Are soccer players worth their weight in gold?

Andrej Pospisil and Sean Dube

Motivation and Goal

Even if you are not a soccer fan, you may have heard the names Cristiano Ronaldo and Neymar. These professional soccer players are attackers, and their clubs spent more than 100 million euros to acquire them. Many may wonder whether midfielders and defenders are more or less expensive than attackers. To address this, we explore whether attackers are more valuable than non-attackers, and if there is a causal effect between player skill and market value.

Data Origin and Overview

EA Sports FIFA is a video game where players can simulate a soccer match with professional teams and players. Each player has six main attributes: Pace (the ability to run fast with and without the ball), Shooting (hitting the ball in an attempt to score a goal), Passing (hitting the ball accurately towards a teammate), Dribbling (maneuvering the ball past players), Defending (taking possession of the ball from an opposing player), and Strength (the ability to stand your ground in a shoulder to shoulder duel). Goalkeepers have their own six main attributes: Reflexes, Handling, Kicking, Positioning, and Speed. Attributes are rated on scale from 1 - 99. The scores are not assigned randomly, but according to continually-updated, collated opinions of many thousands of volunteers who watch the players play in real life and evaluate their skills based on their performance. In addition to these skills, our dataset also contains information about a player's position and market value. Position denotes the role a player takes on the field. Players can be defenders, midfielders or attackers. Market value refers to the amount of money (transfer fee) that a club has to pay to acquire a player. The dataset contains information about the above-described variables for 6798 professional soccer players.

Exploratory overview of relationship between skill and position

First, we will explore the relationship between skill and position. Figure 1 shows the average player skill rating by position. Midfielders have the most balanced profile with similar ratings for each skill. Attackers and defenders are more specialized and have ratings skewed towards skills required for their respective positions. Figure 2 is a correlation half-matrix which presents the strength and direction of correlation between all combinations of player skills. Correlations between skills seem to vary in strength in direction. Strength and Defending have a positive correlation, but negative compared to all other skills. Shooting, Passing, and Dribbling have the highest positive correlations.

Figure 1

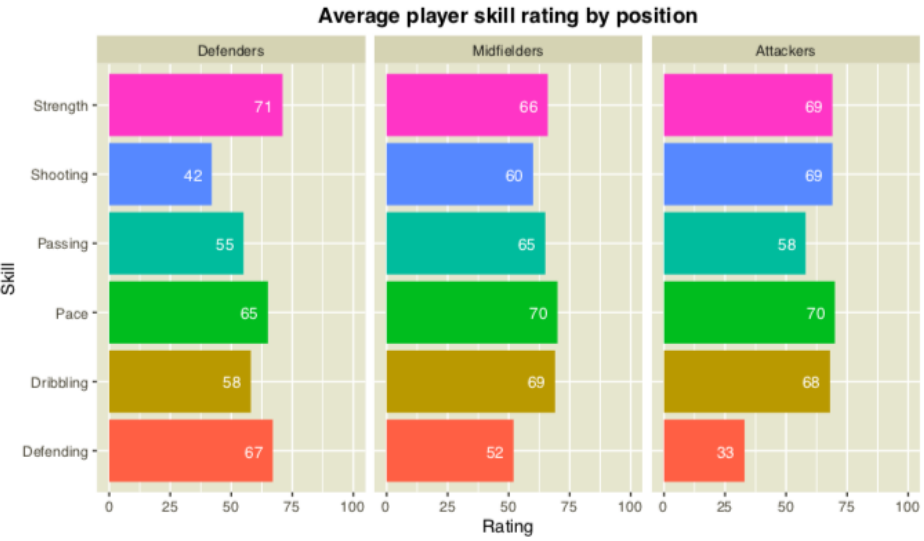

Figure 1. The horizontal axis represents rating on a scale from one to a hundred. The vertical axis shows which skill category is rated. Each subsection of the graph corresponds to players who play in the position indicated by the respective label.

Figure 2

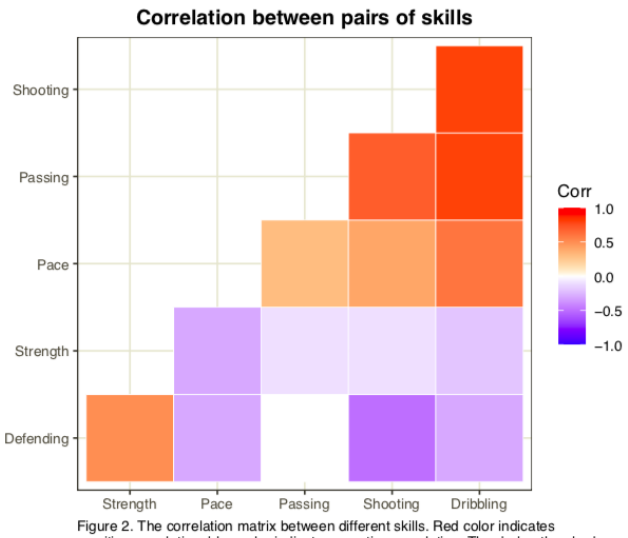

Figure 2. The correlation matrix between different skills. Red color indicates positive correlation; blue color indicates negative correlation. The darker the shade of the color, the stronger the correlation as indicated by the legend on the right.

Analysis of market value by position

Table 1 presents the distribution of player market value by position. Attackers seem to have the highest average market value followed by the midfielders and the defenders, respectively. Players with a market value of 0 are currently without a club and can be acquired for free (free agents). For each position, the distribution of market value is skewed towards the high-end. This makes sense because, relative to the entire soccer population, there exist very few "star" players with high market values. However, here as well we notice the possible discrepancy in market values since the attackers and the midfielders are even more right-skewed than the defenders with respect to market value - the most valuable midfielder is worth 118.5 million euros; the most valuable defender less than half of that at 57 million euros.

Table 1: Distribution of player market values by position (in million euros)

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Attackers	0.05	0.75	1.6	4.96	5.5	86.0
Midfielders	0.00	0.55	1.2	4.11	4.2	118.5
Defenders	0.00	0.48	0.9	2.84	2.9	57.0

Figure 3 is side by side boxplot which depicts the difference in market value by position. The skewness of market values for each position means we need to apply a log transformation to condense the scale of market values. Plotting market values without this transformation would only show outliers. The boxplots show that attackers have the highest market values followed by midfielders and then defenders.

Figure 3

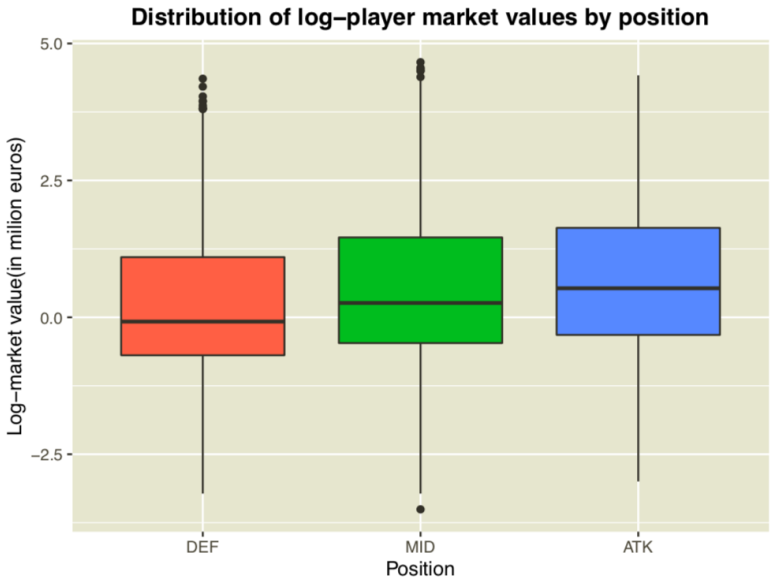

Figure 3. The horizontal axis shows the three categories of soccer players with respect to position: attackers (ATK), midfielders (MID) and defenders (DEF). The vertical axis represents the players' log−market value in millions of euros.

The table and the graph suggest that a difference in market value by position does exist. We constructed a multiple regression model with market value as the response variable. When we included Position as a predictor, i.e. indicator variable, we obtained that the attackers really do have the highest market value. On average, the midfielders are worth 0.8M euros less, while the defenders are worth 1.9M euros less. All of these differences are statistically significant as indicated by the respective p-values of the coefficients for each group of players. However, when we controlled for skill, i.e. when we included the skill variables into the model, we found that the defenders actually have the highest market value. This result, although surprising, does make sense, but it relies on domain knowledge

of soccer for interpretation. The attackers generally have higher overall skill ratings than the defenders. The reason for that is when an attacker scores one goal per game this player is considered excellent even if he misses ten other goal opportunities. Conversely, a defender who prevents ten goal chances, but allows one goal, is considered incompetent. The defenders with high overall skill ratings are rare and more expensive. If you want to buy an attacker with an overall skill rating of 85, there are a hundred players to choose from. If you want to buy a defender with the same overall skill rating, there are maybe ten of them.

Analysis of Causality

Even though the multiple regression model showed statistically significant association between market value and all six skill attributes, this association does not imply causation. A randomized experiment is required for the inference of causality. In an observational study such as this one, one approach that might lead us closer to the inference of a causal relationship is the examination of change of market value and skill over time. If a causal relationship exists, we would expect to see that, when skill changes, the market value changes as well. Conversely, when skill does not change, the market value does not change either. Figure 4 shows change in average player market value by position over the last five years. We notice that the market values of players in all positions have been growing (changing) between 2016 and 2018 and stagnating since. If there existed a causal relationship between market value and skill, we would expect that average skill would be changing (in either direction) between 2016 and 2018 as well, but remaining stable since. Figure 5, however, shows that this was not the case. Overall skill has been stagnating over the entire five year period, with minor deviations between certain years. This indicates that the relationship is likely not causal, i.e. that the players' market values are not increasing because they are becoming more skillful, but due to some other factor. One such factor might be a change in the belief of what is a normal price to pay for a player. Recently, the clubs owned by oil sheikhs have been paying exorbitant transfer fees to acquire players who, given their skill level, would have cost three times less ten years ago. One such transfer was that of Neymar from Barcelona to Paris Saint Germain for 222 million euros in 2017. Neymar's transfer has now become a new benchmark against which all other transfers are evaluated with respect to the balance between the player's skill and the appropriate transfer fee.

Figure 4

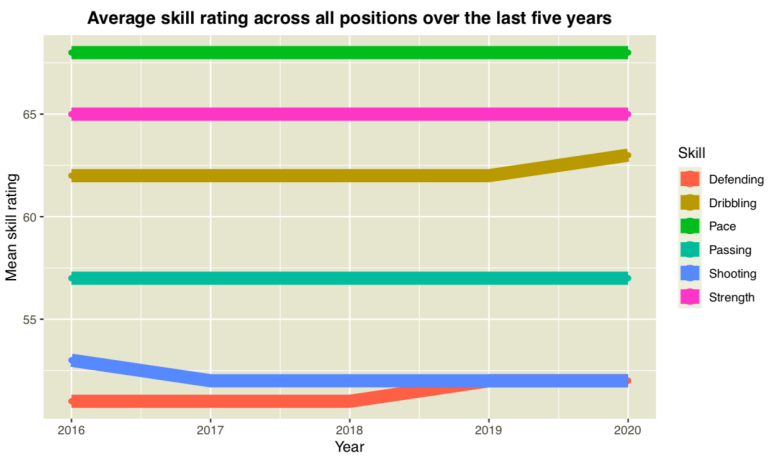

Figure 5. The horizontal axis represents year. The vertical axis represents mean skill rating. Each line corresponds to the skill indicated by the legend on the right.

Figure 5

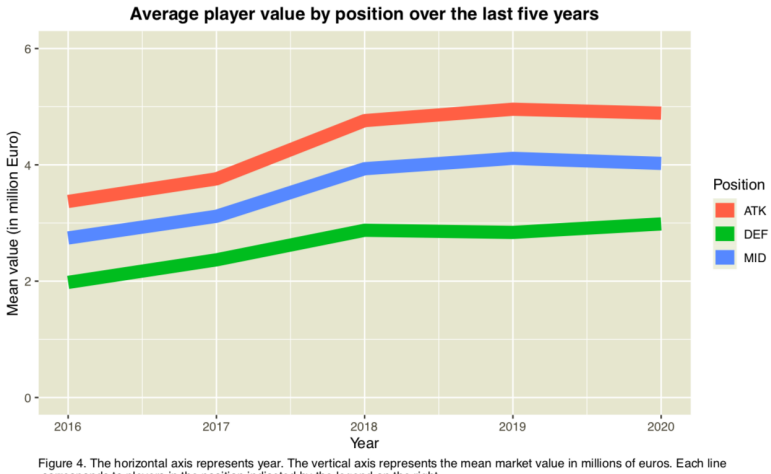

Figure 4. The horizontal axis represents year. The vertical axis represents the mean market value in millions of euros. Each line corresponds to players in the position indicated by the legend on the right.

Limitations

The main limitation of this study is the inability to measure skill objectively. The foundation of this study relies on the assumption that the ratings are consistent between each other. An example of consistency would be that the two players who possess the same percentage of physical duels won have the same rating in Strength. An additional

complication is the fact that skill categories are based on several objective parameters. For example, Pace encompasses running both with and without the ball. If Player A can run without the ball at 8mph and with the ball at 4mph, how should his Pace rating compare to that of Player B who can run without the ball at 7mph and with the ball at 6mph? Therefore, the creators of these ratings should both possess the data on these objective parameters and the formula which aggregates that data into a single number. This is likely not the case. Finally, we have to assume that the ratings are honest in that they only stem from a player's performance, not a rater's affinity for the player.

Conclusion

We were curious if the attackers have a higher market value than the non-attackers and if there is a causal effect between skill and market value. A randomly chosen attacker would likely be more valuable than a randomly chosen defender. However, if you want to buy a player with an overall skill rating of 85, you would have to pay more for a defender than an attacker. Our study also found no evidence of causality between skill and market value. There might be other lurking variables that may help determine if a causal relationship exists.