46 ± 46.5	43.9 ± 49.48	33.1 ± 33.11
Transfer expenditure (EUR million)	40.45 ± 52.92	43.97 ± 56.49	28.35 ± 35.67
Spectators (million)	0.42 ± 0.33	0.49 ± 0.32	0.17 ± 0.24
UEFA club coefficient	20.72 ± 33.77	21.00 ± 34.26	19.73 ± 32.06
League ranking	8.27 ± 5.43	8.20 ± 5.44	8.52 ± 5.38
League			
Premier League	433 (12.3%)	340 (12.5%)	93 (11.7%)
Other English leagues	180 (5.1%)	145 (5.3%)	35 (4.4%)
Ligue 1	432 (12.3%)	344 (12.7%)	88 (11.1%)
Other French leagues	97 (2.8%)	84 (3.1%)	13 (1.6%)
Bundesliga	479 (13.6%)	374 (13.8%)	105 (13.2%)
Other German leagues	113 (3.2%)	87 (3.2%)	26 (3.3%)
Serie A	562 (16.0%)	428 (15.7%)	134 (16.9%)
Other Italian leagues	122 (3.5%)	94 (3.5%)	28 (3.5%)
La Liga	278 (7.9%)	211 (7.8%)	67 (8.4%)
Other Spanish leagues	98 (2.8%)	77 (2.8%)	21 (2.6%)
Other European leagues	647 (18.4%)	487 (17.9%)	160 (20.2%)
South American leagues	48 (1.4%)	35 (1.3%)	13 (1.6%)
Other non-European leagues	23 (0.7%)	13 (0.5%)	10 (1.3%)
Buying club characteristics			
Arrivals	24.21 ± 12.55	24.43 ± 12.84	23.46 ± 11.47
Departures	25.09 ± 13.52	25.28 ± 13.75	24.41 ± 12.68
Transfer income (EUR million)	35.71 ± 44.31	38.69 ± 47.27	25.48 ± 29.98
Transfer expenditure (EUR million)	49.01 ± 52.69	50.85 ± 55.37	42.66 ± 41.68
Spectators (million)	0.43 ± 0.32	0.51 ± 0.30	0.18 ± 0.25
UEFA club coefficient	20.22 ± 33.26	20.48 ± 33.64	19.32 ± 31.91
League ranking	8.28 ± 5.83	8.13 ± 5.66	8.78 ± 5.50

(*Continued*)

Table 2. Continued.

Variable	Full sample (n = 3,512)	Before COVID-19 (n = 2,719)	During COVID-19 (n = 793)
League			
Premier League	662 (18.9%)	491 (18.1%)	171 (21.6%)
Other English leagues	125 (3.6%)	115 (4.2%)	10 (1.3%)
Ligue 1	428 (12.2%)	359 (13.2%)	69 (8.7%)
Other French leagues	21 (0.6%)	18 (0.7%)	3 (0.4%)
Bundesliga	531 (15.1%)	415 (15.3%)	116 (14.6%)
Other German leagues	90 (2.6%)	74 (2.7%)	16 (2%)
Serie A	723 (20.6%)	526 (19.3%)	197 (24.8%)
Other Italian leagues	103 (2.9%)	77 (2.8%)	26 (3.3%)
La Liga	418 (11.9%)	325 (12.0%)	93 (11.7%)
Other Spanish leagues	21 (0.6%)	13 (0.5%)	8 (1.0%)
Other European leagues	274 (7.8%)	211 (7.8%)	63 (7.9%)
South American leagues	45 (1.3%)	38 (1.4%)	7 (0.9%)
Other non-European leagues	71 (2.0%)	57 (2.1%)	14 (1.8%)
Time effects			
Transfer window (summer)	2,903 (82.7%)	2,253 (82.9%)	650 (82%)
Transfer seasons			
2015/16	463 (13.2%)	463 (17.0%)	
2016/17	532 (15.2%)	532 (19.6%)	
2017/18	569 (16.2%)	569 (20.9%)	
2018/19	555 (15.8%)	555 (20.4%)	
2019/20	600 (17.1%)	600 (22.1%)	
2020/21	346 (9.9%)		346 (43.6%)
2021/22	447 (12.7%)		447 (56.4%)

Notes. MP: minutes of playing time. UEFA: Union of European Football Associations.
[a]Performance metrics are calculated per 1,000 min of playing time. The descriptive statistics cover all transfers for which remaining contract duration data is available among the five European major football leagues between 2015/16 and 2021/22. Data are presented as mean ± standard deviation or numbers (%) if they are at a categorical level.

differences compared to the simple linear-regression estimates. For instance, (iv) height is only a comparably precise (positive) predictor for *higher* quantiles (i.e. the 75th and 90th quantiles), while (v) being a South American player (instead of being a player with a European nationality) is only a comparably precise (positive) predictor for *lower* quantiles (i.e. the 10th, 25th, and 50th quantiles). Likewise, (vi) the number of goals is only a comparably precise (positive) predictor for *higher* quantiles (i.e. the 75th and 90th quantiles), while (vii) the number of yellow cards is only a comparably precise (negative) predictor for *lower* quantiles (i.e. until the 50th quantile).

Furthermore, the relation between several predictors and transfer fees is better described by introducing flexible functional forms. For illustrating such non-linear effects, we present some of the most important variables for which we achieved comparably precise estimates in Figure 1.

In line with our expectations and earlier studies, we find an inverted u-shaped relation between (viii) player age and transfer fees. Also, we find that both (ix) remaining contract duration and (x) the number of appearances are positive predictors of transfer fees. Remarkably, all marginal effects are larger for lower quantiles. This can also be observed for (xi) selling-club arrivals ([xii] expenditures), which are negative (positive) predictors of transfer fees. In contrast, the pattern between (xiii) selling-club income and transfer fees is similar across all quantiles (with a sharp increase at low values, a plateau at mid values, and increasing marginal effects at high values). Finally, we observe a u-shaped relation between (xiv) buying-club departures and transfer fees (particularly for the 75th and 90th quantiles), a stepwise relationship between (xv) buying-club expenditure

Table 3. Estimates of linear, generalized additive, and quantile additive models.

	Linear regression	GAM	QAM 10th	QAM 25th	QAM 50th	QAM 75th	QAM 90th
Dependent variable: log-transformed real transfer fees in EUR							
Intercept	15.63(0.09)***	15.43(0.10)***	14.25(0.16)***	14.94(0.11)***	15.5(0.09)***	16(0.08)***	16.37(0.08)***
Player characteristics							
Age	−0.06(0.02)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Height	0.04(0.02)*	0.05(0.02)**	0.03(0.03)	0.04(0.02)	0.04(0.02)**	0.07(0.02)***	0.10(0.02)***
Nationality (ref: Europe)							
Africa	0.04(0.06)	0.07(0.06)	0.04(0.10)	0.08(0.07)	0.06(0.05)	0.04(0.05)	0.02(0.05)
Asia	−0.03(0.14)	−0.01(0.15)	0.06(0.23)	0.09(0.17)	0.07(0.12)	−0.03(0.10)	−0.16(0.11)
North America	0.08(0.29)	0.18(0.29)	0.06(0.40)	0.05(0.30)	0.20(0.25)	0.28(0.23)	0.34(0.24)
South America	0.21(0.06)***	0.21(0.07)***	0.35(0.10)***	0.25(0.07)***	0.15(0.06)***	0.09(0.06)	0.06(0.06)
Position (ref: Attacker)							
Defender	−0.35(0.07)***	−0.37(0.07)***	−0.44(0.11)***	−0.41(0.08)***	−0.34(0.06)***	−0.32(0.06)***	−0.31(0.07)***
Goalkeeper	−0.59(0.12)***	−0.54(0.13)***	−0.67(0.19)***	−0.66(0.15)***	−0.52(0.11)***	−0.53(0.11)***	−0.54(0.12)***
Midfielder	−0.14(0.05)***	−0.17(0.06)***	−0.16(0.08)*	−0.19(0.06)***	−0.16(0.05)***	−0.13(0.05)***	−0.11(0.05)**
Remaining contract duration	0.31(0.02)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Player performance							
UEFA Champions League	0.18(0.06)***	0.20(0.07)***	0.32(0.10)***	0.20(0.07)***	0.15(0.06)***	0.13(0.05)**	0.13(0.06)**
Appearances	0.18(0.08)**	Non-linear	Non-linear	Non-linear	Non-linear*	Non-linear	Non-linear
Substitution on	−0.07(0.04)*	−0.08(0.04)**	−0.10(0.06)*	−0.06(0.04)	−0.06(0.03)*	−0.08(0.03)***	−0.10(0.03)***
Substitution off	−0.01(0.03)	−0.02(0.03)	−0.04(0.04)	−0.05(0.03)	−0.03(0.02)	−0.03(0.02)	−0.01(0.03)
Minutes played	0.04(0.08)	0.07(0.08)	−0.03(0.13)	0.03(0.10)	0.07(0.07)	0.08(0.06)	0.08(0.07)
Points[a]	0.05(0.02)*	0.04(0.03)*	0.05(0.04)	0.04(0.03)	0.03(0.02)	0.03(0.02)	0.03(0.02)
Goals[a]	0.03(0.02)	0.04(0.02)	0.03(0.02)	0.02(0.02)	0.04(0.02)	0.07(0.03)**	0.07(0.03)**
Assists[a]	0.02(0.02)	0.02(0.02)	0.02(0.03)	0.02(0.02)	0.01(0.02)	0.01(0.02)	0.02(0.02)
Yellow cards [a]	−0.06(0.02)***	−0.06(0.02)***	−0.08(0.04)**	−0.06(0.02)**	−0.03(0.02)*	−0.01(0.02)	−0.01(0.02)
Injury proneness	0.01(0.02)	Non-linear	Linear	Non-linear	Non-linear**	Non-linear	Non-linear
Selling-club characteristics							
Arrivals	−0.10(0.07)	Non-linear*	Non-linear**	Non-linear	Non-linear	Linear*	Linear*
Departures	0.12(0.07)*	0.09(0.08)	0.04(0.12)	0.04(0.08)	0.06(0.06)	0.10(0.06)*	0.12(0.07)*
Transfer income	0.14(0.03)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Transfer expenditure	−0.01(0.03)	Non-linear*	Non-linear**	Linear**	Non-linear**	Linear	Linear
Spectators	0.17(0.03)***	Non-linear***	Non-linear***	Non-linear***	Non-linear**	Linear*	Linear
UEFA club coefficient	0.01(0.03)	0.01(0.03)	−0.07(0.05)	−0.01(0.03)	0.01(0.03)	0.02(0.02)	0.01(0.03)
League ranking	0.02(0.02)	0.01(0.02)	−0.01(0.04)	0.01(0.03)	0.01(0.02)	0.01(0.02)	0.01(0.02)
League (ref: Premier League)							
Other English leagues	−0.14(0.11)	−0.13(0.12)	−0.02(0.18)	−0.13(0.13)	−0.19(0.10)*	−0.17(0.09)*	−0.19(0.10)**
Ligue 1	−0.18(0.09)**	−0.20(0.10)**	−0.06(0.15)	−0.13(0.11)	−0.21(0.08)**	−0.23(0.07)***	−0.20(0.08)**

(*Continued*)

Table 3. Continued.

	Linear regression	GAM	QAM 10th	QAM 25th	QAM 50th	QAM 75th	QAM 90th
Other French leagues	−0.4(0.14)***	−0.16(0.15)	0.12(0.23)	−0.01(0.16)	−0.17(0.12)	−0.24(0.11)**	−0.24(0.12)**
Bundesliga	−0.34(0.09)***	−0.31(0.10)***	−0.33(0.15)**	−0.29(0.11)**	−0.27(0.08)***	−0.26(0.08)***	−0.25(0.08)***
Other German leagues	−0.78(0.13)***	−0.60(0.15)***	−0.34(0.22)	−0.56(0.16)***	−0.68(0.13)***	−0.64(0.12)***	−0.58(0.13)***
Serie A	−0.35(0.10)***	−0.30(0.12)***	−0.32(0.19)*	−0.37(0.13)***	−0.31(0.10)***	−0.26(0.09)***	−0.27(0.10)***
Other Italian leagues	−1.21(0.14)***	−0.73(0.16)***	−0.87(0.29)***	−0.64(0.19)***	−0.65(0.14)***	−0.65(0.13)***	−0.6(0.14)***
La Liga	−0.16(0.09)*	−0.22(0.10)**	−0.28(0.15)*	−0.37(0.11)***	−0.26(0.09)***	−0.14(0.08)*	−0.10(0.09)
Other Spanish leagues	−0.58(0.14)***	−0.28(0.16)*	−0.25(0.24)	−0.36(0.17)**	−0.38(0.13)***	−0.28(0.13)**	−0.13(0.14)
Other European leagues	−0.38(0.09)***	−0.26(0.11)**	−0.15(0.16)	−0.18(0.11)	−0.27(0.09)***	−0.32(0.08)***	−0.29(0.09)***
South American leagues	0.41(0.19)**	0.46(0.22)**	0.70(0.28)**	0.54(0.21)***	0.35(0.17)**	0.26(0.19)	0.40(0.20)**
Other non-European leagues	−0.72(0.28)**	−0.16(0.28)	−0.26(0.75)	−0.07(0.38)	−0.05(0.25)	−0.03(0.22)	−0.01(0.23)
Buying-club characteristics							
Arrivals	0.09(0.07)	Non-linear	Non-linear	Linear	Linear	Linear	Linear
Departures	−0.16(0.08)**	Non-linear***	Non-linear*	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Transfer income	0.14(0.02)***	0.06(0.03)**	0.11(0.04)***	0.07(0.03)**	0.05(0.02)**	0.02(0.02)	0.03(0.02)
Transfer expenditure	0.25(0.03)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Spectators	0.19(0.03)***	0.07(0.03)**	0.07(0.05)	0.10(0.03)***	0.08(0.03)***	0.07(0.02)***	0.06(0.03)**
UEFA club coefficient	−0.01(0.02)	Non-linear**	Linear**	Linear*	Non-linear	Non-linear	Non-linear
League ranking	0.02(0.02)	0.01(0.03)	0.01(0.04)	0.02(0.03)	0.01(0.02)	0.01(0.02)	0.01(0.02)
League (ref: Premier League)							
Other English leagues	−0.53(0.11)***	−0.03(0.13)	−0.12(0.19)	−0.05(0.14)	−0.05(0.11)	−0.04(0.10)	−0.05(0.10)
Ligue 1	−0.24(0.08)***	0.01(0.09)	−0.03(0.13)	−0.01(0.10)	−0.01(0.08)	−0.07(0.07)	−0.11(0.08)
Other French leagues	−1.05(0.24)***	−0.46(0.27)*	−0.79(0.37)**	−0.75(0.32)**	−0.42(0.25)*	−0.33(0.22)	−0.36(0.25)
Bundesliga	−0.25(0.08)***	−0.27(0.09)***	−0.33(0.13)**	−0.29(0.10)***	−0.26(0.07)***	−0.23(0.07)***	−0.24(0.07)***
Other German leagues	−1.83(0.14)***	−1.13(0.16)***	−1.33(0.25)***	−1.24(0.19)***	−1.17(0.15)***	−1.11(0.14)***	−0.87(0.17)***
Serie A	−0.06(0.09)	0.07(0.11)	0.25(0.17)	0.17(0.12)	0.07(0.09)	0.02(0.09)	0.03(0.10)
Other Italian leagues	−1.51(0.14)***	−1.02(0.18)***	−3.37(2.28)	−0.69(0.23)***	−0.57(0.16)***	−0.64(0.15)***	−0.69(0.16)***
La Liga	−0.26(0.08)***	−0.01(0.09)	−0.06(0.13)	−0.02(0.10)	−0.02(0.08)	−0.04(0.07)	−0.03(0.08)
Other Spanish leagues	−0.89(0.27)***	−0.03(0.30)	−0.25(0.40)	−0.36(0.36)	−0.01(0.30)	0.10(0.23)	0.03(0.23)
Other European leagues	−0.65(0.1)***	−0.16(0.11)	−0.29(0.16)*	−0.27(0.12)**	−0.24(0.10)**	−0.18(0.09)*	−0.14(0.10)
South American leagues	−0.08(0.17)	0.21(0.19)	0.23(0.27)	0.28(0.20)	0.22(0.16)	0.19(0.16)	0.18(0.16)
Other non-European leagues	0.08(0.14)	0.27(0.16)*	0.39(0.22)*	0.47(0.16)***	0.29(0.12)**	0.16(0.12)	0.13(0.13)
Time effects							
Transfer window (ref: summer)	−0.03(0.05)	0.03(0.05)	0.02(0.08)	0.03(0.06)	0.04(0.05)	0.04(0.04)	−0.01(0.05)
Transfer seasons[b]	0.06(0.03)**	Non-linear	Non-linear	Linear**	Linear**	Linear	Linear

Notes. GAM: generalized additive model. MP: minutes of playing time. QAM: quantile additive model. ref: reference category. UEFA: Union of European Football Associations.

[a]Performance metrics are calculated per 1,000 min of playing time.

[b]Transfer seasons are coded as 1 (season 2015/16) to 5 (season 2019/20). The final models are fitted for the full before-COVID-19 data set (n = 2,719). Model goodness-of-fit: Linear regression (adjusted R^2 = 0.59, $F(60, 2{,}658)$ = 66.0, $P < 0.001$); GAM (adjusted R^2 = 0.67, scale estimate [squared residual standard error] = 0.70, REML = 3,522); QAM (10th to 90th percentile: adjusted R^2 = 0.65, 0.66, 0.67, 0.65, and 0.64, respectively; deviance explained = 81%, 64%, 59%, 69%, and 85%, respectively; REML = 4,860, 3,803, 3,316, 3,283, and 3,670, respectively). The most important additive smooth effects for which we achieved comparably precise estimates are shown in Figure 1. Standard errors are presented in parentheses. Significance levels: $^{*}P < 0.1$, $^{**}P < 0.05$, $^{***}P < 0.01$.

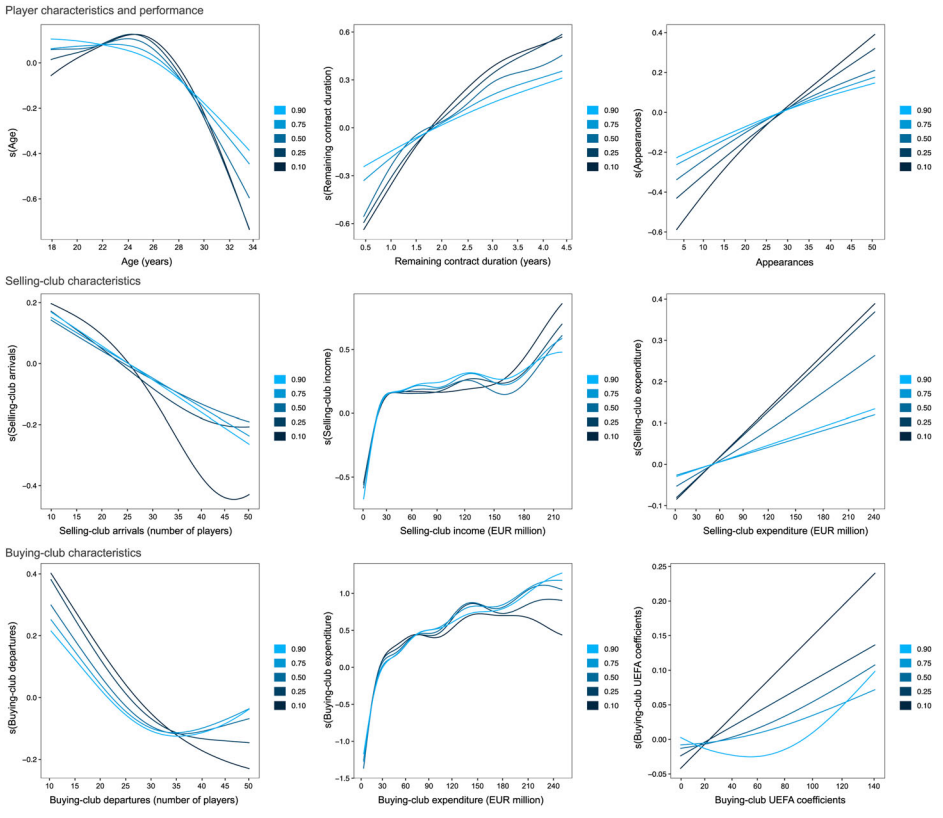

Figure 1. Selected quantile additive smooth effects of non-linear predictors of transfer fees. Displayed are quantile additive smooth effects of non-linear predictors of transfer fees according to the estimations shown in Table 3. For specifications of the variables, see Table 1. In order to avoid misinterpretation based on outliers, all x-axis data point ranges were trimmed excluding any observations below the 1st and beyond the 99th percentile of transfers in the respective category. For arrivals and departures, the upper range was trimmed further to 50 as the maximum for the same purpose. 10% to 90% quantiles are indicated by 0.10 (dark blue) to 0.90 (light blue). The final model is fitted for the full before-COVID-19 data set ($n = 2{,}719$, estimator: residual maximum likelihood [REML]).

and transfer fees, as well as a positive effect of the (xvi) buying-club UEFA coefficient (which seems to have an inverted u-shape, yet non-significant for the 75th and 90th quantiles).[9]

Prediction task

Figure 2 provides an overview of the predictive performance of our three models for the *before*-COVID-19 test data (Figure 2(A)), the *during*-COVID-19 data (Figure 2(B)), and the *during*-COVID-19 data separated by leagues (Figure 2(C)). As indicated by Figure 2 (A), the simple linear regression model explains 57% of the variance in transfer fees

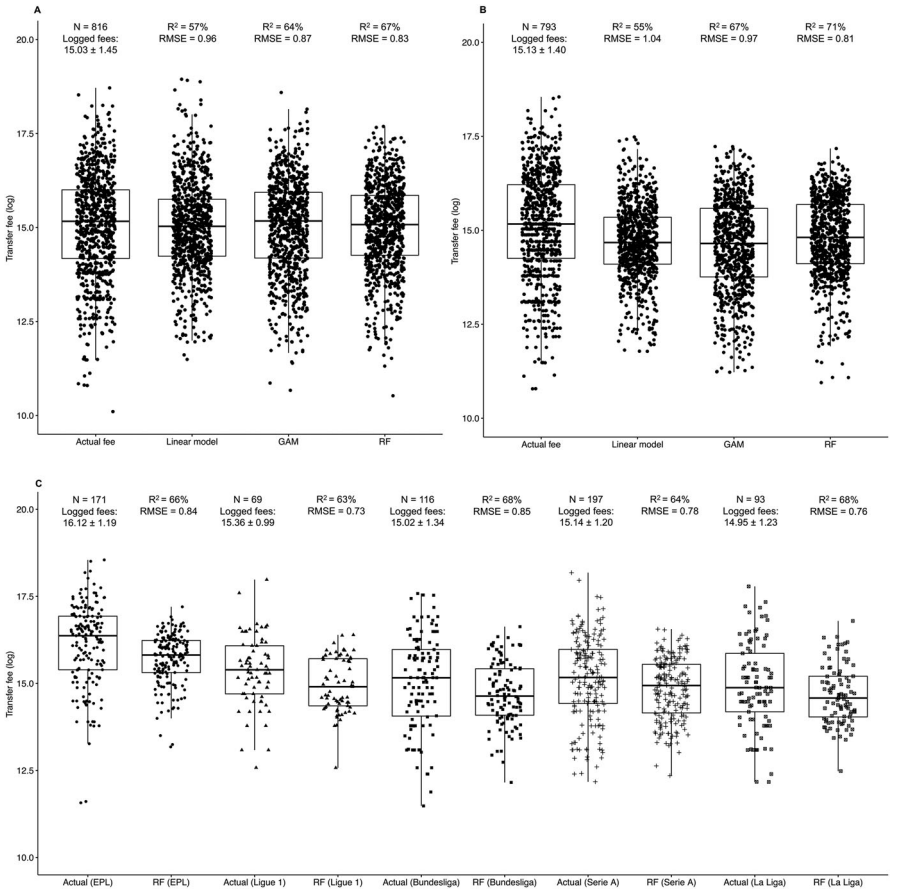

Figure 2. Predicted versus actual transfer fees and model performance. GAM: generalized additive model. EPL: English Premier League. RF: random forest. RMSE: root mean squared error. log: log-transformed transfer fees. A: before-COVID-19 predictions (test set). B: during-COVID-19 predictions. C: during-COVID-19 predictions for each of the big five leagues (classified based on the buying club). Mean test difference between actual log-transformed transfer fees and RF-predicted log-transformed transfer fees are as follows: for A, $t(815) = 0.91$, $P = 0.40$; for B, $t(792) = 12.01$, $P < 0.001$; and for C, EPL, $t(170) = 7.52$, $P < 0.001$; Ligue 1: $t(68) = 5.53$, $P < 0.001$; Bundesliga: $t(115) = 4.30$, $P < 0.001$; Serie A: $t(196) = 5.72$, $P < 0.001$; and La Liga: $t(92) = 3.70$, $P < 0.001$.

(RMSE = 0.96) in our test data. The GAM improves the predictive performance (R^2 = 64%, RMSE = 0.87) already substantially. The RF performs best (R^2 = 67%, RMSE = 0.83).

Figure 3 displays the relative importance of all predictors according to the RF. Amongst the most important club-related variables are expenditures, number of spectators, and the country of the buying club, as well as income and expenditure of the selling club. As expected, the most important player characteristic is the remaining contract duration. This is why we include this variable in all main-model specifications.[10]

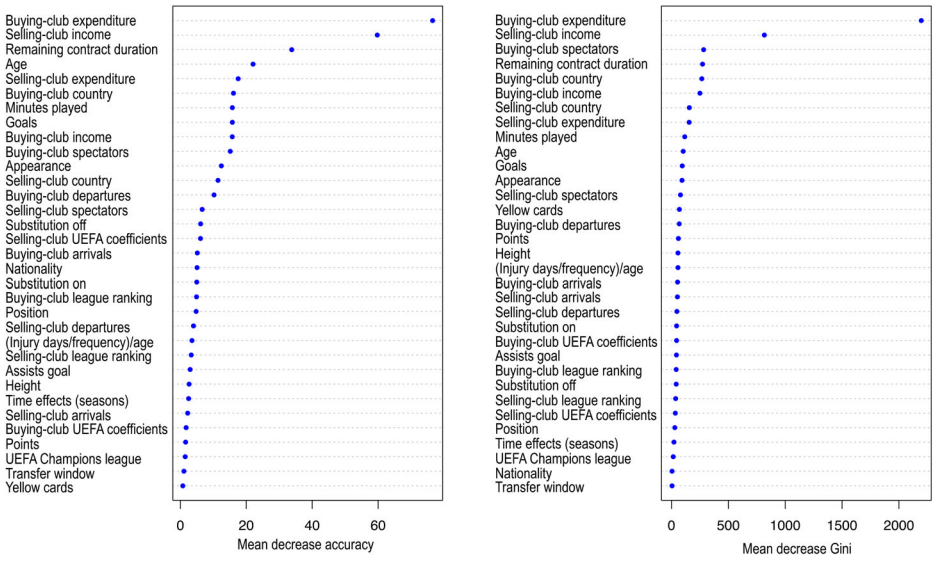

Figure 3. Variable importance according to the random forest estimator. The final model is fitted for the full before-COVID-19 data set (n = 2,719); for the training model, the optimal number of variables sampled for splitting at each node [mtry] is 31. The mean decrease accuracy (MDA, unitless, often termed %IncMSE) measures how much accuracy the model loses by excluding each variable based on the out-of-bag data; the mean decrease Gini (MDG, unitless, often termed IncNodePurity) is based on calculating the loss function per splits of trees. Both the higher the MDA and the MDG, the higher the importance of the variable to the model. UEFA: Union of European Football Associations.

In order to test whether the transfer market has changed since the beginning of the pandemic, we make use of our models trained with *before*-COVID-19 data and predict the transfer fees *during* COVID-19 (Figure 2(B)). Overall, the simple linear regression model explains 55% of the variance in transfer fees (RMSE = 1.04), while the GAM improves the predictive performance to $R^2 = 67\%$ (RMSE = 0.97). The RF model performs best with $R^2 = 71\%$ (RMSE = 0.81). However, while we do not observe any statistically significant mean differences between the (log-transformed) actual fees and the (log-transformed) fees predicted with the RF estimator *before* COVID-19, all models *underestimate* the transfer fees paid *during* COVID-19. Even for the best-performing RF model, there is a significant difference between the predicted ($M = 14.80$, SD = 1.05) and the actual log-transformed fees ($M = 15.13$, SD = 1.40; $t(792) = 12.20$, $P < 0.001$) with an underestimation of about 2.2%. While this underestimation can be observed across leagues, the difference is the largest for the Premier League (2.5%; Figure 2(C)).[11] These findings suggest that clubs paid *higher* transfer fees than predicted during COVID-19.

Lastly, we conducted separate analyses for the most expensive, mid-level priced, and least expensive players (Appendix Figures 4–6, see supplementary material). The results reveal that transfer fees are underestimated for both high-priced (>66th percentile) and mid-level priced transfers (33rd to 66th percentile), while there is an overestimation for low-priced transfers (<33rd percentile). In agreement with previous findings on the Premier League teams' spending patterns, the largest underestimation was observed for the group of high-priced players recruited by Premier League teams.

Discussion

The purpose of our study is to investigate the relevance of a large set of determinants of players' transfer fees in the big five leagues before and during the pandemic. Several of the study's findings are in line with the findings from previous studies employing simple linear regression models. For instance, we observe higher fees paid for attackers (e.g. Feess et al., 2004; Reilly & Witt, 1995), players with UEFA Champions League appearance (e.g. Dobson & Gerrard, 1999; Feess et al., 2004; Gerrard & Dobson, 2000; Reilly & Witt, 1995; they used a similar predictor named full international caps), and players traded from, or to, a Premier League club (Depken & Globan, 2021).

However, introducing flexible functional forms is insightful. In what follows, we select examples that illustrate how moving beyond linearity and modeling quantiles can be revealing for both research and practice.

First, we identify several quantile-dependent predictors. We find that a player's height is only a comparably precise (positive) predictor for higher quantiles. This might help explain previous inconsistencies in studies that do not consider player clusters. Ante (2019), for example, finds positive, while Ruijg and van Ophem (2015) find negative relations of height with transfer fees. Higher fees for South American players are consistently evidenced across our linear model and previous studies that employed linear models without considering specific quantiles (Ante, 2019; Depken & Globan, 2021; Feess et al., 2004). Yet, our QAM analysis reveals that the estimator is only comparably precise for quantiles of transfer fees up to the 50th percentile, implying that debatable player discrimination issues (e.g. see a review conducted by Frick [2007] and recent evidence of customer-based racial and nationality-based discrimination, as revealed by Quansah et al. [2022]) may be particularly relevant for highly priced players. Furthermore, we find a positive time effect although the time estimate lacks precision for higher quantiles. Thus, time may be a less relevant predictor of transfer fees for the most expensive players compared to players in the lower quantiles.

Second, using GAM and QAM allows us to better depict the non-linear relations between several predictors and transfer fees. For instance, our analysis reveals a negative relation between buying-club departures and transfer fees particularly for the 10th percentile. Towards the higher percentiles, the relation resembles a u-shape, implying potential differences in bargaining power compared to the clubs that recruit players from the 10th percentile. Moreover, we may observe a transfer premium paid to clubs with comparably higher transfer-related income. Since the marginal effect increases as income increases (beyond about €160-170 million), transfer income might be interpreted as a signal for a player's evidenced quality in the past.

Age as an inverted u-shaped predictor of transfer fees is well-evidenced (Appendix Table 1, see supplementary material) and our analyses reveal consistently u-shaped relations across quantiles. Regarding the age-relative injury proneness indicator used in the present study, no clear statement can be made other than that the relation seems to be non-linear in nature. Other variables are more important than injury proneness in predicting transfer fees (as indicated by the RF). Particularly including the measure of number of appearances seems important in this context, because the argument put forward by Herberger and Wedlich (2017) to consider injury as a marker for the number of matches played is not convincing, from a conceptual perspective. Thus,

their findings on positive relations between days in injury and market value should be reconsidered against the background of the approach of the present study (controlling for important confounders; relating injuries to a player's age).

Furthermore, several predictors are positively related to transfer fees, but at decreasing margins: remaining contract duration, number of appearances, selling-club income, and buying-club expenditure. Previous authors use the remaining contract duration as a proxy for a selling-club's bargaining power and consistently find positive relations with transfer fees (Coates & Parshakov, 2021; Feess et al., 2004; Garcia-del-Barrio & Pujol, 2020; McHale & Holmes, 2022). We contribute to these findings by showing that margins decrease at higher levels of remaining contract duration (about 1.5-2 years), implying that remaining contract duration is a particularly important predictor when contracts end within less than two years (this is where the slopes are the steepest). Indeed, when the remaining contract duration is relatively short, clubs may want to sell a player and ask for a premium because at the moment when the contract ends, the player is attractive to many clubs (since no fee needs to be paid at all). A buying club may want to recruit a player before this situation occurs to avoid competition with other interested clubs. The club may hence give up bargaining power and pay a higher fee. The results regarding the number of appearances may partly suggest that a player's extraordinary part-time effectiveness (*vs*. full-time presence on the pitch) might add specific value to the team (e.g. a player who is a late substitute and scores goals when needed towards the end of the game).

Third, the RF provides valuable insights into the importance of the predictors of transfer fees. Our findings highlight the importance of a club's financial status (transfer income and expenditure; McHale & Holmes, 2022), size (number of spectators), negotiating power (remaining contract duration; Coates & Parshakov, 2021) as well as a player's experience and potential (e.g. age; Dobson & Gerrard, 1999) and key performance (e.g. goals and minutes played; Carmichael & Thomas, 1993; Ruijg & van Ophem, 2015). The findings may help researchers select relevant variables and statistical tools (note that RF had the highest explanatory power among the tools used) that aim to predict transfer fees in future studies.

Lastly, we provide new insights into how COVID-19 influenced the transfer market. In previous studies, authors largely suggest that club revenues, player market values, player wages, and overall transfer expenses have decreased (Drewes et al., 2021; Parnell et al., 2021; Quansah et al., 2021). Interestingly, we find that the *average* transfer fees during COVID-19 are not lower compared to before COVID-19. Moreover, we find that the clubs (particularly those from the English Premier League) paid higher transfer fees than predicted during COVID-19. This could imply that the clubs may have perceived COVID-19 as a short-term temporary shock and that they may have expected income, such as broadcast revenues, to continue to rise or to at least not fall substantially after the pandemic subsided.[12] Moreover, club managers might have strategically focused their efforts on particular players, thus giving up bargaining power to exactly recruit those players that fit their squad. Looking closely into the models' degree of underestimation, players with highest differences between actual and predicted fees (top-ten differences: €40-88 million) were mostly either traded to English Premier League clubs (7 out of 10), where teams are supported by investors that might be less affected by COVID-19-related financial constraints, such as Chelsea F.C. (signing Kai Havertz and

Romelu Lukaku) and Manchester City (signing Jack Grealish and Rúben Dias); or to teams that are known to have financial problems today, and overspending might have contributed to this (e.g. Juventus F.C. for the Dusan Vlahovic transfer). Indeed, 38 out of 59 transfers with real fees over €25 million were transferred into the English Premier League (top buying clubs: Arsenal F.C. [5], Aston Villa F.C. [5], and Chelsea F.C. [5]), followed by the Serie A (8 transfers).

Conclusion

A football player's transfer is a complex bargaining process involving multiple parties. Our findings offer some practical implications for a football club's management. By inserting the corresponding values for player and club characteristics into our model, a club manager might derive a player-specific predicted transfer fee. This estimated fee could be contrasted with the market value of that specific player. Moreover, such procedure might be particularly useful in combination with industry-derived solutions (e.g. www.realanalytics.org) to predict team performances based on the decision whether a player is signed or not.

However, our study has several limitations. Even though we mostly focus on lagged predictors, we are cautious and refrain from inferring causality from our analysis. For instance, since the selected subsample of transferred players with published fees does not represent a random sample from all players (e.g. non-transferred players), our study could suffer from sample selection issues (Frick, 2007; Ruijg & van Ophem, 2015). Future research may consider this by applying, for instance, the Heckman sample selection estimator to correct the selection bias (Carmichael et al., 1999; Depken & Globan, 2021). For the present sample, however, it is not feasible to identify *every* player who *could* have been possibly traded to one of the teams from the big-five leagues.

Furthermore, we used *conditional* quantile additive models in our analyses. While revealing the relationship between covariates and the unconditional distribution of the dependent variable could be insightful in the present context (e.g. Carrieri et al., 2018), using an unconditional quantile *non-parametric additive* model is challenging.

Finally, while we include a comprehensive set of independent variables in our model (avoiding strong multicollinearity), we cannot rule out that the set is incomplete. For instance, factors such as the pressure from players or agents, particular contractual conditions (which are often not made public), and the urgency to recruit a new player (for several reasons, such as injured players from the own squad), as well as the financial obligations that the selling club must meet (e.g. avoid illiquidity) might be relevant variables. With regards to player-related data, Kharrat et al. (2020) develop so-called plus-minus scores for players, where a player's performance is related to his ability for either changing the net expected goals of a team or changing the results of teams by affecting the expected points of a team. Such scores (see also Liu et al., 2020; McHale & Relton, 2018) and other advanced performance metrics (e.g. McHale & Holmes, 2022) might be used instead of, or in addition to, the variables that we considered.

Notes

1. While the website is often used to scrape *transfer market value* figures, the present study considers *transfer fee* figures (see Quansah et al., 2021, for conceptual differences between the

two). Expert and crowdsource-based knowledge is needed to get these data, because there is no official register. Most transfer fees for players from outside Europe and for less popular players are not publicly accessible otherwise. Coates and Parshakov (2021), for example, relate market value to actual transfer fees and find that market value is a positive predictor (with a particular underestimation of the value of players with national team experience; see also Depken & Globan, 2021, for an alternative approach of considering deviations between the two as the dependent variable, a variable that is called *transfer premium*).

2. Even though this could eventually bias some predictions, we prefer this approach since we end up with a comparably homogenous sample of players.
3. The eligible transfers are scraped using Python packages (e.g. *requests*, *lxml*, *openyxl*), where a web scraping approach is followed (Landers et al., 2016). The web scraping codes are available from the authors on request under a GitHub private repository.
4. Based on previous findings, remaining contract duration is both theoretically and statistically of great importance to be included in the models (Coates & Parshakov, 2021; Feess et al., 2004; Garcia-del-Barrio & Pujol, 2020; McHale & Holmes, 2022). In our own analyses, we find that missing information on the remaining contract duration is not random, but connected to time trends (e.g. there are only 22 out of 577 transfers in 2008/09 with information on remaining contract duration, while there are 463 out of 621 transfers with information on remaining contract duration in 2015/16). Thus, in our prediction analyses, we exclusively consider transfers from the season 2015/16 onwards, that is, the time period for which information about remaining contract duration is mostly available.
5. As a complex contact sport, professional football players have a high injury rate (Hawkins et al., 2001; Pfirrmann et al., 2016). Moreover, studies show that injury history is an important risk factor for another football-related injury (Hägglund et al., 2006).
6. The random sample split was done automatically via an R function (*sample*, among similar others) as part of the cross-validation process. To do so, we set up the size of the training and the test data for the *before*-COVID-19 transfers (70% and 30%, respectively, a widely used split percentage, which fits our models). Then, the function randomly takes out 70% of the data as the training set and the remaining 30% as the test set. To ensure reproducible results, we use the *set.seed* function to generate the same random sequence each time.
7. See Coates and Parshakov (2021), Fort et al. (2019), and Leeds (2014) for applications of standard quantile regression models.
8. Appendix Figure 1 (see supplementary material) shows the trends of season-specific median values of real transfer fees in the corresponding leagues of buying clubs. As could be expected, the highest median of transfer fees can be observed for Premier League clubs.
9. The functional form of other non-linear effects is also in line with expectations. However, since the estimates lack precision, we refrain from discussing these results in the main text. For instance, we observe a significant (inverted u-shaped) relation between injury proneness and transfer fees just for the 50th quantile. Our time trend variable is positive but lacks precision particularly for the higher quantiles (i.e. the 75th and 90th quantiles).
10. For reasons of completeness, we present our models excluding remaining contract duration in Appendix Table 3 and Appendix Figure 3 (see supplementary material). Deviations (if any) between the results of our main specification and these models can simply be explained by other variables serving as rough proxy for contract duration, thus picking up some of the variance when remaining contract duration is excluded.
11. Additional robustness checks are performed by trimming 1% of the lowest and highest transfer fees in the test and the during-COVID-19 data set, respectively. Our main findings remain (Appendix Figure 2, see supplementary material).
12. We thank one of the anonymous reviewers for their insightful comments on this discussion.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Yanxiang Yang http://orcid.org/0000-0001-9478-6224
Joerg Koenigstorfer http://orcid.org/0000-0001-6159-2861
Tim Pawlowski http://orcid.org/0000-0001-5829-963X

References

Ante, L. (2019). *Determinants of transfers fees: Evidence from the five major European football leagues*. https://www.researchgate.net/profile/Lennart-Ante/publication/331929212

Carmichael, F., Forrest, D., & Simmons, R. (1999). The labour market in association football: Who gets transferred and for how much? *Bulletin of Economic Research*, *51*(2), 125–150. https://doi.org/10.1111/1467-8586.00075

Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: Theory and evidence. *Applied Economics*, *25*(12), 1467–1476. https://doi.org/10.1080/00036849300000150

Carrieri, V., Principe, F., & Raitano, M. (2018). What makes you 'super-rich'? New evidence from an analysis of football players' wages. *Oxford Economic Papers*, *70*(4), 950–973. https://doi.org/10.1093/oep/gpy025

Coates, D., & Parshakov, P. (2021). The wisdom of crowds and transfer market values. *European Journal of Operational Research*, *301*(2), 523–534. https://doi.org/10.1016/j.ejor.2021.10.046

Depken, C. A., & Globan, T. (2021). Football transfer fee premiums and Europe's big five. *Southern Economic Journal*, *87*(3), 889–908. https://doi.org/10.1002/soej.12471

Dobson, S., & Gerrard, B. (1999). The determination of player transfer fees in English professional soccer. *Journal of Sport Management*, *13*(4), 259–279. https://doi.org/10.1123/jsm.13.4.259

Dobson, S., Gerrard, B., & Howe, S. (2000). The determination of transfer fees in English nonleague football. *Applied Economics*, *32*(9), 1145–1152. https://doi.org/10.1080/000368400404281

Drewes, M., Daumann, F., & Follert, F. (2021). Exploring the sports economic impact of COVID-19 on professional soccer. *Soccer & Society*, *22*(1-2), 125–137. https://doi.org/10.1080/14660970.2020.1802256

Fasiolo, M., Nedellec, R., Goude, Y., Capezza, C., Wood, S. N., & Fasiolo, M. M. (2020). *Package 'mgcViz'. Visualisations for generalised additive models*. Available from: https://cran.r-project.org/web/packages/mgcViz/index.html

Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., & Goude, Y. (2020). Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, *116*(535), 1402–1412. https://doi.org/10.1080/01621459.2020.1725521

Feess, E., Frick, B., & Muehlheusser, G. (2004). *Legal restrictions on buyout fees: Theory and evidence from German soccer*. https://ssrn.com/abstract = 562445

Feuillet, A., Terrien, M., Scelles, N., & Durand, C. (2020). Determinants of coopetition and contingency of strategic choices: The case of professional football clubs in France. *European Sport Management Quarterly*, *21*(5), 748–763. https://doi.org/10.1080/16184742.2020.1779776

FIFA. (2022). *Global transfer market report 2021*. https://digitalhub.fifa.com/m/2b542d3b011270f/original/FIFA-Global-Transfer-Report-2021-2022-indd.pdf

Fort, R., Lee, Y. H., & Oh, T. (2019). Quantile insights on market structure and worker salaries: The case of major league baseball. *Journal of Sports Economics*, *20*(8), 1066–1087. https://doi.org/10.1177/1527002519851152

Frick, B. (2007). The football players' labor market: Empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, *54*(3), 422–446. https://doi.org/10.1111/j.1467-9485.2007.00423.x

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). Springer.

Garcia-del-Barrio, P., & Pujol, F. (2020). Recruiting talent in a global sports market: Appraisals of soccer players' transfer fees. *Managerial Finance*, *47*(6), 789–811. https://doi.org/10.1108/MF-04-2020-0213

Gerrard, B., & Dobson, S. (2000). Testing for monopoly rents in the market for playing talent–evidence from English professional football. *Journal of Economic Studies*, *27*(3), 142–164. https://doi.org/10.1108/01443580010326049

Gyimesi, A., & Kehl, D. (2021). Relative age effect on the market value of elite European football players: A balanced sample approach. *European Sport Management Quarterly*, 1–17. https://doi.org/10.1080/16184742.2021.1894206. in press.

Hair, J. F. (2009). *Multivariate data analysis*. Prentice Hall.

Hastie, T. J., & Tibshirani, R. J. (2017). *Generalized additive models*. Routledge.

Hawkins, R. D., Hulse, M., Wilkinson, C., Hodson, A., & Gibson, M. (2001). The association football medical research programme: An audit of injuries in professional football. *British Journal of Sports Medicine*, *35*(1), 43–47. https://doi.org/10.1136/bjsm.35.1.43

Hägglund, M., Waldén, M., & Ekstrand, J. (2006). Previous injury as a risk factor for injury in elite football: A prospective study over two consecutive seasons. *British Journal of Sports Medicine*, *40*(9), 767–772. https://doi.org/10.1136/bjsm.2006.026609

Herberger, T. A., & Wedlich, F. (2017). Does selection bias matter in football players' valuation? A crowdsourced valuation approach on players' athletic characteristics. *Journal of Global Sport Management*, *2*(3), 196–214. https://doi.org/10.1080/24704067.2017.1350593

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Kharrat, T., McHale, I. G., & Peña, J. L. (2020). Plus–minus player ratings for soccer. *European Journal of Operational Research*, *283*(2), 726–736. https://doi.org/10.1016/j.ejor.2019.11.026

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(1), 1–26. https://doi.org/10.18637/jss.v028.i05

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492. https://doi.org/10.1037/met0000081

Leeds, M. A. (2014). Quantile regression for sports economics. *International Journal of Sport Finance*, *9*(4), 346–359.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020). Deep soccer analytics: Learning an action-value function for evaluating soccer players. *Data Mining and Knowledge Discovery*, *34*(5), 1531–1559. https://doi.org/10.1007/s10618-020-00705-9

McHale, I. G., & Holmes, B. (2022). Estimating transfer fees of professional footballers using advanced performance metrics and machine learning. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2022.06.033

McHale, I. G., & Relton, S. (2018). Identifying key players in soccer teams using network analysis and pass difficulty. *European Journal of Operational Research*, *268*(1), 339–347. https://doi.org/10.1016/j.ejor.2018.01.018

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

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Parnell, D., Bond, A. J., Widdop, P., & Cockayne, D. (2021). Football worlds: Business and networks during COVID-19. *Soccer & Society*, *22*(1-2), 19–26. https://doi.org/10.1080/14660970.2020.1782719

Pfirrmann, D., Herbst, M., Ingelfinger, P., Simon, P., & Tug, S. (2016). Analysis of injury incidences in male professional adult and elite youth soccer players: A systematic review. *Journal of Athletic Training*, *51*(5), 410–424. https://doi.org/10.4085/1062-6050-51.6.03

Quansah, T., Frick, B., Lang, M., & Maguire, K. (2021). The importance of club revenues for player salaries and transfer expenses – How does the coronavirus outbreak (COVID-19) impact the English Premier League? *Sustainability*, *13*(9), 5154. https://doi.org/10.3390/su13095154

Quansah, T. K., Lang, M., & Frick, B. (2022). *Color blind – Investigating customer-based discrimination in European soccer*. Working Paper, October 4, 2022, Université de Lausanne. http://dx.doi.org/10.2139/ssrn.4237624

Ramos-Filho, L., & Ferreira, M. P. (2021). The reverse relative age effect in professional soccer: An analysis of the Brazilian national league of 2015. *European Sport Management Quarterly*, *21*(1), 78–93. https://doi.org/10.1080/16184742.2020.1725089

Reilly, B., & Witt, R. (1995). English league transfer prices: Is there a racial dimension? *Applied Economics Letters*, *2*(7), 220–222. https://doi.org/10.1080/135048595357302

Ruijg, J., & van Ophem, H. (2015). Determinants of football transfers. *Applied Economics Letters*, *22*(1), 12–19. https://doi.org/10.1080/13504851.2014.892192

Speight, A., & Thomas, D. (1997a). Arbitrator decision-making in the transfer market: An empirical analysis. *Scottish Journal of Political Economy*, *44*(2), 198–215. https://doi.org/10.1111/1467-9485.00053

Speight, A., & Thomas, D. (1997b). Football league transfers: A comparison of negotiated fees with arbitration settlements. *Applied Economics Letters*, *4*(1), 41–44. https://doi.org/10.1080/758521830

Wood, S., & Wood, M. S. (2015). *Package 'mgcv'*. Available from: https://cran.uib.no/web/packages/mgcv/mgcv.pdf