



Expected goals, gamestate, and predictiveness

[Code](#) ▼R SOCCER

Is expected goals (xG) in neutral gamestates a better predictor of future performance than xG across all gamestates?

AUTHOR
Tony ElHabr

PUBLISHED
June 19, 2024

Introduction

❗ What is a 'neutral' gamestate?

The two most common definitions I've seen for "neutral" gamestate are periods of match time when

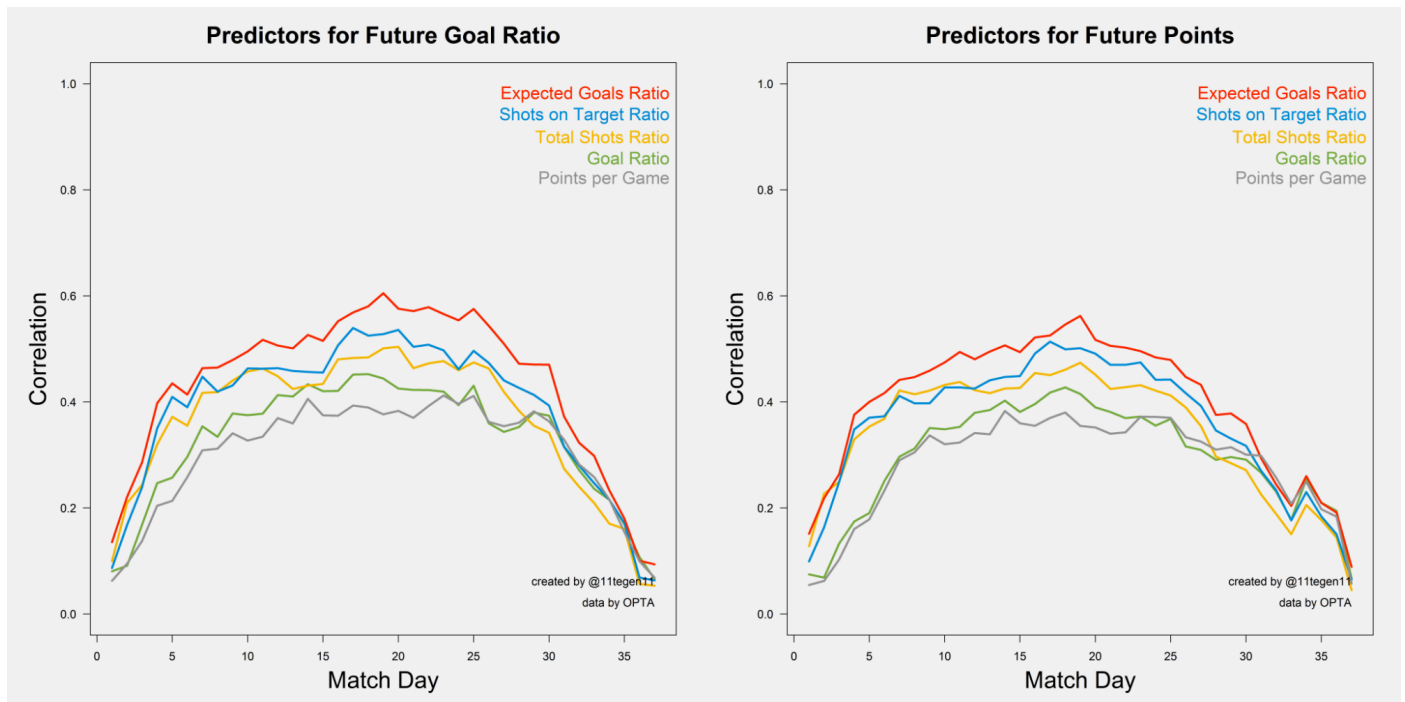
1. the score is tied, i.e. goal difference of 0
2. the absolute goal difference (GD) is 0 or 1

I investigate both.

💡 tldr: The answer to the header question

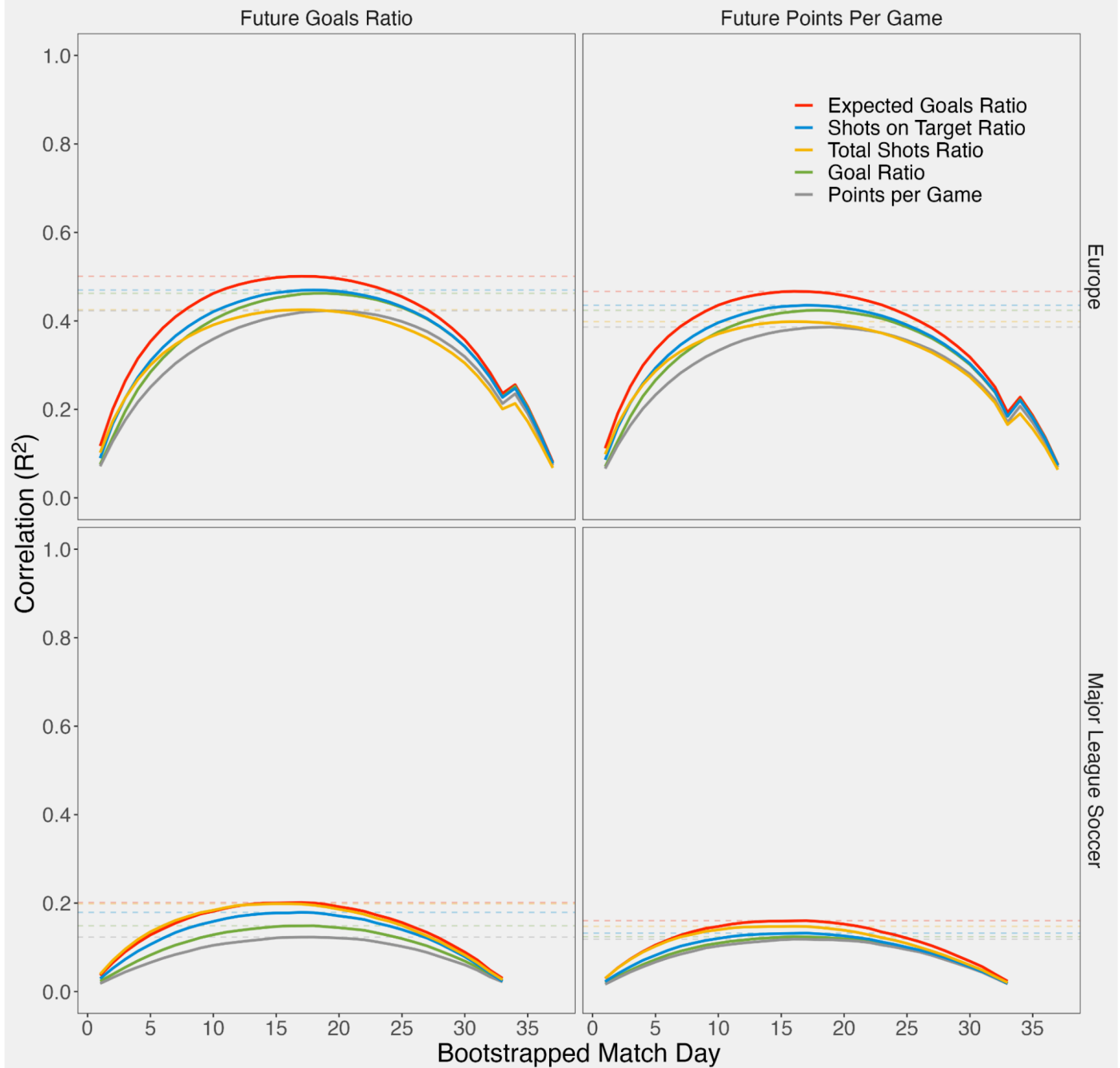


["The best predictor for future performance is Expected Goals"](#), authored by Sander Ijtsma of [11tegen11](#) in 2015, is one of the most influential pieces of writing in the public soccer analytics sphere. The write-up provided compelling evidence for the superiority of [expected goals \(xG\)](#) in terms of forecasting team-season outcomes on a game-by-game basis.¹² Ijtsma drew out curves for running [R² values](#) of a handful of season-to-date ("past") metrics with respect to rest-of-season ("future") performance indicators—goals ratio and [points per game](#)—and found that xG ratio³ tended to be the most predictive.



In 2022, [Eliot McKinley](#) wrote [a cool piece](#) for [American Soccer Analytics \(ASA\)](#) where he replicated and extended Ijtsma's analysis. He used a larger data set of more recent seasons—2018 through 2022— and added a look at [Major League Soccer \(MLS\)](#) in addition to the [the Big 5 European Leagues](#), and adding. Further, Eliot presented a novel aspect where he [bootstrapped](#) the procedure for generating running R^2 values to generate smoother, more interpretable curves.

Predictors for Future Results



i Smaller R^2 values in the MLS



In this post I extend this prior art in two ways:

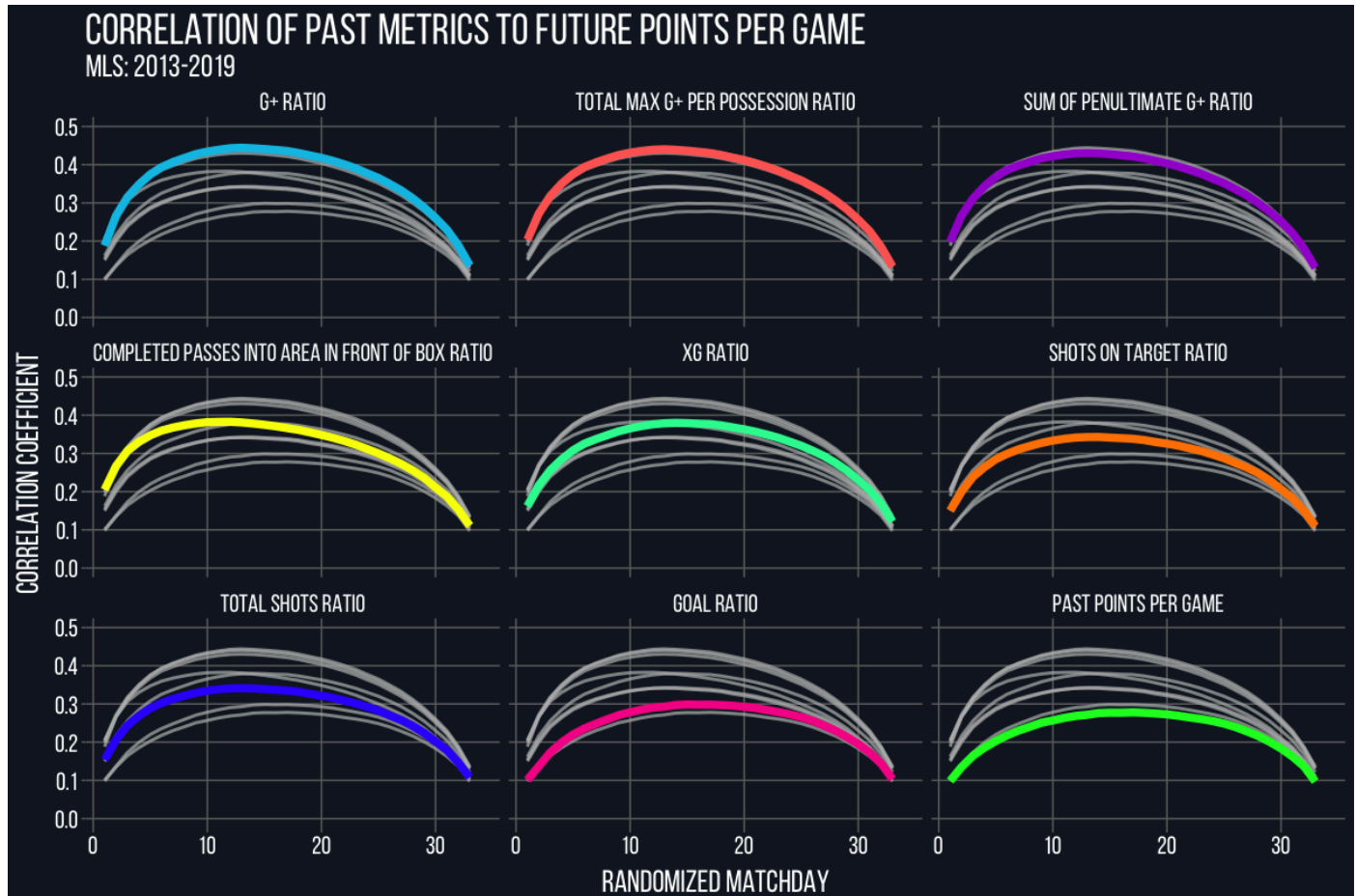
1. I **expand the data set range to 2018 through 2024**, sourcing from a reputed public data source—[FBref](#).⁴ The intention is to make the analysis even more robust and reproducible.
2. I include **additional curves for “gamestate-aware” goals and xG**, in an effort to see if one can achieve metrics that are perhaps even more predictive than xG itself.

On the second point, Ijtsma noted that subsetting the past performance tallies to gamestates when the score is either tied or within 1 goal might abate the potential biases introduced by within-match complacency (manifesting in conservative tactics). In fact, he said he had evaluated these things.

I've looked at this, but I have decided not to include that subanalysis in this post, for the sake of accessibility... I think this method will show that the phenomenon [where teams exert more or less effort in matches against particular opposition] either does not truly exist, or that its effect is so small that correcting for this will allow more noise and thereby weaken the model.

Unfortunately, I couldn't find any subsequent blog post where Ijtsma writes about the role of gamestate in the context of his R^2 analysis, so, alas, I'm here to do exactly that.

Now, my point isn't necessarily to find the absolute best metric possible to use to forecast future team performance. [Tiototal Football](#) (with assistance from Eliot) [showed that](#) comprehensive event-based metrics like goals added (light blue), completed passes into the area in front of the box (yellow), etc. are even more predictive of future points per game than xG, at least in the MLS.



I do look at a few metrics that go beyond Ijtsma's set in the Appendix, but that is not the primary focus of this post.

Data

I use `{worldfootballR}` to retrieve match-level figures and shot logs from FBref. As noted in the intro, we're going to be working with Big 5 and MLS data for the 2018 - 2024 (season-ending) seasons.

- Code Setup
- Data Setup
- Match-level data ingestion

► Gamestate data ingestion

Methods and Analysis

Replicating prior art

Let's begin with replicating Ijtsma's visualization where he evaluated the running R^2 of the five measures of past performance with respect to two measures of future performance. The five measures of past performance are:

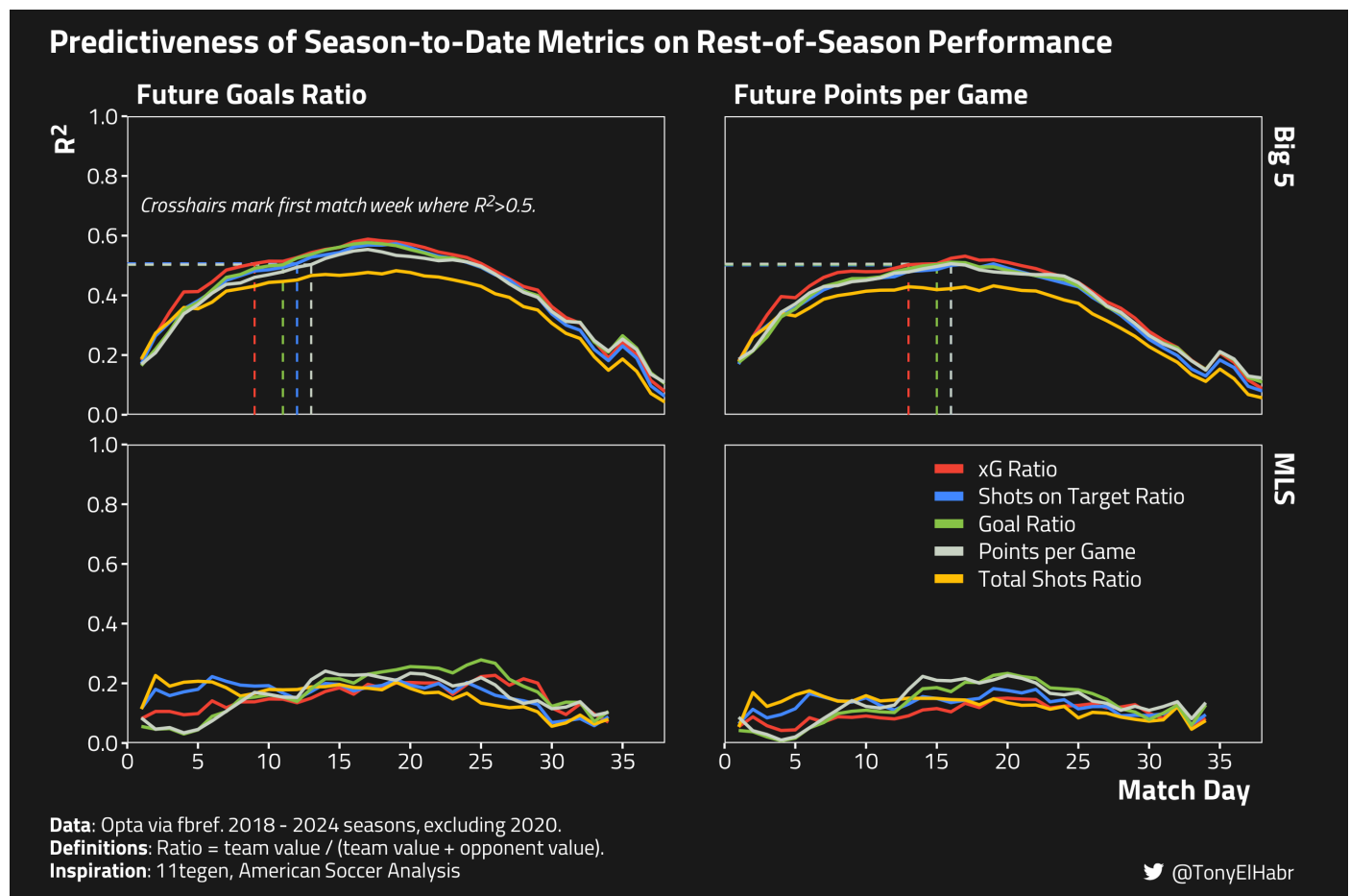
1. points per game
2. goal ratio
3. total shots ratio
4. shots on target ratio
5. xG ratio

The two measures of future performance are:

1. goals ratio
2. points per game

► Calculate rolling R^2 for season-to-date and rest-of-season performance measures

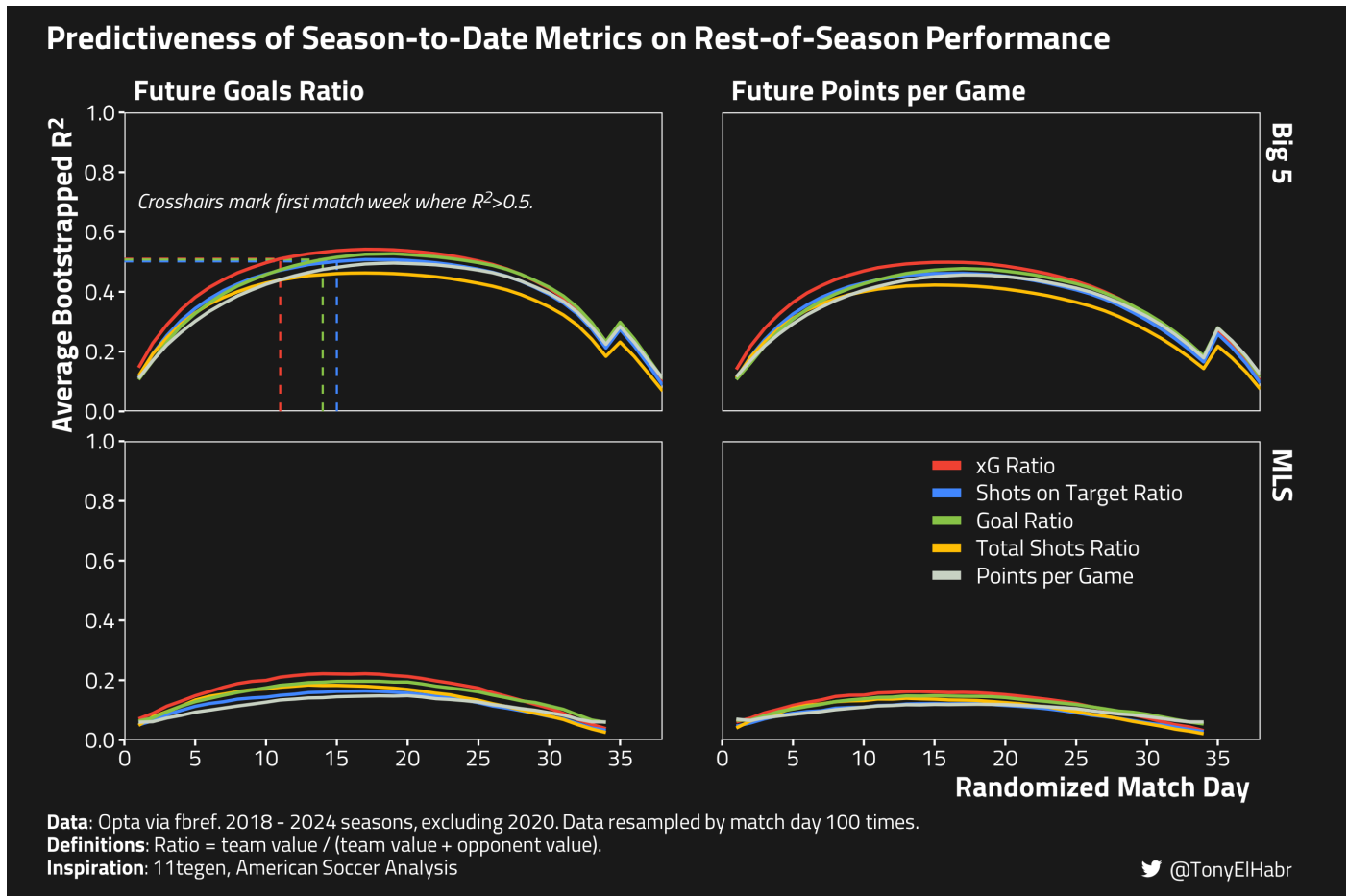
► Plotting the R^2 values



There's a bit of visual noise here that can make it hard to differentiate traces. This can be smoothed out by resampling.

Specifically, if we randomly reorder matchweeks before calculating cumulative in-season measures, and do that 100 times, we end up with a plot that looks like this.

- Calculating bootstrapped R^2 values
- Plotting the bootstrapped R^2 values



Indeed, this looks like the bootstrapped plot from ASA.⁵ Note that,

① Effect of bootstrapping

>

Extending prior art

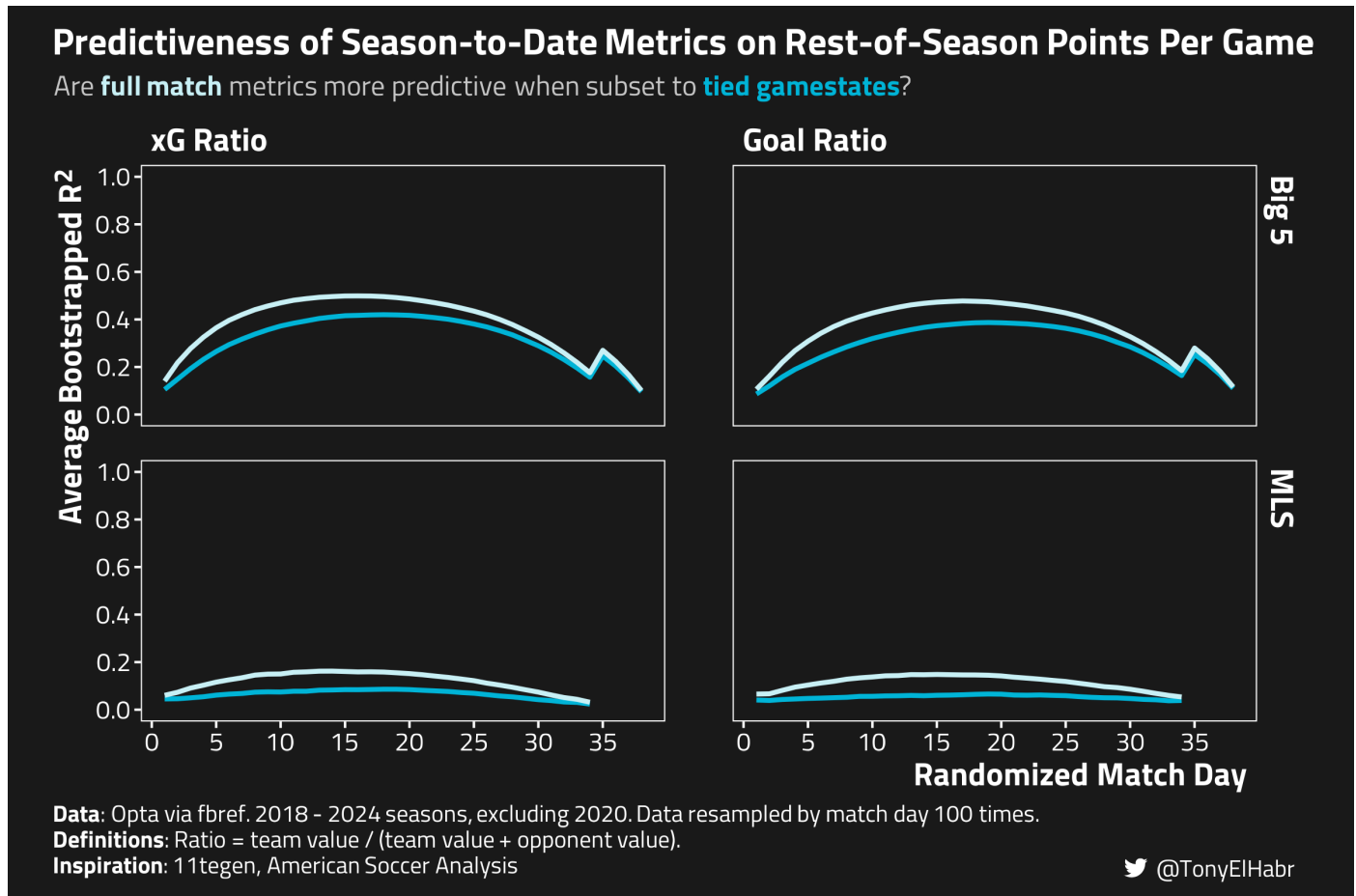
Now we incorporate gamestate-aware measures.⁷ In this case, we're mostly interested game time where the score is close.

A "close" match can be strictly defined as one with a **tied** score, whether it's 0-0, 1-1, etc. The indirect assumption is that teams start to play more conservatively (if at all) when leading, thereby distorting the ratio of shots, goals, and xG that we might otherwise expect given the relative quality of the teams. By excluding events when the game is not tied, we might achieve more "signal" in our measures of performance.

We might also define "close" as periods when the **absolute difference in goals is 0 or 1**, so 0-1, 1-2, 3-2, 5-4, etc. in addition to 0-0, 1-1, etc. This definition indirectly assumes that teams don't start to play more complacently (if at all) when ahead until they have [a 2 goal lead](#). This approach would capture more game time (and have higher tallies of goals, shots, etc.) compared to the `gd = 0` approach, although less time than just using the whole match.

In the plot below, I've split out each the xG and goals measures of past performance into their own facets, paired with tied gamestate analogue (in a brighter blue shade). Note that I'm showing just points per game as the sole measure of future performance.⁸

► Plotting the bootstrapped R^2 values, including the neutral gamestate metrics

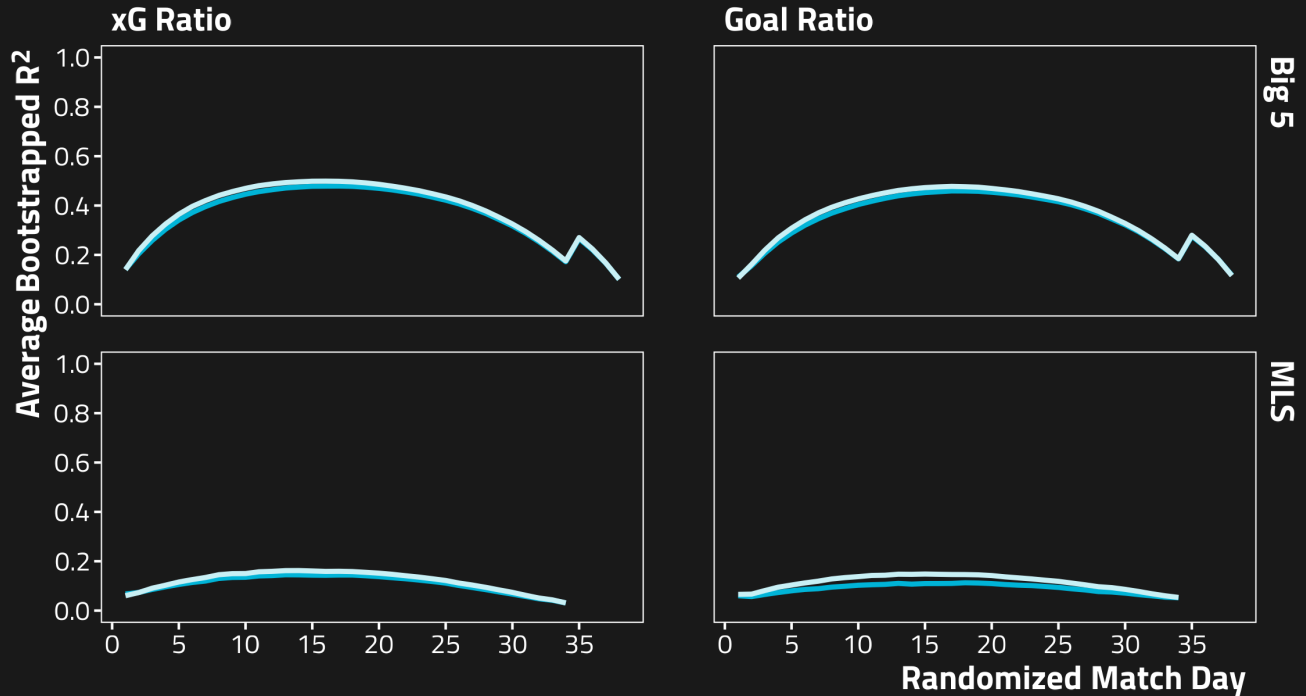


So we see that subsetting xG and goals to tied gamestates **reduces** how predictive they are of future points per game, both for the Big 5 leagues and the MLS. Perhaps this is not surprising, as we're disregarding a lot of data (over 50%!) that is included in full match sums.

Let's see if the same plot using the alternative definition for neutral gamestates—time when the absolute GD is 0 or 1—looks any different.

Predictiveness of Season-to-Date Metrics on Rest-of-Season Points Per Game

Are **full match** metrics more predictive when subset to **gamestates with absolute goal difference ≤ 1** ?



Data: Opta via fbref. 2018 - 2024 seasons, excluding 2020. Data resampled by match day 100 times.

Definitions: Ratio = team value / (team value + opponent value).

Inspiration: 11tegen, American Soccer Analysis

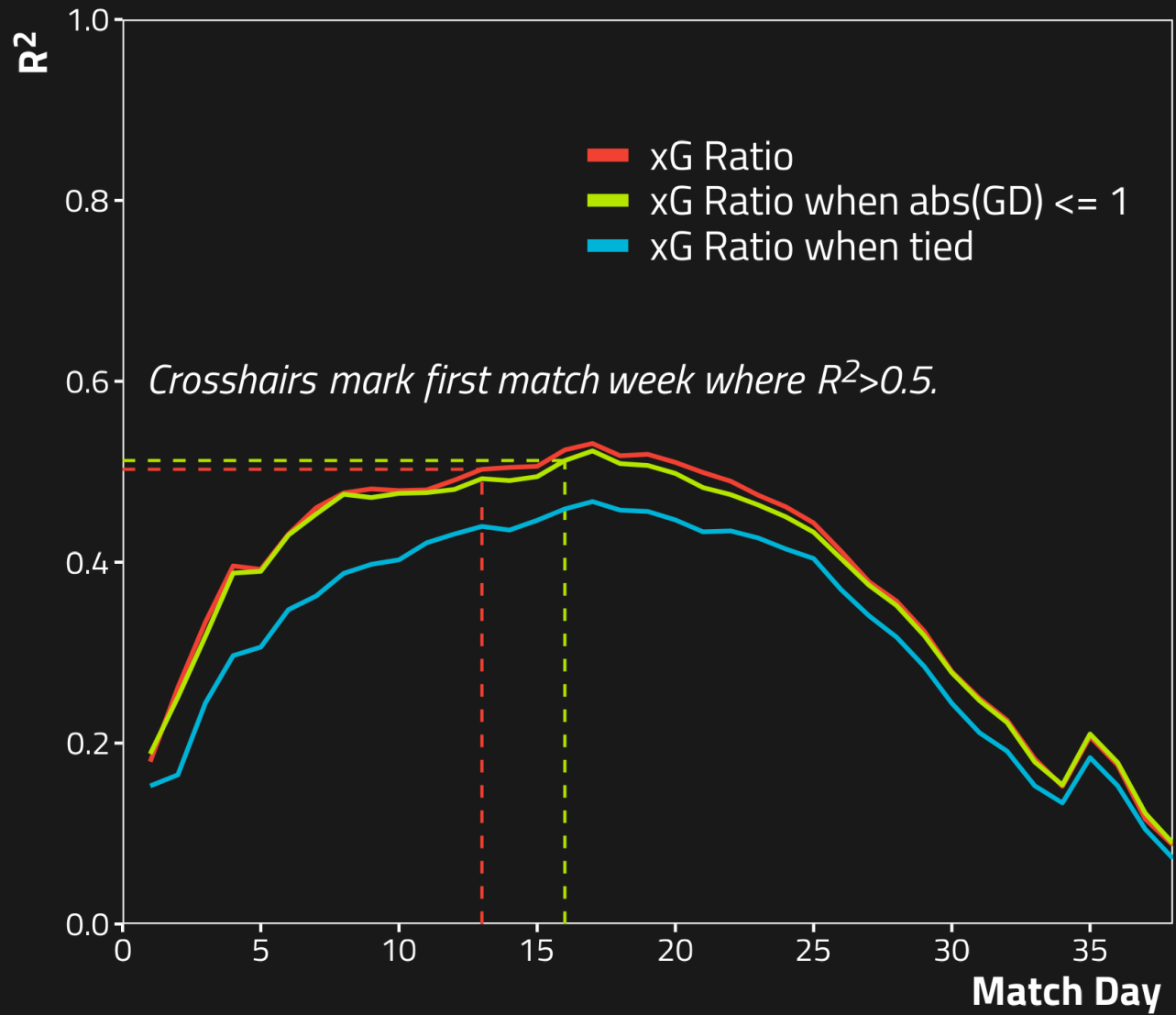
@TonyElHabr

Ah, so using the alternative, more relaxed definition of neutral gamestate, the predictiveness of the season-to-date xG and goal ratios is much closer to the full match analogues. Nonetheless, it's evident that the neutral gamestate xG and goal ratios are no better than the full match ratios, indicating that there is no incremental value to be had with focusing on neutral gamestates. This falls in line with what Ijtsma hypothesized—that the effect of within-match complacency is minimal, and accounting for it (by reducing tallies to neutral gamestates) is more than likely to reduce predictiveness.

Let's put our two neutral gamestate xG traces alongside the full match xG trace for one complete picture.

► Plotting full-match and neutral xG ratios

Predictiveness of Season-to-Date xG on Rest-of-Season Points per Game in the Big 5 leagues



Data: Opta via fbref. 2018 - 2024 seasons, excluding 2020.

Definitions: Ratio = team value / (team value + opponent value).

Inspiration: 11tegen, American Soccer Analysis

@TonyElHabr

Discussion and Conclusion

Ok, so we didn't find that neutral gamestate xG beats out full match xG in terms of rest-of-season predictiveness. On the other hand, if we compare the xG ratio in gamestates with a 0 or 1 absolute score difference to the full match xG, we see that there is not a big difference. This sort of indirectly shows that there is not a lot of value in including the xG in gamestates where the absolute score difference is greater than 1. This backs up the intuition that xG accumulated in lopsided games (either by the trailing or leading team) is not all that meaningful for projecting how either team will perform in the future. Further, in comparing the predictiveness of the two definitions of neutral gamestates, we might say that there is non-trivial value in using the xG accumulated in matches where the score difference is just 1, not just when the teams are tied.

Now, to offer some kind of additional insight that might be useful, I present the game index (i.e. match day) at which my data shows that the season-to-date measures start to provide a reliable signal for forecasting future

performance.⁹ I define the reliability “threshold” as the first week in which the running R^2 value exceeds 0.5. For those reading the footnotes or broadly aware of the literature on xG predictiveness, this is effectively my look at the common adage that xG becomes a reliable indicator of future performance somewhere between the 5 and 15 game mark.¹⁰

► Calculating when past performance measures become reliable indicators of future performance

Earliest match day in which the past metric becomes a reliable measure of **future goals ratio** in the Big 5 leagues.

Predictor	Match Day
xG Ratio	9
Goal Ratio	11
Shots on Target Ratio	12
Points per Game	13
Total Shots Ratio	-

Earliest match day in which the past metric becomes a reliable measure of future **points per game** in the Big 5 leagues.

Predictor	Match Day
xG Ratio	13
Goal Ratio	15
Points per Game	16
Shots on Target Ratio	16
Total Shots Ratio	-

So, does this mean that we should not look at xG in tied gamestates (or leading or trailing gamestates) at all? Eh, I think there's still usefulness for contextualizing xG by gamestate to gain insight into how teams have achieved their results.

Tony@TonyElHabr · [Follow](#)

Expected goal difference by game state in the EPL this season. What stands out?

- 1** Newcastle have been most dominant when leading.
- 2** Man City have been at their best when tied or trailing.
- 3** Spurs have been mediocre when leading, but otherwise very good.

3:00 PM · Oct 28, 2023



131



Share

[Read 4 replies](#)

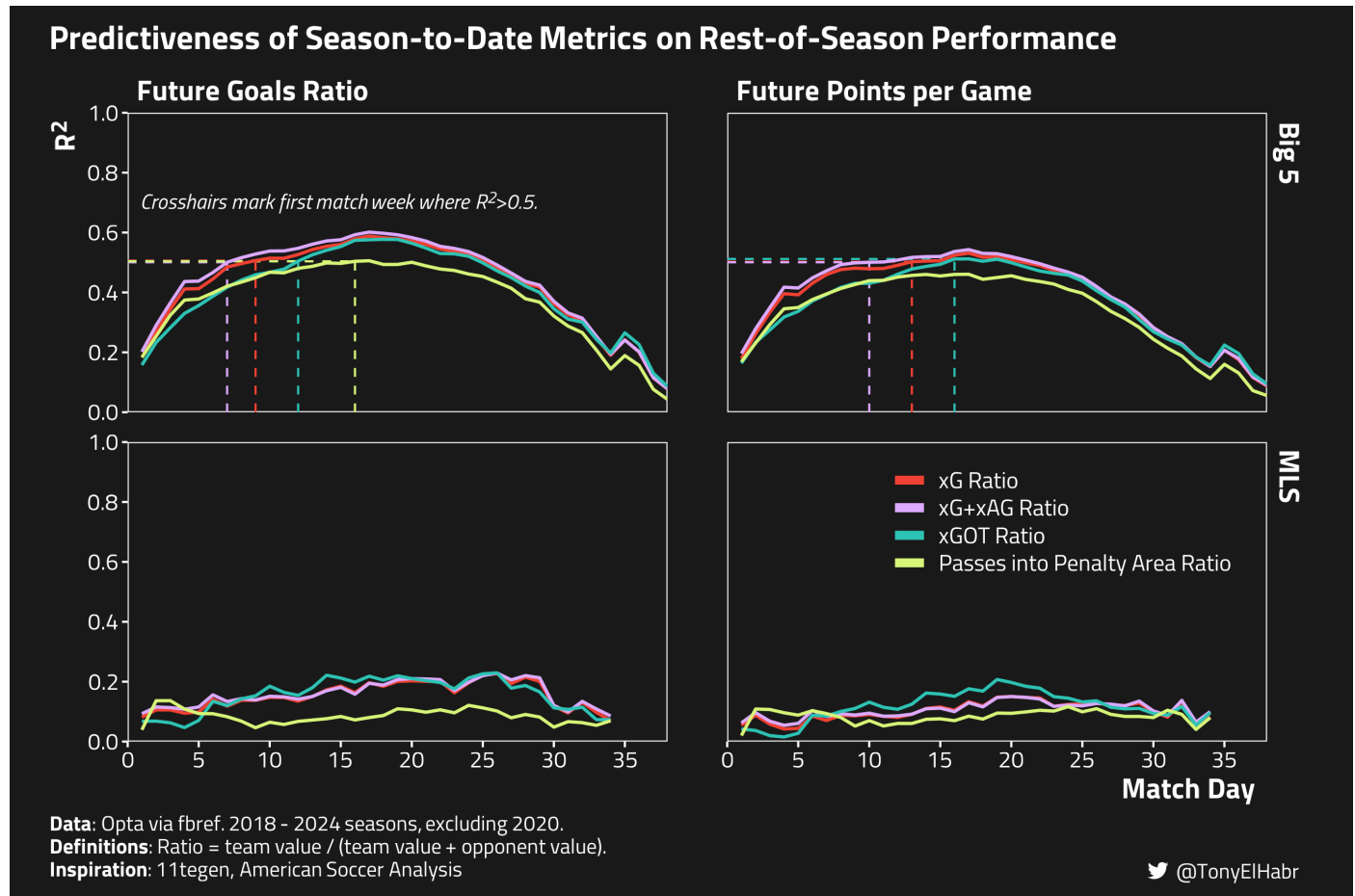
Appendix

Additional Past Metrics

I was interested in evaluating running in-season R^2 values for a few more measures of past performance:

1. passes into the penalty area ratio
2. [expected goals on target \(xGOT\)](#) ratio
3. xG + expected assisted goals¹¹ (xG+xAG) ratio

Below is the same running R^2 plot shown before, just with these new measures, along side our tried and true xG ratio measure.



Aha! It seems that we have identified a measure—xG+xAG ratio (purple above)—that becomes reliable for forecasting rest-of-season performance even earlier than xG ratio. Specifically, it exceeds the reliability threshold at the 7 and 10 game marks for future goal ratio and future points per game respectively. That's a non-trivial improvement compared to the 9 and 13 match day marks for xG ratio.

Alas, xG+xAG metaphorically stands on the shoulders of xG, adding in the component of expected assists (xA) on shots. Assuming that the underlying xA model is well calibrated, perhaps we should not be surprised to see that the composite xG+xAG measure outperforms all others.

[xG+xAG usage](#)

On the other hand, I suppose I am a bit surprised that xGOT ratio does not seem to do nearly as well as xG ratio in terms of forecasting rest-of-season performance. The implication is that **there is value in the xG of off-target shots** for projecting the future. By comparison, shots on target ratio tend to be more predictive than just the

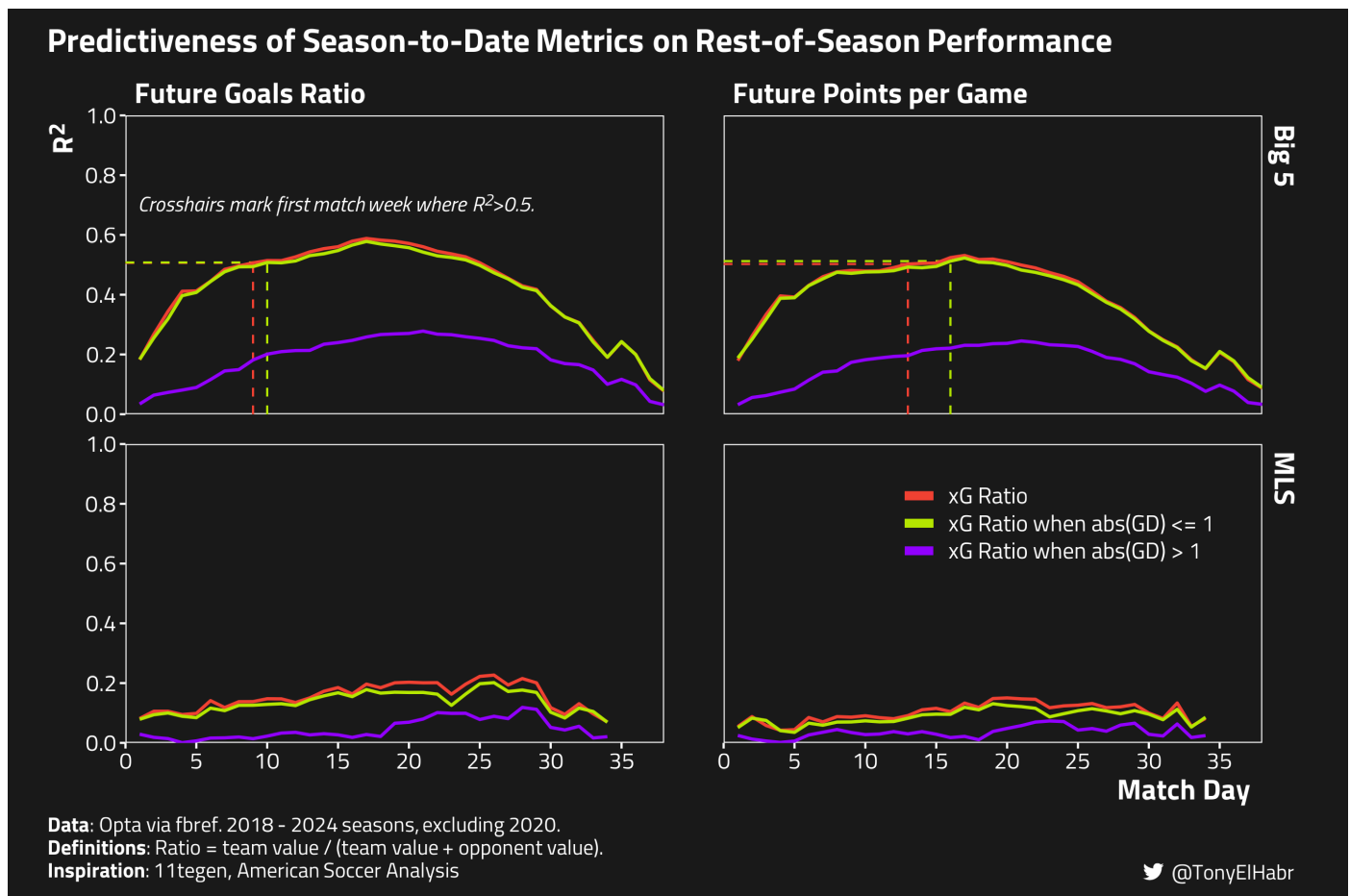
shots ratio, meaning that including off-target shots introduces noise that reduces predictiveness. That's an interesting difference in shots and xG!

Non-Neutral xG Ratio

Finally, I was interested in verifying that there's not a lot of value in non-neutral gamestate xG ratios. The plot below shows the running R^2 values for two measures already visualized:

1. xG ratio
2. xG ratio when the absolute GD is less than or equal to 1 (83.1% of total minutes).

But, in addition to these, I also plot the xG ratio when the absolute goal difference is greater than 1 (16.1% of total minutes)



We see that that the R^2 values for absolute GD > 1 gamestates basically hover around 0 across all match days, for both the Big 5 and the MLS. This isn't all that surprising given how small a percentage of full matches occur in >1 GD gamestates. But it also backs up our intuition that stats (i.e. shots, xG) accumulated in such gamestates do not strongly reflect how the teams will perform for the rest-of-the-season.

Footnotes

1. If you know anything about the broader soccer analytics discourse, you've more than likely heard about one of the Ijtsma's findings regarding the predictiveness of xG. As [Eliot puts](#) it: "If you've ever heard or read someone say that it takes 5-10 (4

- in the article) games for xG to be predictive, they probably read this article or got it second hand from someone who has.”↵
2. Note that there are [some criticisms](#) of the choice to rely on R^2 as a measure of predictiveness. For the purpose of this post, we'll take that criticism on the chin.↵
 3. A “ratio” here is broadly defined in pseudo-code as `team's value / (team's value + opponent's value)`. Ijtsma's notes in a reply to a comment that one might arguably use a difference formula, i.e. `team's value - opponent's value`, as ratios are susceptible to noise when values are themselves fractional. This is most relevant for xG, and not so relevant for shots and goals, which are inherently discrete.↵
 4. Notably, I exclude the COVID-19 influenced 2020 season, so as to reduce a bit of noise in the data. Although bootstrapping games within-season should reduce much of this noise, in practice, [FBref](#) is missing some game-level xG data for 2020 matches, so this choice is sort of just a pragmatic one.↵
 5. I only perform 100 bootstraps instead of 1,000 like Eliot did since I found 100 sufficient for creating visual clarity.↵
 6. This may not be easy to see directly since the bootstrapped and non-bootstrapped curves are not plotted altogether. Nonetheless, one can compare the peaks of the curves against the y-axis in this plot compared to the prior one to verify the suppressing effect of the resampling.↵
 7. If you've been following the code, these measures have already been calculated.↵
 8. There are similar trends with future goal ratio. I've opted for visual brevity here.↵
 9. The tables that follow show non-bootstrapped R^2 values. The bootstrapped R^2 values only exceed 0.5 for a handful of measures, typically in the 15-20 game range.↵
 10. Note that Ijtsma says that xG ratio becomes useful as early as the fourth match week. However, his statement is based purely on a visual assessment of when the R^2 curve generally “settles down” after the expected early season up-tick. Looking at his graph, the R^2 values for xG don't actually exceed the 0.5 threshold until some point between the 10th and 15th match days.↵
 11. FBref's explanation regarding how xAG differs from expected assists (xA): “xA, or expected assists, is the likelihood that a given completed pass will become a goal assist... Players receive xA for every completed pass regardless of whether a shot occurred or not... [Expected Assisted Goals (xAG)] indicates a player's ability to set up scoring chances... Players receive xAG only when a shot is taken after a completed pass.”↵