UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

Predicting Football Results: a Bayesian Approach with a Zero-Modified Poisson Distribution

Vítor Hugo Anastácio Palma

Mestrado em Bioestatística

Dissertação orientada por:
Professor Rui Martins

2025

Agradecimentos

Gostaria de, primeiro que tudo, agradecer profundamente aos meus pais, irmã e avós pelo apoio incalculável que me deram neste, nem sempre fácil, nem sempre difícil percurso, nada daquilo que sou hoje seria possível sem vocês.

Agradeço também ao resto da minha família, também presente nesta jornada, pelas palavras de motivação e constante afeto, que foram essenciais para me manter focado neste objetivo.

Não podia deixar também de agradecer a todos os meus amigos, dos mais antigos, aos que este mestrado me deu. Um agradecimento não só pela inspiração, conselhos e apoio que me deram, mas também e especialmente pelos momentos em que me complicaram a entrega desta tese, que são na realidade os que vou guardar com mais carinho.

Por último, quero agradecer ao meu orientador, o professor Rui Martins, por toda a orientação - passo a redundância -, ferramentas, conhecimento e direção geral que me deu ao longo deste trabalho. Seja ao nível da organização da estrutura da tese, como em dúvidas mais teóricas, sempre me ajudou a iluminar o caminho e por isso um grande obrigado.

Vítor Palma

Resumo

O futebol é um fenómeno mundial, seguido por adeptos exigentes e com uma vontade insaciável de o ver a evoluir constantemente, como tal, seria apenas uma questão de tempo até a estatística entrar em campo. A análise estatística aplicada ao futebol tem vindo a ganhar cada vez mais relevância, tanto para equipas e treinadores, na otimização do desempenho desportivo, mas também para analistas e apostadores, na previsão de resultados.

Com o avanço do conhecimento e da tecnologia, a estatística tem desempenhado um papel cada vez mais fundamental na compreensão do jogo. Atualmente, métricas como os Expected Goals (xG), quilómetros percorridos, passes acertados, duelos ganhos, entre muitas outras, são amplamente utilizadas para avaliar o rendimento das equipas e dos jogadores. Estas abordagens permitiram não apenas melhorar a tomada de decisões estratégicas no futebol, mas também proporcionar ferramentas para a construção de modelos preditivos mais sofisticados.

Contudo, a imprevisibilidade do futebol, influenciada pelo baixo número de golos por jogo e pela elevada frequência de empates, que por natureza, são mais difíceis de prever, representa um grande desafio para os estatísticos. Ao contrário de outros desportos, onde há uma elevada frequência de pontos e menos empates, um único golo pode alterar drasticamente o desfecho de um jogo, tornando mais complexa a modelação estatística.

A distribuição mais utilizada neste tipo de análise é a distribuição de Poisson. No entanto, esta apresenta dificuldades em lidar com conjuntos de dados que exibam inflação ou deflação de zeros, algo bastante frequente no futebol, já que no mesmo campeonato pode tanto haver equipas que marcam em praticamente todos os jogos como equipas com muita dificuldade em fazê-lo. Estes desvios exigem abordagens estatísticas mais flexíveis para capturar com precisão esta dinâmica.

Para ultrapassar esta limitação, foram desenvolvidos modelos como a Poisson inflacionada em zeros, que incorpora um mecanismo adicional para a criação de zeros, ou a Poisson alterada em zeros, que atribui uma massa de probabilidade extra a este valor. Porém, nenhuma destas 2 modificações é adequada para lidar com a deflação de zeros.

É neste contexto que surge o modelo de Poisson modificado em zero, que será o tema central desta tese. Este modelo introduz um parâmetro p, capaz de capturar ambos os tipos de variação na frequência de zeros, permitindo um ajuste mais preciso aos dados. Deste modo, ao contrário dos modelos anteriores, este pode ser utilizado tanto para situações onde há um excesso de golos, como para situações em que estes são menos frequentes do que o esperado numa Poisson padrão.

O estudo é feito com dados da Serie A da época 2022/2023, a primeira liga de futebol italiana, envolvendo 20 equipas, cada equipa defronta as restantes em jogos em casa e fora, perfazendo 38 jornadas e 380 jogos. O conjunto de dados consiste nos golos marcados por ambas as equipas em todos os jogos.

Para enquadrar a metodologia, recorreu-se a um breve enquadramento teórico sobre estatística Bayesiana, que inclui conceitos como priors e likelihood até à criação da distribuição posterior, obtida por métodos de amostragem Markov Chain Monte Carlo (MCMC).

De seguida, apresenta-se uma secção de introdução à Poisson modificada em zeros, onde também está incluída uma breve descrição sobre as três distribuições referidas anteriormente. Este enquadramento permite comparar as vantagens e desvantagens de cada modelo, destacando as características de cada um.

Por último, é apresentado um capítulo detalhado sobre as metodologias utilizadas para obter os resultados, onde são explicados os passos desde a criação da distribuição da Poisson modificada em zero e do modelo, até aos mecanismos usados para gerar as previsões. Este processo inclui a especificação do modelo hierárquico usado, dos seus priors e parâmetros: o ataque e a defesa de cada equipa, um parâmetro que ajusta a vantagem inerente de jogar em casa e os dois parâmetros de modificação de zeros, p1 para a equipa da casa e p2 para a equipa visitante. A introdução destes parâmetros hierárquicos permite que o modelo capte melhor as diferenças individuais entre as equipas, proporcionando previsões mais realistas. Foi também incluída a forma como o MCMC é aplicado, que é uma metodologia particularmente relevante neste estudo, pois permite para além da estimação de distribuições complexas também avaliar a incerteza associada às previsões, algo essencial na modelação estatística de resultados desportivos. Após o período de burn-in e a verificação da convergência das cadeias, são extraídas as estimativas para os parâmetros do modelo, permitindo a sua análise.

Quanto ao processo de previsão foram usados 2 métodos. O primeiro método faz previsões jornada a jornada, onde para prever cada jornada o modelo utiliza os dados de todos os jogos anteriores para estimar os parâmetros. Para, por exemplo, prever a jornada 32, o modelo tem acesso aos dados até à jornada 31. De seguida são retiradas 1000 amostras dos parâmetros estimados, e com eles, para cada jogo, gera-se uma lista com 1000 valores de golos usando a distribuição de Poisson modificada em zeros, tanto para a equipa da casa como para a de fora. Através da comparação dos golos gerados calcula-se a probabilidade de vitória da equipa da casa, empate, ou vitória da equipa visitante. Por fim os resultados mais prováveis são selecionados como os resultados previstos e são usados para atualizar uma tabela classificativa criada para guardar os resultados.

O segundo método foca-se em simular a época inteira a partir de um ponto fixo. Neste caso o modelo utiliza os dados até essa jornada fixa para estimar os parâmetros e depois gera previsões para todas as jornadas restantes, onde ao contrário do primeiro método os valores dos parâmetros não mudam, se a jornada escolhida for a 23ª, então daí até ao final da época todos os jogos serão previstos com os parâmetros estimados para cada equipa na jornada 23. De forma semelhante ao primeiro método são extraídas 1000 amostras dos parâmetros e geradas 1000 simulações de golos para cada equipa em cada jogo. As probabilidades de cada resultado e a introdução na tabela classificativa são também feitas de igual forma.

Cada um dos métodos tem utilidades distintas consoante o objetivo da análise. O primeiro método, ao atualizar os parâmetros jornada a jornada, permite captar a evolução das equipas ao longo da época, tornando-o ideal para previsões em tempo real e para acompanhar a progressão do campeonato à medida que novos jogos são disputados. Já o segundo método, por fixar os parâmetros num determinado ponto da temporada e simular todos os jogos restantes com base nesses valores, oferece uma visão mais global do campeonato, permitindo estimar cenários finais com base na informação disponível até um dado momento, respondendo por exemplo à pergunta: se esta equipa mantiver este nível de performance, qual será a sua classificação final?

Por último, para avaliar o desempenho do modelo, foram realizadas diversas análises, incluindo a medição da sua precisão, a comparação entre os resultados previstos e os dados observados através de várias representações gráficas e tabelas, bem como a avaliação da Poisson Modificada em Zero face aos três modelos de Poisson anteriormente referidos.

O objetivo deste trabalho será determinar se a sua capacidade de lidar com a inflação e deflação de zeros representa, de facto, uma vantagem preditiva significativa.

Palavras-Chave: Futebol, Poisson Modificada em Zero, Estatística Bayesiana, Modelos Preditivos, Inferência Estatística

Abstract

Football is a global phenomenon, and as such, it was only a matter of time until statistics joined the field.

This dissertation explores the application of the Zero-Modified Poisson (ZMP) model within a Bayesian framework for predicting football match outcomes. Traditional Poisson-based approaches often fail to account for irregularities in goal distributions, such as zero inflation or zero deflation, which are common in football scoring patterns. The incorporation of the Zero-Modified Poisson model provides the flexibility to handle varying frequencies of zero outcomes while accommodating unequal means and variances in contrast with the standard Poisson, which assumes equaldispersion.

Using match data from the 2022/23 Italian Serie A, a Bayesian hierarchical model is implemented via the NIMBLE r package, where team-specific parameters for attack, defense, and home advantage were estimated using Markov Chain Monte Carlo (MCMC) techniques.

After the parameter estimation, two simulation methods were employed to generate probabilistic forecasts: a round-by-round and a full-season one. In order to check the performance of the model the first simulation method was replicated with 3 other Poisson models to check if the handling of zero inflation and deflation really gave an advantage in the predictive power of the ZMP

Keywords: Football, Zero-Modified Poisson, Bayesian statistics, Predictive modeling, Statistical inference

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Football is more than just a sport; it's a universal passion that unites billions of fans around the world. Its widespread appeal have also made it an exciting field for statistical analysis, drawing the attention of researchers, analysts, and enthusiasts alike.

Football as a prediction subject is extremely unpredictable. Matches are influenced by a wide range of factors, such as team strategies, individual player performance, weather conditions, injuries, and even referee decisions, unlike other popular sports such as basketball or rugby, where points are scored frequently and draws are rare, the low-scoring nature of football introduces an additional layer of volatility. A single goal can decide the outcome of a match, and the prevalence of draws further complicates prediction efforts. This delicate balance between predictability and chance has spurred the development of innovative statistical and machine learning approaches designed specifically for the sport. Over the years, this pursuit has led to increasingly sophisticated probabilistic models, each refining and expanding upon the limitations of its predecessors.

Early in the evolution of football analytics, (Maher, 1982) introduced a major innovation by modeling match scores as the outcomes of independent Poisson processes. In his framework, each team's expected number of goals is a function of its inherent attacking strength and the opposing team's defensive weakness, with an additional fixed home advantage factor incorporated into the model. Although Maher's model represented a significant step forward in quantifying team performance, it also revealed certain limitations, most notably its tendency to underestimate the likelihood of low-scoring outcomes and draws.

Building on these advancements, (Rue and Salvesen, 2000) developed a dynamic Bayesian generalized linear model that further captured the evolving nature of team performance.By allowing attacking and defensive parameters to vary over time and updating estimates as new match data become available, using Markov Chain Monte Carlo techniques, their framework not only enhanced short-term prediction accuracy but also enabled retrospective analyses, such as evaluating final league standings. This dynamic approach marked a significant shift toward more sophisticated, time-sensitive models in football analytics.

Later, (Baio and Blangiardo, 2010) advanced the field by introducing a Bayesian hierarchical model for the prediction of football matches. Their model improved upon previous Poisson-based approaches by treating attack and defense parameters as random effects with hierarchical priors, thereby capturing variations across teams and seasons with greater flexibility. This method represents a natural progression from the earlier static models to more complex layered Bayesian frameworks.

As these Bayesian methodologies became more prevalent, researchers sought to address another fundamental limitation of Poisson-based models, their inability to handle excess or deficit zero outcomes

properly than would be expected under equidispersion. To solve that issue, the Zero-Inflated Poisson (ZIP) model introduced by (Lambert, 1992) represents a key development in statistical modeling for count data with excess zeros, as highlighted by (Mouatassim and Ezzahid, 2012). This framework combines a standard Poisson distribution for count outcomes with an additional distinct process to account for zero inflation.

Building upon the Zero-Inflated Poisson model, (Conceição et al., 2017) applied a Bayesian Zero-Modified Poisson model in a football context, addressing not only zero-inflation but also zero-deflation. This model introduces an additional parameter to adjust the probability of zero-goal outcomes. Their approach utilized Bayesian inference via MCMC sampling, updating team parameters over time.

Together, these studies, from Maher's foundational Poisson framework, to the hierarchical approach of Baio and Blangiardo, illustrate the evolution of football prediction methodologies. Each of these contributions paved the way for the Bayesian Zero-Modified Poisson models.

This dissertation focuses on predicting the results of the second half of the 2022/23 Italian football league season using the Zero-Modified Poisson model and a Bayesian approach. The study begins with a theoretical introduction to Bayesian statistics, the Zero-Modified Poisson distribution, and the setup of the model for result prediction. Following this, an analysis of the posterior distributions obtained from the Bayesian model is presented in detail. Finally, the results are compared to other similar models, including different variations of zero-modified Poisson models as well as the standard Poisson distribution. The objective is to determine whether adjusting for zero inflation and zero deflation improves predictive accuracy and whether these modifications are justified in the context of football match predictions.

Chapter 2

Bayesian Statistics

2.1 Introduction to Bayesian Statistics

Bayesian statistics offers a different perspective on probability. Rather than treating probability as the frequency of events over repeated trials, it views it as a degree of belief or a measure of uncertainty about an event (Gelman et al., 1995).

This approach enables direct statements about the likelihood of parameters or predictions, for example, a Bayesian credible interval for θ directly quantifies the probability that θ falls within a specified range, making the results intuitive and easy to interpret.

This is different from a frequentist confidence interval, which quantifies uncertainty based on how intervals behave across many repeated samples, instead of giving a direct probability for the actual parameter (Gelman et al., 1995).

To illustrate that, when estimating the average height of adult men in a city, a frequentist confidence interval involves randomly selecting and measuring a sample of, let's say, 100 men to construct an interval. If this process is repeated multiple times with different samples, approximately x% of the resulting intervals will contain the true average height. However, for any individual interval, the true height is either within it or not. In contrast, a Bayesian credible interval directly states that, given the data, there is an x% probability that the true average height falls within the range.

Another key distinction between Bayesian and frequentist approaches lies in how parameters are interpreted. In the frequentist framework, parameters such as population means are regarded as fixed but unknown constants; there is only one true value in the population, and it is estimated from sample data without assigning probability distributions to the parameter itself. Uncertainty about that estimate is summarised through confidence intervals. In contrast, the Bayesian framework treats parameters as uncertain quantities, providing a distribution of plausible values and representing them using probability distributions (Van de Schoot et al., 2014). The corresponding measure of uncertainty is the credible interval, obtained directly from the posterior distribution.

Beyond these differences, Bayesian methods provide a more dynamic and flexible framework by incorporating prior information and continuously updating beliefs as new data become available.

The cornerstone of Bayesian inference, the Bayes' theorem, underpins this process and is expressed as follows:

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{p(y)} \tag{2.1}$$

where:

- $p(\theta \mid y)$ is the posterior probability, representing the updated beliefs about the parameter, θ given the data, y.
- $p(y \mid \theta)$ is the likelihood, representing the plausibility of the data, y, given the parameter, θ.
- $p(\theta)$ is the prior probability, representing prior knowledge or assumptions about the parameter, θ, before observing any data.
- $p(y)$ is the marginal likelihood, a normalizing constant ensuring the posterior is a valid probability distribution.

When the data, y, is collected, the posterior distribution is computed by combining the prior with the likelihood of the observed data and normalizing by the marginal likelihood.

2.2 Priors

Priors are a defining feature of Bayesian inference, setting it apart from other statistical frameworks. They encode knowledge about parameters before observing any data, drawing on past experience or expert opinions. By incorporating prior information, Bayesian inference ensures that the process is not purely data-driven but represents a synthesis of evidence and existing knowledge (Gelman et al., 1995) (Van de Schoot et al., 2014).

To illustrate the concept of priors, let's consider a simple example involving a coin flip. The goal is to estimate the probability of the coin landing on heads. Without any prior information, we assume the coin is fair and assign equal probabilities to heads and tails; this is called a non□informative prior. If we only suspect—without strong evidence—that the coin might favour heads, we can adopt a weakly informative prior that slightly raises the probability of heads while still allowing the data to dominate. When solid prior knowledge indicates that the coin is indeed biased toward heads, an informative prior that assigns a much higher probability to heads is appropriate. As more flips are observed, the posterior distribution updates by combining the chosen prior with the observed data, gradually refining the estimate.

This example highlights how priors affect inference, especially in small-sample settings where they play a significant role in shaping the posterior distribution. As more data is collected, the influence of the prior diminishes, and the posterior becomes increasingly driven by the likelihood.

Priors can be broadly classified into three categories based on the level of informativeness they provide:

i. Informative Priors Informative priors incorporate substantial prior knowledge about a parameter. They are particularly useful when reliable prior information is available, such as from well-established studies or expert consensus. By providing strong guidance, informative priors can improve the precision of the posterior, especially when the data is scarce or noisy.

ii. Noninformative Priors Noninformative priors, are constructed to exert minimal influence on the posterior distribution, ensuring that inferences are predominantly driven by the observed data. These priors often take the form of vague, flat, or uniform distributions and are particularly useful in exploratory analyses or situations where prior knowledge is unavailable.

iii. Weakly Informative Priors Weakly informative priors strike a balance between informative and noninformative priors. They encode broad, plausible constraints on parameters without being overly restrictive. These priors are particularly useful when some general prior knowledge is available but lacks precision. By regularizing the inference, weakly informative priors help prevent extreme or implausible parameter estimates while still allowing the data to play a dominant role.

2.2.1 Conjugate Priors

Conjugate priors are, by definition, informative. What sets them apart is their computational convenience: when a prior is conjugate to the likelihood, the resulting posterior belongs to the same distributional family, which greatly streamlines Bayesian updating.

For example, with a Poisson likelihood to model football goals, assigning a Gamma prior to the rate parameter produces a Gamma posterior. This analytic tractability obviates the need for iterative methods such as Markov chain Monte Carlo, making conjugate priors particularly useful in applications that require rapid or real□time inference.

2.2.2 Selecting Priors

In practice, the selection of priors should be guided by the context and goals of the analysis. When prior knowledge is reliable and well-supported, informative priors can enhance precision. For exploratory work or when little prior knowledge exists, noninformative priors allow the data to speak for itself. Weakly informative priors are useful when some general expectations are available but flexibility is still desired, helping to regularize estimates.

2.3 Likelihood

The likelihood function is a key part of Bayesian inference. It measures how well specific parameter values explain the observed data (Van de Schoot et al., 2014). If y represents the observed data and θ represents the parameters, the likelihood is written as:

$$L(\theta; y) = p(y \mid \theta). \tag{2.2}$$

Unlike a probability distribution, the likelihood is a function of the parameters while the data stays fixed (Gelman et al., 1995). It does not sum or integrate to 1 over the parameters, so it can't be a probability distribution.

The likelihood plays a central role in constructing the posterior distribution, as it's shown in the Bayes' Theorem, it incorporates the observed data into the estimation process and transforms it into a proper probability distribution (Gelman et al., 1995). That role is more prominent as the amount of data increases, in such cases, the posterior becomes similar to the likelihood, making the results more consistent and less dependent on the prior (Van de Schoot et al., 2014).

To ensure valid Bayesian inference, it is critical that the likelihood function is well-specified and matches the characteristics of the data. An incorrectly defined likelihood can result in an invalid posterior distribution, and misleading conclusions. This issue becomes more pronounced with sparse or noisy data, where the model may rely too heavily on the prior to compensate for gaps in the data.

2.4 Posterior Distribution

The posterior distribution represents the updated knowledge about the model parameters after data have been observed. It is the cornerstone of Bayesian inference, blending prior beliefs with empirical evidence to provide a full probabilistic picture of uncertainty.

Because the posterior captures the entire range of plausible parameter values and their probabilities, credible intervals offer a direct probability interpretation.

When a single summary value is needed, Bayesian decision theory selects the point that minimizes the expected posterior loss. Under quadratic loss the optimal estimate is the posterior mean; under absolute loss it is the posterior median;

The posterior distribution underpins a wide range of Bayesian applications, including:

- **Parameter Estimation**: Probabilistic estimates of model parameters are derived directly from the posterior, offering insights into the most likely parameter values and the uncertainty surrounding them.
- **Predictive Inference**: Generating predictions for new data based on the posterior predictive distribution.
- **Model Comparison**: Comparing models using posterior probabilities.
- **Decision Theory**: Supporting optimal decisions under uncertainty.

2.5 Computational Methods

Bayesian analysis relies on computational methods to approximate posterior distributions, which are often intractable due to the complexity of high-dimensional integrals. One of the most widely used techniques for this task is the Markov Chain Monte Carlo, which generates samples from a target distribution by constructing a Markov chain that converges to the desired posterior (Carlo, 2004).

A Markov chain is a sequence of random variables $X_0, X_1, X_2, \ldots$ where the probability of transitioning to the next state depends only on the current state:

$$P(X_{t+1}|X_0, X_1, \ldots, X_t) = P(X_{t+1}|X_t). \tag{2.3}$$

This property, known as *memoryless*, ensures that past states do not influence the evolution of the chain beyond the most recent state. The collection of all possible values the sequence can take is referred to as the *state space* (Carlo, 2004).

Initially, MCMC samples may exhibit significant variability due to the arbitrary starting point of the chain. However, after sufficient iterations, the chain reaches *stationarity*, a phase where the sampling rule stabilizes and the sequence accurately represents the target posterior distribution. Only after achieving stationarity can the samples reliably infer key properties of the posterior.

2.6 Convergence

Convergence ensures that the chain has stabilized and is sampling correctly from the target posterior distribution. Without convergence, the samples fail to represent the posterior accurately, leading to unreliable predictions, and flawed conclusions.

Convergence occurs when the Markov chain transitions from being influenced by its starting conditions to sampling stably from the posterior. Achieving this requires sufficient iterations, as the chain

needs enough time to explore the parameter space thoroughly. Premature stopping or insufficient iterations can result in pseudo-convergence, where the chain appears stable but has not fully explored the posterior.

There are various tools to make sure that convergence was achieved, these tools provide quantitative and visual methods to evaluate the stationarity of the chain and help identify issues such as high autocorrelation, poor mixing, or incomplete exploration of the posterior.

2.6.1 Effective Sample Size (ESS)

In MCMC, samples are not completely independent of one another, each one is generated based on the previous one, creating autocorrelation. This means that each new sample does not provide entirely new information and is somewhat redundant.

The Effective Sample Size quantifies the amount of unique and useful information present in the samples, accounting for this autocorrelation. Rather than simply counting the total number of samples, ESS estimates how many independent samples are effectively represented (Roy, 2020). For example, if there are 1,000 samples and the ESS is 100, this indicates that these 1,000 samples contain as much information as 100 independent samples in a non-MCMC sampling process.

A high ESS suggests that the chain has produced a substantial amount of valuable information and reflects a well-mixed chain that explores the parameter space effectively, contributing to more reliable estimates of the posterior distribution. In contrast, a low ESS indicates high autocorrelation, requiring more samples to achieve reliable results.

2.6.2 Trace Plots

A trace plot is a diagnostic graph that displays the sampled values of a parameter at each iteration of an MCMC chain. The x-axis represents the iteration number, while the y-axis shows the corresponding parameter values.

A good trace plot shows the chain fluctuating freely around a central value, without systematic trends or directional drift over time. These fluctuations are critical because they indicate less dependency between successive samples and also reflect the variability of the posterior distribution and the chain's ability to explore it. "It is often said that a good trace plot should look like a hairy caterpillar." (Roy, 2020).

If a chain appears too flat or overly consistent, it indicates poor mixing or inefficient exploration of the parameter space. In such cases, the chain might be stuck in a narrow region due to high autocorrelation or a poorly tuned algorithm, failing to fully capture the posterior distribution.

2.6.3 Gelman-Rubin Diagnostic

The Gelman-Rubin diagnostic, $\hat{R}$ is a widely used metric to assess the convergence of multiple MCMC chains. It works by comparing the variance between and within chains for each parameter. If the chains have converged to the target distribution, the between-chain variance and the within-chain variance should be similar.

When chains have not converged, they are likely exploring different regions of the parameter space. This results in a higher between-chain variance relative to the within-chain variance.

An $\hat{R}$ value close to 1 (typically less than 1.1) indicates that the chains are sampling from the same target distribution, under these conditions, it can be confidently stated that the parameter estimates are reliable and representative of the posterior distribution (Roy, 2020).

Conversely, a $\hat{R}$ value significantly greater than 1 suggests non-convergence, indicating that the chains have not mixed well and are potentially stuck in different regions of the parameter space.

The focus on the upper confidence interval (C.I.) of the $\hat{R}$ statistic rather than its point estimate is because the upper C.I. provides a conservative estimate of the worst-case scenario for the $\hat{R}$ value, ensuring more reliable diagnostic conclusions.

2.7 Hierarchical Models

Hierarchical models are statistical frameworks designed to analyze data with a nested or grouped structure. Many real-world scenarios involve data organized into clusters, such as matches played by different teams in sports or patients treated at different hospitals in healthcare.

Traditional methods like simple linear regression or ANOVA often fall short in such contexts because they fail to account for variability at multiple levels (Veenman et al., 2024). These methods either ignore differences between groups or treat all observations as independent.

In a sports context, using a simple regression model to predict match outcomes assumes that team performance is identical across all matches, ignoring team-specific strengths, opponent effects, and dynamic game conditions. Without modeling these factors explicitly, predictions fail to capture key performance variations, leading to reduced accuracy.

A key feature of hierarchical models is their ability to incorporate both fixed and random effects:

- **Fixed effects**: These represent consistent trends that apply across all groups or observations. For instance, in a football prediction model, the general advantage of playing at home can be modeled as a fixed effect, capturing its influence on match outcomes uniformly across all teams (Veenman et al., 2024).

- **Random effects**: These capture variability specific to individual groups, allowing for differences between them. For example, team-specific attack and defense strengths in football can be modeled as random effects, reflecting variations in performance across teams while still being influenced by overall trends (Veenman et al., 2024).

Another property of hierarchical models is shrinkage. Shrinkage adjusts individual group estimates toward the overall mean, stabilizing predictions and reducing the impact of extreme values. This is especially valuable in small-sample or noisy data scenarios, where extreme estimates are more likely. By borrowing strength from the overall population, shrinkage ensures predictions are more reliable and less prone to overfitting.

While hierarchical models offer significant advantages, they also present challenges. The inclusion of multiple levels of variability increases computational complexity, making this models more resource-intensive. Additionally, hierarchical models require careful specification of priors, particularly for random effects, as poorly chosen priors can lead to biased estimates.

Chapter 3

Zero-Modified Poisson

3.1 Introduction to Football Predictions

Predicting football match outcomes has become an area of interest within sports analytics, providing valuable information for strategic planning, performance evaluation, and betting markets, which is very appealing for both professional and recreational contexts. That said, there are some challenges that need to be dealt with, in order to improve the quality of the predictions, one of those is the fact that football scores are very different depending on the quality of each team. In the same league it's possible to find teams that have a very low goal count, and teams that score in almost every match. This variability introduces complexities in the modeling process, where traditional models often struggle to adequately capture the differences in goal frequency.

3.2 Standard Poisson Distribution

The Poisson distribution has long been a cornerstone in modeling event counts, including football scores. Its mathematical simplicity and intuitive appeal make it a popular choice among statisticians, and it is the most widely used approach for modeling the occurrence of random events over a fixed interval of time or space, commonly referred to as count data.

It assumes that events occur independently and that the mean and variance of the data are equal. This simplicity makes it computationally efficient and straightforward to interpret, providing a solid baseline for analysis.

A random variable X follows a Poisson distribution with parameter λ if its probability mass function (PMF) is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots, \ \lambda > 0. \tag{3.1}$$

where:

- k is the number of occurrences of the event in a fixed interval.
- λ is the expected number of events in that interval.
- e is Euler's number.

The Poisson distribution is uniquely defined by the single parameter λ, which represents both the mean and variance of the distribution (equidispersion)(Zhao et al., 2020):

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda. \tag{3.2}$$

While effective in some contexts, the Poisson distribution has notable limitations when applied to football score data. A fundamental issue is its assumption that the mean and variance are equal, which often does not align with real-world football data, since matches frequently exhibit significant goal variability, leading to both overdispersion and underdispersion. These discrepancies could result in inaccurate predictions, as the Poisson model underestimates or overestimates the frequency of zero scores.

3.3 Zero-Inflated Poisson (ZIP) Model

Count data often exhibit an excess of zeros, which standard probability models, such as the standard Poisson distribution, may fail to accurately capture. The ZIP model, first introduced by (Lambert, 1992), addresses that zero inflation by introducing a dual-process mechanism, distinguishing structural zeros, and random zeros.

This distinction allows the ZIP model to provide a more nuanced representation of count data with zero inflation, which is very important in the context of football, where goalless matches frequently occur.

3.3.1 Conceptual Framework of the ZIP Model

With a Zero-Inflated Poisson model the zero-generating process consists of two distinct components (Mouatassim and Ezzahid, 2012):

- **A structural zero process**: A mechanism that only produces zeros.
- **A count process with possible random zeros**: A standard Poisson distribution, that generates counts, some of which are zeros.

To account for the two sources of zeros, the ZIP model has the parameter p, which determines whether an observation comes from the structural zero process or the count process. The PMF of a ZIP model is given by:

$$P(Y = 0) = p + (1 - p)\, e^{-\lambda}, \qquad 0 < p < 1,\ \lambda > 0 \tag{3.3}$$

$$P(Y = x) = (1 - p)\, \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 1, 2, \ldots,\ 0 < p < 1,\ \lambda > 0 \tag{3.4}$$

where:

- p represents the probability of the zero belonging to the structural zero process.
- (1-p) represents the remaining probability, so the zero belonging to the count process

However, despite its flexibility, the ZIP model has limitations. It assumes that all deviations in zero frequency are due to inflation, making it ineffective for cases of zero deflation, where fewer zeros are observed than expected. As a result, the Zero-Inflated Poisson may not be the best choice for datasets with diverse zero patterns that do not conform to a simple inflation mechanism.

3.4 The Zero-altered Poisson Model

The Zero-Altered Poisson ((Klugman et al., 2012)) model similar to the ZIP is designed for count data that exhibit an excess of zeros compared to what the standard Poisson distribution would predict. In this model, the probability mass function of the standard Poisson distribution is directly modified by increasing the probability assigned to zero. In other words, instead of letting the Poisson process naturally generate zeros at its usual rate, the ZAP model forces a higher probability for a zero outcome and then redistributes the remaining probability mass among the positive counts so that the total probability still sums up to one.

Every observation, including zeros, is assumed to come from one modified process. This distinguishes it from the Zero-Inflated Poisson model, which assumes that zeros arise from two different sources. In the ZAP model, all zeros are treated uniformly as part of a single, reweighted probability function.

Mathematically, the Zero-Altered Poisson model can be described as follows: a fixed probability, p_0, is assigned to the zero outcome, ensuring that the likelihood of zero is higher than the probability predicted by a standard Poisson distribution. For outcomes greater than zero, the original Poisson probabilities are scaled down proportionally so that together with the fixed zero probability they sum to one. This reweighting mechanism increases the mass on the left side of the distribution (i.e., at zero) while preserving the overall shape of the Poisson distribution for positive counts.

The PMF of the ZAP distribution is given by:

$$p(x) = \begin{cases} p_0, & x = 0, \\ \frac{(1-p_0)}{(1-e^{-\lambda})} f(x), & x = 1, 2, \ldots \end{cases} \tag{3.5}$$

where $\lambda > 0$ and $0 \leq p_0 \leq 1$.

- $f(x)$ represents the standard Poisson PMF
- p_0 is the fixed probability assigned to zero.

This distribution is also commonly referred to as Zero-Modified Poisson model, which is the same name of the distribution this work focus ono. Due to this naming ambiguity, it was suggested to me to refer to the model discussed in this section as the Zero-Altered Poisson model, distinguishing it from the distribution used in the main analysis.

3.5 Zero-Modified Poisson Model

The Zero-Modified Poisson (ZMP) model (Conceição et al., 2014), and the focus of my dissertation, is a generalization of the standard Poisson that introduces a modification parameter p. This parameter allows for both zero inflation and deflation by dynamically adjusting the probability of zero outcomes, unlike the ZIP and ZAP models.

Other advantage is that, while the Poisson distribution assumes only equidispersion, the ZMP also assumes over and underdispersion.

3.5.1 Mathematical Formulation of Zero-Modified Poisson

The PMF of the ZMP model is defined as:

$$P_{ZMP}(Y = y) = (1 - p)I(y) + pP_P(Y = y) \tag{3.6}$$

where $0 < p < \frac{1}{1 - e^{-\mu}}$ This equation combines the probability of zero and non-zero occurrences into a single unified expression.

- **Random Variable** Y: The random variable Y represents the count outcome, which can take any non-negative integer value:
$$Y \in \{0, 1, 2, \ldots\}$$

- **Indicator Function** $I(y)$: The indicator function $I(y)$ is defined as:
$$I(y) = \begin{cases} 1, & \text{if } y = 0 \\ 0, & \text{if } y > 0 \end{cases} \tag{3.7}$$
The purpose of $I(y)$ is to single out the zero case, so that the Poisson probability of a zero outcome is modified by the additional term $(1-p)I(y)$. For non-zero values ($y > 0$), $I(y) = 0$, meaning that this term contributes nothing, and the probability is determined solely by the Poisson component scaled with the zero-modification parameter.

- **Poisson Probability** $P_P(Y = y)$: The term $P_P(Y = y)$ represents the probability mass function of a standard Poisson distribution with mean μ.

3.5.2 Mean and Variance

Unlike the Poisson distribution, the ZMP with the modification parameter p allows for both overdispersion ($\text{Var}(Y) > E[Y]$) and underdispersion ($\text{Var}(Y) < E[Y]$), making it a more versatile choice for modeling real-world count data. This flexibility is particularly valuable in applications like football score modeling, where goal distributions often exhibit higher variability than a standard Poisson process predicts.

Expected Value and Dispersion Properties

The expected value and variance of a ZMP distributed random variable Y are given by

$$E[Y] = p\mu \tag{3.8}$$

where, μ represents the mean of the underlying Poisson component, and p scales the expected value according to the zero modification.

$$\text{Var}(Y) = p\left(\mu + (1 - p)\mu^2\right) \tag{3.9}$$

Where it demonstrates that the variance can vary independently of the mean, allowing the model to capture both overdispersed and underdispersed data. Specifically, when $p < 1$, the variance is greater than the mean (overdispersion), when $p = 1$ the variance is equal to the mean (equidispersion) and when $p > 1$, the variance is less than the mean (underdispersion).

3.5.3 How p Modifies Zero Frequencies

The parameter, p, allows the ZMP model to dynamically adapt to scoring patterns, accommodating datasets where zero anomalies are present.

Depending on it's value, the model determines the type of zero modification it is handling:

- **When** $p = 0$ **then:**

$$P_{ZMP}(Y = 0) = 1 \tag{3.10}$$

  - The probability of $Y = 0$ is 1. This represents a degenerate distribution where all the mass is at zero, meaning all observations are 0.

- **When** $0 < p < 1$ **then:**

$$(1 - p)(1 - P_{Po}(Y = 0)) > 0 \tag{3.11}$$

  hence

$$P_{ZMP}(Y = 0) > P_{Po}(Y = 0) \tag{3.12}$$

  - The frequency of zeros is greater than predicted by the standard Poisson distribution. This means the modified Poisson model adjusts for zero inflation so it will increase the probability of zero counts, resulting in a Zero-Inflated Poisson (ZIP) distribution.

- **When** $p = 1$ **then:**

$$P_{ZMP}(Y = 0) = P_{Po}(Y = 0) \tag{3.13}$$

  - There is no modification to the count of zeros. The modified distribution reverts to the standard Poisson distribution.

- **When** $1 < p < \frac{1}{1-P_{Po}(Y=0)}$ **then:**

$$(1 - p)(1 - P_{Po}(Y = 0)) < 0 \tag{3.14}$$

  hence

$$P_{ZMP}(Y = 0) < P_{Po}(Y = 0) \tag{3.15}$$

  - The frequency of zeros is less than predicted by the standard Poisson distribution. This means the modified Poisson model adjusts for zero deflation so it will decrease the probability of zero counts, resulting in a Zero-Deflated Poisson (ZDP) distribution.

- **When** $p = \frac{1}{1-P_{Po}(Y=0)}$ **then:**

$$P_{ZMP}(Y = 0) = 0 \tag{3.16}$$

  - There are no zero occurrences in the observations. This represents a Zero-Truncated Poisson (ZTP) distribution.

The maximum value of p ensures that the PMF remains valid, meaning all probabilities are non-negative and sum to 1

$$0 \leq p \leq \frac{1}{1 - P_{Po}(Y = 0)} \tag{3.17}$$

where $P_{Po}(Y = 0)$ is the probability of zero under the standard Poisson distribution.

3.5.4 Parameter p in a football context

At first glance, it may seem that a higher p is correlated with a stronger team, while a lower p indicates a weaker team. That's not always the case, p does not measure goal-scoring ability directly, it measures how much the actual zero-goal frequency deviates from the Poisson expectation.

In other words, a weak team, with a small μ, can have an higher p if its zero frequency is lower than predicted by the Poisson distribution, compared to a strong team with a large μ but a higher-than-expected zero frequency.

Considering two teams:

- **Strong Team (S)** with a high mean goals per game μ_S.
- **Weak Team (W)** with a low mean goals per game μ_W.

The standard Poisson model predicts the probability of scoring zero goals as:

$$P(Y = 0) = e^{-\mu} \tag{3.18}$$

Strong Team (S)

- Mean goals per game: $\mu_S = 2.5$
- Poisson expected probability of zero goals:

$$P(Y = 0) = e^{-2.5} \approx 8.2\% \tag{3.19}$$

- Supposing the actual observed probability is 12%, meaning this team goes scoreless slightly more often than expected.
- This suggests zero inflation ($p < 1$) or at least near Poisson behavior ($p \approx 1$).

Weak Team (W)

- Mean goals per game: $\mu_W = 0.8$
- Poisson expected probability of zero goals:

$$P(Y = 0) = e^{-0.8} \approx 44.8\% \tag{3.20}$$

- Supposing the actual observed probability is 38%, meaning this team goes scoreless less often than what the Poisson predicted.
- This suggests zero deflation ($p > 1$).

Even though the weak team still has more scoreless games than the strong team, it has a higher p because it deflates zeros relative to the Poisson expectation.

Chapter 4

Model Implementation

4.1 Dataset

The Italian Football League, commonly known as Serie A, is one of the most prestigious and competitive football leagues in Europe. The league consists of 20 teams that compete in a double round-robin format, with each team playing 38 matches - 19 at home and 19 away - resulting in a total of 380 games throughout the season. Points are awarded based on the outcome of each match: three points for a win, one point for a draw, and none for a loss. At the end of the season, the team with the most total points is crowned the champion, while the three lowest-ranked teams are relegated to Serie B, the second tier of Italian football.

This study analyzes data from the 2022/23 Serie A season. During this season, Juventus was penalized with a 10 point deduction due to financial irregularities related to past transfer dealings. For the purposes of this analysis, this penalty has been disregarded, as it did not affect the team's performance in individual matches.

The match data used in this study was sourced from ZeroZero (Zerozero, 2024), a portuguese website mainly focused in football. ZeroZero compiles detailed historical and contemporary data on match results, team performances, and player statistics from various leagues around the world.

4.2 Implementation of the Bayesian ZMP Model

The Zero-Modified Poisson model was implemented using the NIMBLE package in R, a framework designed for Bayesian modeling and MCMC simulations. Since NIMBLE does not natively have a ZMP distribution function, a custom implementation was necessary.

The implementation comprises two key components: dZMPC and rZMPC.

The dZMPC function computes the probability mass function for the zero-modified Poisson distribution.

For $x = 0$,

$$P(Y = 0) = (1 - p) + p \cdot e^{-\mu}, \tag{4.1}$$

which combines the standard Poisson probability of zero, $e^{-\mu}$, with an extra mass $(1 - p)$.

For $x > 0$:

$$P(Y = x) = p \cdot \frac{\mu^x e^{-\mu}}{x!}, \tag{4.2}$$

which is a scaled version of the standard Poisson probability mass function, where the scaling factor p adjusts the weight given to non-zero counts.

Before computing the PMF, the function performs validity checks on the parameters μ, p, and x. These checks ensure that all parameters are non-negative, and that the parameter p does not exceed the allowable maximum $1/(1-e^{-\mu})$, ensuring that the probabilities remain well-defined.

The rZMPC function provides a method to generate random samples from the zero-modified Poisson distribution. It first calculates the probability of a zero score using the same formula. It then draws a random number to decide whether to return a zero or, if not, to sample from a standard Poisson distribution repeatedly until a non-zero value is obtained. This approach ensures that the generated data follows the modified distribution correctly.

Integration with NIMBLE

Finally, the custom distribution is registered with NIMBLE using the `registerDistributions` function. This step enables the model to recognize and utilize the dZMPC distribution.

4.2.1 Model Specification

For each match g, the goals scored by the home team ($y_1[g]$) and the away team ($y_2[g]$) were modeled using the ZMP distribution with parameters μ (expected goals) and p (zero-modification parameter):

$$y_1[g] \,\Big|\, \mu_1[g],\, p_1[\text{hometeam}[g]] \;\sim\; \text{ZMP}\big(\mu_1[g],\, p_1[\text{hometeam}[g]]\big). \tag{4.3}$$

$$y_2[g] \,\Big|\, \mu_2[g],\, p_2[\text{awayteam}[g]] \;\sim\; \text{ZMP}\big(\mu_2[g],\, p_2[\text{awayteam}[g]]\big). \tag{4.4}$$

where:

- μ: Represents the expected number of goals before accounting for zero modification.
- p: Adjusts for the zero modification in the data, it's specific for each team.

4.2.2 Linear Predictors

The expected number of goals for the home team (4.5), and for the away team (4.6) were modeled using linear predictors based on team-specific strengths and home advantage. These equations were first introduced by Maher (1982) and are widely used in football score modeling:

$$\log(\mu_1[g]) = \text{home} + \text{att}[\text{hometeam}[g]] + \text{def}[\text{awayteam}[g]] \tag{4.5}$$

$$\log(\mu_2[g]) = \text{att}[\text{awayteam}[g]] + \text{def}[\text{hometeam}[g]] \tag{4.6}$$

where:

- home, captures the overall advantage of playing at home.
- att, is the team-specific attacking strength, the higher the value, the stronger the attack.
- def, is the team-specific defensive strength, the lower the value, the better the defense.

The home team's expected goals, μ_1, depends on home advantage, the home team's attack, and the away team's defense. Likewise, the away team's expected goals, μ_2, depends on the away team's attack and the home team's defense.

4.2.3 Priors for Zero-Modification Parameters

The zero-modification parameters p_1 and p_2 are each assigned a uniform prior on the interval $[0, 3]$, given by:

$$p_1[j] \sim \text{Uniform}(0, 3), \quad p_2[j] \sim \text{Uniform}(0, 3), \quad j = 1, \ldots, n_{\text{teams}} \tag{4.7}$$

This broad range allows for the possibility of both zero-inflation ($p < 1$) and zero-deflation ($p > 1$), or even no modification ($p = 1$) if the data suggest it. Although the ZMP distribution requires $p \leq \frac{1}{1-e^{-\mu}}$ to remain valid for each match-specific μ, incorporating this restriction directly into the model proved challenging due to implementation constraints in R, so any draws that exceed this condition are ignored in the prediction process.

The choice of a uniform prior reflects weak prior knowledge (Piegorsch et al., 2002): by assigning equal probability to all values in $[0, 3]$, the posterior distribution for p is determined primarily by the observed data. This design preserves flexibility in capturing a range of zero-inflation or zero-deflation behaviors without imposing a strong constraint at the prior stage.

4.2.4 Hierarchical Modeling of Latent Parameters

In this model, the attack and defense strengths of each team are treated as latent parameters, which are unknown values that are not directly observed, instead they are inferred from the data.

These parameters are modeled hierarchically to allow for team-specific variations while sharing information across all teams.

The raw attack strength, att.star$[t]$, and raw defense strength, def.star$[t]$, which are the uncentered, initial values of the parameters for each team t, are drawn from normal distributions with global mean and precision parameters:

$$\text{att.star}[t] \sim \mathcal{N}(\mu_{\text{att}}, \tau_{\text{att}}), \quad \text{def.star}[t] \sim \mathcal{N}(\mu_{\text{def}}, \tau_{\text{def}}) \quad \text{for } t = 1, \ldots, n_{\text{teams}} \tag{4.8}$$

- $\mu_{\textbf{att}}$: Represents the league-wide average attacking strength.
- $\mu_{\textbf{def}}$: Represents the league-wide average defensive strength.
- $\tau_{\textbf{att}}$: Precision parameter (inverse of variance) for attack strengths, controlling how much teams can deviate from the league-wide average.
  - A **high precision** value implies that teams' attack strengths are very similar to each other.
  - A **low precision** value allows teams' attack strengths to vary more significantly.
- $\tau_{\textbf{def}}$: Precision parameter for defense strengths, with the same interpretation as τ_{att}.

Centering for Identifiability The raw attack and defense strengths are not directly used in the linear predictors. Instead, they are centered to ensure that the model is identifiable. This is done by subtracting the mean of all raw attack or defense strengths across teams from each team's raw strength:

$$\text{att}[t] = \text{att.star}[t] - \text{mean}(\text{att.star}), \quad \text{def}[t] = \text{def.star}[t] - \text{mean}(\text{def.star}) \tag{4.9}$$

This ensures that the sum of the centered parameters across all teams is zero:

$$\sum_{t=1}^{n_{\text{teams}}} \text{att}[t] = 0, \quad \sum_{t=1}^{n_{\text{teams}}} \text{def}[t] = 0 \tag{4.10}$$

Why is Centering Necessary?

1. **Identifiability**: Without centering, the global mean parameters (μ_{att} and μ_{def}) and the team-specific raw parameters (att.star[t] and def.star[t]) would not have a fixed reference point. This would lead to multiple combinations of values explaining the same outcome, making it impossible to find the true value of the parameters.

2. **Interpretability**: Centered parameters have a clear interpretation:

   - If att[t] > 0, team t has above-average attacking strength.
   - If def[t] < 0, team t has above-average defensive strength.

Why is this an Hierarchical Approach? Because team-specific parameters (att.star[t] and def.star[t]) are modeled as random variables that depend on shared global parameters (μ_{att}, μ_{def}, τ_{att}, and τ_{def}).

Since it's an hierarchical approach the shrinkage effect ensures that the posterior estimates of the raw parameters are influenced by the global means, pulling extreme values toward the league-wide averages unless they are strongly supported by the data.

4.2.5 Priors for Parameters

The model employs weakly informative priors for the global parameters and the home advantage, ensuring minimal constraints and allowing the data to predominantly influence the estimates.

- **Global Parameters for Attack and Defense (μ_{att}, μ_{def}, τ_{att}, τ_{def}):**
  - The league-wide means of attacking and defensive strengths (μ_{att} and μ_{def}) are modeled with normal priors, with mean and precision parameters:

    $$\mu_{\text{att}}, \mu_{\text{def}} \sim \mathcal{N}(0, 0.001) \tag{4.11}$$

    These priors are centered at zero, indicating no strong prior belief about the average attacking and defensive abilities across teams. The high variance ($\sigma^2 = \frac{1}{0.001} = 1000$) ensures a broad prior, meaning that there's not a strong commitment to the mean value.
  - The precision parameters (τ_{att} and τ_{def}) are assigned gamma priors:

    $$\tau_{\text{att}}, \tau_{\text{def}} \sim \text{Gamma}(0.01, 0.01) \tag{4.12}$$

    These values result in a very broad gamma distribution, reflecting the uncertainty of a weakly informative prior

- **Home Advantage (home):**
  The home advantage parameter is also modeled with a normal prior, with mean and precision parameters:

  $$\text{home} \sim \mathcal{N}(0, 0.0001) \tag{4.13}$$

  This broad prior reflects significant uncertainty about the size of the home advantage, with a variance of $\sigma^2 = \frac{1}{0.0001} = 10000$. This ensures that the data predominantly determines the magnitude of this effect.

These ensure that the global trends in attack, defense, and home advantage are adequately captured without imposing any restrictive assumptions.

4.3 MCMC Sampling Techniques

To estimate the posterior distributions of the model parameters, MCMC sampling was implemented using the NIMBLE framework. The primary goal was to infer the key parameters of the Zero-Modified Poisson model, the team-specific attack and defense strengths, home advantage, and zero-modification probabilities.

4.3.1 Round-Specific Sampling

The MCMC sampling process was conducted round by round, starting from round 20, which corresponds to the first round of the second half of the season, to round 38.

For each round being predicted, the dataset was subseted to include only the matches played up to that round. This ensured that posterior distributions were made using the most up-to-date information available at the time, reflecting recent form and trends, simulating how real-world predictions would be carried out during a live season.

For example, when inferring the posterior distribution for round 24, only the data from rounds 1 to 23 was used.

In the end, there are 19 MCMC outputs to analyze, one for each round.

This round-specific approach improves prediction accuracy by maximizing the amount of information used for each round, instead of relying on a fixed dataset for all predictions. Additionally, it enabled analysis of how the parameters changed over time as more data became available.

4.3.2 MCMC Configuration

After preparing the dataset for a given round, the following steps were taken to configure and run the MCMC process:

- **Team indexing**: Each team was assigned a unique index to facilitate parameter tracking and ensure consistent model inputs.
- **Monitored parameters**: The parameters monitored during the MCMC run included $\mu 1$, $\mu 2$, att, def, home, $p1$, $p2$.
- **Initial values for the chains**: Three chains were used, each one initialized with slightly different values for the parameters. Using multiple chains with different starting points ensures thorough exploration of the parameter space. Convergence across these 3 chains provides strong evidence that the posterior distributions are reliable.
- **Number of iterations**: A total of 80000 iterations were performed for each chain. Each iteration produces a sample of the parameters being estimated, consequently the MCMC creates a chain of samples that, over time, converge to the target posterior distribution, that way a large number of iterations increases the reliability of the estimated posterior.
- **Burn-in period**: The first 35000 iterations were discarded as burn-in. This process eliminates the influence of the arbitrary initial values by allowing the Markov chain to stabilize. This is useful

because during the early stages of the chain, the samples are heavily influenced by the starting values, which may not represent the true posterior distribution. By discarding this initial segment, it's ensured that only the stabilized samples are retained for inference.

- **Thinning**: Every 10th sample was retained from the MCMC chain. This helps reduce autocorrelation, ensuring more independence for the retained samples, improving the reliability of the posterior distribution. This approach also helps manage computational and storage demands by reducing the overall size of the chain. While this simplifies working with large datasets, it's crucial to ensure that the number of retained samples is sufficient to capture the posterior distribution accurately and provide robust inferences.
- **Size of each chain**: 4500 samples

4.4 Convergence Diagnostics

Once the MCMC run was completed, several diagnostic tools were used to assess whether the chains had or not converged to the true posterior distributions. They were, trace plots, the Gelman-Rubin diagnostic and the ESS.

4.5 Prediction Process

After confirming convergence, predictions were made with 2 methods

First Method

Round by round using the following steps:

1. **Sampling from the posterior distributions**: For each round, 1000 samples were randomly selected from the MCMC output for each parameter.
2. **Computing the expected number of goals**: Using the selected parameter samples, the expected number of goals for the home (4.14) and away (4.15) teams in each match was computed as:

$$\mu_1 = \exp(\text{home} + \text{att}[\text{hometeam}] + \text{def}[\text{awayteam}]) \tag{4.14}$$

$$\mu_2 = \exp(\text{att}[\text{awayteam}] + \text{def}[\text{hometeam}]) \tag{4.15}$$

3. **Generating goal outcomes**: For each sampled pair of μ and p, 1000 goals were generated using the ZMP distribution.

   So when predicting let's say a Inter vs Napoli, one μ and one p were randomly selected for Inter, and another pair was selected for Napoli. Then, 1000 goal outcomes were generated for each team (4.16).

   Since this process was repeated across 1000 sampled parameter values, a total of 1 million goals were generated per team in each match simulation. Essentially meaning that each match would be simulated 1 million times.

$$y_{\text{home}} \sim \text{ZMP}(\mu_1, p_1), \quad y_{\text{away}} \sim \text{ZMP}(\mu_2, p_2) \tag{4.16}$$

4. **Calculating win probabilities**: By comparing the goals scored by the home and away teams, the probability of each possible outcome (Home Win, Away Win, or Draw) was calculated. For instance, if the home team scored more goals than the away team in 60% of the simulated matches, the predicted probability of a home win would be 0.6.

5. **Generating the standings table**: The outcome with the highest probability was considered the predicted result.

   Using the predicted results, a standings table was created. The process was the follow: selecting the MCMC output corresponding to the first round, generating predictions, and adding the results to the table. Then, select the MCMC output for the next round, made the predictions for that one, and the new results added to those from the previous rounds. This process continued sequentially until the final round, ensuring that only one round was simulated at a time.

This entire process was repeated 1,000 times to account for the variability inherent in Bayesian models. In the end the model created 1000 simulated standing tables that ended up averaged in a single one.

Second Method

In this method, predictions were made using a full-season simulation approach, with the results for all remaining matches in the season simulated from the perspective of a specific round.

This approach provides a comprehensive view of the season's potential trajectory and final standings, offering probabilistic insights into all matches with only the data available from one specific round.

The first method focuses on short-term predictions, where each MCMC run forecasts the outcomes of individual matchdays, progressively incorporating more observed data as the season advances. Using this method leaves open the question of how well the model performs when tasked with predicting the final league standings at various points during the season.

To address this, is introduced a full-season prediction analysis, where the data from each MCMC is used to simulate all remaining matchdays, starting from the half of the season and continuing up to the last round.

In the early MCMCs, this approach can be considered long-term forecasting, as a significant portion of the season remains to be played. But, in the later MCMCs, the predictions will start being similar to the ones in the first method since, most of the season has already been played, and only a few matchdays remain to be predicted.

Predictions:

For each specific MCMC output, and similarly to the first method, 1000 samples of each parameter were randomly selected for each team in a match. For each pair of sampled parameters, 1000 goals were generated, which were then used to calculate the probability of the match outcome. The main difference in this method is that, instead of predicting only the matches for the specific round corresponding to the MCMC under analysis, the model simulated the results for the entire season.

This entire process was repeated 1000 times, resulting in 1000 predicted tables for each MCMC. Each one of these 19 tables were then averaged into a single one displayed the mean ranking of each team for any given round. For example, if Napoli finished in first place in all 1000 predicted tables of the 3th MCMC output, the compiled table would show a value of 1 in Napoli's row under the first-place column.

4.6 Predictive Performance Measures and Analysis

In probabilistic forecasting, evaluating the quality of predictions is essential to understanding a model's performance. This section explores the metrics used to assess predictive accuracy and reliability, focusing on **Percentage of Correct Predictions**, **Brier Score**, and **Geometric Mean**.

4.6.1 Percentage of Correct Predictions

The simplest and most intuitive measure is the percentage of correct predictions. This metric is calculated by comparing the predicted results of all 190 games with their observed outcomes to determine the proportion of matches correctly predicted by the model.

4.6.2 Brier Score and Geometric Mean

The **Brier Score** (Egidi and Torelli, 2021) measures the accuracy of probabilistic predictions by calculating the squared error between the predicted probabilities and the actual outcomes. This score evaluates how well the model assigns high probabilities to correct outcomes while penalizing overconfidence in incorrect predictions. Lower probabilities for incorrect predictions result in smaller penalties.

Key Characteristics:

- **Low Brier Scores** indicate better predictive performance, reflecting smaller deviations between forecasts and observed results.
- Scores range from **0** (perfect predictions) to **1** (worst predictions).

Calculation:

For each match, there are three possible outcomes:

$$\text{Home Win (H)}, \quad \text{Draw (D)}, \quad \text{Away Win (A)}.$$

The Brier Score for each match is computed as:

$$(\hat{p}_H - o_H)^2 \; + \; (\hat{p}_D - o_D)^2 \; + \; (\hat{p}_A - o_A)^2, \tag{4.17}$$

where:

- $\hat{p}_H, \hat{p}_D, \hat{p}_A$ are the predicted probabilities of each outcome.
- $o_H, o_D, o_A \in \{0, 1\}$ indicate the actual outcome, with 1 for the observed outcome and 0 for the others.

For example, if the observed outcome is a home win, $o_H = 1$, while $o_D = o_A = 0$.

The **Geometric Mean** (Egidi and Torelli, 2021) evaluates the overall predictive performance of the model by combining the probabilities assigned to the actual observed outcomes across all matches. This metric rewards confident, correct predictions and penalizes poor confidence in accurate results.

Key Characteristics:

- Higher Geometric Mean values indicate better predictive performance, with a maximum score of **1** representing perfect predictions.

Calculations:

The geometric mean for M matches is given by:

$$\text{GeoMean} = \left(\prod_{m=1}^{M} p_m \right)^{1/M}, \tag{4.18}$$

where:

- p_m is the predicted probability of the observed outcome for match m.
- M is the total number of matches.

Interpretation:

- **High p_m:** Indicates confident, correct predictions, contributing positively to the geometric mean.
- **Low p_m:** Reflects poor confidence in the correct outcome, negatively impacting the geometric mean.

The Brier Score and Geometric Mean are complementary metrics that evaluate the predictive performance of the model from different perspectives. While both aim to assess accuracy, they emphasize different aspects of the model's probabilistic predictions.

Key Differences:

- The **Brier Score** is more sensitive to overconfidence in incorrect predictions, making it useful for identifying cases where the model may be poorly calibrated.
- The **Geometric Mean** highlights the model's ability to provide strong confidence in correct predictions, providing a measure of overall reliability.

Complementary Insights:

- A low Brier Score combined with a high Geometric Mean suggests that the model is well-calibrated and consistently confident in its predictions.
- If the Brier Score is low but the Geometric Mean is also low, the model may be accurate overall but lacks confidence in its correct predictions.
- A high Geometric Mean but high Brier Score indicates strong confidence in correct outcomes but poor calibration in probabilistic assignments to other outcomes.

Chapter 5

Results and Discussion

5.1 Exploratory Analysis

Metric	Home	Away
Total Games	380	
Win	161	119
Draws	100	100
Mean Goals Scored	1.41	1.15
Clean Sheets	116	93

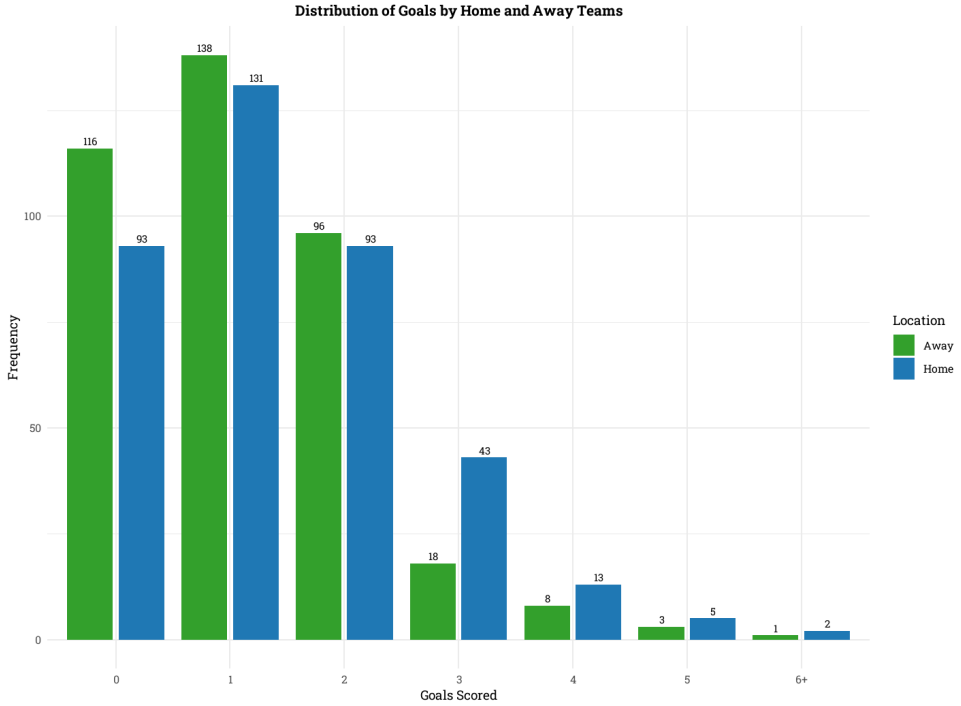


Figure 5.1: Summary statistics of total games, wins, draws, mean goals, and clean sheets for home and away teams, alongside the distribution of goals scored

As expected, Figure 5.1 shows a clear advantage for the Home Team. Clubs playing at their own stadium win more frequently than when they are the visiting side. They also score more goals and achieve more clean sheets, with a mean of 1.41 goals per game and 116 matches without conceding. Conversely, when playing away, teams score an average of 1.15 goals and manage 93 clean sheets.

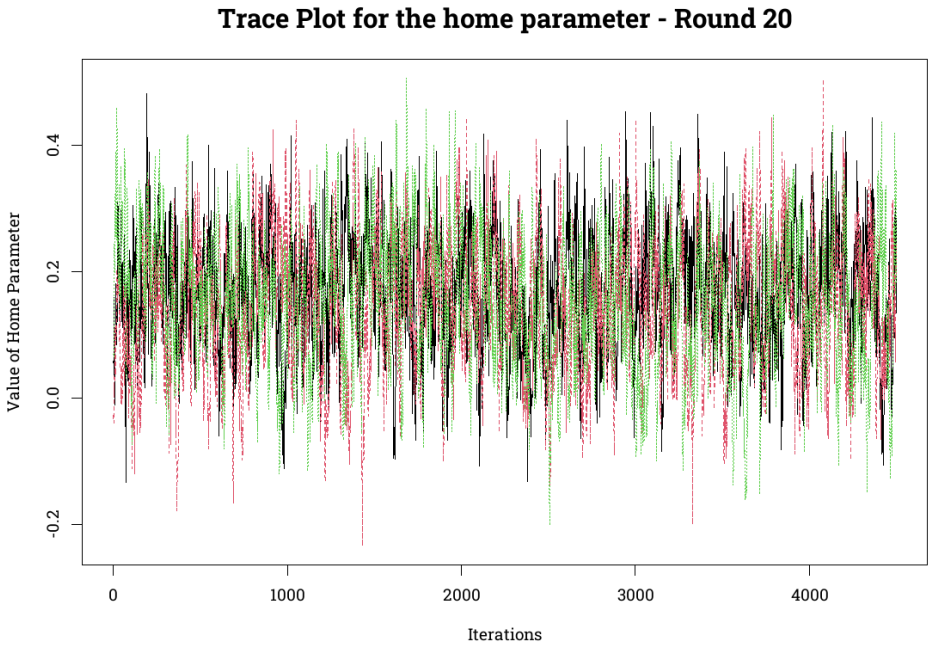
As implied by the clean sheet data, Away Teams fail to score more often than Home Teams and tend to have lower scores, such as 1 or 2 goals per game. Although the difference in performance is not significant for low scores, the real disparity emerges with 3 or more goals, where Home Teams clearly outperform their opponents.

5.2 Convergence Analysis

Trace Plots

The trace plots presented in Figure 5.2, corresponding to the home parameter in round 20 and Lazio's attack parameter for round 36, indicate good convergence of the MCMC chains. Despite being initialized from different starting values, all three chains show good mixing and overlap, with no visible patterns or trends, throughout the iterations.

The parameter values oscillate randomly around a central range, suggesting that the chains have reached stationarity and are effectively exploring the posterior distribution.



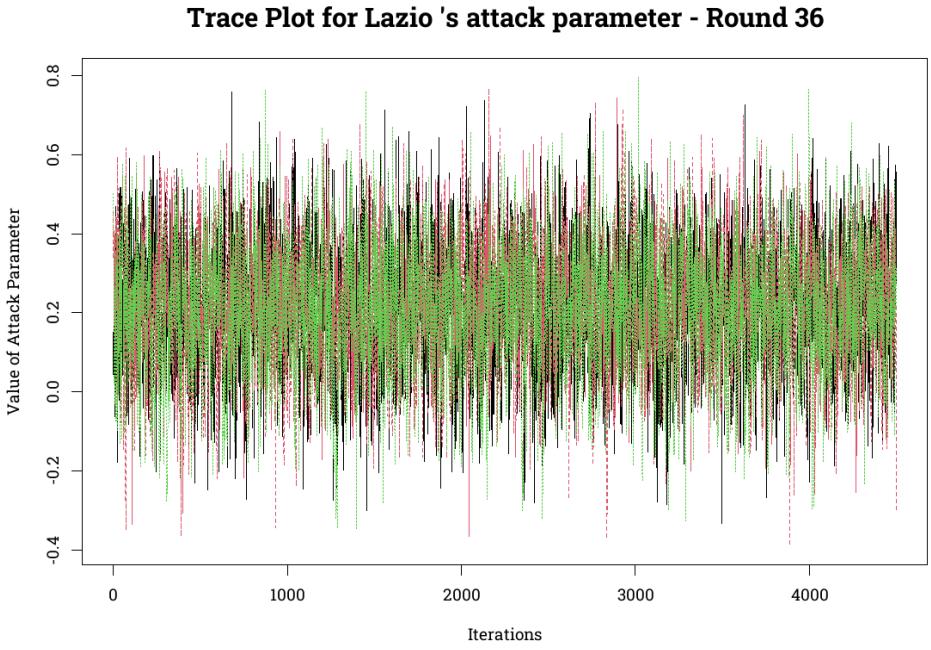


Figure 5.2: MCMC Trace Plots for round's 20 home parameter and Lazio's round's 36 attack parameter, showing the evolution of sampled values across iterations, in the 3 chains

Effective Sample Size

The effective sample size (Fig. 5.3) provides a quantifiable measure of convergence. Typically, an ESS value of at least 1000 is considered sufficient for most applications. Although the home parameter in round 20 seems to fall slightly below the typical threshold of 1000, all the other tested parameters exceed this value significantly, suggesting efficient sampling and minimal autocorrelation in the chains.

```
Effective Sample Size for Round 20
home : 933.4403
att[Bologna] : 3182.026
def[Empoli] : 11557.17
p1[Cremonese] : 6099.652
p2[Napoli] : 9218.498

Effective Sample Size for Round 29
home : 1361.317
att[Bologna] : 4727.795
def[Empoli] : 12581.07
p1[Cremonese] : 6425.433
p2[Napoli] : 10006.65

Effective Sample Size for Round 38
home : 2017.882
att[Bologna] : 4287.348
def[Empoli] : 12612.42
p1[Cremonese] : 7080.764
p2[Napoli] : 8928.681
```

Figure 5.3: Effective Sample Size for all model parameters in different rounds.

Gelman-Rubin Diagnostic

The first part of the analysis (Fig. 5.4) focuses on the home parameters for Rounds 20 and 34. For Round 20, the point estimate of the diagnostic is $\hat{R} = 1.0217$, with an upper confidence interval of 1.0736. This indicates that while the chains are close to convergence, it's almost at the limit of the ideal range, which was expected seeing the results from the ESS. By Round 34, the diagnostic improves significantly, with a point estimate of $\hat{R} = 1.0011$ and an upper confidence interval of 1.0036, reflecting excellent convergence for the home parameter.

The second section (Fig. 5.4) details the results for the $P1$ parameters in Round 20. Here, all point estimates of $\hat{R}$ are very close to 1.0, ranging from 1.0005 to 1.0098, and their corresponding upper confidence intervals fall between 1.0014 and 1.0288. These values confirm that the chains for this $P1$ parameter also converged effectively, as none of the diagnostics approach the critical threshold of 1.1.

The final part (Fig. 5.5) of the results provides a summary of the maximum upper confidence intervals, which captures the worst-case scenario for the $\hat{R}$ value. These include the home parameter (1.0736), attack parameters (1.0309), defense parameters (1.0061), and the $P1$ and $P2$ parameters (1.0399 and 1.0198, respectively). All these values are comfortably below 1.1.

```
Gelman-Rubin Diagnostic for HOME parameters:

Round 20 Results:
     Point est. Upper C.I.
[1,]   1.021731   1.073614

Round 34 Results:
     Point est. Upper C.I.
[1,]   1.001108   1.003632

Gelman-Rubin Diagnostic for P1 parameters:

Round 20 Results:
      Point est. Upper C.I.
 [1,]   1.000514   1.001428
 [2,]   1.003760   1.012713
 [3,]   1.000809   1.001522
 [4,]   1.004846   1.014119
 [5,]   1.001765   1.006083
 [6,]   1.000847   1.002387
 [7,]   1.000127   1.001030
 [8,]   1.000504   1.002354
 [9,]   1.012475   1.039870
[10,]   1.003129   1.011596
[11,]   1.001199   1.002976
[12,]   1.001563   1.003266
[13,]   1.002957   1.008929
[14,]   1.000466   1.000617
[15,]   1.010060   1.026494
[16,]   1.008560   1.024194
[17,]   1.006874   1.022395
[18,]   1.006797   1.020308
[19,]   1.003992   1.008040
[20,]   1.009755   1.028844
```

Figure 5.4: Gelman-Rubin Diagnostics for convergence assessment of home and p1 parameters. It presents point estimates and upper confidence intervals (C.I.).

```
Maximum Upper CI of Gelman-Rubin Diagnostic for Each Parameter Group:

$home
[1] 1.073614

$att
[1] 1.030907

$def
[1] 1.006091

$p1
[1] 1.03987

$p2
[1] 1.019827
```

Figure 5.5: Maximum Upper Confidence Intervals of Gelman-Rubin Diagnostics across all parameters.

Overall, these results show that the MCMC sampling works well and performs reliably across all parameters. Both the visual and quantitative diagnostics provide clear evidence that the chains have successfully converged, giving confidence in the accuracy of the results. This strong convergence is essential for ensuring that predictions are dependable, as it confirms that the posterior distributions have been thoroughly explored.

5.3 Parameter Analysis

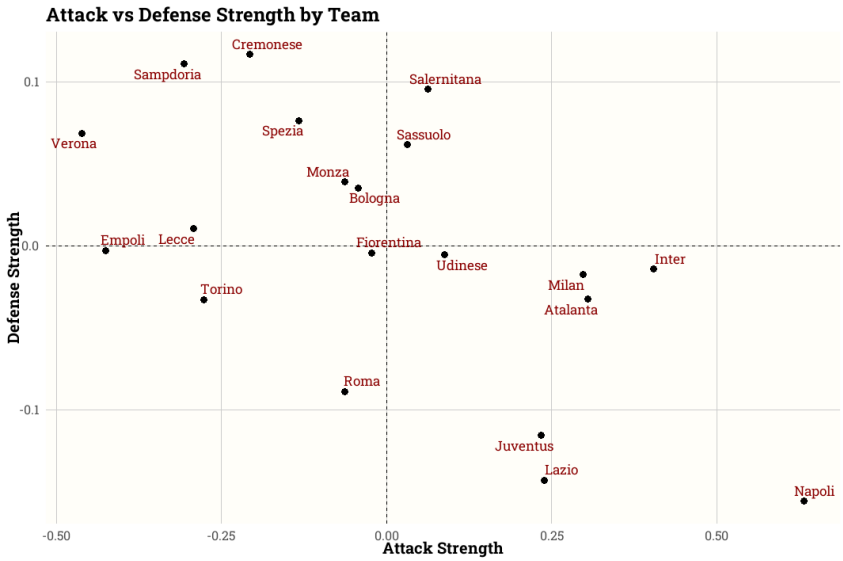


Figure 5.6: Estimated attack and defense parameter values, where higher attack strength indicate stronger offensive performance, and higher defense strength values indicate weaker defensive capabilities.

The estimated parameters provide key insights into team performance, helping to explain both the model's predictions and the observed trends in the season. Each point in Figure 5.6 and in Figure 5.7 represent the mean parameter value from all the MCMCs for each team.

Looking at the relationship between attack and defense, the results show a clear pattern. Teams with high attacking strength and strong defense, Napoli being a clear evidence, are naturally the most dominant. They not only score more goals but also concede fewer, which translates into consistently strong performances throughout the season. On the other end of the spectrum, weaker teams (in the top left corner), struggle both offensively and defensively, making them predictable candidates for relegation. Teams located in the top-right quadrant tend to have higher goal-scoring averages but are also more prone to conceding goals, whereas teams in the bottom-left typically score fewer goals while maintaining stronger defensive capabilities.

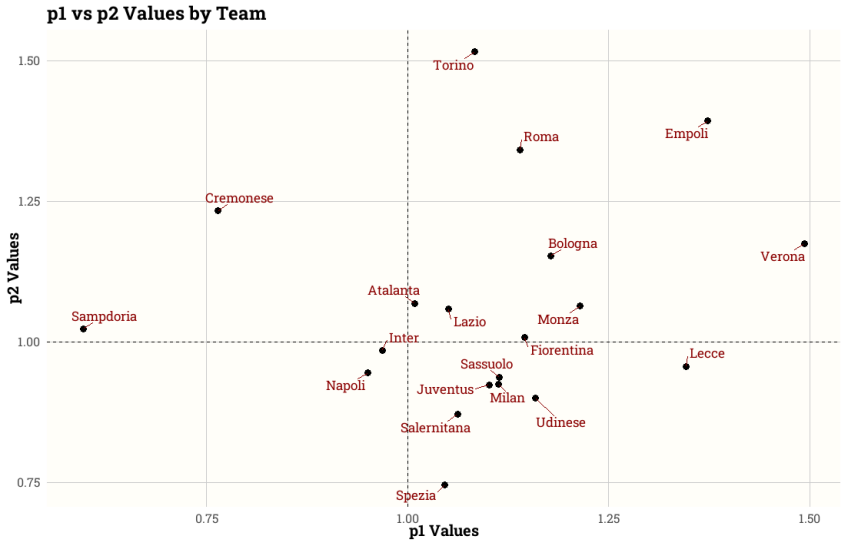


Figure 5.7: Estimated zero modification parameter values, p1 corresponds to the home team value, and p2 to the away team.

Figure 5.7 illustrates the parameter p for teams playing at home (p1) versus playing away (p2). Most teams show clear zero deflation when playing at home, meaning they fail to score less frequently than a basic Poisson model would predict. This makes sense since home advantage provides teams with an inherent advantage, hence being more prone to score at least 1 goal. However, a few teams go against this trend and actually display zero inflation at home, meaning they end up scoreless more often than expected.

When looking at away games, the situation is more balanced. 11 the teams still manage to deflate zeros, and the remaining 9 exhibit zero inflation. This suggests that while some teams maintain an attacking presence regardless of location, others are more likely to not score when playing away.

5.4 Predictive Performance Analysis

5.4.1 First Method

Brier Score and Geometric Mean

The Brier Score (Fig. 5.8) exhibits some variability across rounds, with the highest values observed in rounds 24 (0.672) and 30 (0.759), indicating less accurate predictions during these points in the season, these peaks especially in round 30 indicate rounds with surprising results, and in fact neither Napoli, Milan, Inter and Juventus managed to win their games in that specific round. The lowest values appear in rounds 31 (0.535) and 38 (0.445), reflecting the most accurate forecasts. The overall trend shows a gradual decline in the Brier Score, as indicated by the dashed trend line, suggesting improved prediction accuracy as the season progresses. The average Brier Score across all rounds is 0.616, reflecting moderate predictive performance.

Similarly, the highest Geometric Mean values (Fig. 5.8) are seen in rounds 31 (0.401) and 38 (0.456), while the lowest values are found in rounds 24 (0.328) and 30 (0.295). The dashed trend line shows a gradual increase in the Geometric Mean, signaling increasing confidence in correct predictions, reinforcing the notion that predictions become more reliable as more data is incorporated. The overall Geometric Mean for the season is 0.359, indicating moderate reliability in predictions.

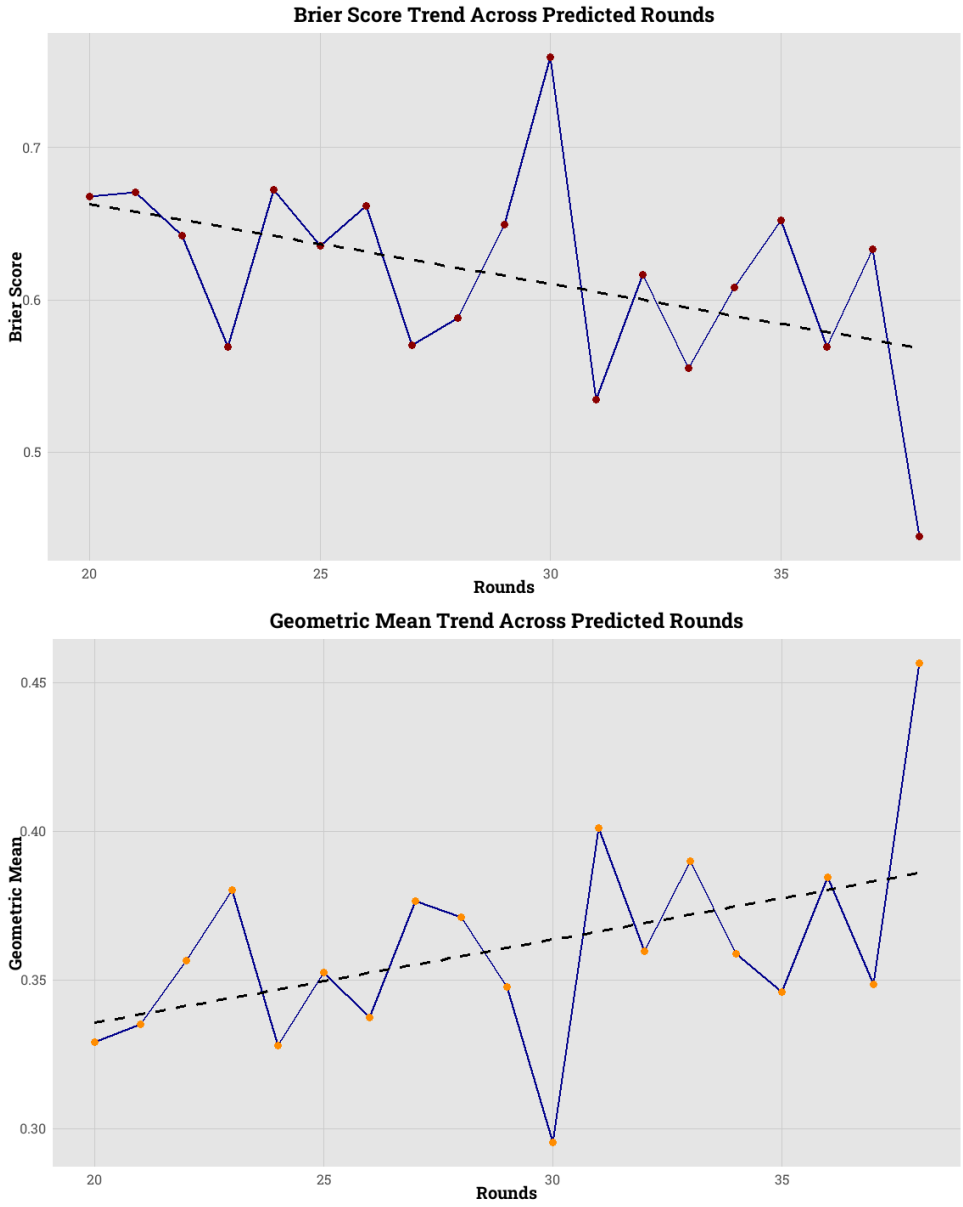


Figure 5.8: Trends of Brier Score and Geometric Mean Across Predicted Rounds. The top plot shows the variation in the Brier Score, with lower values indicating better calibration. The bottom plot displays the trend of the Geometric Mean, which evaluates predictive performance.

Boxplot Interpretation

Figure 5.9 shows the distribution of the predicted points for the second half of the season combined with the observed points from the first half, achieved by each team across 1000 simulations. Observed points are represented as red dots, and teams are ordered by their median simulated points.

Each box represents the interquartile range (IQR) of the simulated points. The whiskers extend to the most extreme simulated values within 1.5 times the IQR, and any points beyond the whiskers are considered outliers, representing rare but predicted scenarios generated by the model. The red dots provide a clear comparison to the model's predictions, showing how closely the predictions align with reality.

Together, these boxplots highlight both the central tendencies of each team's expected performance over the simulated seasons and the extent of their variability. Teams on the left side of the graph consistently display low point totals across simulations. This consistency is reflected in the narrow variability shown in the boxplots, confirming their poor performance and likelihood of relegation. On the other

hand, teams on the right show significantly higher point totals and similarly tight distributions, indicating the model's high confidence in the expected performance of both top and bottom teams in the league.

Mid-table teams exhibit wider variability in their simulated points, reflecting the increased uncertainty in predicting their performance due to closely matched abilities among teams in this range.

An interesting trend emerges when comparing the model's predicted points to the observed values. For top-ranked teams, the model tends to overpredict their points, while for lower-ranked teams, it tends to underpredict them—an outcome that seems contrary to the typical shrinkage effect.

However, as shown in Figure 5.10, the second half season goals (which is what is influenced by shrinkage) tell a different story. Teams with the lowest observed goals tend to have an overprediction of goals, whereas teams with higher observed goals generally exhibit underpredictions, with Napoli and Juventus as an exception. Additionally, Sassuolo, Fiorentina, and Bologna appear to have scored significantly more than predicted. This discrepancy may be partly due to an improved performance from these teams in the second half of the season, and since the model incorporates all previous games, the combined effect of poor past performances and shrinkage, results in a predicted goal count that is much lower than what actually occurred.

Even though there is some shrinkage in the goal predictions, it does not seem sufficient to alter the outcomes, as the top teams still secure positive results when directly compared to other. This explains the substantial disparity between the predicted and observed points.

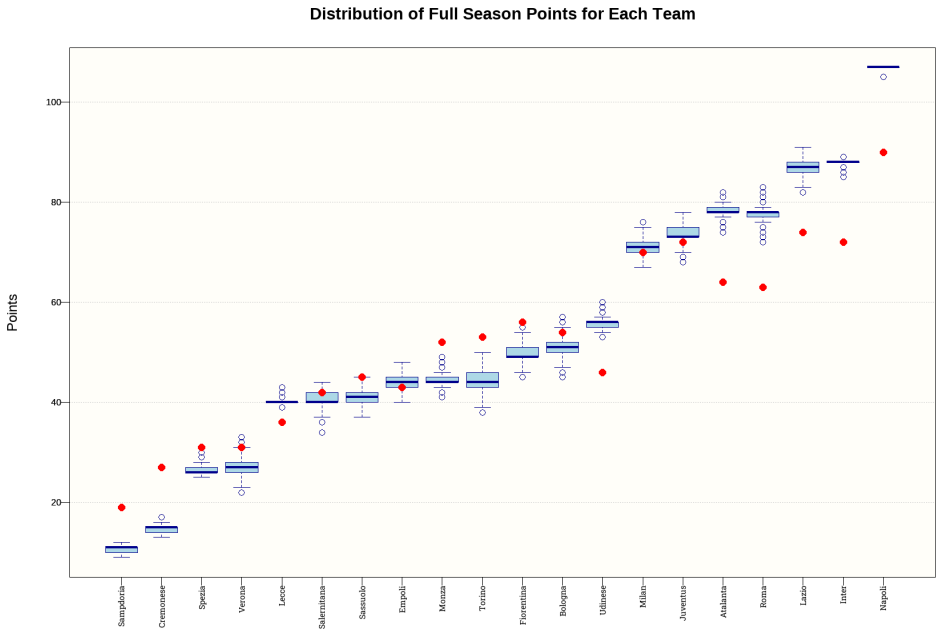


Figure 5.9: The boxplots display the spread of simulated points across the 1000 season iterations, while the red dots represent actual observed points.

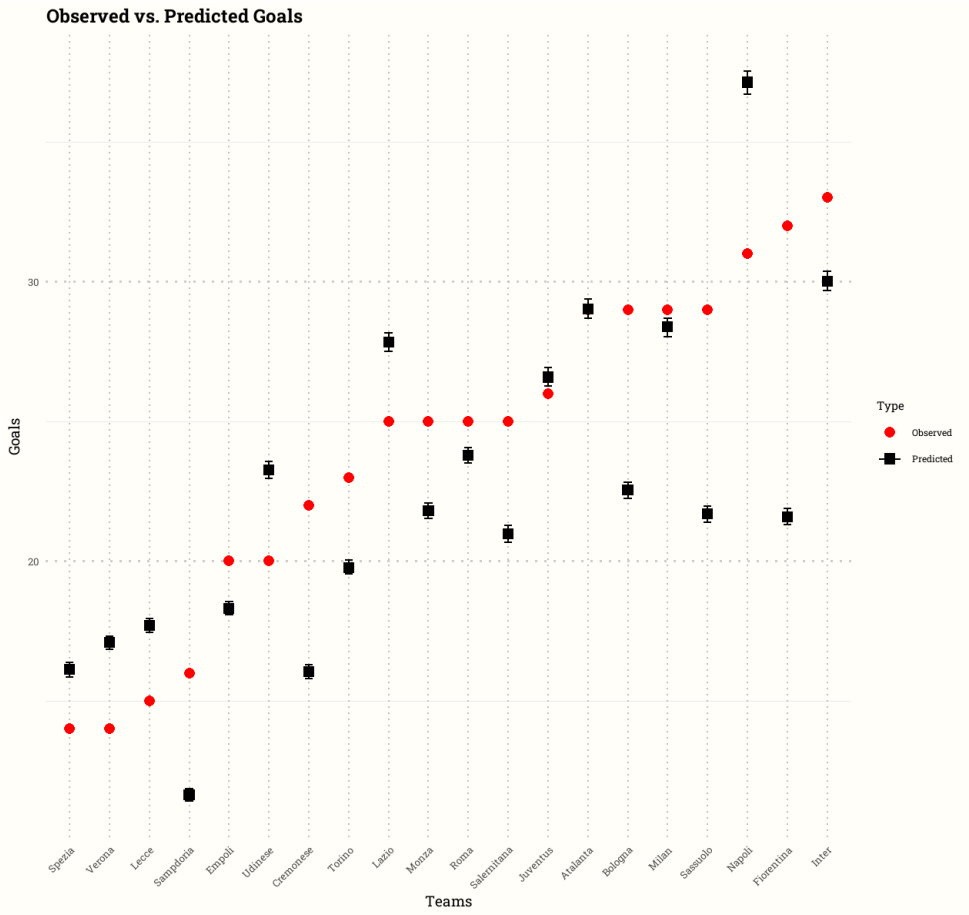


Figure 5.10: Observed vs. Predicted Goals for Each Team. The red dots represent the actual number of goals scored by each team, while the black boxplots indicate the distribution of predicted goals over the 1000 season iterations.

Final League Standings

Table 5.1 provides a comprehensive summary of the simulated final standings across 1000 runs of the model. Each row corresponds to a team, showing its observed ranking, mean predicted position and the standard deviation of that position across all simulations. The numbered columns represent the probabilities of a team finishing in each specific position by the end of the season. These rankings were calculated by combining the observed points from the first half of the season with the model's predicted points for the second half.

Napoli, for instance, is the standout performer with a mean classification of 1 and 0 standard deviation, indicating that in all 1000 simulations, they were consistently predicted to finish in first place. Inter, has a mean classification of 2.443 with a standard deviation of 0.557. This reflects that, on average, Inter is predicted to finish between second and third place, with some variability across simulations. The probabilities in the table confirm this, as Inter has significant chances of finishing in both the second and third spots, fighting for those positions with Lazio.

Further down the table, teams start to show much larger spreads in their predicted rankings with bigger standard deviations. This indicates that their final positions are less certain, with a wide range of possibilities.

At the bottom of the table, teams like Cremonese and Sampdoria have mean classifications of 19 and 20, respectively, with a standard deviation of 0.002. The table shows that only in 2 out of the 1000 runs they ended up out of the predicted positions.

A team highlighted in green is indicative that the model correctly predicted their observed position, only happening in 7 out of 20 teams, with high incidence in 5 of the 6 bottom teams.

Table 5.1: Summary of simulated final league standings over 1000 model runs. The table presents the mean final classification for each team, along with the standard deviation, indicating variability in rankings across the different runs. The columns represent the probability of each team finishing in a specific position. Green cells indicate that the observed ranking was correctly predicted by the model.

Team	Obs. Classification	Mean Classification	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Napoli	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inter	4	2.443	0.557	0	0.557	0.443	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lazio	2	2.557	0.557	0	0.443	0.557	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Atalanta	6	4.488	0.812	0	0	0	0.532	0.448	0.020	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Roma	7	4.646	0.873	0	0	0	0.458	0.447	0.086	0.009	0	0	0	0	0	0	0	0	0	0	0	0	0
Juventus	3	6.053	1.198	0	0	0	0.010	0.099	0.719	0.172	0	0	0	0	0	0	0	0	0	0	0	0	0
Milan	5	6.813	0.597	0	0	0	0	0.006	0.175	0.819	0	0	0	0	0	0	0	0	0	0	0	0	0
Udinese	12	8.009	0.034	0	0	0	0	0	0	0	0.991	0.009	0	0	0	0	0	0	0	0	0	0	0
Bologna	9	9.341	0.335	0	0	0	0	0	0	0	0.008	0.656	0.323	0.013	0	0	0	0	0	0	0	0	0
Fiorentina	8	9.731	0.367	0	0	0	0	0	0	0	0.001	0.327	0.615	0.054	0.003	0	0	0	0	0	0	0	0
Torino	10	11.929	0.878	0	0	0	0	0	0	0	0	0.007	0.05	0.402	0.201	0.262	0.050	0.022	0.006	0	0	0	0
Empoli	14	11.985	0.787	0	0	0	0	0	0	0	0	0.001	0.009	0.297	0.410	0.268	0.010	0.005	0	0	0	0	0
Monza	11	12.163	0.826	0	0	0	0	0	0	0	0	0	0.003	0.234	0.376	0.372	0.014	0.001	0	0	0	0	0
Lecce	16	14.841	0.832	0	0	0	0	0	0	0	0	0	0	0	0	0.016	0.326	0.459	0.199	0	0	0	0
Salernitana	15	14.878	0.968	0	0	0	0	0	0	0	0	0	0	0	0.007	0.050	0.347	0.250	0.346	0	0	0	0
Sassuolo	13	15.123	0.963	0	0	0	0	0	0	0	0	0	0	0	0.003	0.032	0.253	0.263	0.449	0	0	0	0
Verona	17	17.395	0.605	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.605	0.395	0	0
Spezia	18	17.605	0.605	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.395	0.605	0	0
Cremonese	19	19.002	0.002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.998	0.002
Sampdoria	20	19.998	0.002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.002	0.998

Champions League, European Spots, and Relegation

Table 5.2 provides a detailed breakdown of the probabilities of teams qualifying for the Champions League, any European competitions, and relegation across the 1000 simulations. The top four teams qualify for the Champions League, while the top seven secure places in European competitions. At the other end, the bottom three teams face relegation.

Napoli, Inter, and Lazio have a 100% probability of securing Champions League qualification in all simulations. This dominance reflects their strong performance in the first half of the season and their continued momentum in the model's predictions. Atalanta, with a 53.9% chance, emerges as a borderline contender, showcasing their potential to fight for a top-four finish. Roma follows closely at 45.1%, and finally Juventus, stands out as an underperformer, with only a 1% probability of qualification, despite their historical stature as a top team in the league.

For any of the European competition spots (positions 1 to 7), the top seven best-rated teams were guaranteed to qualify in 1000 of the 1000 simulations, with no other team having any chance of securing a spot.

At the bottom of the table, Cremonese and Sampdoria stand out with a 100% probability of relegation. This result emphasizes their poor performances across both observed and simulated games, which leave no room for recovery. Spezia, with a 62.6% chance of relegation, and Verona, at 37.4%, remain in precarious positions.

In reality, the four teams with the most points were Napoli, Lazio, Inter, and Juventus, that contrary to what the model predicted managed to save a spot in the champions league. Completing the top seven as the model predicted were Milan, Atalanta and Roma. The three relegated teams were also correctly predicted being Sampdoria, Cremonese, and Spezia.

Table 5.2: Probabilities of teams qualifying for the champions league, european competitions, and relegation, based on 1000 model runs.

Team	Champions League	Team	European Competitions	Team	Relegation
Napoli	1.000	Napoli	1.000	Cremonese	1.000
Inter	1.000	Inter	1.000	Sampdoria	1.000
Lazio	1.000	Lazio	1.000	Spezia	0.626
Atalanta	0.539	Atalanta	1.000	Verona	0.374
Roma	0.451	Roma	1.000	---	---
Juventus	0.010	Juventus	1.000	---	---
---	---	Milan	1.000	---	---

Comparison with Other Poisson Models

To assess how well the ZMP model performs, it was compared to three alternative models: the Standard Poisson, the ZIP, and the ZAP. This comparison helps understand how different models handle key aspects of football data, particularly in dealing with the variability of zeros in football. The goal is to evaluate predictive accuracy and how well each model reflects real-world match results.

As previously discussed, each of these models has its own characteristics and assumptions regarding goal distribution. The question is whether the ZMP model's ability to handle both zero inflation and zero deflation provides a meaningful advantage over existing approaches.

Results and Insights

Table 5.3 compares the percentage of accurate predictions and the Brier Score of the 4 models. These metrics provide a look at how well each predicts football match outcomes and how calibrated their probability estimates are.

The ZAP achieves the highest percentage of accurate predictions, correctly predicting 51.33% of the matches, outperforming the other models in terms of raw accuracy. However, it has the highest Brier Score at 0.623, indicating that its probabilistic predictions are less reliable and overconfident compared to the others. The Standard Poisson model, despite being less accurate with a prediction rate of 48.80%, demonstrates the best Brier Score at 0.614. This low score means that the Standard Poisson model provides the most well-calibrated probabilistic predictions, ensuring that its confidence levels align closely with actual match outcomes. The ZMP model falls between these extremes, with an accuracy of 49.47% and a Brier Score of 0.616, showing that while it benefits from its flexibility to account for zero inflation and deflation, it does not significantly outperform the other models. The ZIP, designed to address zero inflation, has the lowest accuracy at 48.27% but performs better than the ZAP in terms of the Brier Score.

Interestingly it seems that the models with the highest accurate predictions don't have the best Brier scores, since this metric evaluates not only whether predictions are correct but also the quality of the probability estimates, penalizing overconfidence in incorrect predictions, this suggests that they often assign high probabilities to outcomes that do not align with the observed results while the the Poisson is more conservative in their estimates.

Table 5.3: Comparison of Zero-Modified Poisson, Zero-Altered Poisson, Zero-Inflated Poisson, and Standard Poisson Models across different metrics. The table presents the percentage of accurate predictions and the Brier Score for each model, highlighting their predictive performance.

Metric	ZMP	ZAP	ZIP	Standard Poisson
Accurate Predictions (%)	49.47	51.33	48.27	48.80
Brier Score	0.616	0.623	0.619	0.614

Table 5.4 extends the analysis by introducing one additional metric, the F1-Score.

F1-Score is a measure of a model's ability to correctly classify match outcomes while balancing precision and recall. Precision answers the question: "When the model predicts for example, an home win, how often is it correct?" On the other hand, recall, answers: "Out of all the actual home wins, how many did the model correctly identify?" This balance is particularly relevant in the context of football match predictions, where certain outcomes, such as draws, occur less frequently and may be harder to predict accurately. By incorporating both precision and recall, F1-Score ensures that models are not only evaluated based on how often they make correct predictions but also on how effectively they capture the characteristics of each match outcome.

Table 5.4: Comparison of F1-Scores for Zero-Modified Poisson, Zero-Altered Poisson, Zero-Inflated Poisson, which evaluates the balance between precision and recall across models.

Metric	Class	ZMP	ZAP	ZIP	Standard Poisson
F1-Score	Home Win	0.605	0.627	0.613	0.610
	Draw	0.286	0.286	0.216	0.216
	Away Win	0.487	0.472	0.432	0.491
F1-Score Average	---	0.459	0.462	0.386	0.439

For Home Wins, all models perform reasonably well, with F1-Scores above 0.6, reflecting their ability to capture this outcome with some efficiency. In contrast, as it was expected draws present a significant challenge for all models. Both the ZMP and the ZAP achieve an F1-Score of 0.286, with the standard Poisson performing slightly worse, while the ZIP display a noticeable difference at 0.114. Away Wins show a more moderate performance, with Standard Poisson achieving the highest F1-Score at 0.491, closely followed by the ZMP at 0.487. The ZAP and the ZIP perform slightly worse, with scores of 0.472 and 0.432, respectively, suggesting that these models have a harder time capturing the less frequent away win outcomes.

The averaged F1-Score provides a summary of each model's overall classification performance across all three classes. The ZAP achieves the highest average at 0.462, closely followed by the ZMP at 0.459. The Standard Poisson, while simpler in design, performs respectably with a score of 0.439, demonstrating its robustness despite its simplicity. The ZIP, on the other hand, lags behind with a average of 0.386, largely due to its poor performance in predicting draws.

The stronger F1-Scores for Home Wins can be attributed to the inherent home advantage in football, where home teams win more often. This makes Home Wins easier to predict since they follow a more consistent pattern, resulting in higher precision and recall across all models.

Predicted Zeros

All three models aim to adjust the probability of zero counts, so it's going to be assessed how well they capture the proportion of matches where a team failed to score, compared to the actual proportions observed in the second half of the season (Fig. 5.5).

The most noticeable difference is that while the Poisson, ZMP, and ZAP models produce similar zero proportions, the ZIP model stands out with a significantly higher proportion of zeros, confirming its expected bias toward overestimating zero outcomes. In terms of averages, the ZAP model is the closest to the observed one, with a proportion of zeros of 0.308, though the difference from the ZMP and the standard Poisson is not substantial.

The observed proportions exhibit greater variability across teams than any of the models. This occurs because of shrinkage, the values tend to be drawn toward the global mean, smoothing out team-specific variations.

An interesting finding is that, although the ZAP model is designed to handle zero inflation by increasing the probability of zeros compared to the Poisson model through the parameter p, it exhibits an unexpected behavior in this case. On average, the observed zero counts are lower than those predicted by the Poisson model, and most teams appear to display *zero deflation* rather than inflation. As a result, the ZAP model ends up predicting a lower proportion of zeros than the Poisson model, reflecting the data rather than forcing zero inflation. This suggests that the ZAP model might also have some capacity to adjust to zero-deflated data. When zeros are underrepresented, the parameter p decreases, reducing the probability of zero outcomes and reallocating probability mass to non-zero counts.

Unlike the ZIP model, which assumes a distinct structural process for excess zeros, the ZAP model modifies probabilities directly, treating all zeros uniformly. While not explicitly designed for zero deflation, this structure appears to allow the ZAP model to approximate zero-deflated patterns in certain datasets.

Table 5.5: Comparison of observed and model predicted proportion of zero-goal matches. The table presents the observed proportion of matches with zero goals for each team and compares them with estimates from the Zero-Modified Poisson, Zero-Altered Poisson, Zero-Inflated Poisson, and Standard Poisson models. The last row provides the average proportion across all teams.

Team	Obs. Porportion	ZMP	ZAP	ZIP	Poisson
Spezia	0.526	0.469	0.474	0.520	0.412
Sampdoria	0.474	0.580	0.580	0.629	0.491
Lecce	0.421	0.369	0.354	0.467	0.394
Udinese	0.421	0.291	0.267	0.379	0.294
Atalanta	0.368	0.213	0.205	0.314	0.227
Empoli	0.368	0.321	0.325	0.455	0.390
Verona	0.368	0.366	0.367	0.486	0.416
Cremonese	0.263	0.451	0.449	0.532	0.412
Inter	0.263	0.228	0.214	0.313	0.221
Juventus	0.263	0.259	0.239	0.348	0.258
Salernitana	0.263	0.361	0.363	0.432	0.330
Bologna	0.211	0.269	0.263	0.392	0.317
Lazio	0.211	0.223	0.220	0.330	0.243
Monza	0.211	0.294	0.292	0.407	0.326
Napoli	0.211	0.183	0.177	0.249	0.152
Torino	0.211	0.304	0.292	0.440	0.3544
Milan	0.158	0.229	0.200	0.315	0.231
Roma	0.158	0.229	0.224	0.370	0.293
Sassuolo	0.158	0.324	0.328	0.410	0.321
Fiorentina	0.105	0.316	0.329	0.414	0.326
Average	0.282	0.314	0.308	0.410	0.320

5.4.2 Second Method

Correct Predictions

This line chart (Fig. 5.11) depicts the progression of the percentage of correct predictions made by the model for each MCMC.

Initially, the correct prediction percentage remains relatively stable and moderate, with fluctuations around 50%.

Towards the latter rounds, a significant increase in accuracy is evident. This surge aligns with the low number of games remaining to predict and the substantial amount of data that the model can now utilize. By the final rounds, the model's predictions become much more reliable, achieving a peak in

correct predictions.

The average accurate predictions is 51.81%

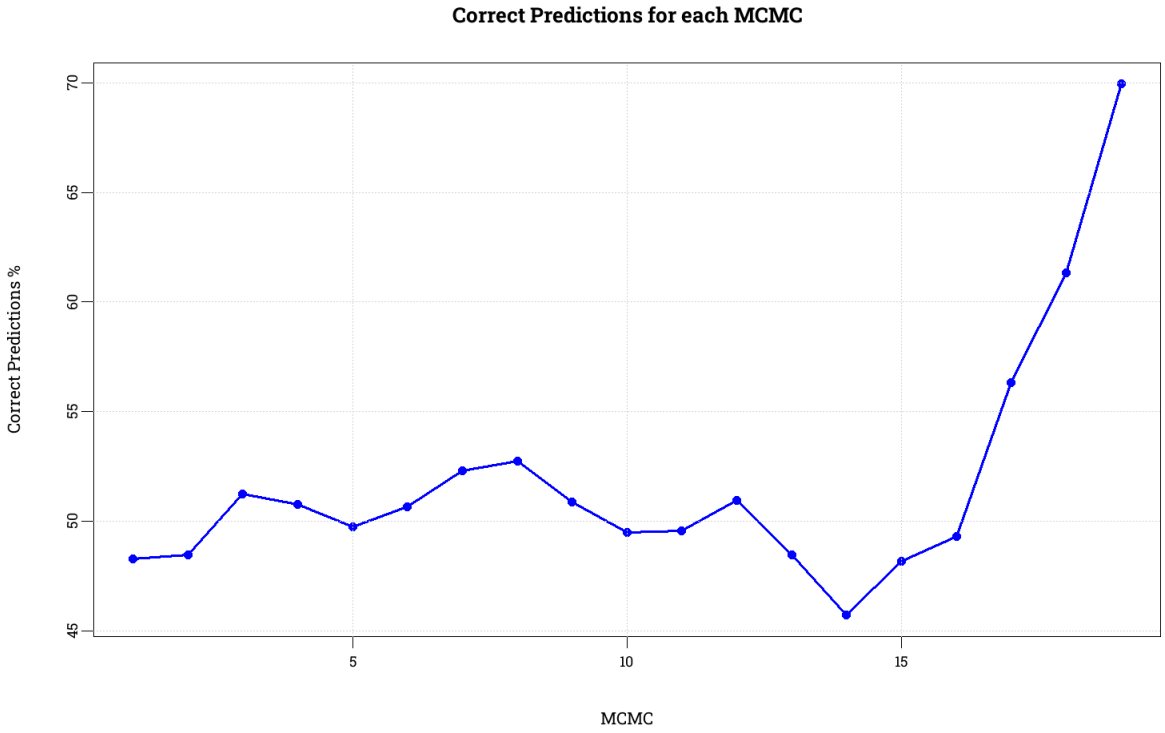


Figure 5.11: Percentage of correct predictions across the 19 MCMC iterations.

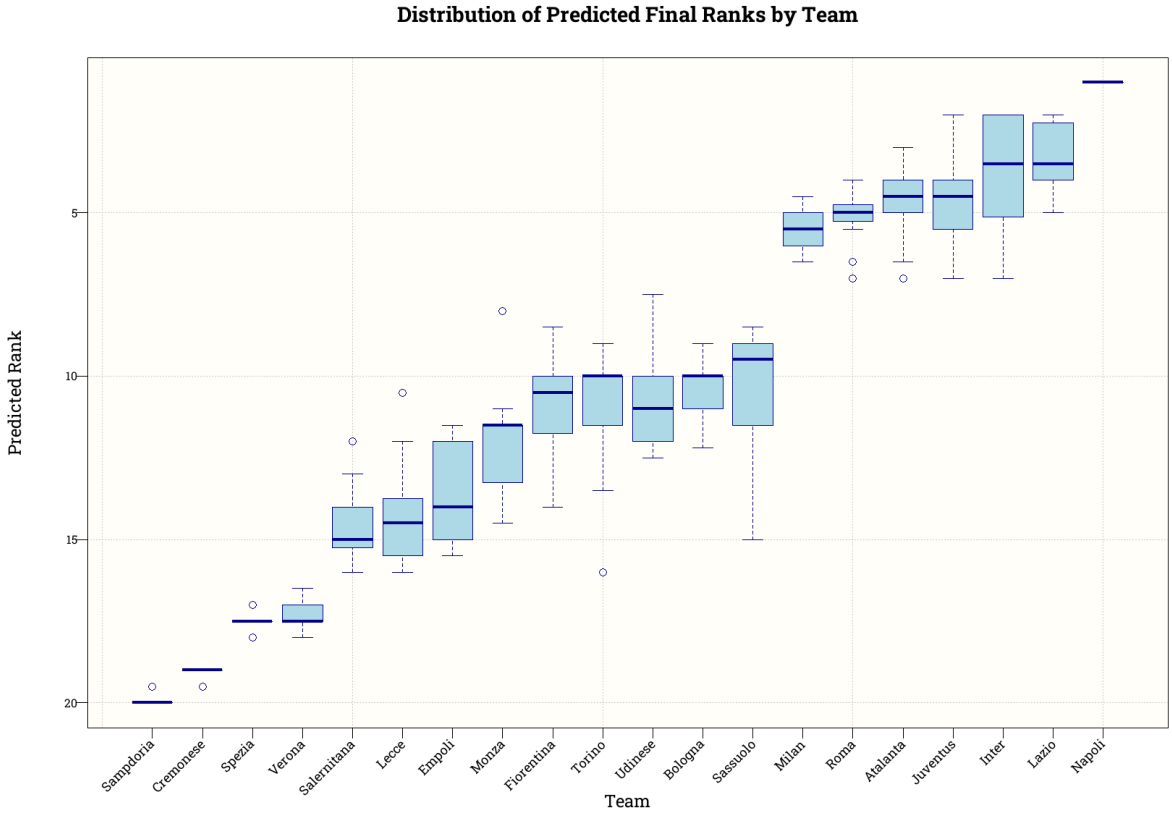


Figure 5.12: Distribution of predicted final rankings for each team. The boxplots represent the variability in the 19 MCMC runs.

In Figure 5.12, each boxplot illustrates the range of mean rankings achieved by a team across all MCMC runs. Each one of them is built with 19 values, one for each MCMC, that value represents the average ranking of the team in that specific run.

Most teams display some degree of variability in their predicted rankings, with the exception of the bottom four teams and Napoli. These teams remain relatively stable in their positions, this stability indicates consistent performance throughout the season. Their point differences from the rest of the clubs were so significant that there was little room for variability in their rankings, leaving no margin for changes regardless of additional observations.

The rest of the teams can clearly be divided into two groups, with Sassuolo and Milan being the turning point. Within these groups, there was a moderate level of variability, highlighting the challenge of predicting standings for teams that are closely matched in quality. Teams with narrower boxplots stabilize in their rankings faster than those with greater variability.

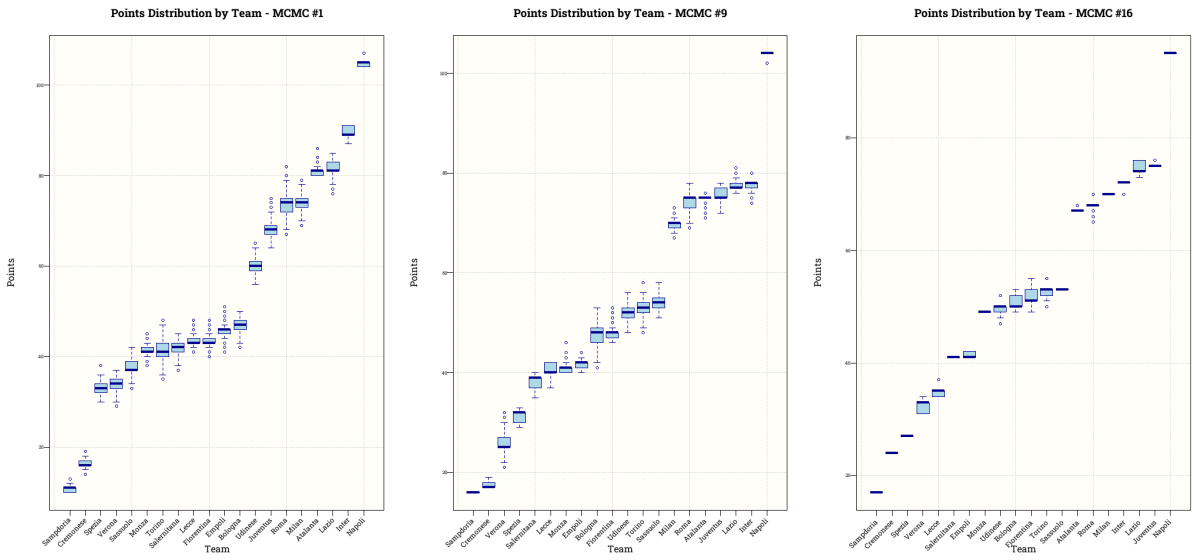


Figure 5.13: The boxplots show the evolution of the predicted distribution of season points for each team at three different MCMC iterations.

The three panels in Figure 5.13 illustrate the points distribution by team at three distinct MCMC runs: 1, 9 and 16. A clear trend emerges, as the more data is added to the model and the less games are left to be played, the trend is that the variability in team points decreases, although nothing seems to change from MCMC 1 to 9, it's clear in MCMC 16. This is because the number of remaining games is limited, and the points still up for grabs are minimal, leaving little room for variability.

By MCMC 16, there's only 3 games left and so the point variability is basically none.

Heatmaps

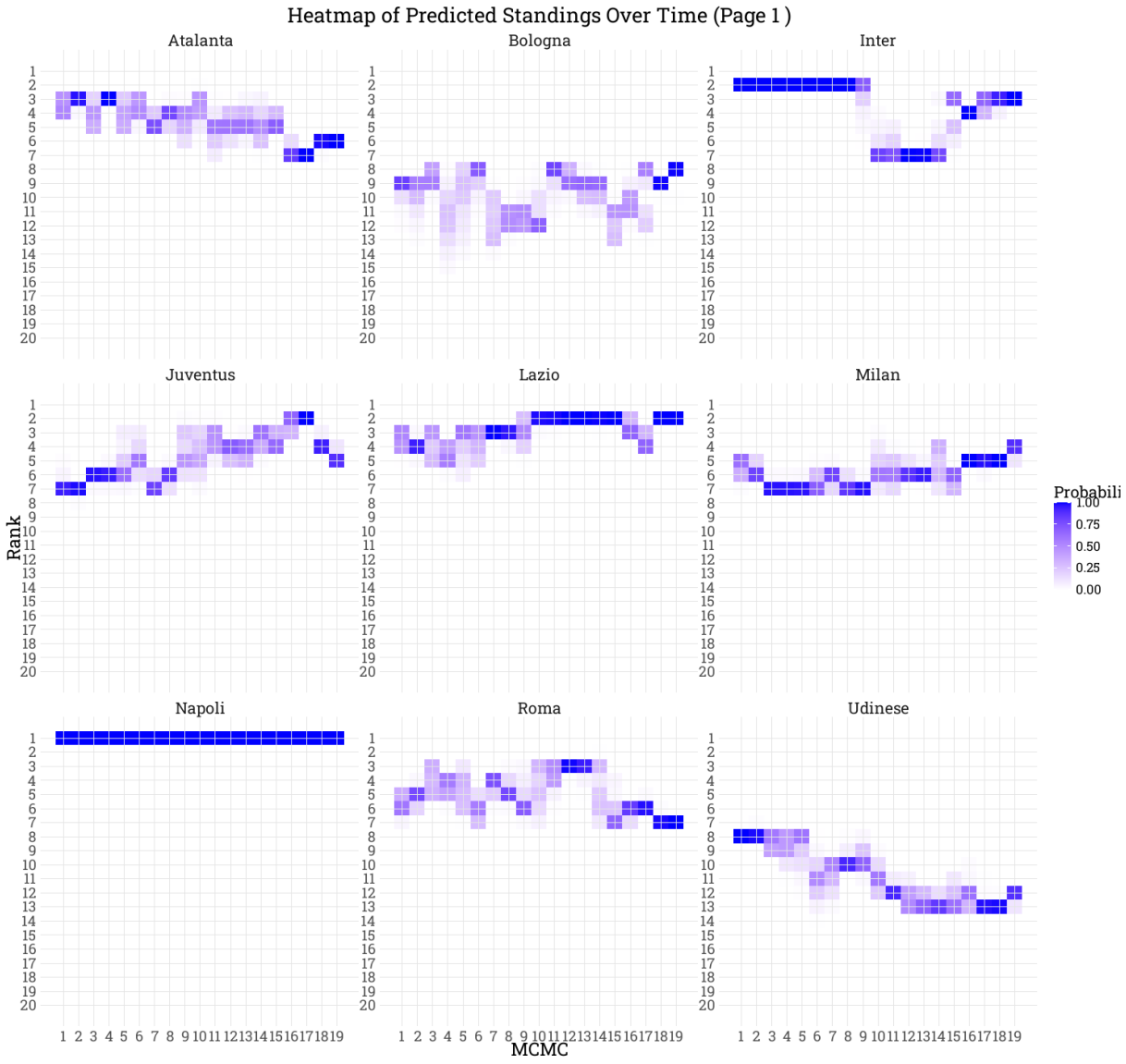


Figure 5.14: Each subplot represents the heatmap of the evolving probability of a team's predicted rank throughout the various MCMCs. Darker shades indicate higher probabilities of a team finishing in a particular rank.

The heatmap analysis (Fig. 5.14) illustrates the dynamic evolution of the predicted final rankings across the MCMCs. Each cell's intensity reflects the proportion of times a team occupied a specific rank during the 1000 simulations, providing a visual representation of the predictions. Interestingly in the first MCMCs the predicted rankings don't seem to change that much, appearing stabilized, it's only after a couple of MCMCs that the predictions exhibit considerable uncertainty, with probabilities distributed across multiple ranks, highlighting the difficulty in precisely pinpointing the standings.

As the runs progress, the predictions stabilize, confirming the insights derived from the boxplots.

Chapter 6

Conclusion

In the introduction, It was traced the evolution of statistical approaches in football prediction, moving from Maher's early Poisson-based methods to more recent Bayesian formulations. This progression moved in the direction of models that could capture the dynamic and unpredictable nature of football. And the Zero-Modified Poisson model, with the zero-modification parameter p, it's just a symptom of that. It's nuanced handling of zero counts should fit like a glove in a sport like football.

The results show a more sobering scenario. The Zero-Modified Poisson, achieved a very modest performance, with an unimpressive Brier Score and Geometric Mean, correctly predicting just 49.47% of the game outcomes. Although this is better than a blind guess (33%), informed football predictions are not blind, and a person with some football knowledge would perform if not the same, better than the model. The Zero-Modified Poisson may help highlight patterns that might otherwise be overlooked by a human, but on it's own doesn't have a satisfactory predictive power.

When compared with similar models, despite the handling of zero inflation and deflation, the ZMP performance is similar to that of the other 3 tested models, except in predicting zeros and draws, where there was clearly an underperformance of the Zero-Inflated Poisson model.

All models showed a big disparity in the correct prediction of the 3 outcomes, being much higher for home wins than for draws or away wins. This can be explained from the dominance of home wins in the dataset, home teams benefit from familiar environments, crowd support, and reduced travel fatigue, which naturally leads to that. Consequently, the model, trained on this dataset, is more inclined to predict home wins more often than other outcomes. That combined with the inclusion of a home advantage parameter amplifies this tendency.

As the model predicts home wins more frequently, and since home wins are indeed the most observed outcome, its predictive accuracy for home wins is consequently higher, this cyclical effect, affects the predictions of the other two outcomes.

Regarding the simulation methods, the round-by-round approach offers significant advantages by updating the model as new data become available. It captures the evolving nature of a football season so it's the one that makes more sense and overall the most useful, despite in this case having slightly lower accuracy compared to the second method. The full-season simulation provides the forecasting of the entire remainder of the season from any given point, it's ideal for strategic planning, as it answers the question of how current performance might translate into final standings, but it's less useful for predicting specific results.

It is worth considering that the dataset used, the 2022/23 season of the Italian Serie A, may not have been the optimal choice. Exploring results using data from other leagues could show more interesting results.

Future work should address the challenges of predicting draws and away wins (especially draws), and in general seek to improve the overall predictive power of the model. This, could be achieved by incorporating additional factors such as injuries, player transfers, recent form, or even weather conditions. With the ever growing computational power, prediction models can only improve, whether being the one used in this work, or not.

On a more personal note. Every day brings new and innovative techniques, yet no model will ever predict every outcome with absolute certainty, and that's good, it's the reason people love football, and why statisticians will always strive to create better models, and against all odds, try to predict the unpredictable.

Although the predictions were far from perfect, it might seem trivial to some, but It was really enjoyable for me playing around with this numbers. I was able to quantify something that is often seen as a subjective matter, the ”quality” of a team. Quantify it, visualize it, and draw conclusions from it. Predictive models not only deepen the understanding of the world around us, but can also provide a fun perspective of what could have been, what is, and what might be.

References

Aki, V., Andrew, G., Daniel, S., Bob, C., and Paul-Christian, B. (2020). Rank-normalization, folding, and localization: An improved rhat for assessing convergence of mcmc. *Bayesian Analysis*, 6:16.

Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37(2):253–264.

Bertoli, W., Conceição, K. S., Andrade, M. G., and Louzada, F. (2019). Bayesian approach for the zero-modified poisson–lindley regression model.

Carlo, C. (2004). Markov chain monte carlo and gibbs sampling. lect. *Notes EEB*, 581(3).

Conceição, K. S., Andrade, M. G., and Louzada, F. (2014). On the zero-modified poisson model: Bayesian analysis and posterior divergence measure. *Computational Statistics*, 29:959–980.

Conceição, K. S., Suzuki, A. K., Andrade, M. G., and Louzada, F. (2017). A bayesian approach for a zero modified poisson model to predict match outcomes applied to the 2012–13 la liga season. *Brazilian Journal of Probability and Statistics*, pages 746–764.

Egidi, L., Pauli, F., and Torelli, N. (2018). Combining historical data and bookmakers' odds in modelling football scores. *Statistical Modelling*, 18(5-6):436–459.

Egidi, L. and Torelli, N. (2021). Comparing goal-based and result-based approaches in modelling football outcomes. *Social Indicators Research*, 156(2):801–813.

Friendly, M. and Meyer, D. (2015). *Discrete data analysis with R: visualization and modeling techniques for categorical and count data*. CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Geyer, C. J. (2011). Introduction to markov chain monte carlo. *Handbook of markov chain monte carlo*, 20116022(45):22.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*, volume 715. John Wiley & Sons.

Kurtsmajlaj, V. (2022). Dissecting poisson based prediction models in association football: A comprehensive look at methodology, assumptions, and accuracy using data from the main european leagues (2011–2022). Master's thesis.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

Loeys, T., Moerkerke, B., De Smet, O., and Buysse, A. (2012). The analysis of zero-inflated count data: Beyond zero-inflated poisson regression. *British Journal of Mathematical and Statistical Psychology*, 65(1):163–180.

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.

Mouatassim, Y. and Ezzahid, E. H. (2012). Poisson regression and zero-inflated poisson regression: application to private health insurance data. *European actuarial journal*, 2(2):187–204.

Piegorsch, W. W. et al. (2002). Encyclopedia of environmetrics. *Chichester: John*.

Reid, N. (2010). Likelihood inference. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5):517–525.

Roy, V. (2020). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, 7(1):387–412.

Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):399–418.

Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., and Van Aken, M. A. (2014). A gentle introduction to bayesian analysis: Applications to developmental research. *Child development*, 85(3):842–860.

Veenman, M., Stefan, A. M., and Haaf, J. M. (2024). Bayesian hierarchical modeling: An introduction and reassessment. *Behavior Research Methods*, 56(5):4600–4631.

Zerozero (2024). Zerozero.pt: Football statistics and results. `https://www.zerozero.pt/`.

Zhao, J., Zhang, F., Zhao, C., Wu, G., Wang, H., and Cao, X. (2020). The properties and application of poisson distribution. In *Journal of physics: conference series*, volume 1550, page 032109. IOP Publishing.

Appendices

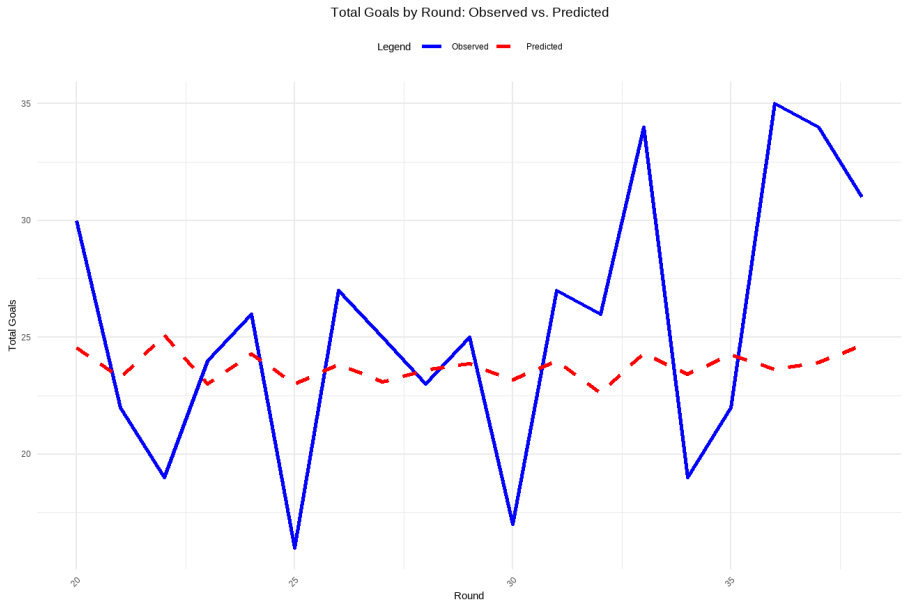


Figure 6.1: First Method: Number of goals observed vs predicted in each round

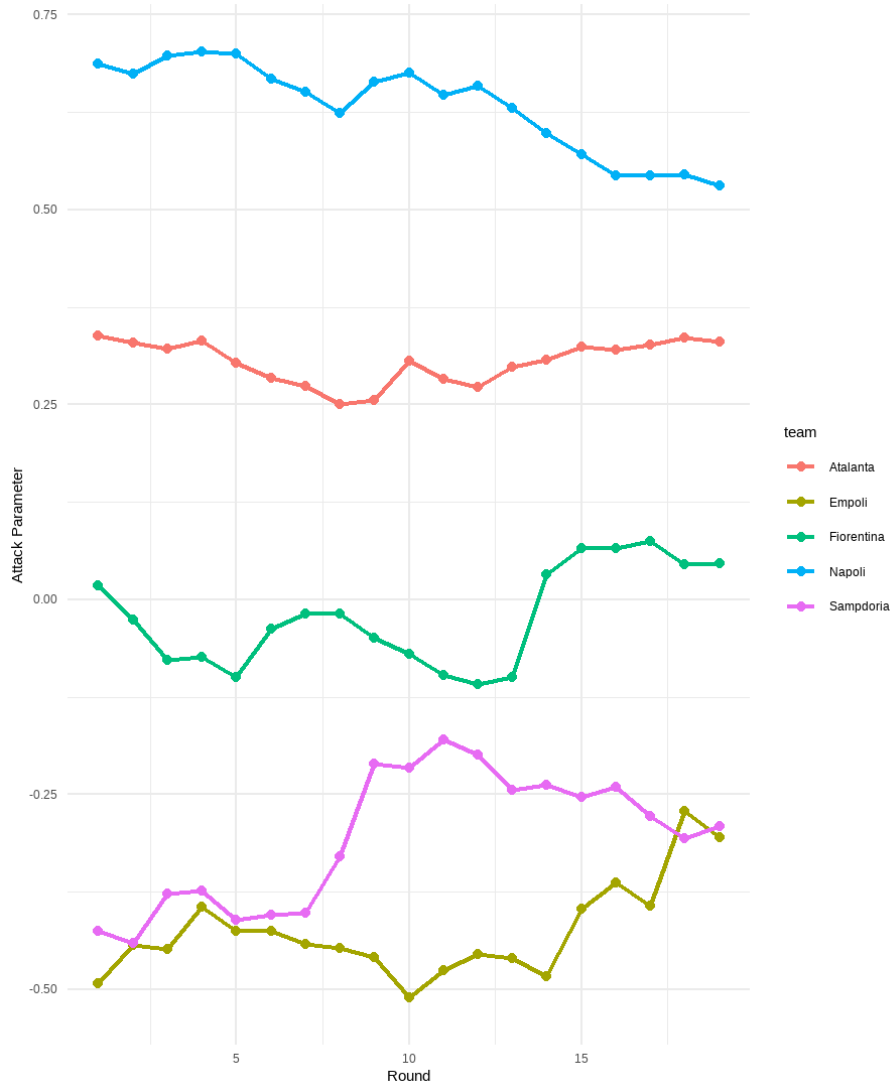


Figure 6.2: First Method: Variation of the attack parameter through rounds

```
--- Round 16 ---
                   mean    sd   2.5%    50%  97.5% Rhat n.eff
home              0.139 0.071 -0.003  0.140  0.276    1  1762
att[Atalanta]     0.320 0.166 -0.007  0.322  0.640    1  5188
att[Bologna]     -0.089 0.189 -0.476 -0.083  0.272    1  3454
att[Cremonese]   -0.191 0.215 -0.640 -0.182  0.208    1  5392
att[Empoli]      -0.363 0.228 -0.848 -0.351  0.050    1  3769
att[Fiorentina]   0.065 0.178 -0.288  0.066  0.410    1  4534
att[Inter]        0.476 0.158  0.158  0.478  0.775    1  5966
att[Juventus]     0.224 0.170 -0.117  0.227  0.552    1  4847
att[Lazio]        0.211 0.171 -0.128  0.213  0.541    1  5225
att[Lecce]       -0.330 0.236 -0.832 -0.317  0.102    1  3956
att[Milan]        0.228 0.167 -0.111  0.231  0.543    1  3679
att[Monza]       -0.072 0.185 -0.447 -0.067  0.278    1  4370
att[Napoli]       0.544 0.149  0.248  0.547  0.831    1  5527
att[Roma]        -0.074 0.185 -0.451 -0.069  0.278    1  2644
att[Salernitana]  0.056 0.186 -0.317  0.059  0.409    1  5566
att[Sampdoria]   -0.241 0.247 -0.761 -0.227  0.205    1  6178
att[Sassuolo]     0.013 0.184 -0.352  0.017  0.368    1  4705
att[Spezia]      -0.142 0.215 -0.592 -0.135  0.266    1  6584
att[Torino]      -0.300 0.213 -0.749 -0.291  0.093    1  1963
att[Udinese]      0.117 0.180 -0.241  0.120  0.469    1  3799
att[Verona]      -0.454 0.246 -0.969 -0.436 -0.010    1  3183
def[Atalanta]    -0.009 0.093 -0.202 -0.006  0.164    1 12752
def[Bologna]      0.053 0.091 -0.134  0.056  0.221    1 11397
def[Cremonese]    0.138 0.074 -0.016  0.142  0.270    1  8313
def[Empoli]       0.008 0.091 -0.175  0.010  0.180    1 13047
def[Fiorentina]  -0.003 0.096 -0.204  0.001  0.173    1 12890
def[Inter]       -0.066 0.099 -0.273 -0.062  0.120    1 11767
def[Juventus]    -0.142 0.111 -0.379 -0.135  0.052    1  9494
def[Lazio]       -0.197 0.116 -0.445 -0.188  0.004    1  8486
def[Lecce]        0.026 0.096 -0.174  0.029  0.198    1 12886
def[Milan]       -0.057 0.094 -0.256 -0.054  0.120    1 12097
def[Monza]        0.051 0.089 -0.132  0.054  0.217    1 10958
def[Napoli]      -0.217 0.122 -0.481 -0.208 -0.009    1  7745
def[Roma]        -0.094 0.105 -0.319 -0.088  0.097    1 10003
def[Salernitana]  0.115 0.078 -0.049  0.119  0.257    1  9364
def[Sampdoria]    0.150 0.070  0.002  0.155  0.278    1  7586
def[Sassuolo]     0.079 0.085 -0.094  0.083  0.232    1 10551
def[Spezia]       0.114 0.079 -0.051  0.119  0.254    1  9969
def[Torino]      -0.032 0.097 -0.232 -0.029  0.150    1 12625
def[Udinese]     -0.003 0.094 -0.198  0.000  0.172    1 13150
def[Verona]       0.085 0.085 -0.089  0.089  0.239    1 10596
```

Figure 6.3: First Method: Summary of the attack, defense and home parameters in round 35

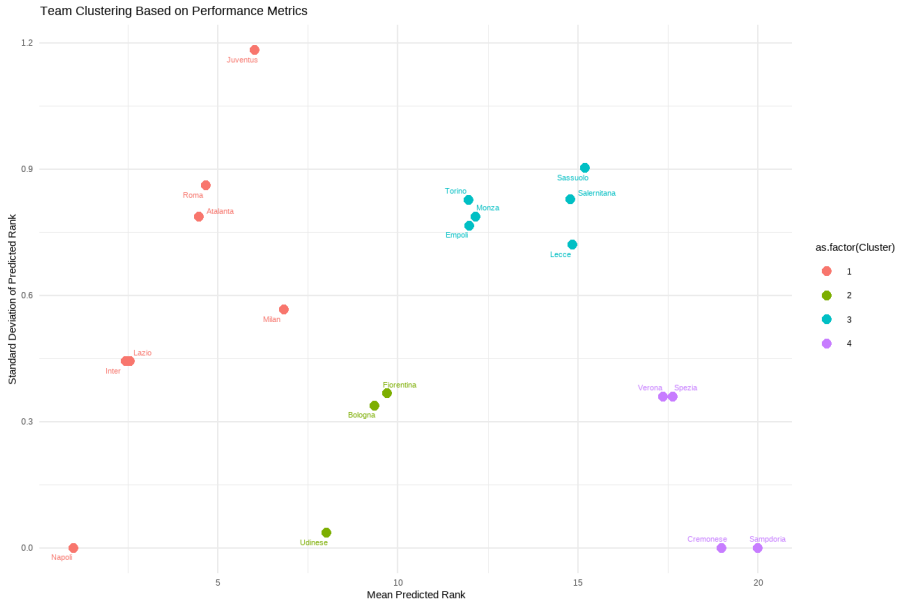


Figure 6.4: First Method: Teams divided in 4 clusters, constructed using mean predicted points, goals and classification

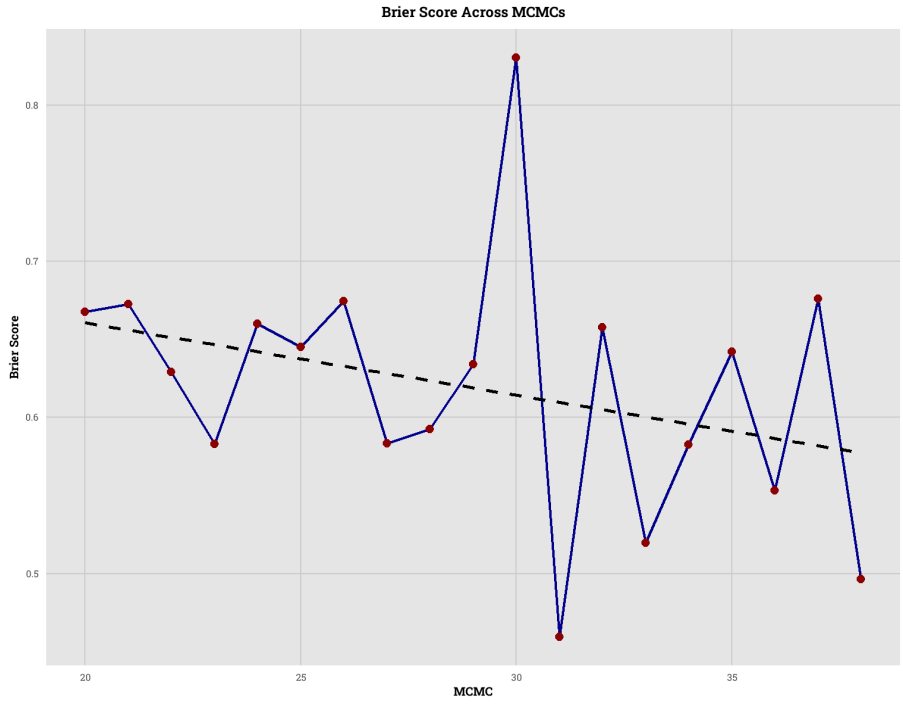


Figure 6.5: Second method: Brier Score

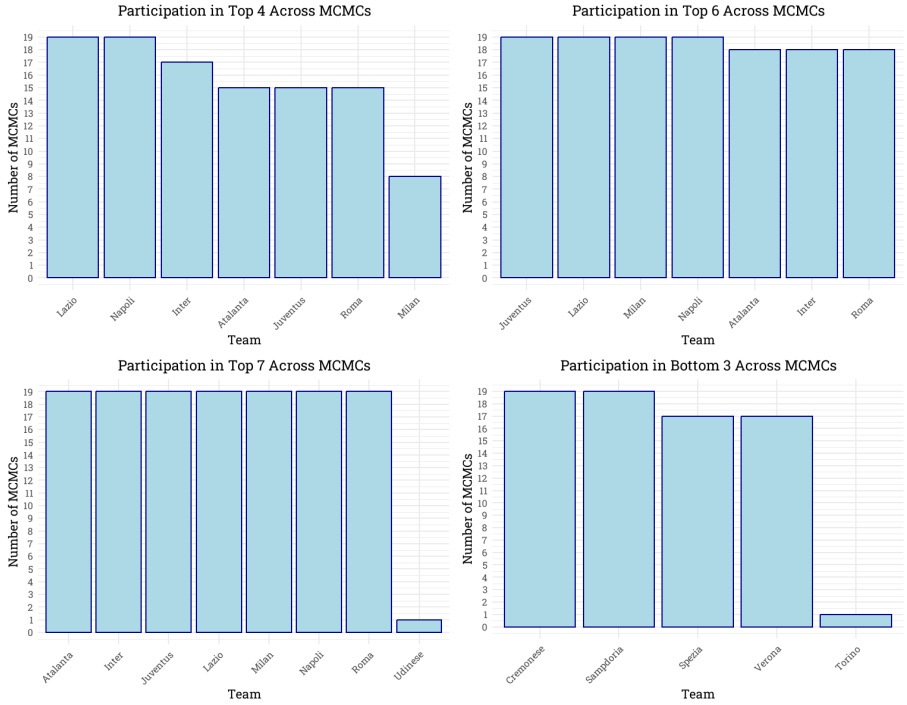


Figure 6.6: Second method: Number of times a team appear in a certain position

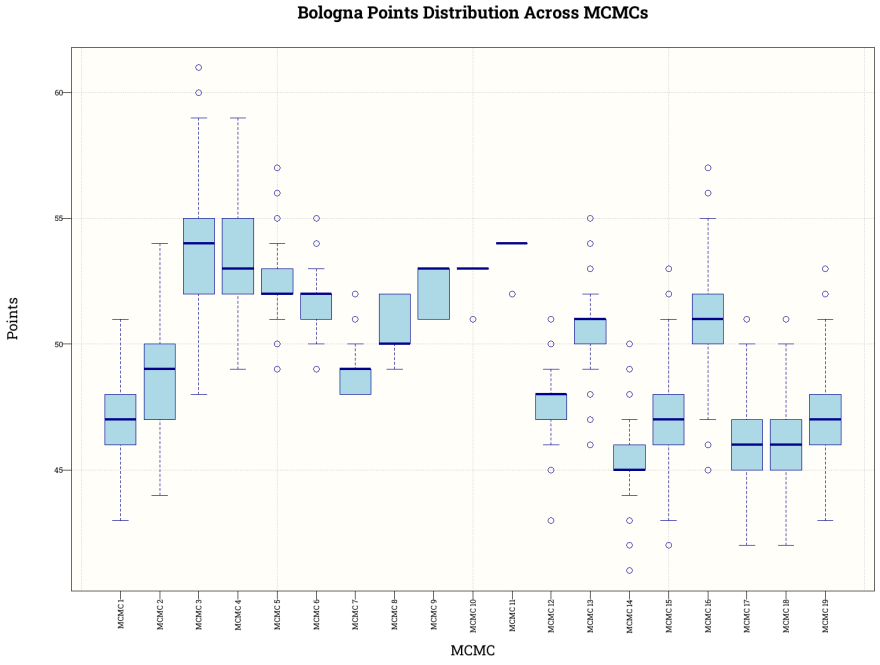


Figure 6.7: Second method: Bologna distribution of points across MCMCs