

Electronic Appendix

Manuscript title: Predicting transfer fees in professional European football before and during COVID-19 using machine learning
Authors: Yanxiang Yang¹, Joerg Koenigstorfer¹, and Tim Pawlowski²
Journal title: European Sport Management Quarterly, accepted November 25, 2022

Table of Content

- Appendix Table 1.** Literature review on determinants of football transfer fees.
Appendix Table 2. Number of transfers in the summer and winter transfer windows (n = 7,918).
Appendix Table 3. Model performance excluding the remaining contract duration.
Appendix Table 4. Estimates of linear, generalized additive, and quantile additive models excluding remaining contract duration.
- Appendix Figure 1.** Trends of season-specific median values of real transfer fees in the corresponding leagues of the buying clubs (n = 7,918).
Appendix Figure 2. Predicted versus actual transfer fees and model performance for trimmed data.
Appendix Figure 3. Variable importance according to the random forest estimator excluding the remaining contract duration.
Appendix Figure 4. Predicted versus actual transfer fees and model performance for the higher (>66th percentile) transfer fees.
Appendix Figure 5. Predicted versus actual transfer fees and model performance for the mid (between 33rd and 66th percentile) transfer fees.
Appendix Figure 6. Predicted versus actual transfer fees and model performance for the lower (<33rd percentile) transfer fees.

¹ Technical University of Munich, Chair of Sport and Health Management, Campus D Uptown Munich, Georg-Brauchle-Ring 60/62, 80992 Munich, Germany. Phone: +49 89 289 24559, email: yanxiang.yang@tum.de, joerg.koenigstorfer@tum.de.

² University of Tuebingen, Chair of Sport Economics, Sport Management and Media Research, Institute of Sports Science, Waechterstrasse 67, 72074 Tuebingen, Germany. Phone: +49 7071 297 6544, email: tim.pawlowski@uni-tuebingen.de.

Appendix Table 1. Literature review on determinants of football transfer fees.

Authors (year)	Data: N; league; season; source	Dependent variable; analysis	Player characteristics	Player performance	Significant independent variables		
					Selling-club characteristics	Buying-club characteristics	Time effects
Carmichael & Thomas (1993)	N = 214; EPL; Season 1990/91; Football yearbook	Log (transfer fee); simple linear (OLS) regression	Age ² (-)	Career games played (+)	Club goal difference (+); Club in first (+) or second division (+) compared to fourth; Club's league position last season (-)	Club's average attendance last season (+); Club goal difference (+); Club in first (+), second (+), or third division (+) compared to fourth; Clubs' league position last season (-)	NA
Reilly & Witt (1995) ^a	N = 2 02; EPL; Season 1991/92; Football yearbook	Log (transfer fee); simple linear (OLS) regression	Age (+); Forward (+); Number of previous clubs (-)	Appearances last season (+); Goals current season (+); Full international caps (+)	Club in first (+), second (+) or third (+) division, compared to fourth	Club in first (+) or second (+) division, compared to fourth; Log of league receipts current season in £ thousands (+)	NA
Speight & Thomas (1997b)	N = 404 (thereof 217 from seasons 1978/79–1991/92; 187 from season 1990/91); EPL; Football league and yearbook	Log (transfer fee); simple linear (OLS) regression	Age (+); Age ² (-)	Appearances last season (+)	Club's average attendance last season (+); Club in first (+) or second (+) division compared to fourth; Club's last end-of-season position (+) and its square (-)	Club's average attendance last season (+); Club in first (+), second (+) or third (+) division compared to fourth	NA
Speight & Thomas (1997a)	N = 164; EPL; Seasons 1985/86–1989/90; Football league and yearbook	Log (arbitrated transfer fee)	Age (+); Age ² (-)	Career goals (+); Number of games last season (+); Full international caps (+)	Club goal difference (-)	Club average attendance last season (+); Club goal difference last season (+); Club's league position last season (-); Club in third (-) or fourth division (-) compared to first	Time effects of year 1985/86–1987/88
Dobson & Gerrard (1999) ^b	N = 1,350; EPL; Seasons 1990/91–1996/97; Football yearbook	Log (transfer fee); simple linear (OLS) regression	Age (+); Age ² (-); Number of previous clubs (-)	Career league appearances (+) and its square (-); Career goal rate (+); Full international caps (+) and its square (-); U21 caps (+) and its square (-); League appearances last season (+); Goals last season (+)	Club goal difference (+); Club's league position last season (-)	Club goal difference (+); Club's league position (-); Club's league position last season (-) and its square (+); Club in first (+) and second (+) division compared to fourth	Transfers in Feb (-) and Oct (-) compared to Dec; Transfers in season 1993/94 (-) and 1994/95 (-) compared to 1990/91
Carmichael et al. (1999)	N=240; EPL; Season 1993/94; Football yearbook	Transfer fee; simple linear (OLS) and Heckman two-step model	Age (+); Age ² (-)	League appearances for current club (+) and for other clubs (+) before transfer; Goals in season 1991/92 interacted with forward (+) and mid-field (+); Played for England at least once (+) or for a country other than England (+)	Club in first (-), second (-), or third (+) divisions compared to EPL excluded category	NA	NA
Feess et al. (2004)	N=239; German Bundesliga; Seasons 1994/95–1999/2000; NA	Log (transfer fee); simple linear (OLS) and Heckman model	Age (+); Age ² (-); Forward (+); South American (+); Semi-professional (-); From other countries vs. Europe (+); Remaining contract years (+)	Career games played (+) and its square (-); Full international caps (+); In national team (+);	NA	Club qualified for European cup competition (+)	NA
Ruijg & van Ophem (2015)	N = 373, EPL; Season 2011/12; Transfermarkt + Sky Sports	Log (transfer fee); Heckman two-step and simple linear (OLS) model	Age (+); Age ² (-); Height (-)	Average minutes played (+); Average red cards (+); Percentage of substitute (-); Percentage of golden substitute (+); Number of games played (-)	NA	NA	NA
Ante (2019)	N = 389; Big five; Season 2018/19 summer; Transfermarkt	Log (transfer fee); simple linear (OLS) regression	Age ² (-); Height (+); South American (+); Popularity (likes on Facebook, +)	Yellow cards (-); Number of fouls (+); Minutes played (+)	NA	NA	NA
Garcia-del-Barrio & Pujol (2020)	N = 1083; Big five; Seasons 2010/11–2014/15; Transfermarkt	Transfer fee; simple linear (OLS) regression	Age at contract end (+); Year of experience (+) and its square (-); Remaining contract duration (+)	Media visibility status: individual visibility index (+) and share of media visibility within players' team (+)	NA	Media visibility status (+)	Transfer in season 2012/13 (+), 2013/14 (+), 2014/15 (+) compared to 2010/11; League differences: EPL (+)
Depken & Globan (2021)	N = 5,760; Big five; Seasons 2005/06–2017/18; Transfermarkt	Transfer fee (premium); simple linear (OLS) regression and Heckman selection model	Age 20-24 (+); South American (+); Asian (-)	NA	TV contracts (+)	NA	Transfer seasons since 2013/14 (+); League difference: EPL (+)
Coates & Parshakov (2021)	N = 3,324; Big five; Seasons 1996/97–2015/16; Transfermarkt + EA Sports	Transfer fee; simple linear (OLS) and quantile regressions	Market value (+); Age (+); Time left on current contract (+)	Goals per 1,000 minutes (+); Assists per 1,000 minutes (-); In national team (+); Market value interacted with in national team (+); FIFA rating (+); In national team interacted with FIFA rating (-)	NA	NA	NA
McHale & Holmes (2022) ^c	N = 1,946; 69 leagues; August 11, 2016–September 29, 2020; Transfermarkt + Instat + sofifa	Log (transfer fee); simple linear (OLS) regression, machine learning (glmnet, xgbTree, xgbDART)	Age (-); Age ² (-); height (+); contract remaining (weeks, +); on-loan in last year (dummy, -)	Percentage of minutes played in last year (+); International caps in last three years (+); Goal Impact Metric (GIM) performance ratings (+); Tackles won per 90 minutes in last three years (-)	Median purchase price in the last three years (+)	Median purchase price in the last 3 years (+)	NA

Notes. EPL: English Premier League. ^aOne study re-examining the racial dimension using 29 transfers in EPL during the season 2001/02 is excluded. ^bSeveral significant interactions are not summarized (see Gerrard & Dobson, 2000). ^cTo be consistent, the results are summarized for the linear models; see further results obtained from advanced machine learning models in the original paper. Only the significant determinants are summarized, therefore “NA” refers to either non-significant determinants or determinants that are not assessed. Manuscripts with the following characteristics are not included: (i) bachelor and master theses, (ii) studies focusing on market values of players, (iii) non-English literature, and (iv) studies on non-professional football leagues. References are provided in the list of references at the end of the Electronic Appendix.

PREDICTING TRANSFER FEES

Appendix Table 2. Number of transfers in the summer and winter transfer windows (n = 7,918).

Season	Summer	Winter
2008/09	484	93
2009/10	399	73
2010/11	364	113
2011/12	495	116
2012/13	420	106
2013/14	473	88
2014/15	460	85
2015/16	501	120
2016/17	513	121
2017/18	573	111
2018/19	568	95
2019/20	583	99
2020/21	323	67
2021/22	388	87

Notes. The data collection was done until 2 February, 2022 (i.e. the day when the European big five's winter transfer window was closed for all leagues). Any transfer that took place after March 1, 2020 was classified as during-COVID-19 transfer (n = 865).

PREDICTING TRANSFER FEES

Appendix Table 3. Model performance excluding the remaining contract duration.

Model	R ² for test data set	RMSE for test data set
Simple linear regression	0.51	1.05
Generalized additive model (GAM)	0.62	0.92
Random forest (RF)	0.67	0.86

Notes. GAM: generalized additive model. RF: random forest. RMSE: root mean squared error. The final models are fitted for the full before-COVID-19 data set excluding remaining contract duration (n = 4,149). Model performance is displayed for the randomly split before-COVID-19 test data.

When we re-estimate all models using the full sample (i.e. including transfers without contract duration information), we find that the simple linear model and the GAM have a somewhat smaller predictive power. The RF again outperforms the other models (see also Appendix Figure 2).

PREDICTING TRANSFER FEES

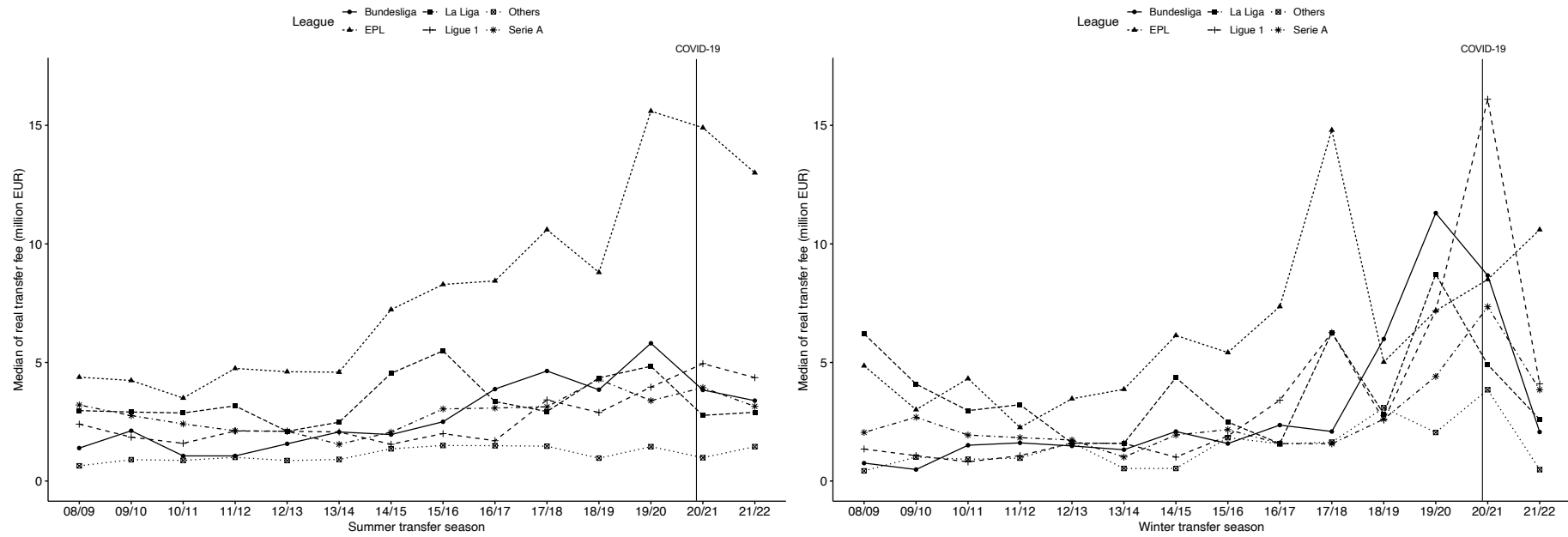
Appendix Table 4. Estimates of linear, generalized additive, and quantile additive models excluding remaining contract duration.

	Linear regression	GAM	QAM 10th	QAM 25th	QAM 50th	QAM 75th	QAM 90th
Dependent variable: log-transformed real transfer fees in EUR							
Intercept	15.69(0.09)***	15.30(0.11)***	14.20(0.18)***	14.94(0.12)***	15.40(0.09)***	15.82(0.08)***	16.18(0.09)***
<i>Player characteristics</i>							
Age	0.11(0.02)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Height	0.07(0.02)***	0.08(0.02)***	0.07(0.04)*	0.08(0.03)***	0.08(0.02)***	0.08(0.02)***	0.09(0.02)***
<i>Nationality (ref: Europe)</i>							
Africa	0.08(0.06)	0.08(0.06)	0.14(0.10)	0.10(0.07)	0.05(0.05)	0.01(0.04)	-0.02(0.05)
Asia	0.02(0.14)	0.13(0.14)	0.28(0.22)	0.14(0.16)	0.10(0.12)	0.03(0.1)	-0.07(0.11)
North America	0.08(0.27)	-0.11(0.26)	-0.50(0.48)	-0.03(0.33)	0.04(0.23)	-0.01(0.19)	-0.02(0.22)
South America	0.28(0.06)***	0.17(0.07)***	0.23(0.11)**	0.18(0.07)**	0.11(0.06)**	0.11(0.05)**	0.07(0.05)
<i>Position (ref: Attacker)</i>							
Defender	-0.39(0.07)***	-0.42(0.07)***	-0.63(0.12)***	-0.50(0.08)***	-0.37(0.06)***	-0.29(0.06)***	-0.26(0.07)***
Goalkeeper	-0.65(0.11)***	-0.59(0.12)***	-0.77(0.20)***	-0.72(0.14)***	-0.54(0.11)***	-0.50(0.10)***	-0.52(0.11)***
Midfielder	-0.14(0.05)***	-0.13(0.06)**	-0.18(0.09)*	-0.20(0.06)***	-0.14(0.05)***	-0.08(0.05)*	-0.05(0.05)
<i>Player performance</i>							
UEFA Champions League	0.28(0.06)***	0.28(0.07)***	0.35(0.11)***	0.28(0.08)***	0.21(0.06)***	0.17(0.06)***	0.23(0.06)***
Appearances	0.18(0.07)***	Non-linear	Non-linear	Non-linear**	Non-linear	Linear*	Linear
Substitution on	-0.07(0.03)**	-0.07(0.04)**	-0.15(0.06)**	-0.09(0.04)**	-0.04(0.03)	-0.04(0.03)	-0.06(0.03)*
Substitution off	-0.03(0.03)	-0.03(0.03)	-0.05(0.05)	-0.04(0.03)	-0.02(0.02)	-0.02(0.02)	-0.03(0.02)
Minutes played	0.07(0.07)	0.08(0.07)	0.01(0.14)	0.05(0.09)	0.13(0.06)**	0.11(0.06)**	0.06(0.07)
Points ^a	0.05(0.02)**	0.05(0.03)**	0.08(0.05)*	0.06(0.03)**	0.04(0.02)**	0.03(0.02)	0.03(0.03)
Goals ^a	0.05(0.02)**	0.04(0.02)*	0.05(0.02)**	0.04(0.02)*	0.04(0.02)*	0.06(0.03)**	0.06(0.03)**
Assists ^a	0.03(0.02)	0.01(0.02)	0.02(0.03)	0.01(0.03)	0.01(0.02)	0.01(0.02)	0.03(0.02)
Yellow cards ^a	-0.04(0.02)*	-0.04(0.02)*	-0.05(0.04)	-0.03(0.03)	-0.02(0.02)	-0.02(0.02)	-0.04(0.02)*
Injury proneness	0.01(0.02)	Non-linear*	Non-linear	Non-linear*	Non-linear***	Non-linear**	Non-linear
<i>Selling-club characteristics</i>							
Arrivals	-0.14(0.07)**	Non-linear	Non-linear	Non-linear	Non-linear**	Linear***	Linear**
Departures	0.17(0.07)**	0.13(0.07)*	0.02(0.12)	0.10(0.08)	0.14(0.06)**	0.18(0.06)***	0.15(0.06)**
Transfer income	0.18(0.03)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Transfer expenditure	-0.03(0.03)	Non-linear*	Linear	Linear**	Linear**	Non-linear**	Non-linear**
Spectators	0.21(0.03)***	Non-linear**	Non-linear***	Non-linear**	Non-linear	Linear	Linear*
UEFA club coefficient	-0.06(0.03)**	-0.05(0.03)*	-0.11(0.06)**	-0.07(0.04)*	-0.04(0.03)	-0.03(0.02)	-0.04(0.03)
League ranking	-0.01(0.02)	-0.01(0.02)	-0.04(0.04)	-0.02(0.03)	0.01(0.02)	0.01(0.02)	0.02(0.02)
<i>League (ref: Premier League)</i>							
Other English leagues	-0.34(0.11)***	-0.24(0.12)*	-0.26(0.20)	-0.35(0.13)***	-0.29(0.10)***	-0.19(0.09)**	-0.16(0.10)*
Ligue 1	-0.16(0.09)*	-0.10(0.10)	-0.09(0.17)	-0.14(0.11)	-0.11(0.08)	-0.10(0.07)	-0.08(0.08)
Other French leagues	-0.44(0.14)***	-0.12(0.16)	-0.06(0.27)	-0.11(0.17)	-0.18(0.13)	-0.17(0.12)	-0.13(0.14)
Bundesliga	-0.40(0.09)***	-0.27(0.10)***	-0.30(0.18)*	-0.29(0.12)**	-0.2(0.09)**	-0.18(0.08)**	-0.19(0.08)**
Other German leagues	-0.83(0.14)***	-0.34(0.15)**	-0.24(0.25)	-0.31(0.17)*	-0.34(0.12)***	-0.35(0.12)***	-0.37(0.12)***
Serie A	-0.39(0.10)***	-0.30(0.11)***	-0.43(0.20)**	-0.36(0.13)***	-0.23(0.10)**	-0.18(0.09)**	-0.19(0.09)**
Other Italian leagues	-1.19(0.13)***	-0.64(0.15)***	-0.75(0.27)***	-0.83(0.17)***	-0.67(0.13)***	-0.54(0.11)***	-0.51(0.13)***
La Liga	-0.11(0.09)	-0.12(0.10)	-0.19(0.17)	-0.26(0.12)**	-0.14(0.09)	-0.08(0.08)	-0.01(0.08)
Other Spanish leagues	-0.54(0.14)***	-0.10(0.15)	-0.13(0.26)	-0.31(0.17)*	-0.19(0.13)	-0.08(0.12)	0.03(0.13)
Other European leagues	-0.39(0.09)***	-0.21(0.11)**	-0.14(0.18)	-0.22(0.12)*	-0.21(0.09)**	-0.20(0.08)**	-0.21(0.09)**
South American leagues	-0.03(0.16)	0.20(0.17)	0.62(0.27)**	0.20(0.18)	0.14(0.14)	0.10(0.13)	0.13(0.14)
Other non-European	-0.72(0.23)**	0.03(0.25)	0.20(0.40)	-0.06(0.27)	-0.05(0.22)	0.03(0.19)	-0.01(0.20)
<i>leagues</i>							
<i>Buying-club characteristics</i>							
Arrivals	0.17(0.07)**	Non-linear	Linear**	Linear***	Linear	Linear	Linear
Departures	-0.23(0.07)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Transfer income	0.12(0.02)***	0.02(0.03)	0.07(0.04)	0.04(0.03)	0.01(0.02)	0.01(0.02)	0.01(0.02)
Transfer expenditure	0.27(0.03)***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***	Non-linear***
Spectators	0.22(0.03)***	0.07(0.03)**	0.06(0.05)	0.06(0.04)	0.07(0.03)***	0.07(0.02)***	0.07(0.03)***
UEFA club coefficient	-0.01(0.02)	0.01(0.02)	-0.02(0.03)	0.01(0.02)	0.01(0.02)	0.01(0.02)	0.02(0.02)
League ranking	0.04(0.02)*	0.01(0.02)	0.03(0.04)	0.02(0.03)	0.01(0.02)	0.01(0.02)	-0.01(0.02)
<i>League (ref: Premier League)</i>							
Other English leagues	-0.7(0.11)***	-0.03(0.13)	-0.05(0.20)	-0.13(0.14)	-0.06(0.11)	-0.02(0.1)	-0.03(0.10)
Ligue 1	-0.29(0.08)***	0.13(0.09)	0.16(0.14)	0.11(0.10)	0.08(0.07)	0.04(0.07)	0.01(0.07)
Other French leagues	-1.20(0.25)***	-0.27(0.28)	-0.35(0.4)	-0.44(0.30)	-0.33(0.26)	-0.34(0.21)	-0.45(0.28)
Bundesliga	-0.31(0.08)***	-0.18(0.09)**	-0.26(0.15)*	-0.24(0.10)**	-0.17(0.08)**	-0.13(0.07)**	-0.15(0.07)**
Other German leagues	-1.96(0.14)***	-1.01(0.16)***	-1.23(0.26)***	-1.15(0.19)***	-1.14(0.15)***	-0.91(0.17)***	-0.60(0.15)***
Serie A	-0.13(0.09)	0.03(0.11)	0.12(0.18)	0.06(0.12)	0.02(0.09)	0.01(0.08)	0.02(0.09)
Other Italian leagues	-1.70(0.13)***	-0.92(0.16)***	-1.79(0.47)***	-0.73(0.21)***	-0.55(0.14)***	-0.57(0.13)***	-0.53(0.14)***
La Liga	-0.38(0.08)***	0.01(0.08)	-0.11(0.15)	0.01(0.09)	0.02(0.07)	0.02(0.07)	-0.01(0.07)
Other Spanish leagues	-1.20(0.28)***	-0.18(0.27)	-0.37(0.42)	-0.54(0.28)*	-0.36(0.34)	0.01(0.27)	0.22(0.27)
Other European leagues	-0.78(0.10)***	-0.19(0.11)*	-0.21(0.18)	-0.28(0.12)**	-0.24(0.10)**	-0.19(0.09)**	-0.19(0.09)**
South American leagues	-0.25(0.17)	0.25(0.19)	0.16(0.30)	0.25(0.23)	0.31(0.16)*	0.24(0.14)*	0.20(0.16)
Other non-European leagues	-0.13(0.14)	0.38(0.15)*	0.34(0.25)	0.4(0.17)**	0.41(0.13)***	0.36(0.11)***	0.30(0.12)**
<i>Time effects</i>							
Transfer window (ref: summer)	-0.04(0.05)	0.09(0.05)*	0.04(0.09)	0.08(0.06)	0.12(0.05)***	0.09(0.04)**	0.03(0.04)
Transfer seasons (ref: 2015/16)	0.09(0.03)***	-0.01(0.03)	0.01(0.05)	0.03(0.03)	0.02(0.02)	-0.01(0.02)	-0.03(0.02)

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 minutes 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 excluding remaining contract duration (n = 3,284). Model goodness-of-fit: Linear regression (adjusted R² = 0.54, F(59, 3,224) = 64, P < 0.001); GAM (adjusted R² = 0.64, scale estimate [squared residual standard error] = 0.79, REML = 4,426); QAM (10th to 90th percentile: adjusted R² = 0.62, 0.63, 0.63, 0.63, and 0.61, respectively; deviance explained = 79%, 60%, 54%, 67%, and 84%, respectively; REML = 6,268, 4,861, 4,214, 4,147, and 4,593, respectively). Standard errors are presented in parentheses. Significance levels: * P < 0.1, ** P < 0.05, *** P < 0.01.

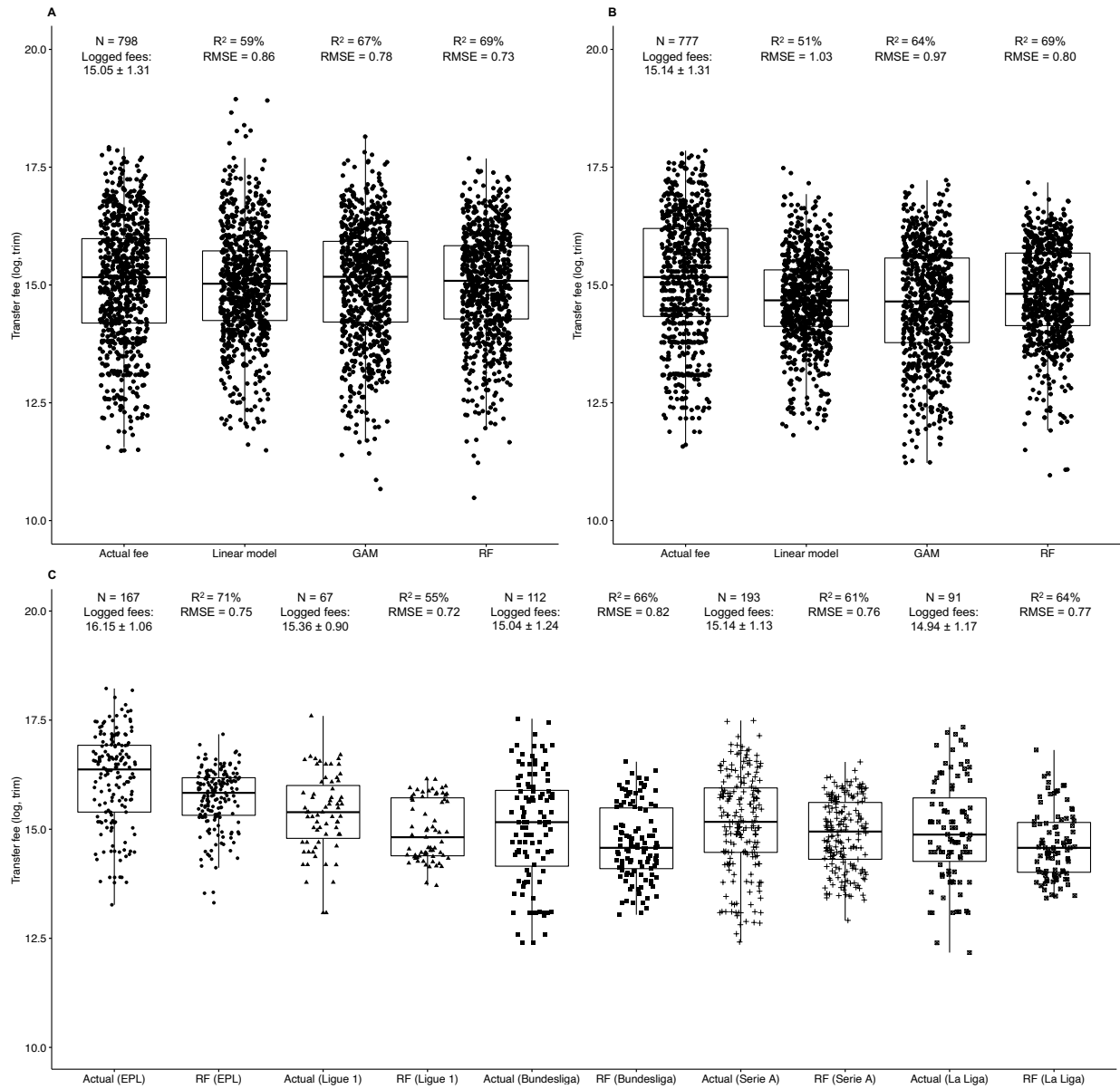
PREDICTING TRANSFER FEES

Appendix Figure 1. Trends of season-specific median values of real transfer fees in the corresponding leagues of the buying clubs ($n = 7,918$).



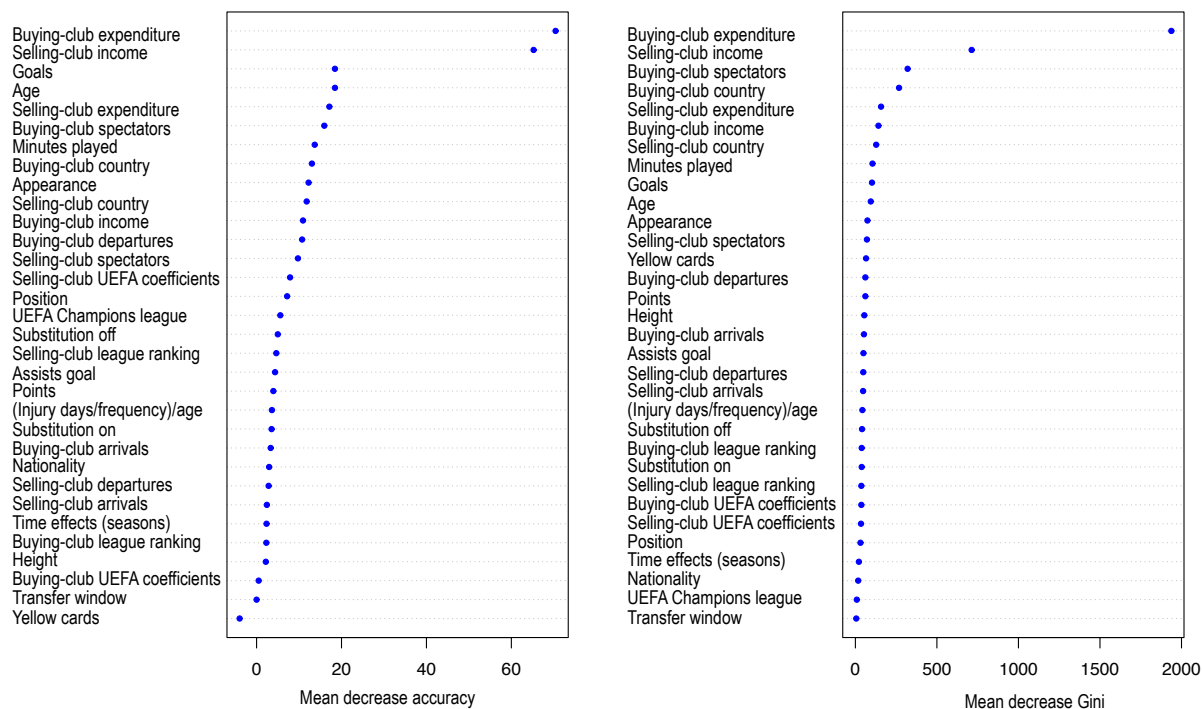
Notes. EPL: English Premier League. Displayed are season-specific median values of real transfer fees in the corresponding leagues of the buying clubs. Others include transfers from one of the big five leagues to buying clubs of (1) one of the big five lower leagues, (2) other European leagues, or (3) other non-European leagues.

Appendix Figure 2. Predicted versus actual transfer fees and model performance for trimmed data.



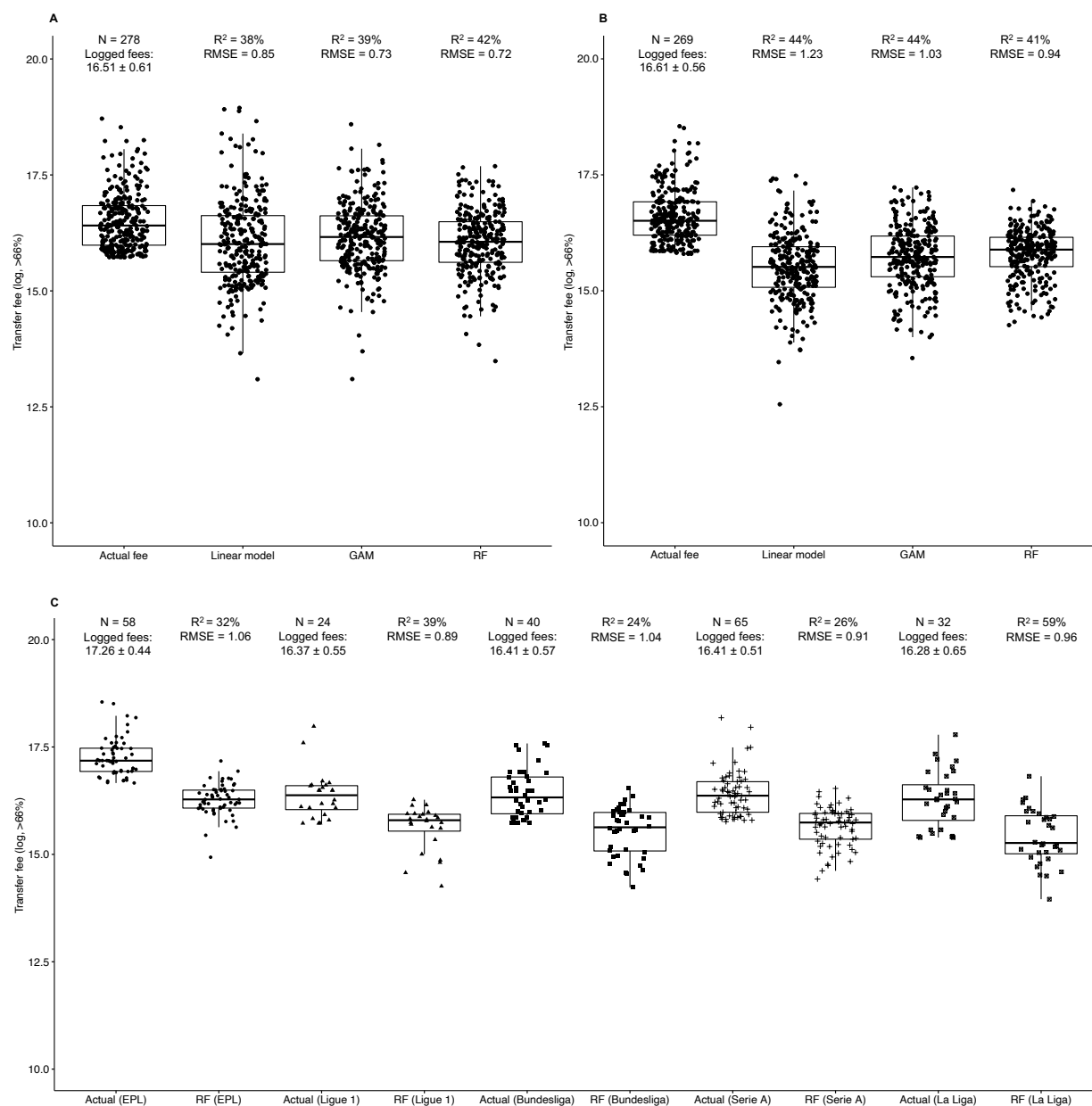
Notes. GAM: generalized additive model. EPL: English Premier League. RF: random forest. RMSE: root mean squared error. log: log-transformed transfer fees. trim: trimmed data set (i.e. we only kept observations of transfer fee values from the 1st up to the 99th percentile). A: Before-COVID-19 predictions (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, $t(797) = 1.67$, $P = 0.10$; for B, $t(776) = 12.30$, $P < 0.001$; and for C, EPL, $t(166) = 9.68$, $P < 0.001$; Ligue 1, $t(66) = 5.44$, $P < 0.001$; Bundesliga, $t(111) = 5.02$, $P < 0.001$; Serie A, $t(192) = 5.06$, $P < 0.001$; and La Liga, $t(90) = 3.94$, $P = 0.0002$.

Appendix Figure 3. Variable importance according to the random forest estimator excluding the remaining contract duration.



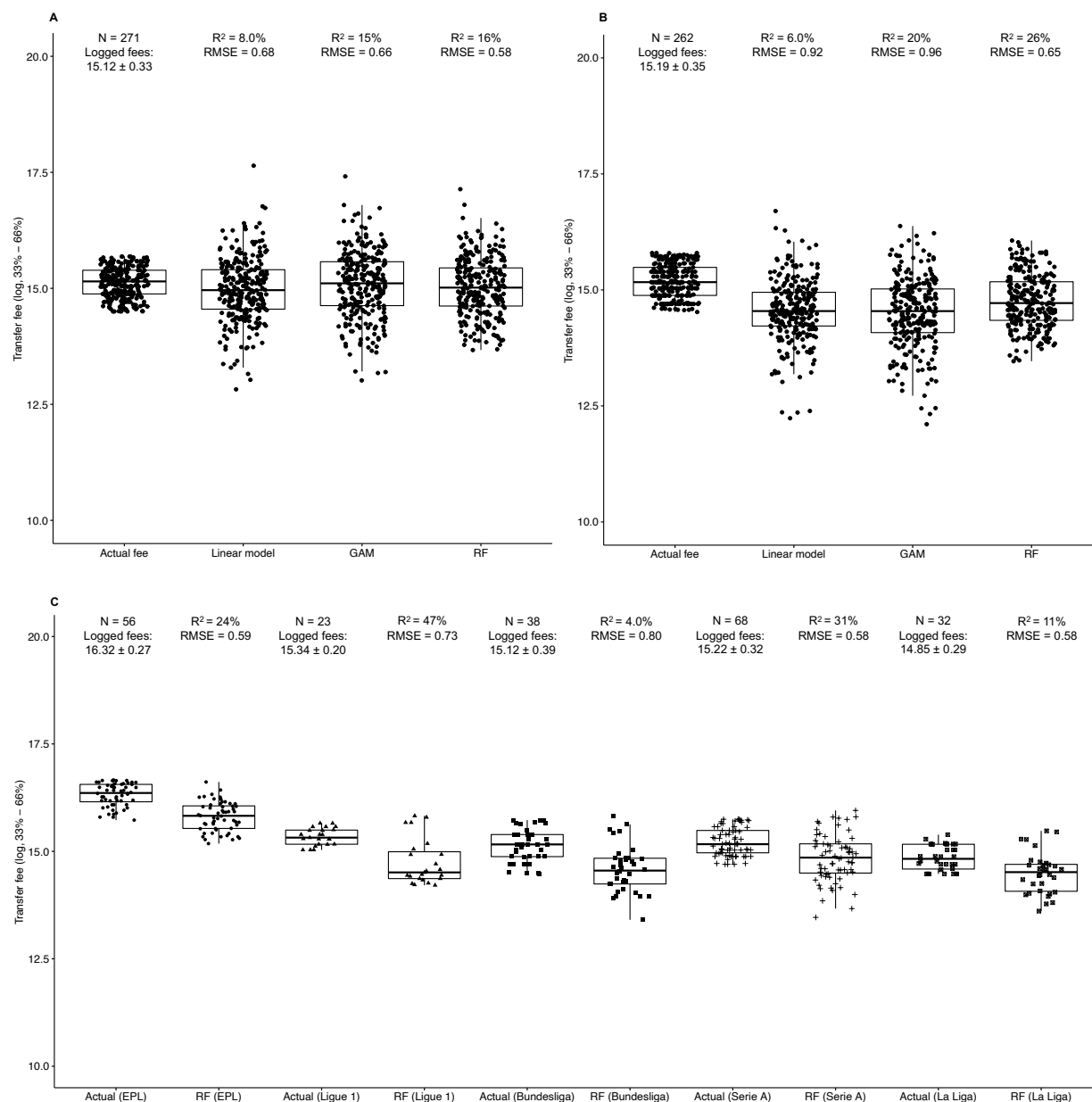
Notes. The model is fitted for the full before-COVID-19 data set ($n = 3,284$). For the training model, the optimal number of variables sampled for splitting at each node [mtry] is 30. The mean decrease accuracy (MDA, unitless, often termed %IncMSE) measures how much accuracy the model loses by excluding 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.

Appendix Figure 4. Predicted versus actual transfer fees and model performance for the higher (>66th percentile) transfer fees.



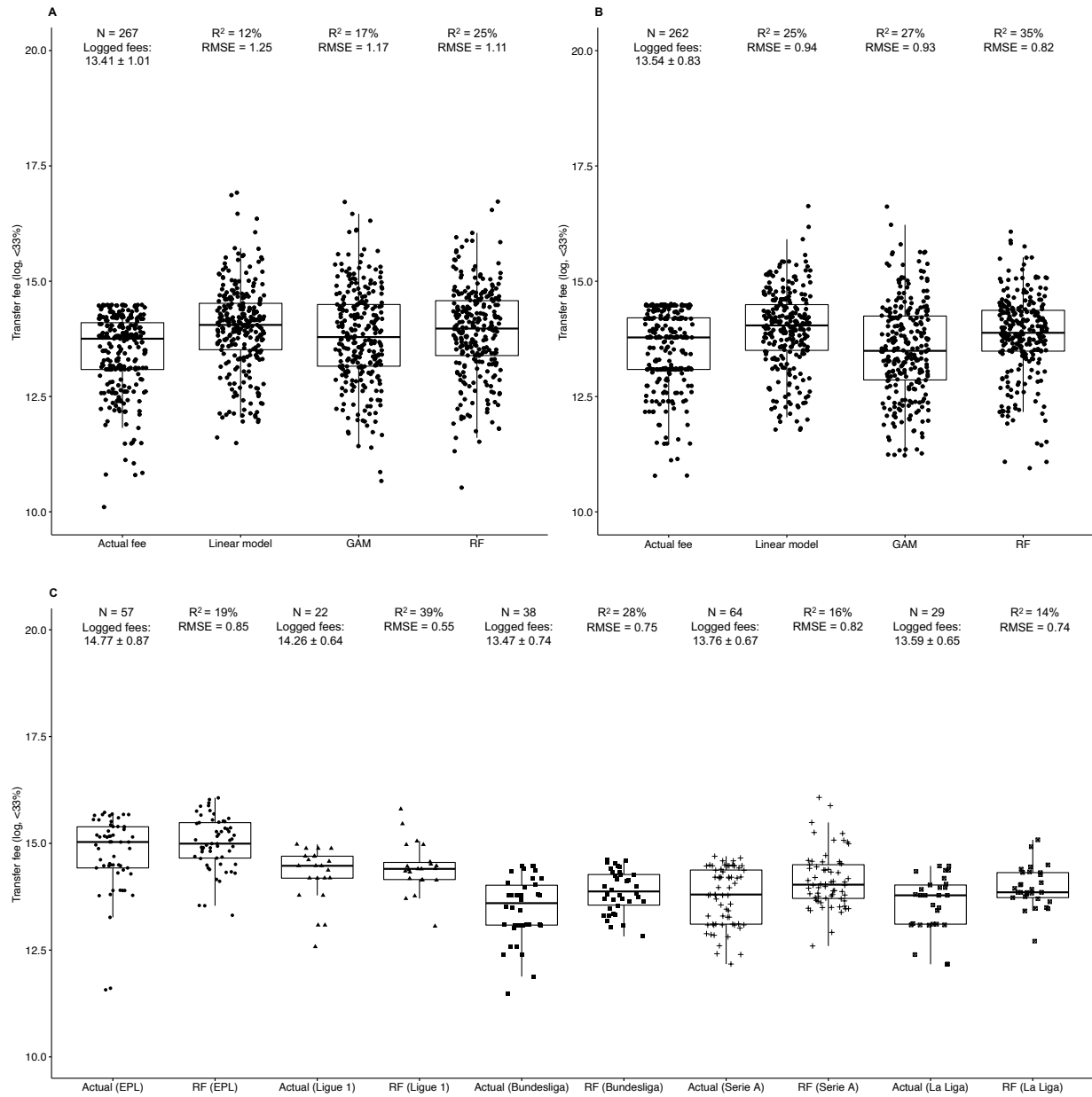
Notes. GAM: generalized additive model. EPL: English Premier League. RF: random forest. RMSE: root mean squared error. log: log-transformed transfer fees. A: Before-COVID-19 predictions (test set). B: During-COVID-19 predictions. C: During-COVID-19 predictions for each of the big five leagues (classified 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, $t(277) = 14.15$, $P < 0.001$; for B, $t(268) = 28.79$, $P < 0.001$; and for C, EPL, $t(57) = 19.67$, $P < 0.001$; Ligue 1, $t(23) = 7.95$, $P < 0.001$; Bundesliga, $t(39) = 9.67$, $P < 0.001$; Serie A, $t(64) = 12.88$, $P < 0.001$; and La Liga, $t(31) = 11.10$, $P < 0.001$.

Appendix Figure 5. Predicted versus actual transfer fees and model performance for the mid (between 33rd and 66th percentile) transfer fees.



Notes. 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, $t(270) = 2.59$, $P = 0.01$; for B, $t(261) = 13.73$, $P < 0.001$; and for C, EPL, $t(55) = 12.14$, $P < 0.001$; Ligue 1, $t(22) = 6.73$, $P < 0.001$; Bundesliga, $t(37) = 5.62$, $P < 0.001$; Serie A, $t(67) = 6.91$, $P < 0.001$; and La Liga, $t(31) = 4.22$, $P = 0.0002$.

Appendix Figure 6. Predicted versus actual transfer fees and model performance for the lower (<33th percentile) transfer fees.



Notes. 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, $t(266) = -8.15$, $P < 0.001$; for B, $t(261) = -6.00$, $P < 0.001$; and for C, EPL, $t(56) = -2.21$, $P = 0.03$; Ligue 1, $t(21) = -1.56$, $P = 0.13$; Bundesliga, $t(37) = -4.03$, $P < 0.001$; Serie A, $t(63) = -4.23$, $P < 0.001$; and La Liga, $t(28) = -3.17$, $P = 0.004$.

Electronic Appendix 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>
- 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>
- Feess, E., Frick, B., & Muehlheusser, G. (2004). Legal restrictions on buyout fees: Theory and evidence from German soccer. <https://ssrn.com/abstract=562445>
- 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, 142-164.

<https://doi.org/10.1108/01443580010326049>

- 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>. In press.
- 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>