

World Cup: Five-Factor Pathway Analysis

This document outlines the mathematical framework used to evaluate the team identities of the 2026 World Cup contenders. Rather than comparing teams against one another, the model analyzes each team internally to identify its strongest and weakest pathways to victory.

Model Overview

This five-factor pathway model evaluates a team's ability to suppress opponents, generate scoring bursts, execute late, survive extended matches, and protect a lead. It identifies each team's primary and secondary pathways by measuring how each factor deviates from its average level. Factors 10 percent points above the average are strengths, while those 10 points below are weaknesses. The raw results are then translated into each contender's strengths, weaknesses, opportunities, and threats for the 2026 World Cup.

Control / Suppression Score

$$\text{Control Score} = 0.6 \cdot (1 / (1 + \text{SoT allowed})) + 0.4 \cdot (1 / (1 + \text{Goals allowed}))$$

This expression measures defensive suppression by capturing shots on target allowed and their outcomes. It rewards teams that limit opponent shots while preventing those chances from becoming goals, reflecting defensive control.

Volatility / Burst Score

$$\text{Burst Score} = \text{Efficiency} \cdot \log(1 + \text{Pressure})$$

This expression measures attacking volatility by combining finishing efficiency with shot volume. It rewards teams that convert chances efficiently while generating consistent pressure, producing sudden scoring bursts.

Late-Game Execution Score

$$\text{Late Score} = 0.6 \cdot (\text{Late Goal Share}) + 0.4 \cdot (\text{Late Conversion})$$

This expression measures late-game execution by capturing a team's tendency to score after the 60th minute and convert those moments into wins. It rewards teams that generate late goals and sustain them.

Extended-Game Survival Score

$$\text{Survival Score} = 0.5 \cdot (\text{ET Survival}) + 0.5 \cdot (\text{Penalty Survival})$$

This expression measures extended-game survival by capturing a team's ability to win beyond regulation time. It rewards teams that navigate extra time and penalties, maintaining performance in high-leverage moments.

Game-State Protection Score

$$\text{Protection Score} = 0.7 \cdot (\text{Lead Retention}) + 0.3 \cdot (1 / (1 + \text{Post-Lead GA}))$$

This expression measures game-state protection by capturing a team's ability to convert early advantages into wins and limit damage after taking the lead. It rewards teams that maintain control once ahead, limiting opponent momentum.

Variable Definitions, Computations, & Data Collection

Independent Variables

SoT allowed = Opponent shots on target per game

Goals allowed = Opponent goals per game

Efficiency = Goals / Shots on target

Pressure = Shots on target per game

Late Goals = Goals scored after 60 minutes

Late Conversion = Wins when scoring after 60 minutes

ET Survival Rate = Wins in extra time / total extra time games

Penalty Survival Rate = Wins in pens / total pen games

Lead Games = games where team scored first

Lead Retention = Wins when scoring first / lead games

Post-Lead GA = goals conceded after taking the lead

Normalization & Smoothing

Factor scores are normalized within each team and scaled relative to that team's highest and lowest values.

Smoothing then pulls extreme scores slightly toward the team's average, preventing limited sample sizes from producing exaggerated strengths or weaknesses.

Data Collection

Match-level data was collected from FBref.com for each team across the most recent continental tournaments, the 2024 Copa América and UEFA European Championship. The dataset includes match context (Stage, opponent, result), core performance metrics (goals, shots on target, and situational game-state variables such as late goals, lead status, and extra time or penalty outcomes).

This structure allows the analysis to move beyond aggregate statistics and instead capture how teams perform across different game states and pressure environments within a tournament setting.

Model Assumptions

Factor scores are evaluated relative to each team, not across teams.

One tournament reflects recent form, not a complete or fixed identity.

Game-state and situational metrics capture performance under pressure, not overall quality.

Pathway factors represent tendencies, not guarantees of success.

The model estimates how teams win, not which team will win.