Predicting real-world market values of football players based on data from the FIFA 19 video game

Jay Kanchankoti

07/12/2021

Introduction

Player contracts are generally evaluated based on their performances in games, popularity with fans and sponsors, skills/abilities and many other factors. Generally, player performances are calculated through physical devices and sensors. However, this can be a costly and time-consuming process (Markovits et al., 2016). As an alternative, there is lots of data on player performances available for free online.The FIFA video game franchise hires expert reviewers to watch and rate players on various dimensions such as speed, passing, shooting, etc. to make their in-game players seem as realistic as possible. Singh & Lamba (2019)were able to predicte real market valuation of players through online datasets with a correlation of more than 0.5. This evidence supports the ability for us to study valuation of players based on FIFA ratings.

Market valuations will be ethically web-scraped from Transfermarkt.com. This website is known to provide accurate estimates of market value (Müller et al., 2017). In order to avoid the confounding variable of the league a team plays in, only teams from the UEFA Champion's League 2018 are considered. These values will be predicted through ratings of FIFA 2019. The research question is thus, "How can real-world market value of UEFA Champion's League players be predicted through the FIFA 19 video game?"

Methods

Variable Selection

Football players can be roughly divided into four types - defensive players, midfielders, offensive players and goalkeepers. Defensive players would be relatively better at defensive skills such as sliding tackles and standing tackles compared to offensive players who would be better at finishing and performing skill moves. A linear regression model would not be able to accurately predict the valuation for both a striker and a defender based on these parameters. For instance, they may both have the same market valuation but score extremely different on skills such as finishing and sliding tackle. Therefore, to get accurate predictions of market valuations, this study would focus on only valuing offensive players.

This assumption allows us to group variables based on the type of players - specifically into defensive skills, goalkeeping skills and midfielding skills. Some variables, such as agression and composure, can be grouped into mental attributes. The purpose of grouping these variables is to use analysis of covariance (ANCOVA) with partial F-test, to easily be able remove them if they are insignificant to the model. This method will compare two models, the complete one (with 'n' predictors) and a reduced one which has all the variables in the complete minus the group of variables selected ('n' - 'k' variables). Reducing the number of variables in a model would increase the unexplained variation of the response thus increase the Residual Sum Squared (RSS). The partial F-test checks whether the change in RSS due to the dropped variables is significant compared to the RSS of the complete model. If the p-value of the test statistic is more than 0.05, we do not reject the null hypothesis and conclude that the dropped group of variables are insignificant to the model.

The F-test statistic for a significant relationship is shown below:

$$P(F = \frac{RSS(\text{ drop })/k}{RSS(\text{ full })/n-p-1} < F(k, n-p-1)) < 0.05$$

Similarly to the partial F-test, t-test statistic will be used to remove individual predictors that cannot be grouped. In a sense, the t-test also checks how significant the removal of a single predictor is on increasing the unexplained variation of the response. If the removal significantly increases RSS, the p-value of the t-test will be less than 0.5, the null hypothesis will be rejected, thus, the predictor will not be removed. The p-value cutoff for the t-test is considered to be 0.5 instead of 0.05 as many predictors might have to be removed which would oversimplify the model and reduce its predictive accuracy.

Lastly, predictors will be removed from a model based on their correlation with other predictors. If they have high correlation, they may affect the sign of the coefficients, incorrectly lead to highly significant F-test values for a model which is not significant on individual predictors, or would lead to high standard deviations for coefficients. This multi-collinearity will be assesed using Variance Inflation Factor (VIF). It assess the variance of a predictor explained by other predictors. If the VIF is more than 5, it indicates severe multi-collinearity and that predictor will be removed from the original model.

$$VIF(X_j) = \frac{1}{1-R_j^2}$$

Model Validation

This requires two independant sets of data. One dataset will be used to train a model while the other will be used to test the model. However, since we already have all the information on UEFA Champion's League players, we cannot find new information Instead, the existing dataset will be split in two equal halfs, to test and train the model. If the coefficients for the predictors in both models are similar, they have a similar adjusted coefficient of determination ($Adj-R^2$) and none of the assumptions of a linear model are violated in the test and training models, the models will be said to be validated.

Model Violations & Diagnostics

There are four assumptions needed while considering linear regression. First, the population errors should have a mean of 0. Second, the population errors are uncorrelated to each other. Third, the population errors have constant variance around the conditional mean. Lastly, the population errors must be normally distributed with a mean of 0 and have a covariance matirx based on the second and third assumption. The following equation sums up these assumptions:

$$\mathbf{Y} \mid \mathbf{X} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\right)$$

, where $\mathbf{X}\boldsymbol{\beta}$ refers to the conditional mean in the first assumption and $\sigma^2\mathbf{I}$ refers to the covariance matrix for the second, third and fourth assumption.

To check whether these assumptions hold true, residual versus predictor plots, residual versus fitted values plots and Normal quantile-quantile (QQ) plots will be used. To obtain residuals and check these plots, a base model will be created using all the variables in the dataset. Assumptions will hold if no discernible patterns are observed in these plots. Essentially , in the residual plots there will be a look out for any systematic patterns such as curvi-linear relationships, large clusters of residuals separate from the rest of the points or fanning pattern of residuals. The Normal QQ plots will compare the normality of the residuals compared to standard normal deviations and violations. These residual plots asses two additional conditions - whether the mean response is a single function of the linear combination of predictors and whether conditional mean of each predictor is a linear function of another predictor. Model violations can be attempted to be corrected through transformations on the predictors using the powertransform function in R. Lastly, if multiple reduced models must be compared, diagnostics such as adjusted coefficients of determination, AIC_c and BIC will be used to select the best model.

Results

Numerical summarues of each variable is listed in the Appendix - Table A1.

The Normal QQ plot of the base model revealed violation of the normality assumption of the residuals. As can be seen in Fig. 1, the tail ends of the Normal QQ Plot deviate quite a lot from the diagonal line that represents the standard normal deviation. Additionally, some of the individual residual vs predictor plots do not have a linear relationship as seen in Fig. 2 of the residual versus Finishing plot. All the points are clustered towards the right.

Fig. 1 Normal QQ Plot of Base Model

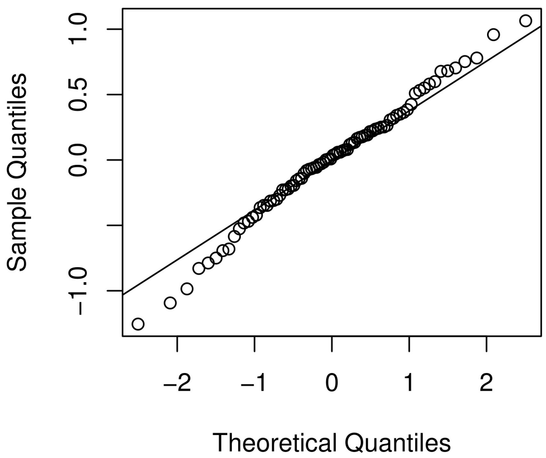
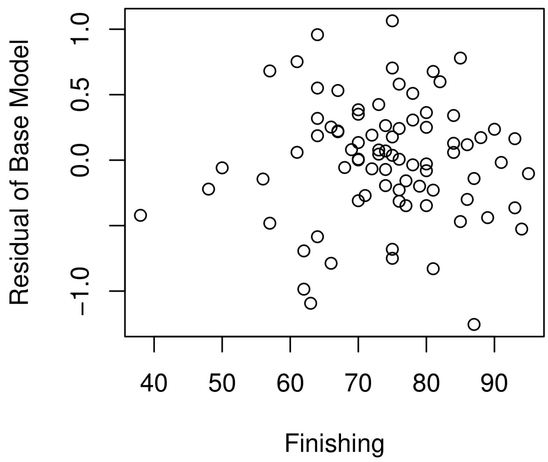


Fig. 2 Residual versus Finishing Plot



A list of all the transformations to correct the violated assumptions are provided in the Appendix, Table A2. While, these transformations were effective in correcting the residual versus predictor plots, they were not as effective at fixing tail ends of the Normal QQ Plot. This was corrected later after the variable selection methodology.

The ANCOVA Partial F-test were conducted on four groups of variables. These groups were formed based on formulas used by FIFA to get an overall score for each player:

Group 1: Goalkeeper attributes - Goalkeeper Diving, Goalkeeper Handling, Goalkeeper Kicking, Goalkeeper Positioning and Goal Keeper Reflxes

Group 2: Defender Attributes - Standing Tackle, Sliding Tackle, Marking, Interception and Heading Accuracy

Group 3: Mental Attributes - Aggresion, Positioning, Vision and Composure

Group 4: Midfielder Attributes - Long Passing, Shot Power, Crossing and Long Shots

The results indicated that each of these group were insignificant and could be removed from the model. The p-values of the ANCOVA Partial F-test are shown in the table below (Table 1) to be more than 0.05. The table also includes an F-test of the model after removing all insignificant individual variables using t-test methodology. These individual variables which had a p-value of more than 0.5 are Curve, Acceleration, Height, Dribbling, Balance, Agility, Weight, Volleys. The RSS does not increase by a lot after reducing the model. The number of predictors removed in each model is shown by Df or degrees of freedom. The F-test statistic its p-value are also given.

Table 1: ANCOVA Results

Model	**Res. Df**	**RSS**	**Df**	**F**	**P-value**
Base Model	41	21.908			
Reduced Model without Group 1	46	25.462	5	1.3301	0.2707

Table 1: ANCOVA Results

Reduced Model without Group 2	51	28.093	5	0.9507	0.4577
Reduced Model without Group 3	55	28.471	4	0.3781	0.9519
Reduced Model without Group 4	59	30.055	4	0.7650	0.5526
Model after t-test reductions	67	31.438	5	0.3392	0.9471

Multi-collinearity checks through VIF revealed that Reaction had the highest VIF at 13.6802 and was removed first. Removing it significantly reduced the VIF of other predictors. Similarly, Ball Control had the highest VIF in the new model and was removed followed by Jumping. Model diagnostics will look at changes in adjusted R squared, AIC_c and BIC as shown in Table 2 below. Based on these results it is reasonable to choose the last model as it has a high adjusted R squared while not compromising a lot on AIC_c and BIC relative to other models.

Table 2: Model Diagnostics

Model	SSres	Rsq	Rsq_adj	AIC	AIC_c	BIC
Reduced Model	31.4378053	0.8662384	0.8382882	-245.5591	-241.9324	-191.8640
Model without Ball Control	31.9314908	0.8641379	0.8381642	-244.9882	-241.8094	-194.3990
Model without Reaction	35.8537089	0.8474496	0.8209191	-227.8722	-225.1090	-180.3890
Model without Jumping	36.0723248	0.8465194	0.8224011	-228.8692	-226.4901	-184.4919

Fig. 3 shows the Normal QQ Plot with tail ends not deviating as much in the final model compared to the base model. A test on problematic observations revealed that there were several outliers, leverage points and influential observations. These are highlighted in a response versus predicted response plot given in Fig. 4.

Fig. 3 Final Normal QQ plot

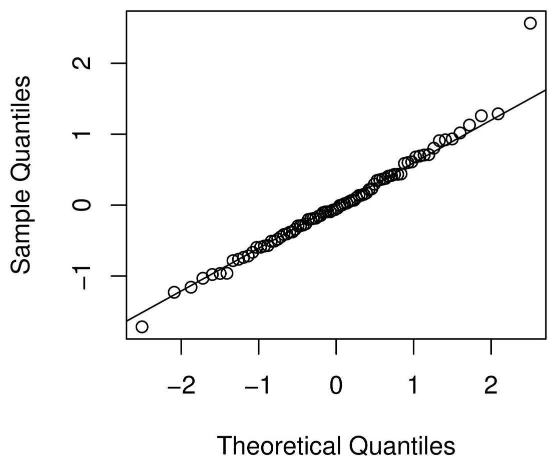
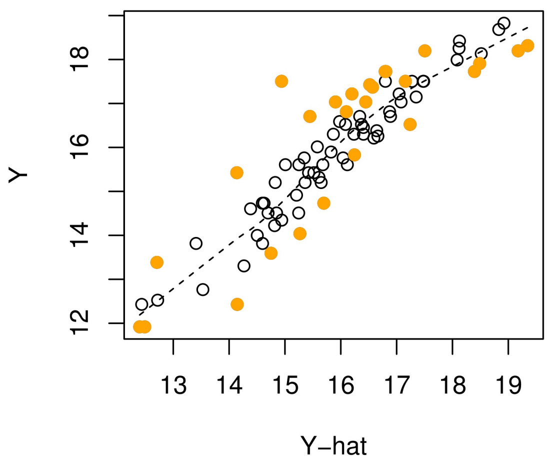


Fig. 4 Final Y vs. Y−hat



Lastly, when the model was tested on the testing dataset, it resulted in similar estimates for coefficients, and the same variables were significant. However, the Adjusted R squared for the training dataset was 0.8224 while that for the test was 0.7018. All assumptions were met for the model in the testing dataset as revealed by plots in the Appendix.

Discussion

Following is the final model selected to best answer the research question, "How can real-world market value of UEFA Champion's League players be predicted through the FIFA 19 video game?":

$$LogT.Value = Age + International.Reputation + Weak.Foot + Skill.Moves + Finishing + ShortPassing$$
$$+FKAccuracy + SprintSpeed + Stamina + Strength + Penalties$$

Model validation led to satisfactory results while there were some limitations. One of them was the several problematic observations seen in the response versus fitted values plot. Since, the data was artificially created by reviewers, there could be a possibility that their bias in rating players may have led to some of the problematic observations. However, one cannot be sure of the exact reasons of why these observations appear. Thus, even though these observations cannot be removed or fixed, it is equally important to acknowledge their presence.

Secondly, with only 82 data points to work with, for both the test and training dataset, it is quite a small data set compared to the thousands of players who play professional football. This might explain the high variance in the final model. Acknowledging the higher variance and irregular normality of residuals, the individual variables selected were allowed a higher p-value of 0.5. Another limitation to be considered is that the transformations applied on the variables could be too specific to the training dataset. This could explain the lower adjusted R-squared that resulted when validating the model on the testing data set. An Adj. R-squared of 0.7018 on the testing data set is significantly lower than a score of 0.8224 on the training data set.

These problems/limitations cannot be solved by going back and changing the model as it defeats the point of testing a model. Thus, the most that can be done is acknowledge the presence of these problems. Future studies should look at creating models for defenders or midfielders from the same dataset. This is easily possible as all the information is already available for such a study. Moreover, the final model attained in this study should be tested with data sets from other leagues such as the Bundesliga in Germany, or the Indian Super League in India or any other football league. This would be a real test of the model as it would reveal if the same characteristics are used to value players around the world. Lastly, to improve on this study and further more to come, the number of data points considered can be increased by looking at bigger leagues such as the English Premiere league held in United Kingdom. It is one of the biggest and most watched leagues apart from the UEFA Champion's League.

References

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

UEFA Champions League 18/19 - participating teams. Transfermarkt. (2019). Retrieved December 10, 2021, from https://www.transfermarkt.com/uefa-champions-league/teilnehmer/pokalwettbewerb/CL/saison_id/2018.

Markovits, Andrei & Green, Adam. (2016). FIFA , the video game: a major vehicle for soccer's popularization in the United States. Sport in Society. 20. 1-19. 10.1080/17430437.2016.1158473.

Singh, P., & Lamba, P. S. (2019). Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players. Journal of Discrete Mathematical Sciences and Cryptography, 22(2), 113–126. https://doi.org/10.1080/09720529.2019.1576333

FIFA 20 attributes guide - all players attributes explained. | FIFA Ultimate Team. (2021, May 12). Retrieved December 15, 2021, from https://fifauteam.com/fifa-20-attributes-guide/#:~:text=The%20aggression%20level%20of%20a,or%20commitment%20to%20a%20match.

Appendix

Table A1: Numerical Summaries

X	
minimum	1
median (IQR)	456 (230.00, 668.00)
mean (sd)	446.57 ± 262.37
maximum	912
Value	
minimum	150,000.00
median (IQR)	9,000,000.00 (2,200,000.00, 19,500,000.00)
mean (sd)	16,705,606.06 ± 21,297,643.75
maximum	110,500,000.00
T.Value	
minimum	150,000.00
median (IQR)	8,000,000.00 (2,500,000.00, 25,000,000.00)
mean (sd)	19,018,030.30 ± 26,874,522.34
maximum	150,000,000.00
Age	
minimum	16
median (IQR)	24 (21.00, 28.00)
mean (sd)	24.19 ± 4.63
maximum	34
International.Reputation	
minimum	1
median (IQR)	1 (1.00, 3.00)
mean (sd)	1.84 ± 1.06
maximum	5
Weak.Foot	
minimum	2
median (IQR)	3 (3.00, 4.00)
mean (sd)	3.39 ± 0.70
maximum	5

Table A1: Numerical Summaries	
Skill.Moves	
minimum	2
median (IQR)	3 (3.00, 4.00)
mean (sd)	3.24 ± 0.69
maximum	5
Height	
minimum	165.10
median (IQR)	180.34 (175.26, 185.42)
mean (sd)	181.11 ± 6.41
maximum	200.66
Weight	
minimum	121
median (IQR)	165 (157.00, 176.00)
mean (sd)	166.79 ± 17.15
maximum	223
Crossing	
minimum	23
median (IQR)	64 (50.00, 70.00)
mean (sd)	60.76 ± 13.84
maximum	90
Finishing	
minimum	35
median (IQR)	74 (66.00, 80.00)
mean (sd)	72.93 ± 10.88
maximum	95
HeadingAccuracy	
minimum	34
median (IQR)	65 (55.00, 75.00)
mean (sd)	64.64 ± 13.94
maximum	93
ShortPassing	
minimum	49
median (IQR)	73 (65.00, 78.00)
mean (sd)	71.55 ± 9.17
maximum	91
Volleys	
minimum	35
median (IQR)	69 (60.00, 76.00)
mean (sd)	67.67 ± 12.65
maximum	90
Dribbling	
minimum	48
median (IQR)	75 (68.00, 81.00)
mean (sd)	74.58 ± 8.69
maximum	97
Curve	
minimum	25
median (IQR)	66 (56.00, 76.00)
mean (sd)	65.45 ± 13.55
maximum	93
FKAccuracy	
minimum	27

Table A1: Numerical Summaries

median (IQR)	57 (45.00, 70.00)
mean (sd)	57.85 ± 15.50
maximum	94
LongPassing	
minimum	21
median (IQR)	62 (52.00, 71.00)
mean (sd)	60.47 ± 13.65
maximum	89
BallControl	
minimum	44
median (IQR)	76 (69.00, 82.00)
mean (sd)	75.47 ± 8.76
maximum	96
Acceleration	
minimum	29
median (IQR)	75 (68.00, 79.00)
mean (sd)	73.58 ± 10.73
maximum	94
SprintSpeed	
minimum	32
median (IQR)	75 (69.00, 79.00)
mean (sd)	73.79 ± 10.12
maximum	95
Agility	
minimum	33
median (IQR)	75 (68.00, 80.00)
mean (sd)	73.25 ± 10.43
maximum	94
Reactions	
minimum	43
median (IQR)	74 (64.00, 80.00)
mean (sd)	72.37 ± 11.11
maximum	96
Balance	
minimum	24
median (IQR)	70 (64.00, 78.00)
mean (sd)	69.70 ± 12.25
maximum	95
ShotPower	
minimum	47
median (IQR)	76 (68.00, 81.00)
mean (sd)	73.98 ± 9.54
maximum	95
Jumping	
minimum	34
median (IQR)	67 (59.00, 75.00)
mean (sd)	66.67 ± 11.57
maximum	95
Stamina	
minimum	34
median (IQR)	70 (63.00, 77.00)
mean (sd)	69.72 ± 11.45

Table A1: Numerical Summaries

maximum	93
Strength	
minimum	29
median (IQR)	68 (59.00, 78.00)
mean (sd)	67.61 ± 13.89
maximum	95
LongShots	
minimum	35
median (IQR)	70 (61.00, 76.00)
mean (sd)	68.78 ± 11.17
maximum	94
Aggression	
minimum	28
median (IQR)	59 (46.00, 69.00)
mean (sd)	58.47 ± 15.71
maximum	93
Interceptions	
minimum	11
median (IQR)	33 (24.00, 45.00)
mean (sd)	36.85 ± 16.31
maximum	86
Positioning	
minimum	41
median (IQR)	75 (68.00, 81.00)
mean (sd)	73.99 ± 10.94
maximum	95
Vision	
minimum	43
median (IQR)	70 (62.00, 77.00)
mean (sd)	69.19 ± 10.66
maximum	94
Penalties	
minimum	40
median (IQR)	66 (61.00, 75.00)
mean (sd)	67.18 ± 10.08
maximum	90
Composure	
minimum	43
median (IQR)	73 (64.00, 80.00)
mean (sd)	71.64 ± 10.95
maximum	96
Marking	
minimum	11
median (IQR)	38 (28.00, 50.00)
mean (sd)	39.75 ± 14.68
maximum	83
StandingTackle	
minimum	11
median (IQR)	34 (25.00, 46.00)
mean (sd)	36.95 ± 15.71
maximum	87
SlidingTackle	

Table A1: Numerical Summaries	
minimum	12
median (IQR)	29 (20.00, 40.00)
mean (sd)	32.02 ± 15.21
maximum	90
GKDiving	
minimum	1
median (IQR)	10 (8.00, 13.00)
mean (sd)	10.15 ± 3.41
maximum	27
GKHandling	
minimum	1
median (IQR)	10 (8.00, 13.00)
mean (sd)	10.35 ± 3.35
maximum	25
GKKicking	
minimum	1
median (IQR)	10 (8.00, 13.00)
mean (sd)	10.08 ± 3.65
maximum	31
GKPositioning	
minimum	1
median (IQR)	9 (7.00, 12.00)
mean (sd)	9.82 ± 3.59
maximum	33
GKReflexes	
minimum	1
median (IQR)	11 (8.00, 14.00)
mean (sd)	10.73 ± 3.91
maximum	37
LogT.Value	
minimum	11.92
median (IQR)	15.89 (14.73, 17.03)
mean (sd)	15.77 ± 1.60
maximum	18.83

Table A2: Transformations

Variable	**Transformation**
Age	log(x)
Crossing	x^2
Finishing	x^2
Heading Accuracy	x^2
Short Passing	x^3
Volleys	x^2
Dribbling	x^2
Curve	x^2
Long Passing	x^2
Ball Control	x^3
Acceleration	x^2
Sprint Speed	x^2
Agility	x^3

Table A2: Transformations	
Reactions	x^2
Balance	x^2
Shot Power	x^2
Stamina	x^2
Long Shots	x^2
Interceptions	sqrt(x)
Positioning	x^3
Vision	x^2
Composure	x^2
Sliding Tackle	sqrt(x)
GKDiving	sqrt(x)
GKKicking	sqrt(x)
GKPositioning	sqrt(x)
GKReflexes	sqrt(x)