

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

Yanxiang Yang ¹, Joerg Koenigstorfer ¹, & Tim Pawlowski ²

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

Background

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).

Previous studies commonly rely on simple linear regressions and explore a rather limited set of variables for comparably small samples (Carmichael & Thomas, 1993; McHale & Holmes, 2022).

Research Questions

- What are the key determinants of transfer fees?
- How can we make comparably accurate predictions for such fees?
- Did the COVID-19 pandemic affect the relevance of common predictors and the accuracy of predictions based on pre-COVID-19 evidence?

Aims

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. Moreover, we provide a simple test of whether the transfer market has changed since the beginning of the COVID-19 pandemic.

Methods Sample **Tariables Player characteristics** Time effects transfer Mmarkt Crowd-sourcing community Inclusion criteria Reliable database (Herm et al., 2014 Season 08/09 – season 21/22 Transfers with fee (no free, loan transfers) Widely used in sport management (Feuillet et al., 2020 Season 15/16 – season 21/22 Transfers within the European Big five Nationality: Europe, Asia, Africa, South Transfer window (summer, winter) America. North America Between the seasons 08/09 and 21/22 (until Position: Defender, goalkeeper, attacker, midfielder Web scraping with Python Remaining contract length (days) **Player performance** Selling & buying club Five sets of variables characteristics **UEFA Champions League** Player characteristics Data processing Player performance (injury) Appearances Missing information: imputed manually Arrivals of players Duplicates: removed (e.g., same transfer while either Selling-club characteristics Substitution on Departures of players as incoming or outgoing in two leagues). Buying-club characteristics Substitution off Transfer income Minutes played Transfer expenditure Points (/1000 MP) Spectators Goals (/1000 MP) Full data set (*n*=**7918** transfers) **UEFA** club coefficients Assists goal (/1000 MP) Robustness checks League ranking Yellow cards (/1000 MP) Models on full set (n=7918) Leagues (13 types) Models on full set from 2015/16 on Player injury history: (Injury days/injury Final models (*n*=**3512** transfers) frequency)/age Note. All variables refer to previous season Predictions for each of the big five Since season 2015/16 or player career history With contract length data Trimming 1% and 99% of transfer fees Modelling Logged transfer fee = $\beta_0 + \beta_1*(player\ characteristics) + \beta_2*(player\ performance) +$ β_3 *(selling-club characteristics) + β_4 *(buying-club characteristics) + β_5 *(time effects) + ε **Supervised Machine Learning Framework** Go beyond linear functional forms of OLS (machine learning; James et al., 2013) Random data splitting (n = 3,512): training (n = 1,903), testing (n = 816), during COVID-19 (n = 793) 10-fold cross-validation Evaluating model performance: testing R² and RMSE (Root Mean Square Error) Ordinary Least Squares Generalized additive model (GAM) Quantile additive model (QAM) Interpretability Random Forest

Results Moving Beyond Linearity: Quantile Additive Non-linear Effects of Predictors 0.75 0.50 Selling-club characteristics 0.75 0.75 Buying-club expenditure (EUR million) Buying-club UEFA coefficients **Random Forest-based Variable Importances Predicted vs. Actual Transfer Fees During COVID-19** Buying-club expenditure Selling-club income RMSE = 0.81 Remaining contract duration 15.13 ± 1.40 . • • European Big Five Leagues: Predicted vs. Actual Transfer Fees During COVID-19 $R^2 = 68\%$ N = 197 $R^2 = 64\%$ N = 93 $R^2 = 68\%$ Logged fees: 15.14 ± 1.20 RMSE = 0.78 Logged fees: Logged fees: RMSE = 0.84 RMSE = 0.73Logged fees: 15.02 ± 1.34 RMSE = 0.85 Logged fees: RMSE = 0.7614.95 ± 1.23 15.36 ± 0.99 16.12 ± 1.19

Conclusions

Premier League

 We showcase how moving beyond linearity and modeling quantiles can be revealing for both research and practice.

Actual (Bundesliga) RF (Bundesliga)

 The 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.

Selected References

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
FIFA. (2022). Global transfer market report 2021. https://doi.org/10.1080/00036849300000150
FIFA. (2022). Global transfer market report 2021. https://digitalhub.fifa.com/m/2b542d3b011270f/original/FIFA-Global-Transfer-Report-2021-2022-indd.pdf
McLaud Color of the color

Flexibility

 McHale, I. G., & Holmes, B. (2022). Estimating transfer fees of professional footballers using advanced performance metrihttps://doi.org/10.1016/j.ejor.2022.06.033
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). Springer.









Actual (La Liga)



TUM SCHOOL OF MANAGEMENT POSTER COMPETITION

Yanxiang Yang, Joerg Koenigstorfer, Tim Pawlowski

presented at the TUM School of Management Research Fest 2022 a poster on

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

Joachim Henkel

Joachin Huld

Vice Dean of Research & Innovation TUM School of Management