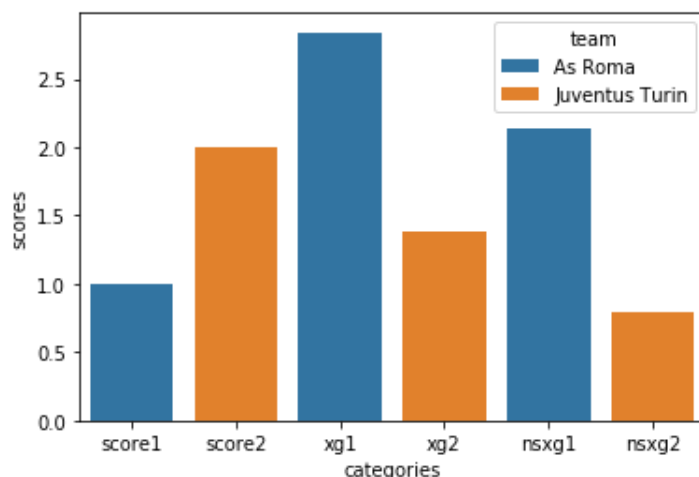


How the Course of the Game Influences Expected Goals in Soccer Matches

By Michael Oberst

In soccer analytics, it has become fashionable to rely on so-called “expected goals” metrics (or “xG”, for short). Expected goals have been invented as a way of evaluating the chances of a team not just by counting them, but also by considering their quality. That is, shots from inside the goal area have a higher xG value than shots from far behind, shots from a one-on-one situation a higher value than shots under pressure, shots by foot count more than headers, and so on. Expected goals are calculated such that their value represents the probability of scoring a goal – for instance, a shot having an xG score of 0.2 has a probability of 20 percent of being converted. Expected goals are often used to decide which team is better.¹ For example, on January 12 in 2020, AS Roma lost the league match against Juventus by 1-2 in Italian Serie A. However, when you look at the xG scores assigned by the website FiveThirtyEight,

Roma led 2.84 to 1.38!² So Roma actually was the better team, right?



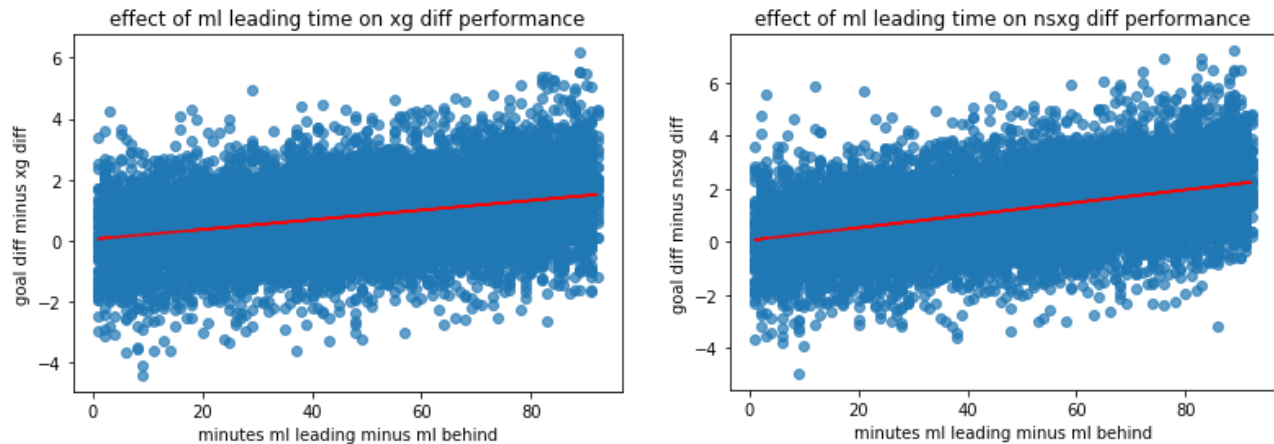
Well, this may be the case, but the conclusion would be too hasty. There are a couple of reasons why expected goals do not tell the full story of a match. Some of them are straightforward: perhaps the striker is off form, or the goal keeper makes exceptionally good saves. Both factors can contribute to the xG score looking better for a team than the actual

final scoreline. Or a team creates many dangerous situations that nonetheless do not result in a shot, which keeps the xG score low. For this reason, FiveThirtyEight predicts matches by using not only expected goals, but also metrics called “adjusted goals” (which adjusts the actual score for the minute when a goal was scored or for dismissals) and “non-shot expected goals” (or just “non-shot”, which considers actions like “passes, interceptions, take-ons and tackles” happening close to the opponent’s goal).³

However, there is also a less obvious reason why we should not trust expected goals alone. Consider what happened in the match between Juve and Roma: while Juve scored two early goals in the 3rd and

10th minute which gave them a comfortable lead, Roma managed to score one goal only in the 68th minute. When Juve was ahead 2-0 by the latest, they had no great interest in scoring yet another goal.⁴ All they needed to win was to prevent Roma from scoring more than one goal. It is debatable how good they have been at that – Roma's xG score of 2.84 is respectable – but, for the rest of the match, the strategic situation for Juve was such that it was more important for them not to concede a goal than maximizing the goal difference. For Roma, it was quite the converse: if they wanted to reach at least a draw, they needed to score two goals, so they had to attack even if this could mean that it increases the likelihood of receiving one goal more than of scoring one. And the fact that Roma scored a goal in the second half means that they had a realistic chance of avoiding a loss until the very end. In brief, the correct strategy depends on the scoreline: if a team is behind, they need to focus on scoring a goal; if a team is leading, they must give priority to defence.

To be sure, it is yesterday's news that teams adapt their strategy to the current scoreline. However, the effect on the xG score appears to have been unexplored so far. I have used data from FiveThirtyEight and built a web scraper to get the minute of a match when a goal was scored from [weltfussball.de](https://www.weltfussball.de) (which contains a comprehensive list of soccer results) to analyse more than 10,000 matches from leagues all over the world between 2016 and the pandemic-induced suspensions in March 2020 for which FiveThirtyEight provides xG and non-shot scores.⁵ To determine the connection between the scoreline and expected goals, I looked at the effect that the length of the period during which a team leads (call this the "leading period") has on xG values. For this purpose, I determined the team that were in the lead for a longer period than the opponent – call this the "main leader", which is not necessarily the winner – took the length of the time during which the main leader led, and, if the opponent also was in the lead in that match, subtracted the length of the leading period of the opponent from the one of the main leader.⁶ I also looked at the goal difference, the expected goal difference, and the difference of these differences. The goal difference in the match between Juventus and Roma is 1 for Juve, the xG difference -1.46, and the difference of these differences is 2.46 – in other words, the main leader Juventus has a goal difference which is 2.46 better than the xG difference suggests. Call this the "xG diff performance". Finally, I performed a simple linear regression that determines the effect that the length of the period during which the main leader leads has on the xG diff performance. I did the same for FiveThirtyEight's non-shot expected goals score, where Juve's non-shot diff performance is 2.34 goals per 90 minutes. The results of the linear regressions look like this:



One can easily see that there is a correlation – and a fairly big one. For expected goals, the effect is 1.43 goals per 90 minutes. That is, a main leader that scores the winning goal just before the full time whistle has an xG diff performance of only 0.05. But a team that is leading for 90 minutes is expected to overperform its xG difference by 1.48 goals. The effect is even more impressive for non-shot: it is 2.14 goals per 90 minutes.⁷ As it happens, Juve led for exactly 90 minutes.⁸ Does this mean that Juve’s remarkable overperformance is merely, or at least in large part, the result of the fact that Roma was behind for almost the whole match and a corresponding shift of strategy?

Well, things are more complicated than that. For there is also a meaningful correlation between the goal difference and the xG and the non-shot diff performances. The Pearson correlation coefficient is 0.60 for xG and 0.71 for non-shot. In the event of a draw, the main leader overperforms the xG score by only 0.11 goals, but if they win by 1, the overperformance is 0.64, it is 1.21 if they win by 2, and 2.64 if they win by 5. This is even more striking if you look at the non-shot expected goals score: in a draw, the overperformance is 0.21, it is 1.01 if the main leader wins by 1, 1.77 if they win by 2, and 3.64 if they win by 5. A team that scores more goals apparently is just more likely to have “got lucky”, which is why they overperform their xG or non-shot values by a higher margin. But a team that scores more goals is also more likely to lead for a longer period of time: the correlation between the length during which the main leader is leading and the goal difference is 0.46. In a match won by one goal, the main leader leads for 45 minutes on average, but if the match is won by five goals, the main leader leads for 75 minutes. Thus, the goal difference is a confounding variable that needs to be dealt with.

I addressed this issue by performing a linear regression for every match result individually and then forming an average of the coefficients weighted by the number of matches of every result. That is, I looked at, for all matches ending 1-0, how a team performs that leads for, say, 10 minutes as compared to a team that leads for 80 minutes, and did this for the other results, too. It turns out that the effect of

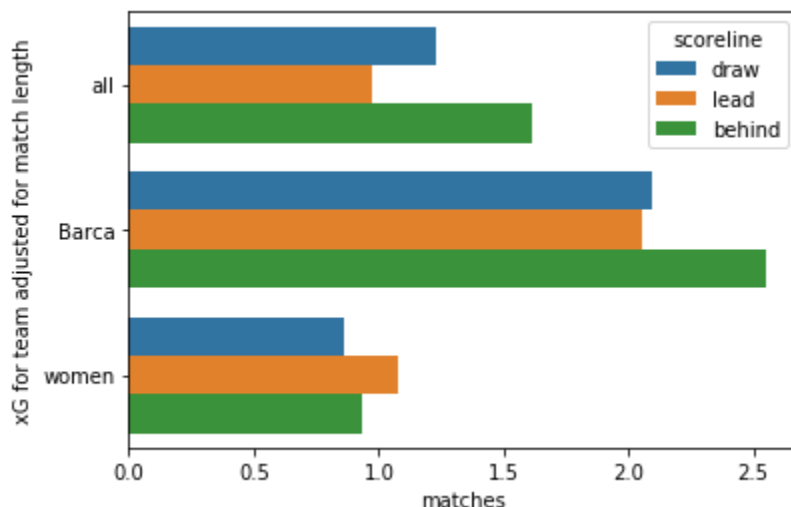
the leading time is smaller, but does not disappear: the weighted average of the regression coefficients is 0.63 goals per 90 minutes for expected goals and 1.05 for non-shot expected goals. While this may not look like much, the (weighted) standard deviations of xG and non-shot diff performance are 0.99 and 1.03, respectively, which is quite large and obtains independently of the leading time.⁹ That is, even if you are the leader for 90 minutes like Juventus, you could easily either underperform as compared to the expected goals or (as is the case with Juve) vastly overperform.¹⁰

This divergence arguably can in large part be explained by “luck” (i.e., good or bad shot conversion), but is also in line with what should be expected based on strategic reactions to the scoreline. For even if a team focuses on attack because they are behind, they might fail to create many shots if the leader does not admit them. Or perhaps the team succeeds in creating good opportunities, but the leader makes use of open spaces and gets even better chances. Another factor that can contribute to a main leader underperforming their xG goals is that at some point their opponent may just give up. But this is not what appears to have happened in the match Juventus versus Roma. Roma had a realistic prospect of equalizing until the very end – not the least because they got one back – and they created chances worth more than two goals, far better than Juve. In matches like this, the strategic effect may well amount to one expected goal per 90 minutes, although there is no easy way to estimate the effect size. In brief, it is plausible to assume that the effect resulting from strategic reactions to a changed scoreline can vary largely from game to game, even though the average effect is that the main leader modestly overperforms their xG and non-shot values.

The data from FiveThirtyEight do not contain information about *when* teams have their shots or what is the individual xG value of these shots. For this reason, I used freely available data from **STATSBOMB**, which provide just that.

Unfortunately, the free data they offer is not very representative: they

provide all league matches from FC Barcelona from 2019 back to 2004 and some women league matches in England and the USA.¹¹ Still, the Statsbomb data offer useful insights into what happens to expected goals when one team goes ahead. I looked at the xG values reached during the time a team was leading,



behind, or in a draw, divided this through the length of these periods, multiplied it by match length, and formed the difference to the xG score of the whole match. Finally, I averaged these scores of all matches weighted by the length of the respective leading periods. It turns out that there is a significant difference between periods during which the scoreline is a draw, when a team is leading, and when a team is behind.¹² In the event of a draw, a team has 0.03 more expected goals per minute than during the whole match. But this value goes up to 0.15 if a team is behind and down to -0.18 when a team is leading.¹³ This clearly shows that a team creates more and better chances if they are behind and less so if they are leading. Conversely, it also means that a team admits more and better shots if they are leading and less so if they are behind.

Finally, let's find a way to determine the leading time of a match by just looking at the result and the xG or non-shot scores. Since we usually do not know the main leader without knowing the leading time, let's perform the investigation for the winner and not the main leader as above (which means that draws are excluded from this analysis). To begin with, if you just look at all matches and perform a linear regression with the xG and the non-shot diff performance as input and the leading time as output, the xG diff performance is 8:11 minutes per goal and the non-shot diff performance 9:59 minutes per goal. That is, if you score exactly as many goals as the xG or the non-shot values suggest (i.e., the diff performance is 0), then you are expected to have a leading time of 37:56 or 31:05 minutes, respectively. But these scores go up to 46:08 and 41:04 if you overperform by one goal. However, we have learned that the goal difference is a confounding factor – if you form the weighted average of the coefficients obtained by regressions for single results in the same way as above, the effect size is much smaller: 3:25 minutes per goal for xG and 6:14 for non-shot performance.

In summary, I have investigated the effect of changing scorelines on (non-shot) expected goal metrics, both theoretically and empirically. It turns out that, on purely game-theoretic considerations, teams that are ahead and those that are behind alike do not focus on maximizing the goal difference, but rather on either avoiding goals of the team behind or scoring goals against the leader. This can explain the results of my empirical analysis, which has found that, the longer a team leads in total, the more the team overperforms as compared to xG and non-shot scores on average, even after adjusting for the result. Make no mistake: the variance is high, so the leading time can hardly be used for predicting the xG/non-shot diff performance. Still, when looking at expected goals, don't forget that they do not tell the full story of the match.

¹ See, e.g., <https://understat.com/>.

² See <https://projects.fivethirtyeight.com/soccer-predictions/serie-a/>.

³ See <https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work/>.

⁴ This of course depends on a team's strengths. If a team is good at attack, but lousy at defence, then it is probably better for them not to stop attacking when they are leading early in the game. There may also be a variety of other reasons why the actual behaviour of a team might differ from the way described above.

⁵ The data from FiveThirtyEight can be downloaded from <https://data.fivethirtyeight.com/#soccer-spi>. The leagues are A-League (Australia), Série A (Brazil), Super League (China), Barclays Premier League, League Championship (both England), Ligue 1 (France), Bundesliga, 2. Bundesliga (both Germany), Serie A, Serie B (both Italy), Primera Division Torneo Clausura (Mexico), Liga (Portugal), Primera Division (Spain). For a small number of matches, the data from weltfussball.de could not be assigned to the matches.

⁶ This includes most draws, but excludes matches without goals and matches in which the leading times of each team are equal. – Since the additional time is not included in the data, I added one minute as additional time for the first half (in most matches) and two minutes for the second half.

⁷ The R^2 values are 0.12 and 0.19, respectively, but they are close to 0 when one adjusts the regression for the result (see below).

⁸ Granted, the true leading time may be different due to imperfections in my estimation of the additional times, which is not provided by weltfussball.de.

⁹ I established the last point by slicing the leading time column into six groups depending on whether the main leader led for max. 15, 30 minutes, etc., and determining the standard deviation and the correlation between leading time and xG and non-shot performance within each group.

¹⁰ Of those main leaders winning 2-1, 32 percent underperformed their xG score, but only 2 percent overperformed it by a wider margin than Juve (that is, by 2.46 goals). The situation is similar with respect to the non-shot values: 25 percent underperformed and 6 percent overperformed by a higher margin than Juve (i.e., 2.34 goals).

¹¹ Additionally, they also provide free data of the last men and women world cups and, in the time since I downloaded the data, appear to have added male Champions League matches. But unlike league matches, in these matches teams are often focused on yielding a certain result (like winning by more than one goal). Hence, I consider it best not to include these competitions in my analysis.

¹² I tested for significance by performing a series of paired t-tests adjusted by the Holm-Sidak correction.

¹³ This statistic is potentially distorted by the unique dominance of Barcelona. If you just look at their matches, the value for a draw is -0.03, -0.13 when leading, and 0.14 when behind. But for women, the values are 0.14, -0.27, and 0.17, respectively. However, even if the effect may in fact be larger, the general conclusion remains the same.