Bundesliga Pythagorean Prediction*

John Zhang

October 13, 2025

Table of contents

*Code and data are available at: https://github.com/Clearsky21z/Bundesliga_Pythagorean_Prediction

1 Introduction

2 Methodology

2.1 Pythagorean Expectation for Soccer

We use a four-parameter variant of the Pythagorean expectation adapted to the 3-points-per-win system:

$$\text{frac}_i = \frac{\text{GF}_i^{\,b}}{\text{GF}_i^{\,c} + \text{GA}_i^{\,d}}, \qquad \widehat{\text{PTS}}_i = a \cdot \text{frac}_i \cdot \text{PLD}_i,$$

where i indexes a team–season observation; $a > 0$ is a scale parameter mapping the win fraction to points; $b, c, d > 0$ control the non-linearity and the relative influence of scoring vs. conceding.

2.2 Estimation on Historical Seasons

Let $y_i = \text{PTS}_i$ and $\hat{y}_i(a, b, c, d)$ be the model prediction from the formula above. We estimate (a, b, c, d) by minimizing the mean absolute error (MAE) across all team–season observations in the **2010–11 to 2023–24** pool:

$$(a, b, c, d) \;=\; \underset{a,b,c,d}{\arg\min} \; \frac{1}{N} \sum_{i=1}^{N} |\, y_i - \hat{y}_i(a, b, c, d) \,|\,.$$

Optimization uses the Nelder–Mead simplex algorithm (derivative-free) with multiple starting values (including Beggs' EPL coefficients) to reduce sensitivity to local minima. The resulting (a, b, c, d) are treated as **Bundesliga-specific coefficients** and reused for forecasting in new seasons.

2.3 Out-of-Sample Validation (Historical)

We assess generalization via leave-one-season-out (LOSO) validation. For each season s in the estimation pool, we refit (a, b, c, d) on all other seasons and evaluate on s. Performance is summarized by the median MAE and Pearson correlation r between actual and predicted EoS points across held-out seasons.

3 Result

4 Discussion

A Appendix