Predicting Transfer Fees in Professional European Football Before and During COVID-19

Previous title: *Accuracy of a Prediction Model of Transfer Fees for European Soccer Players Traded during the COVID-19 Pandemic*

Yanxiang Yang [1], Tim Pawlowski [2], Joerg Koenigstorfer [1]

[1] Technical University of Munich, Germany [2] University of Tuebingen, Germany

Yanxiang Yang, PhD candidate

Advisor: Prof. Joerg Koenigstorfer

Chair of Sport and Health Management

TUM School of Management

Technical University of Munich

Germany

European Football Transfer Market and Transfer Fees

Talented players are traded on the international market ($6 billion, FIFA 2021).

Transfer date		
Season 17/18 - Aug 3, 2017		
Barcelona	Neymar →	Paris SG
LaLiga	competition	Ligue 1
First Tier	League type	First Tier
Ernesto Valverde	Coach	Unai Emery
Pep Segura	Manager	Antero Henrique
Market value at time of transfer: $110.00m		
Age at time of transfer: 25 years 05 months 29 days		
Remaining contract duration at Barcelona: 03 Years 10 Months 27 Days (Jun 30, 2021)		
Fee: $244.20m		

***Industry*: FIFA 2022**: (Men's football)

- International transfers INCREASE
- Transfer fees DECREASE
- Back to pre-pandemic

***Academia*: Key determinants of transfer fees?**

Player characteristics
Age, height, remaining contract length

Player performance
Goals, minutes played, injury

Selling- and buying-club characteristics
Market size (spectators), performance

Other variables
Time effects

- How to make comparably accurate predictions?
- Did the COVID-19 affect the relevance and accuracy of common predictors?

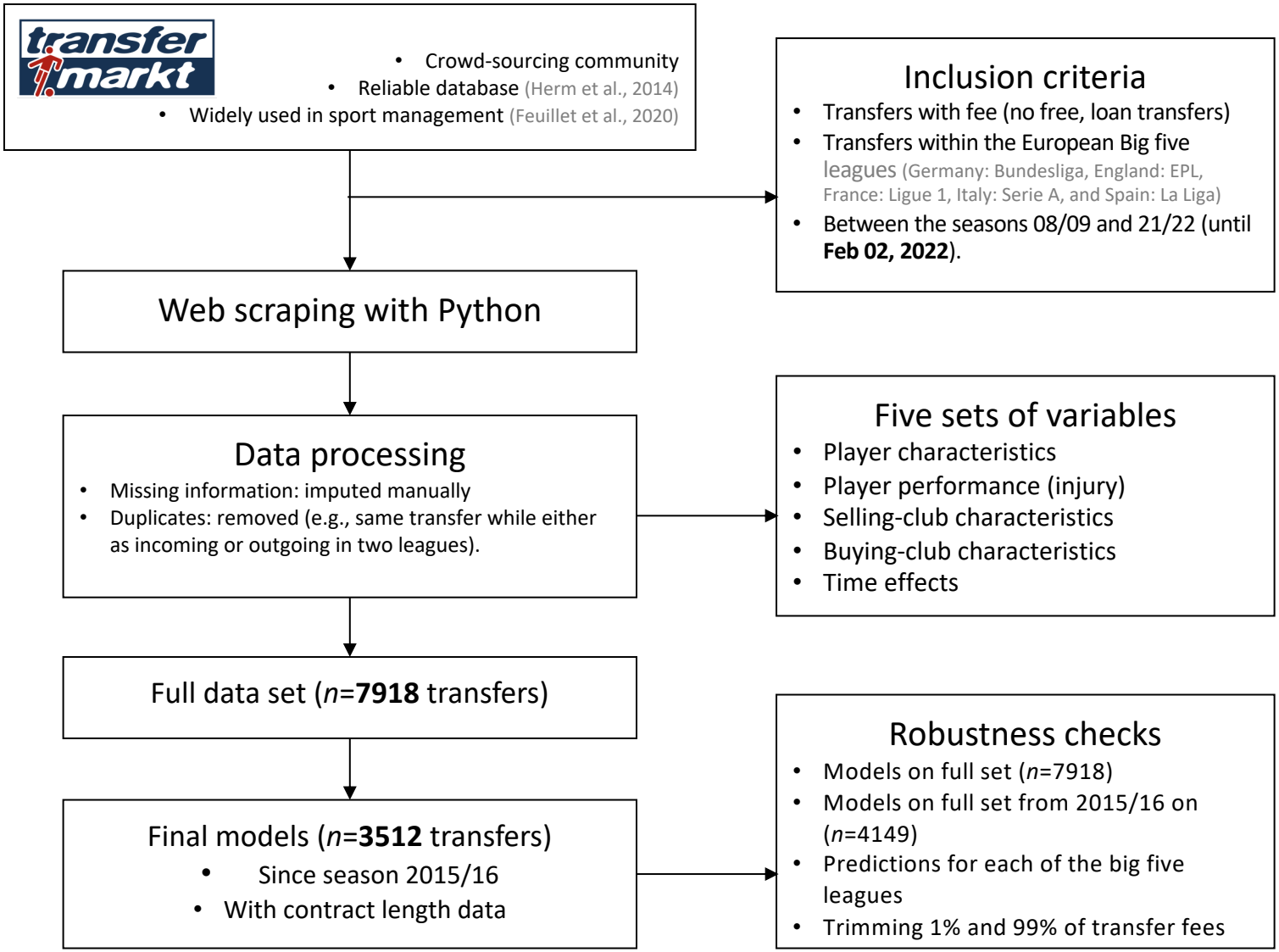
Literature: Determinants of Transfer Fees

Authors	Data: N; source	DV; analysis	Player characteristics	Player performance	Selling-club	Buying-club	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
...	...	...	...	...	...	...	...
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 * in national team (+); FIFA rating (+); In national team * FIFA rating (-)	NA	NA	NA

Gaps & Aims: To extend findings from earlier efforts exploring the factors associated with transfer fees:

- *Variables*: Enlarging the feature space (injury, contract duration)
- *Sample*: Analyzing a larger number of, and more recent transfers
- *Linearity*: Building models beyond linear functional forms (machine learning)
- *COVID-19*: Testing hypothesis that transfer market has changed during COVID-19

Sample


transfermarkt
• Crowd-sourcing community
• Reliable database (Herm et al., 2014)
• Widely used in sport management (Feuillet et al., 2020)
Inclusion criteria
• Transfers with fee (no free, loan transfers)
• Transfers within the European Big five leagues (Germany: Bundesliga, England: EPL, France: Ligue 1, Italy: Serie A, and Spain: La Liga)
• Between the seasons 08/09 and 21/22 (until **Feb 02, 2022**).
Web scraping with Python
Data processing
• Missing information: imputed manually
• Duplicates: removed (e.g., same transfer while either as incoming or outgoing in two leagues).
Five sets of variables
• Player characteristics
• Player performance (injury)
• Selling-club characteristics
• Buying-club characteristics
• Time effects
Full data set (*n*=**7918** transfers)
Final models (*n*=**3512** transfers)
• Since season 2015/16
• With contract length data
Robustness checks
• Models on full set (*n*=7918)
• Models on full set from 2015/16 on (*n*=4149)
• Predictions for each of the big five leagues
• Trimming 1% and 99% of transfer fees

Measures: List of Variables

1. Player characteristics

- Age
- Height
- Nationality: Europe, Asia, Africa, South America, North America
- Position: Defender, goalkeeper, attacker, midfielder
- Remaining contract length (days)

5. Time effects

- Season 08/09 – season 21/22
- Season 15/16 – season 21/22
- Transfer window (summer, winter)

3&4. Selling & buying club characteristics

- Arrivals of players
- Departures of players
- Transfer income
- Transfer expenditure
- Spectators
- UEFA club coefficients
- League ranking
- Leagues (13 types):
  - 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

2. Player performance

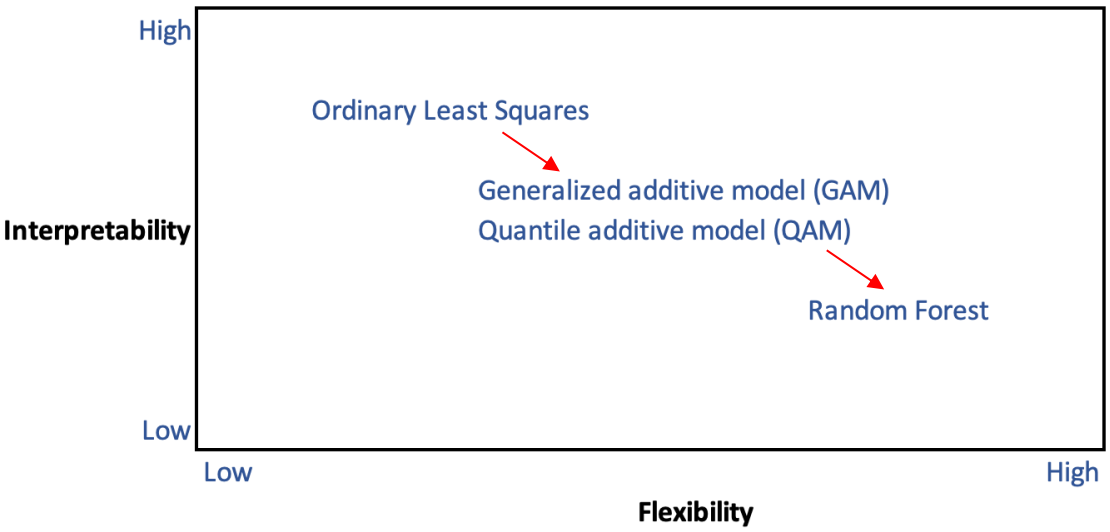
- UEFA Champions League
- Appearances
- Substitution on
- Substitution off
- Minutes played
- Points (/1000 MP)
- Goals (/1000 MP)
- Assists goal (/1000 MP)
- Yellow cards (/1000 MP)
- Player injury history: (Injury days/injury frequency)/age

***Note**. All variables refer to previous season or player career history*

Modelling

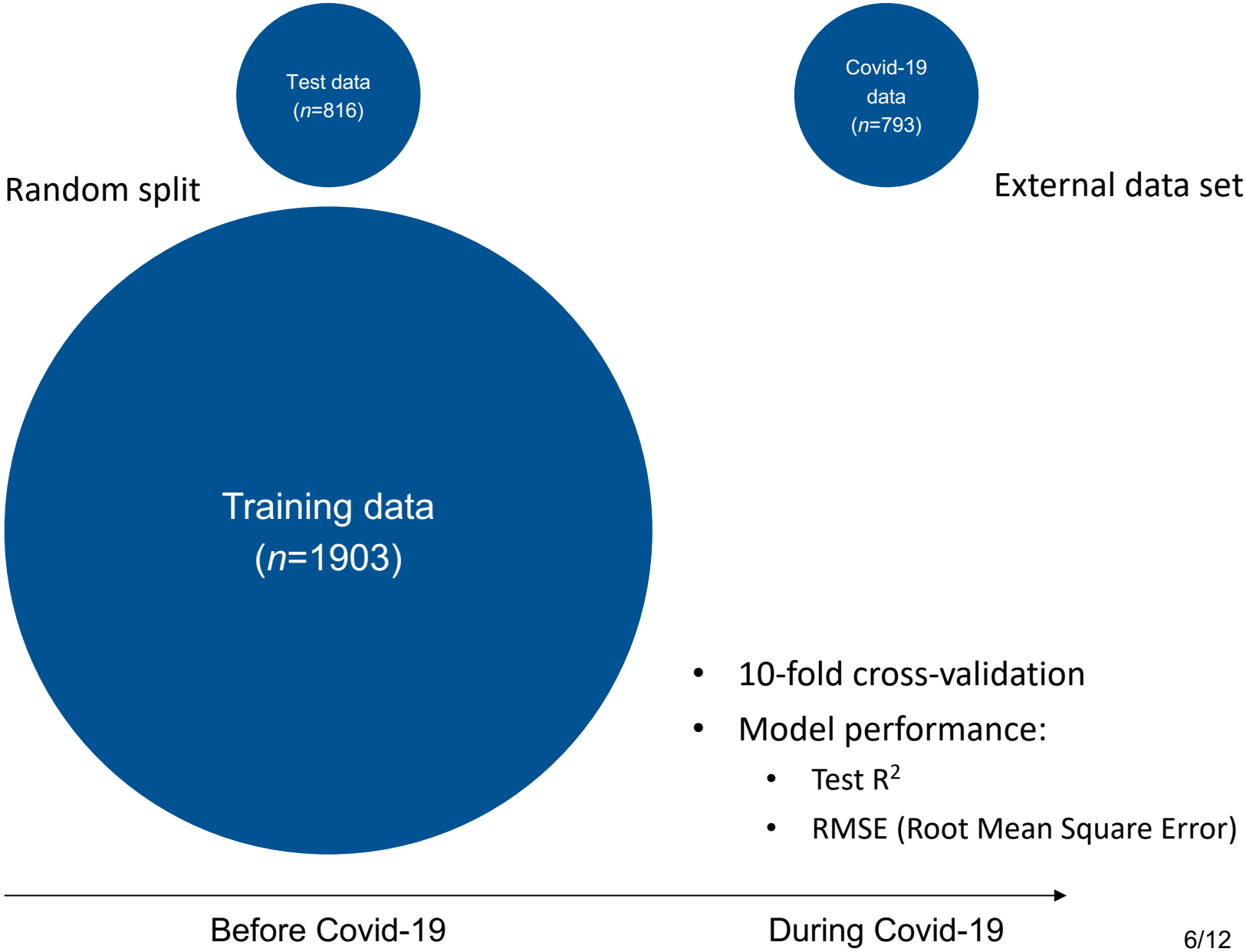
$$\text{Logged transfer fee} = \beta_0 + \beta_1*(\text{player characteristics}) + \beta_2*(\text{player performance}) + \beta_3*(\text{selling-club characteristics}) + \beta_4*(\text{buying-club characteristics}) + \beta_5*(\text{time effects}) + \varepsilon$$

Supervised machine learning framework



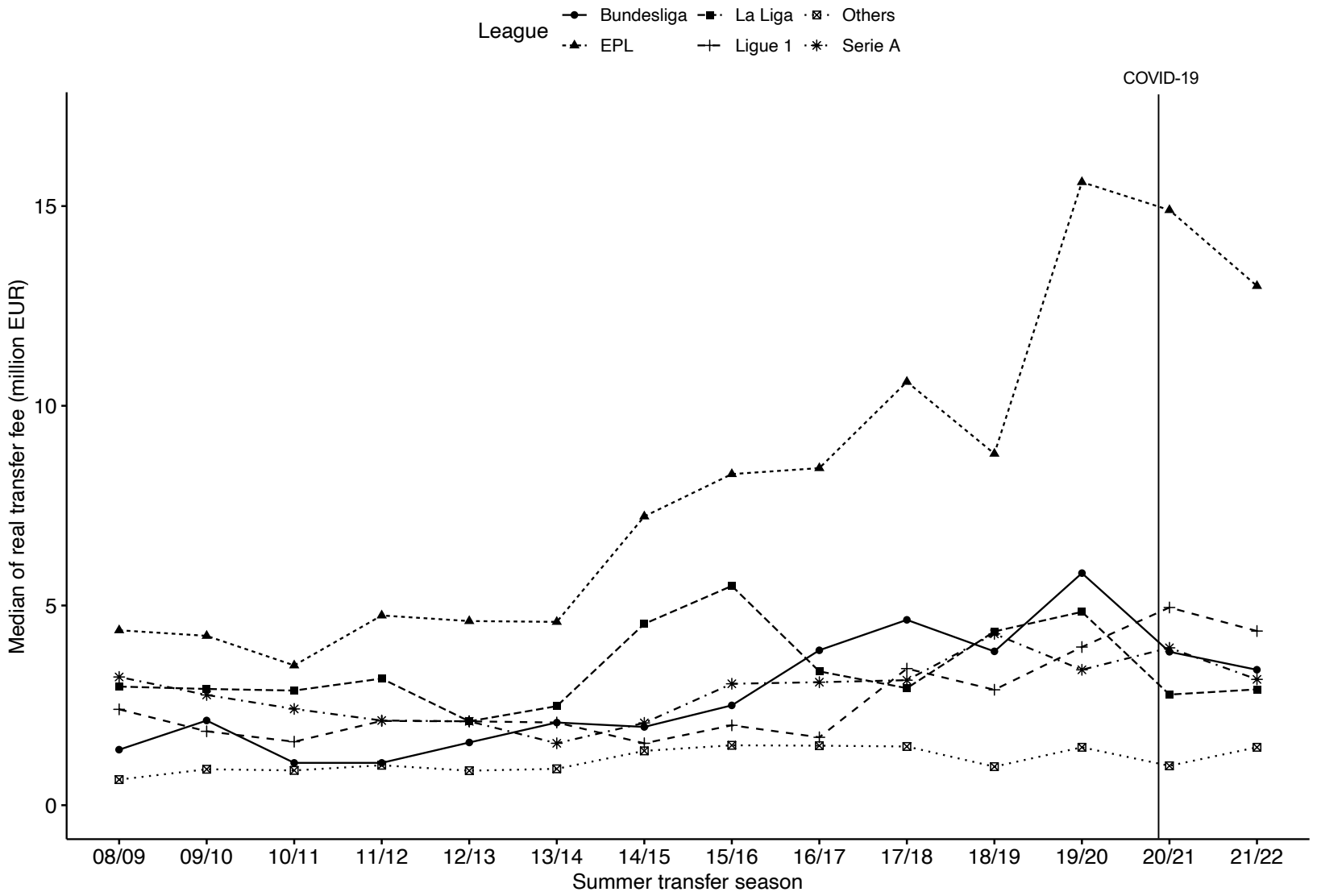
Serving the goal to go beyond linear functional forms (machine learning)

Data Splitting (n = 3512)



- 10-fold cross-validation
- Model performance:
  - Test R^2
  - RMSE (Root Mean Square Error)

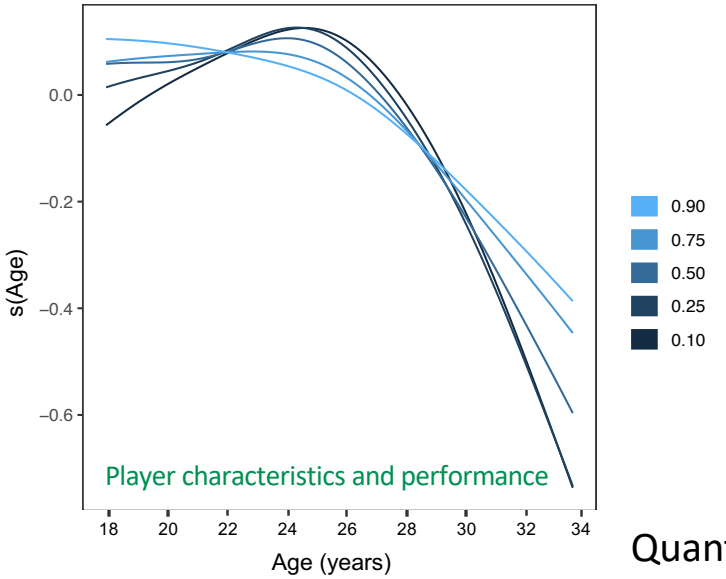
Descriptive: Trends of Median Transfer Fees (Summer)



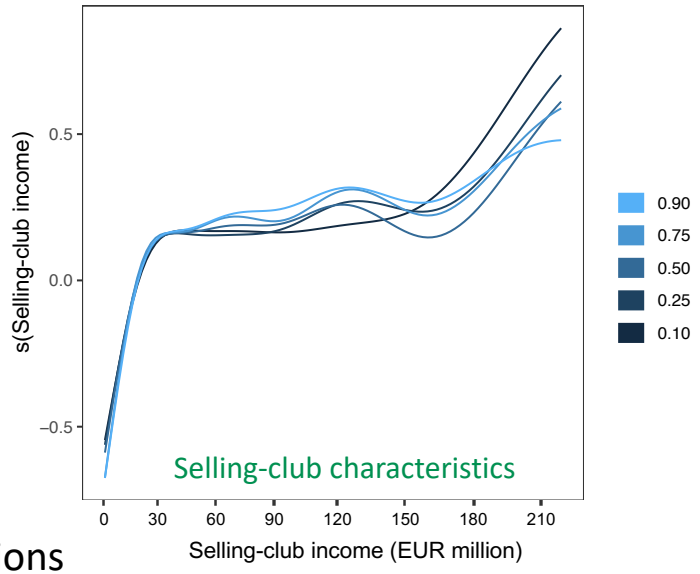
Note. Mean and SDs were calculated while not shown.

Quantile Additive Non-linear Effects of Predictors (3/9)

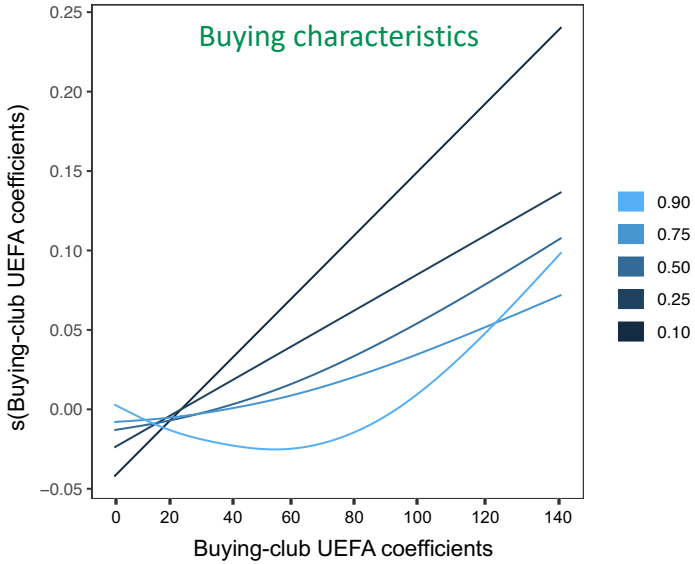
Precise non-linear relations



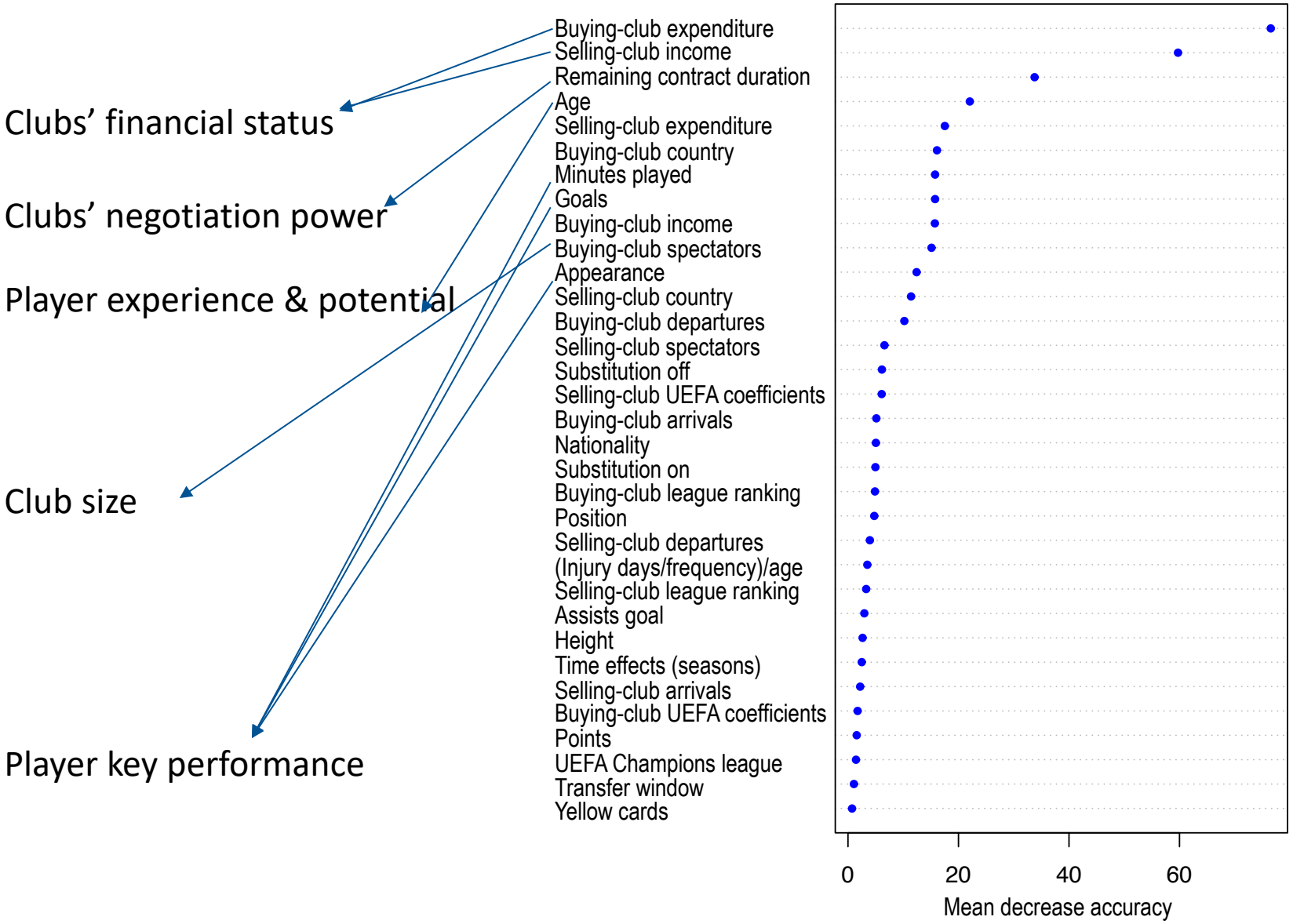
Positive yet decreasing margines



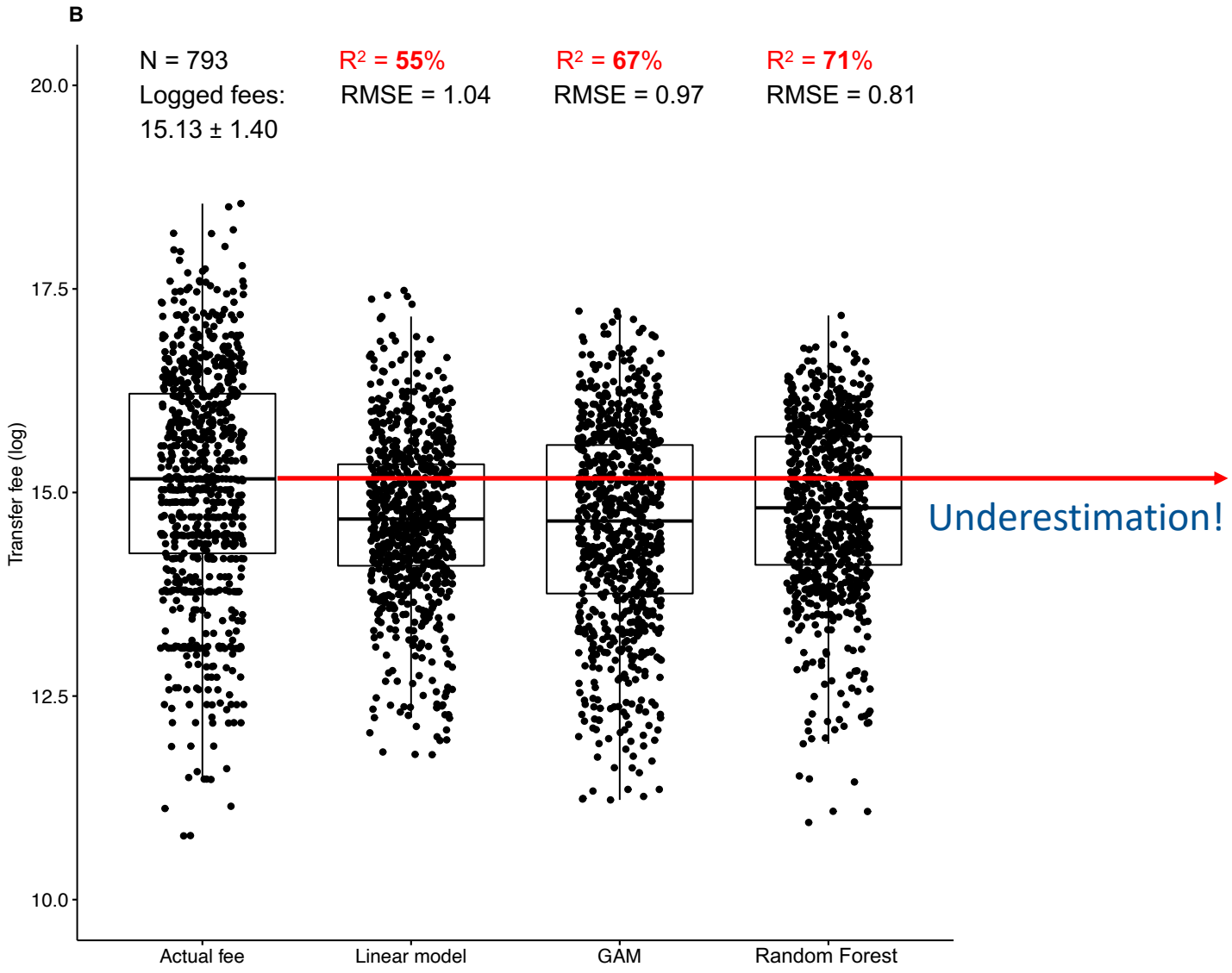
Quantile-dependent relations



Random Forest-based Variable Importances

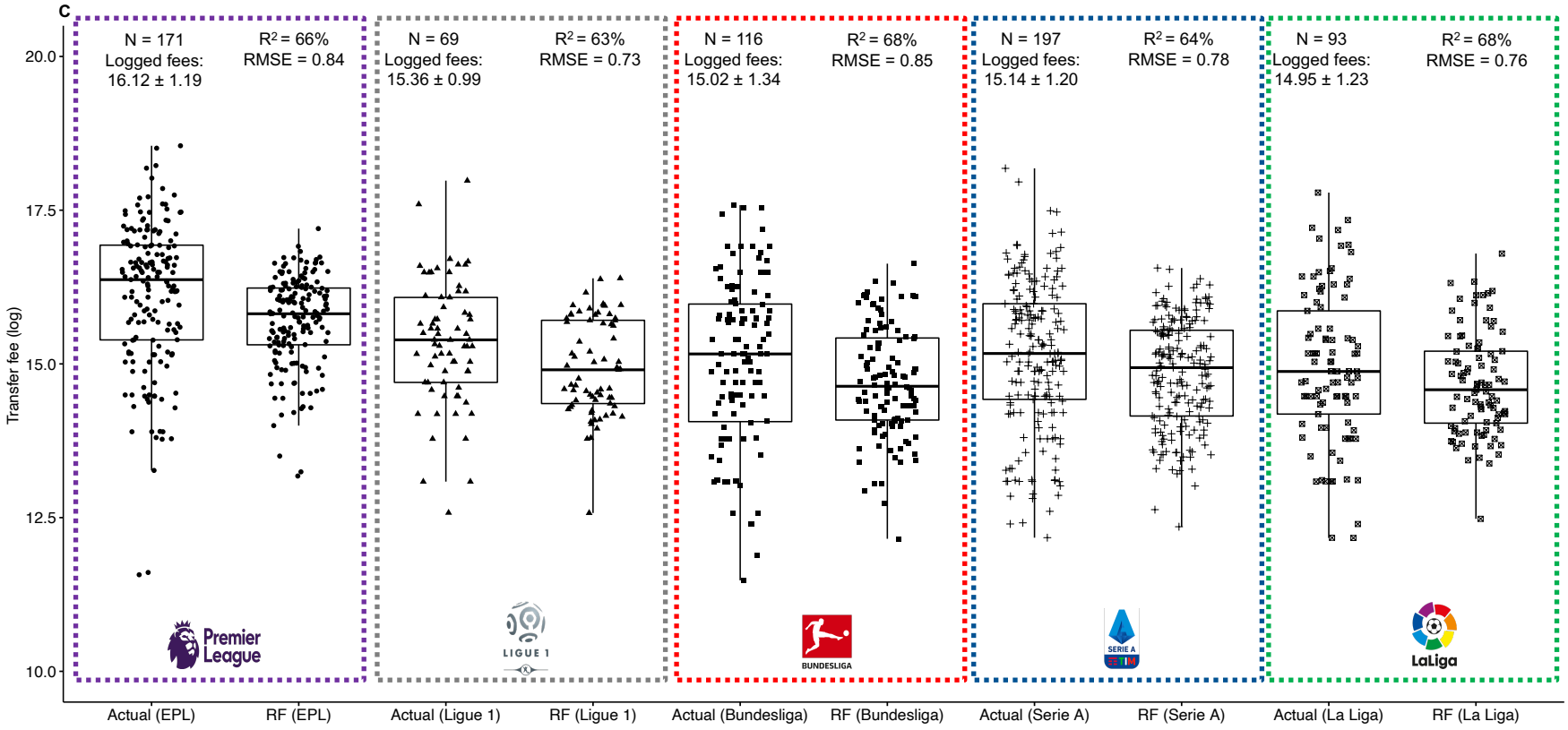


Predicted vs. Actual Transfer Fees During COVID-19



Note. Non-significant mean differences between predicted and actual fees

Robustness Checks: European Big Five Leagues



Underestimation: clubs paid *higher* transfer fees than predicted during COVID-19

We conducted a series of robustness checks (e.g., sample, contract length, trimming); our main findings remain.

Discussion

Contribution, Outlook, & Limitation:

- *Theoretically*: Moving beyond linearity is insightful:
  - Precise nonlinear relationships
  - Variable importance
  - Make predictions (vs. correlational)
- *Prediction*: No evidence of cooling-off effect of the transfer market during COVID-19
- Enlarging variable space was appropriate (e.g., Player injury history, Player remaining contract length; Club UEFA coefficients)

- *Outlook*: AI aiding decision making and predictions in football
- *Limitations*:
  - Causality
  - Sample selectivity & incomplete data

Thank you very much for your feedback!

yanxiang.yang@tum.de