Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Simulation-Based Calibration Applied to Sports Modeling

Igor Patrício Michels[1], Luiz Max Carvalho[2]
FGV/EMAp, RJ

This work evaluates five sports prediction models using Simulation-Based Calibration (SBC) [3]: the Bradley–Terry models BT_1 [1] and BT_2, and three Poisson models—$Poisson_1$ [2], $Poisson_2$, and the player-based $Poisson_{v2}$. The models are defined as follows:

Bradley-Terry Model 1 (BT_1):

$$p_{ij} = \frac{\exp(\theta_i)}{\exp(\theta_i) + \exp(\theta_j)} \tag{1}$$

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{2}$$

$$\theta_i \sim \mathcal{N}(0, 1) \tag{3}$$

Bradley-Terry Model 2 (BT_2):

$$p_{ij} = \frac{\exp(\theta_i + \gamma)}{\exp(\theta_i + \gamma) + \exp(\theta_j)} \tag{4}$$

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \tag{5}$$

$$\theta_i \sim \mathcal{N}(0, 1) \tag{6}$$

$$\gamma \sim \mathcal{N}(0, 1) \tag{7}$$

where p_{ij} denotes the win probability for team i, θ_i its strength, y_{ij} the match result, and γ the home advantage effect.

Poisson Model 1 ($Poisson_1$):

$$\lambda_{ij} = \frac{\theta_i}{\theta_j} \tag{8}$$

$$g_{ij} \sim \text{Poisson}(\lambda_{ij}) \tag{9}$$

$$\theta_i \sim \mathcal{N}^+(1, 1) \tag{10}$$

Poisson Model 2 ($Poisson_2$):

$$\lambda_{ij} = \frac{\theta_i + \gamma}{\theta_j} \tag{11}$$

$$g_{ij} \sim \text{Poisson}(\lambda_{ij}) \tag{12}$$

$$\theta_i \sim \mathcal{N}^+(1, 1) \tag{13}$$

$$\gamma \sim \mathcal{N}^+(1, 1) \tag{14}$$

Poisson Model with Players ($Poisson_{v2}$):

$$\lambda_{ij} = \frac{\sum_{p \in P_i} \theta_p}{\sum_{q \in P_j} \theta_q} \tag{15}$$

$$g_{ij} \sim \text{Poisson}(\lambda_{ij}) \tag{16}$$

$$\theta_p \sim \mathcal{N}^+(1, 1) \tag{17}$$

where λ_{ij} is the goal rate, g_{ij} the scored goals, P_i the players of team i, and $\mathcal{N}^+$ a normal distribution truncated at zero.

Bradley–Terry models were run with 50 SBC simulations; Poisson models with 100 and longer sampling (1000 iterations). The player-level model also varied squad size (11–36 players).

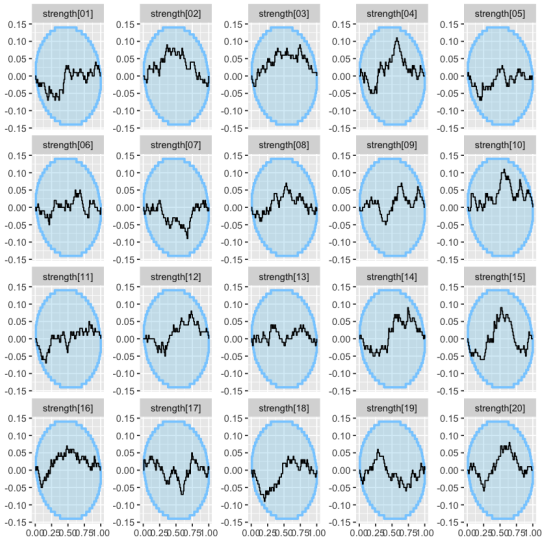
SBC results showed strong contrasts: BT_1 and BT_2 displayed substantial miscalibration (60% and 71.4% deviations). Poisson models were far better calibrated: $Poisson_1$ and $Poisson_2$ deviated in only 5% and 4.7% of parameters, with ECDF differences mostly inside the confidence bands (Figure 1). $Poisson_{v2}$ also showed good calibration (3.3–9.4%), regardless of squad size.

Applied to Brazilian Championship data (2014–2023; >3,800 matches; >1,500 players), Poisson models again outperformed Bradley–Terry models. $Poisson_1$ and $Poisson_2$ provided robust team-level estimates, while $Poisson_{v2}$ additionally captured player contributions, albeit with sensitivity

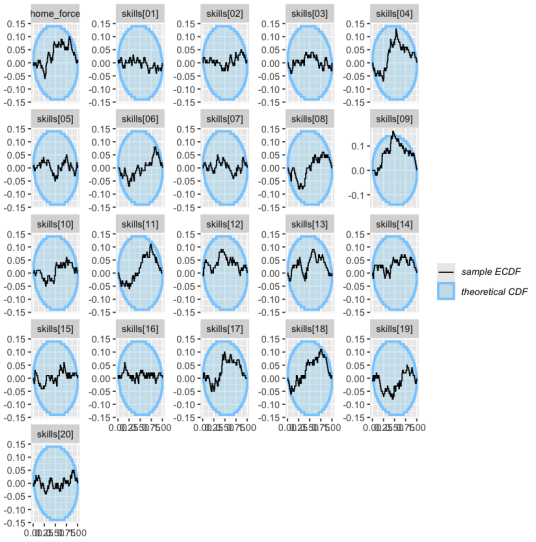
[1]igorpmichels@gmail.com
[2]lmax.fgv@gmail.com

to limited playtime. The choice between $Poisson_2$ and $Poisson_{v2}$ depends on whether inference focuses on teams or players.

Figure 1 compares ECDF differences for the Poisson models, with 95% confidence bands indicating expected variation under perfect calibration. Lines exiting the bands reflect deviations; panel (a) corresponds to $Poisson_1$, and panel (b) to $Poisson_2$.



(a) Model $Poisson_1$ ECDF differences.



(b) Model $Poisson_2$ ECDF differences.

Figure 1: Empirical Cumulative Distribution Function (ECDF) differences for Poisson models. Source: Author's elaboration based on SBC results.

References

[1] R. A. Bradley and M. E. Terry. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". In: **Biometrika** 39.3/4 (1952), pp. 324–345. URL: `http://www.jstor.org/stable/2334029`.

[2] M. J. Maher. "Modelling association football scores". In: **Statistica Neerlandica** 36.3 (1982), pp. 109–118. DOI: `https://doi.org/10.1111/j.1467-9574.1982.tb00782.x`.

[3] S. Talts and et al. **Validating Bayesian Inference Algorithms with Simulation-Based Calibration**. 2020. URL: `https://arxiv.org/abs/1804.06788`.