Abstract

In paired comparison sports data analysis, practitioners and researchers have identified the varying abilities of teams due to injuries, team psychology, and team improvement in the course of sequential competitions. The most commonly used framework to describe the score difference or the match outcome is mainly based on an appropriate transformation of the difference in abilities of the home team and the visiting team. Under such consideration, the abilities of teams can be further modeled with dynamic effects in the frequentist or Bayesian perspective. By integrating these features into a model formulation, we propose more general dynamic models for the abilities of teams. In addition, some criteria are developed to select a better predictive model for playoffs among competing models. The practicality of our proposal is also investigated by the data from the 2009-2010 season to the 2018-2019 season of the National Basketball Association.

KEY WORDS: Paired comparisons; Dynamic abilities; Mixed effects models; Model selection; Proportion of correct predictions; Prediction mean squared error.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

How to assess the abilities of sports teams has been of great interest to researchers and practitioners. National Collegiate Athletic Association (NCAA) established a ranking system reflecting the abilities of teams to select teams for playoffs. Predictions of future outcomes can be made by the abilities of participating teams, which are highly concerned by practitioners.

Paired comparison models have been commonly used for sports events. A season of basketball matches in NBA league can be regarded as a series of paired comparisons. The advantage of paired comparisons is reducing the effects of confounding. For example, two teams share the same referee in a match, whereas one team may play with several different referees throughout the whole season, and there may be judgment biases among referees. Existing Paired comparison models for sports events characterize the score difference or outcome to be related to the home team's ability and visiting team's ability by a linear model or generalized linear model, respectively.

Previous studies proposed a variety of paired comparison models for sports events, including random/fixed effects models with/without dynamic effects on the abilities. In the spirit of existing models, we further propose two flexible models under different cases, and many existing models can be unified in the proposed models. We connect Bayesian and frequentist viewpoints by mixed effects models. The dynamic scheme of abilities is more general by considering the fixed dynamic scheme and random processes for the abilities simultaneously. We provide model selection criteria to select a better model and setup for the prediction purpose. Two measures of predictive ability are used to compare the predictive performances of competing models.

In section 2, several existing paired comparison models are introduced. Section 3 describes the proposed models under different setups. Section 4 introduces the estimation method, which consists of the least squares method, maximizing observed likelihood, and maximizing the posterior likelihood. The measures of goodness-of-fit and predictive ability are also introduced. Section 5 presents an application to the National Basketball Association.

Chapter 2

Existing Paired Comparison Models

Let m be the number of matches; T the number of teams; Y_i the score difference of match i, $i = 1, \ldots, m$; a_k and b_k the home ability and visiting ability of team k respectively, $k = 1, \ldots, T$; h_i and v_i the home team and visiting team in match i respectively; and t_i the time of match i.

The first paired comparison model for sports events proposed by [1] did not consider the dynamic effects and the home ability and visiting ability were considered to be the same. That is, $a_k = b_k \triangleq \alpha_k, \forall i = 1, \ldots, m$, and $k = 1, \ldots, T$, which leads to the following model:

$$Y_i = \alpha_{h_i} - \alpha_{v_i} + \varepsilon_i,\ i = 1, \ldots, m. \tag{2.1}$$

[2] improved model (2.1) by considering the home court advantage θ, i.e., $a_k \triangleq \alpha_k + \theta$ and $b_k \triangleq \alpha_k$, which leads to the following model:

$$Y_i = \theta + \alpha_{h_i} - \alpha_{v_i} + \varepsilon_i,\ i = 1, \ldots, m. \tag{2.2}$$

[3] considered team-specific home court advantages θ_k, i.e., $a_k \triangleq \alpha_k + \theta_k$ and $b_k \triangleq \alpha_k$, which leads to the following model:

$$Y_i = \theta_{h_i} + \alpha_{h_i} - \alpha_{v_i} + \varepsilon_i,\ i = 1, \ldots, m. \tag{2.3}$$

The first model considering the dynamic abilities was proposed by [4]. The model suggested that there is a deviation of performance $S_k(t_i)$ from a team's underlying ability in each game. In the formulation of (2.2), let $a_k(t_i) \triangleq \theta + \alpha_k(t_i)$ and $b_k(t_i) \triangleq \alpha_k(t_i)$. The model leads to

$$\alpha_k(t_i) = \alpha_k + S_k(t_i), \tag{2.4}$$

where $S_k(t_i)$ follows a random process. [5] proposed a model for ordered categories:

$$P(Y_i \leq r) = F(\theta_r + \alpha_{hi}(t_i) - \alpha_{vi}(t_i)),\ r = 1, \ldots, k \tag{2.5}$$

with $\alpha_k(t_i)$ following a random process and $\alpha_k(0) = \alpha_k$ for all k. [4] and [5] both assumed that the dynamic effects on ability depend on some random processes, whereas [6] proposed a fixed dynamic scheme for the dynamic evolution of ability.

Let $\lambda_1, \lambda_2 \in [0, 1]$ and $Y_i = 1$ if the home team won, and $Y_i = 0$ if the visiting team won. $t_i^{(-1)}$ denotes the time of the previous home match in which h_i was also the home team. $t_i'^{(-1)}$ denotes the time of the previous away match in which v_i was also the visiting team. [6] proposed the following dynamic Bradley-Terry model:

$$P(Y_i = 1 | Y_{i-1} = y_{i-1}, \ldots, Y_1 = y_1) = \frac{\exp\{a_{h_i}(t_i) - b_{v_i}(t_i)\}}{1 + \exp\{a_{h_i}(t_i) - b_{v_i}(t_i)\}}$$

with

$$\begin{cases} a_{h_i}(t_i) = \lambda_1 \gamma_1 y(t_i^{(-1)}) + (1 - \lambda_1) a_{h_i}(t_i^{(-1)}), \\ b_{v_i}(t_i) = \lambda_2 \gamma_2 (1 - y(t_i'^{(-1)})) + (1 - \lambda_2) b_{v_i}(t_i'^{(-1)}), \end{cases} \tag{2.6}$$

where $y(t_i)$ denotes the outcome of the match at time t_i. [6] assumed that all teams started with the same home and visiting underlying abilities $\gamma_1 \bar{r_h}$ and $\gamma_2 \bar{r_v}$ respectively, where $\bar{r_h}$ and $\bar{r_v}$ are the average win rates of home matches and away matches over the previous regular season respectively.

Chapter 3

Proposed Models for Dynamic Effects

Based on existing models, a paired comparison model for sports events can be formulated as

$$Y_i = a_{h_i}(t_i) - b_{v_i}(t_i) + \varepsilon_i,\ i = 1, \ldots, m. \tag{3.1}$$

The main issue is how to model the dynamic evolutions of abilities $a_{h_i}(t_i)$ and $b_{v_i}(t_i)$ for $i = 1, \ldots, m$. Let Y_i denote the score difference of match i. $Y_h(t_i)$ and $Y_v(t_i)$ denote the scores of the home team and the visiting team of match i respectively. For the regression formulation, we define the following notations:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix},\ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix},\ a = \begin{pmatrix} a_1 \\ \vdots \\ a_T \end{pmatrix},\ b = \begin{pmatrix} b_1 \\ \vdots \\ b_T \end{pmatrix},\ \gamma_1 = \begin{pmatrix} \gamma_{11} \\ \vdots \\ \gamma_{1T} \end{pmatrix},\ \text{and}\ \gamma_2 = \begin{pmatrix} \gamma_{21} \\ \vdots \\ \gamma_{2T} \end{pmatrix}.$$

3.1 Background

In the spirit of [6], we propose a more general model for abilities (hereinafter referred to M1):

$$\begin{cases} a_{h_i}(t_i) = \lambda_1 \gamma_{1h_i} y_h(t_i^{(-1)}) + (1-\lambda_1) a_{h_i}(t_i^{(-1)}), \\ b_{v_i}(t_i) = \lambda_2 \gamma_{2v_i} y_v(t_i'^{(-1)}) + (1-\lambda_2) b_{v_i}(t_i'^{(-1)}). \end{cases} \tag{3.2}$$

The underlying abilities a_k and b_k are fixed unknown parameters for $k = 1, \ldots, T$. The dynamic Bradley-Terry model is a special case of model M1 under the following three conditions: (i) underlying abilities are set to be the average win rates of home matches and away matches over the previous regular season respectively. (ii) $\gamma_{11} = \cdots = \gamma_{1T}$ and $\gamma_{21} = \cdots = \gamma_{2T}$. (iii) The score difference is replaced with the outcome.

In model M1, the underlying abilities a and b are involved in the updating scheme with scores. Thus the effects of underlying abilities will decrease as the season goes on. However, model (2.4) provides a different aspect: The effects of underlying abilities should be the same throughout the season. The dynamic effects are explained as the deviations of the actual performances from the underlying abilities. Considering the feature, we propose a model in which the underlying abilities are not involved in the updating scheme (hereinafter referred to M2):

$$\begin{cases} a_{h_i}(t_i) = a_{h_i} + \gamma_{1h_i} S_{h_i}(t_i) \quad \text{and} \quad b_{v_i}(t_i) = b_{v_i} + \gamma_{2v_i} S_{v_i}(t_i), \\ S_{h_i}(t_i) = (1-\lambda_1) S_{h_i}(t_i^{(-1)}) + \lambda_1 y_h(t_i^{(-1)}), \\ S_{v_i}(t_i) = (1-\lambda_2) S_{v_i}(t_i'^{(-1)}) + \lambda_2 y_v(t_i'^{(-1)}), \end{cases} \tag{3.3}$$

where $S_{v_i}(0) = 0$ and $S_{h_i}(0) = 0$.

There are two meaningful cases of model M2: $\lambda_1 = \lambda_2 = 1$ and $\lambda_1 = \lambda_2 = 0$ (hereinafter referred to M2-i and M2-ii respectively). The former implies that the dynamic effects only count on the result of the previous match. That is,

$$S_{h_i}(t_i) = y_h(t_i^{(-1)}) \quad \text{and} \quad S_{v_i}(t_i) = y_v(t_i'^{(-1)}). \tag{3.4}$$

The case of $\lambda_1 = \lambda_2 = 0$ means that there is no dynamic effect, i.e.,

$$S_{h_i}(t_i) \equiv 0 \quad \text{and} \quad S_{v_i}(t_i) \equiv 0. \tag{3.5}$$

The roles of λ_1 and λ_2 are the weights of averaging the previous match result and historic results. λ_1 and λ_2 can be considered as covariates that we can design. One can set λ_1 and λ_2 for arbitrary value in $[0, 1]$. Several values of λ_1 and λ_2 are tested in chapter 5.

The above models are in the frequentist framework. The abilities are updated by fixed parameters and historical data. In the Bayesian framework, the abilities are updated by some random processes. The simplest case is the first-order random walk model:

$$\begin{cases} a_{h_i}(t_i) = a_{h_i}(t_i^{(-1)}) + u_h(t_i), \\ b_{v_i}(t_i) = b_{v_i}(t_i'^{(-1)}) + u_v(t_i), \end{cases} \tag{3.6}$$

where $u_h(t_i)$'s $\overset{i.i.d.}{\sim} N(0, \sigma_h^2)$, $u_v(t_i)$'s $\overset{i.i.d.}{\sim} N(0, \sigma_v^2)$ and they are mutually independent for $i = 1, \ldots, m$. Considering both factors simultaneously, we further propose models M1R, M2R, M2R-i, and M2R-ii from M1, M2, M2-i, and M2-ii respectively, in which the abilities follow a random process. In the case of first-order random walk, M1R and M2R are proposed as follows respectively.

$$\begin{cases} a_{h_i}(t_i) = \lambda_1 \gamma_{1h_i} y_h(t_i^{(-1)}) + (1 - \lambda_1) a_{h_i}(t_i^{(-1)}) + u_h(t_i), \\ b_{v_i}(t_i) = \lambda_2 \gamma_{2v_i} y_v(t_i'^{(-1)}) + (1 - \lambda_2) b_{v_i}(t_i'^{(-1)}) + u_v(t_i). \end{cases} \tag{3.7}$$

$$\begin{cases} a_{h_i}(t_i) = a_{h_i} + \gamma_{1h_i} S_{h_i}(t_i) \quad \text{and} \quad b_{v_i}(t_i) = b_{v_i} + \gamma_{2v_i} S_{v_i}(t_i), \\ S_{h_i}(t_i) = (1 - \lambda_1) S_{h_i}(t_i^{(-1)}) + \lambda_1 y_h(t_i^{(-1)}) + u_h(t_i), \\ S_{v_i}(t_i) = (1 - \lambda_2) S_{v_i}(t_i'^{(-1)}) + \lambda_2 y_v(t_i'^{(-1)}) + u_v(t_i). \end{cases} \tag{3.8}$$

The underlying abilities a, b, and parameters of dynamic effects γ_1, γ_2 can be fixed unknown parameters or random parameters. From the frequentist viewpoint, if the sample size is large enough, consistency of the estimation guarantees the estimated parameters will converge to the true parameters. Form the Bayesian viewpoint, by assuming the parameters follow some prior distributions, it can reduce the number of parameters to be estimated, which is an advantage when the sample size is small. We cover the above two viewpoints by considering the following 4 different cases:

Case 1 . a, b and γ_1, γ_2 are fixed;

Case 2 . a, b are fixed and γ_1, γ_2 are random;

Case 3 . a, b are random and γ_1, γ_2 are fixed; and

Case 4 . a, b and γ_1, γ_2 are random.

In each of the above cases, if a and b are random, we assume that a_k's $\overset{i.i.d.}{\sim} N(\mu_a, \sigma_a^2)$ and b_k's $\overset{i.i.d.}{\sim} N(\mu_b, \sigma_b^2)$. If γ_1 and γ_2 are random, we assume that γ_{1k}'s $\overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2)$ and γ_{2k}'s $\overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2)$ and all the random parameters are mutually independent. If there is any random parameter, the normality assumption ε_i's $\overset{i.i.d.}{\sim} N(0, \sigma_0^2)$ is required due to concerns about estimation.

3.2 Regression Model Formulation

All proposed models can be rewritten as the following form:

$$Y = X\beta + \varepsilon, \tag{3.9}$$

where

$$\beta = \begin{pmatrix} a \\ b \\ \gamma_1 \\ \gamma_2 \\ u_h \\ u_v \end{pmatrix} \text{ with } u_h = \begin{pmatrix} u_h(t_1) \\ \vdots \\ u_h(t_m) \end{pmatrix}, \text{ and } u_v = \begin{pmatrix} u_v(t_1) \\ \vdots \\ u_v(t_m) \end{pmatrix}.$$

Model M1 is chosen as an example to show how to obtain the regression formulation. In an arbitrary match i (or at time t_i), the home ability and visiting ability of team h_i and team v_i under model M1 can be rewritten as the follows respectively:

$$\begin{cases} a_{h_i}(t_i) = & (1-\lambda_1)^{K_1} a_{h_i} + \left[\lambda_1 \sum_{j=0}^{K_1-1} (1-\lambda_1)^j y_h(t_i^{(-j-1)})\right] \gamma_{1h_i}, \\ b_{v_i}(t_i) = & (1-\lambda_2)^{K_2} b_{v_i} + \left[\lambda_2 \sum_{j=0}^{K_2-1} (1-\lambda_2)^j y_v(t_i^{\prime(-j-1)})\right] \gamma_{2v_i}, \end{cases} \tag{3.10}$$

where K_1 and K_2 denote the number of home matches and away matches h_i and v_i had played before time t_i respectively. From (3.10), one can design the covariate matrix X and obtain the regression model formulation of model M1. Proceeding in the same way, one can derive the regression formulation of all models. Due to the problem of identifiability, we assume that $\sum_{k=1}^{T} b_k = 0$ in model M2, M2-i, M2-ii, M2R, M2R-i, and M2R-ii. If there are random effects in the model, mixed effects model formulation has more advantages in estimation.

Let β_R denote the random parameters, and β_F denotes the fixed parameters, i.e.,

$$\textit{Case 2. } \beta_R = \begin{pmatrix} \gamma_1 - \mu_1 \mathbb{1}_T \\ \gamma_2 - \mu_2 \mathbb{1}_T \end{pmatrix}, \beta_F = \begin{pmatrix} a \\ b \\ \mu_1 \\ \mu_2 \end{pmatrix};$$

$$\textit{Case 3. } \beta_R = \begin{pmatrix} a - \mu_a \mathbb{1}_T \\ b - \mu_b \mathbb{1}_T \end{pmatrix}, \beta_F = \begin{pmatrix} \mu_a \\ \mu_b \\ \gamma_1 \\ \gamma_2 \end{pmatrix}; \text{ and}$$

$$\textit{Case 4. } \beta_R = \begin{pmatrix} a - \mu_a \mathbb{1}_T \\ b - \mu_b \mathbb{1}_T \\ \gamma_1 - \mu_1 \mathbb{1}_T \\ \gamma_2 - \mu_2 \mathbb{1}_T \end{pmatrix}, \beta_F = \begin{pmatrix} \mu_a \\ \mu_b \\ \mu_1 \\ \mu_2 \end{pmatrix}.$$

Due to the problem of identifiability, we assume that $\mu_b = 0$ in model M2, M2-i, M2-ii, M2R, M2R-i, and M2R-ii. By similar procedures, one can derive the mixed effects model formulation:

$$Y = X_R \beta_R + X_F \beta_F + \varepsilon. \tag{3.11}$$

Chapter 4

Estimation and Model Selection

To write down the estimation approach explicitly, we first define some notations. Let X_{γ_1} and X_{γ_2} be the covariate matrix of γ_1 and γ_2 respectively. Let $\gamma = (\gamma_1^T, \gamma_2^T)^T$, $X_\gamma = (X_{\gamma_1}, X_{\gamma_2})$, and $Y^* = Y - X_F\beta_F$.

4.1 Estimation

By the model formulation (3.11):

$$Y = X_R\beta_R + X_F\beta_F + \varepsilon \triangleq X_F\beta_F + \varepsilon^*, \tag{4.1}$$

where $\varepsilon^* = \varepsilon + X_R\beta_R$ and $\varepsilon^* \sim N_m(0, \sigma_0^2 I_m + X_R Var(\beta_R) X_R^T)$. The estimation approach of β_F and λ is proposed to minimize the sum of squares

$$SS(\lambda, \beta_F) = (Y - X_F\beta_F)^T(Y - X_F\beta_F). \tag{4.2}$$

Least square method can be applied in this minimization.

$$\min_{\lambda,\beta_F} SS(\lambda, \beta_F) = \min_{\lambda} \min_{\beta_F} SS(\lambda, \beta_F) = \min_{\lambda}(Y - X_F\hat{\beta}_F)^T(Y - X_F\hat{\beta}_F),$$

where $\hat{\beta}_F = (X_F^T X_F)^{-1} X_F^T Y$ is the least square estimator of β_F.

We illustrate the rest of the estimation by case 2, in which $Var(\beta_R) = \text{diag}(\sigma_1^2 I_T, \sigma_2^2 I_T)$. The observed likelihood function of $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ is

$$L(\sigma_0^2, \sigma_1^2, \sigma_2^2 | X, Y) = (2\pi)^{-\frac{m}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y - X_F\beta_F)^T \Sigma^{-1} (Y - X_F\beta_F)}, \tag{4.3}$$

where $\Sigma = \sigma_0^2 I_m + \sigma_1^2 X_{\gamma_1} X_{\gamma_1}^T + \sigma_2^2 X_{\gamma_2} X_{\gamma_2}^T$ and the observed log-likelihood function is

$$\begin{aligned}&-\frac{m}{2}\log(2\pi) - \frac{m}{2}\log\sigma_0^2 - \frac{1}{2}\log|I_m + \frac{\sigma_1^2}{\sigma_0^2} X_{\gamma_1} X_{\gamma_1}^T + \frac{\sigma_2^2}{\sigma_0^2} X_{\gamma_2} X_{\gamma_2}^T| \\ &-\frac{1}{2\sigma_0^2}(Y - X_F\beta_F)^T (I_m + \frac{\sigma_1^2}{\sigma_0^2} X_{\gamma_1} X_{\gamma_1}^T + \frac{\sigma_2^2}{\sigma_0^2} X_{\gamma_2} X_{\gamma_2}^T)^{-1}(Y - X_F\beta_F).\end{aligned} \tag{4.4}$$

The estimation approach for $(\sigma_0^2, \sigma_1^2, \sigma_2^2)$ is proposed to maximize the observed log-likelihood function, i.e.,

$$(\hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2) = \underset{(\sigma_0^2, \sigma_1^2, \sigma_2^2)}{\operatorname{argmax}}\, l(\sigma_0^2, \sigma_1^2, \sigma_2^2 | X, Y).$$

To estimate the predictors $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of γ_1 and γ_2, it is proposed to maximize the posterior log-likelihood. That is,

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \underset{(\gamma_1, \gamma_2)}{\operatorname{argmax}} \log L(\gamma_1, \gamma_2 | \hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, X, Y), \tag{4.5}$$

where

$$\begin{aligned} & \log L(\gamma_1, \gamma_2 | \sigma_0^2, \sigma_1^2, \sigma_2^2, X, Y) \propto \log f_Y(y|\sigma_0^2, \gamma_1, \gamma_2)\pi(\gamma_1, \gamma_2|\sigma_1^2, \sigma_2^2) \\ = \; & -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log\sigma_0^2 - \frac{1}{2\sigma_0^2}|(Y - X_F\beta_F - X_{\gamma_1}\gamma_1 + X_{\gamma_2}\gamma_2)|^2 \\ & (-m)\log(2\pi) - \frac{m}{2}(\log\sigma_1^2 + \log\sigma_2^2) - \frac{1}{2}(\frac{1}{\sigma_1^2}|\gamma_1|^2 + \frac{1}{\sigma_2^2}|\gamma_1|^2) \\ \propto \; & -(Y^* - X_\gamma\gamma)^T(Y^* - X_\gamma\gamma) - \gamma^T W \gamma, \end{aligned}$$

and

$$W = \begin{pmatrix} \frac{\sigma_0^2}{\sigma_1^2} I_T & 0 \\ 0 & \frac{\sigma_0^2}{\sigma_2^2} I_T \end{pmatrix}.$$

The posterior log-likelihood is maximized when

$$\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = (X_\gamma^T X_\gamma + W)^{-1} X_\gamma^T Y^* = E[\gamma | \sigma_1^2, \sigma_2^2, \sigma_v^2, X, Y]. \tag{4.6}$$

4.2 Model Selection

With all the models and cases, how to select the best model is an important issue. Good predictions of outcomes and differences in scores are important for practitioners. In the following paragraphs, we first introduce two measures of predictive ability to measure the performance of a model on the prediction purpose. Then several criteria are proposed based on goodness-of-fit and predictive ability measured by regular season data.

The measures of predictive ability are proportion of correct predictions and prediction mean squared error, hereinafter denoted by PCP and PMSE respectively. Let (X^0, Y^0) be a future run.

$$\text{PCP}(\mathcal{M}) = P(sign(Y^0) \cdot sign(X^0 \hat{\beta}_{\mathcal{M}}) > 0) + 0.5P(X^0 \hat{\beta}_{\mathcal{M}} = 0), \qquad (4.7)$$

and

$$\text{PMSE}(\mathcal{M}) = E(Y^0 - X^0 \hat{\beta}_{\mathcal{M}})^2, \qquad (4.8)$$

where $\hat{\beta}_{\mathcal{M}}$ denotes the estimate of β under model $\mathcal{M}$. To estimate the PCP and PMSE of model $\mathcal{M}$, the playoffs data are considered as future runs, and the probability is estimated by the empirical distribution of playoffs data.

The Bayesian Information Criterion in [7] is a common approach in model selection. With the observed log-likelihood function (4.4), the BIC value of a model $\mathcal{M}$ is derived by

$$\text{BIC}(\mathcal{M}) = -2 \log L(\mathcal{M}) + p_{\mathcal{M}} \log m, \qquad (4.9)$$

where $p_{\mathcal{M}}$ denotes the number of parameters in the model $\mathcal{M}$. [7] suggested to choose the model with the smallest BIC value.

To get an analogue of prediction by regular season data, cross validation is a commonly used approach. However, it can not be applied to the proposed models since the estimation of dynamic abilities depends on historical data. Thus we split the regular season data into training data and testing data by a given time. Estimation is done by training data, and measures of predictive ability can be derived by applying the estimated parameters to the testing data. To be more explicit, we introduce the following notations:

$$Y = \begin{pmatrix} Y^{tr} \\ Y^{te} \end{pmatrix}, \ X = \begin{pmatrix} X^{tr} \\ X^{te} \end{pmatrix},$$

where (X^{tr}, Y^{tr}) and (X^{te}, Y^{te}) denote the training data and testing data respectively. Let $\hat{\beta}^{tr}_{\mathcal{M}}$ be the estimate of β by the training data under model $\mathcal{M}$. The PCP and PMSE of model $\mathcal{M}$ are estimated by

$$\widehat{\text{PCP}}(\mathcal{M}) = \frac{1}{S}\sum_{i=1}^{S}\left[I(sign(Y_i^{te}) \cdot sign(X_i^{te}\hat{\beta}^{tr}_{\mathcal{M}}) > 0) + 0.5I(X_i^{te}\hat{\beta}^{tr}_{\mathcal{M}} = 0)\right], \tag{4.10}$$

and

$$\widehat{\text{PMSE}}(\mathcal{M}) = \frac{1}{S}\sum_{i=1}^{S}(Y_i^{te} - X_i^{te}\hat{\beta}^{tr}_{\mathcal{M}})^2, \tag{4.11}$$

where S is the number of matches in testing data. The model $\mathcal{M}_{\text{PCP}}$ and $\mathcal{M}_{\text{PMSE}}$ with highest $\widehat{\text{PCP}}$ and smallest $\widehat{\text{PMSE}}$ are chosen, i.e.,

$$\mathcal{M}_{\text{PCP}} = \underset{\mathcal{M}}{\text{argmax}}\, \widehat{\text{PCP}}(\mathcal{M}) \tag{4.12}$$

and

$$\mathcal{M}_{\text{PMSE}} = \underset{\mathcal{M}}{\text{argmin}}\, \widehat{\text{PMSE}}(\mathcal{M}). \tag{4.13}$$

Chapter 5

An Application to National Basketball Association

The proposed models are applied to 2009–2010 season to 2018–2019 season of the National Basketball Association. The data are available via an API provided in [8]. The data consist of the index of every match sorting by calendar time, the home teams and visiting teams of every match, and scores of the home teams and away teams in every match. The regular season is used to fit the proposed models, and the playoffs data are treated as future runs to estimate the proportion of correct predictions and prediction mean squared error. Model M1 and M2 with $\lambda \in \Lambda$ are also considered in this section, where $\Lambda = \{k(0.1, 0.1)^T : \ k = 0, \ldots, 9\}$. Hereinafter we denote $\mathcal{M}_\cdot = \{$M $\cdot$ including the cases $\lambda \in \Lambda\}$, for $\cdot = 1, 2$, and $\mathcal{M}_T = \mathcal{M}_1 \cup \mathcal{M}_2$.

We investigate the dynamic effects of abilities in model M1R, M2R, M2R-i, and M2R-ii with a first-order random walk. The means of the ratios $\hat{\sigma}_h^2/\hat{\sigma}_0^2$ and $\hat{\sigma}_v^2/\hat{\sigma}_0^2$ over ten seasons are shown in Table 6.3. In these NBA data, the variances of u_h and u_v are small compared to the estimated variances of ε. Moreover, the predictions of outcomes are almost the same whether considering u_h and u_v or not. To simplify the presentation, we do not consider models M1R, M2R, M2R-i, and M2R-ii in this investigation based on these facts.

The proportion of correct predictions of dynamic Bradley-Terry model (2.6) proposed in [6] (hereinafter referred to DBT model) and its modification, which replaces the outcome with score difference (hereinafter referred to DBTS model) are listed in table 6.1 compared to the proposed models. DBT model produces predictions that the home team will win for every match. DBTS model has a lower PCP than the proposed models. One step of estimating the parameters is to minimize the sum of squares function (4.2). Figure 6.2 and 6.3 show that the non-convexity of the sum of squares function under model M1 and M2 makes the minimization difficult. Thus we do not recommend DBT model and DBTS model.

Table 6.4 shows the mean of BIC values over ten seasons under different models and cases. BIC suggests that model M2-ii in the case that abilities are random should be selected, which means that there is no dynamic effect. The conclusion also holds if we look at the BIC values season by season.

We compare the PCP of $\mathcal{M}_{\text{PCP}}$, $\mathcal{M}_{\text{PMSE}}$ and model $\mathcal{M}_{\text{BIC}}$ (without dynamic effect) for ten seasons to see if there is any evidence of dynamic effect. In most of the seasons except 2015-2016 season, the models with dynamic effects can have higher PCP than model $\mathcal{M}_{\text{BIC}}$. This can be evidence that there are dynamic effects in most of the seasons except 2015-2016 season. Moreover, in 2010–2011 season to 2012-2013 season we successfully select the models with dynamic effect with higher PCP than model $\mathcal{M}_{\text{BIC}}$. Over the ten seasons, the PCP of selected models are comparable to the highest PCP of all models except 2013-2014 season and 2017-2018 season.

Table 6.4 shows the mean of PCP's over ten seasons under different models and cases. We can see that most of the models are comparable except model M1 under case 1, M1 under case 3, and model M2 under case 2.

Chapter 6

Conclusion and Discussion

In the application to NBA, table 6.2 shows that the fixed dynamic scheme inspired by [6] performs poorly in the sense of prediction with PCP 0.586 and DBT model tends to produce meaningless predictions that the home team always wins. Estimation of the weight λ is also difficult (see Figure 6.2) and such estimation is based on the sense of goodness-of-fit, which may not correspond to the predictive ability. The fact that u_h and u_v are inapparent shows that the first-order random walk assumption on the home ability and visiting ability (cf. [4] and [5]) has no contribution to the dynamic effects. To sum up, most of the existing approaches to estimate the dynamic effects are based on the goodness-of-fit with some specific models. By such approaches, either there is no evidence of dynamic effects or the dynamic effects may produce poor predictions.

In the aspect of regression, λ should play the role as designed covariate. Assigning given values to λ avoids the difficulties in estimation, and the proposed model selection criteria can suggest the best λ in the sense of predictive ability. In the applications to the NBA, the proposed model selection criteria can select the model with nearly the highest PCP in most of the seasons.

All the proposed models are parametric models that model the score difference to be the difference in home ability and visiting ability. A more general semiparametric model may characterize the relation between score difference and abilities better. In the applications to the NBA, the format of playoffs and regular season are different, which may decrease the PCP and increase PMSE for our models. This problem is still needed to be solved in future research.

Reference

[1] Stefani, R. T. (1977). Football and basketball predictions using least squares. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(2), 117-121.

[2] Stefani, R. T. (1980). Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics*, 10(2), 116-123.

[3] Clarke, S. R., and Norman, J. M. (1995). Home ground advantage of individual clubs in english soccer. *The Statistician*, 44(4), 509.

[4] Harville, D. (1977). The use of linear-model methodology to rate high school or college football teams. *Journal of the American Statistical Association*, 72(358), 278-289.

[5] Fahrmeir, L., and Tutz, G. (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, 89(428), 1438-1449.

[6] Cattelan, M., Varin, C., and Firth, D. (2012). Dynamic Bradley-Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1), 135-150.

[7] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461464.

[8] Bresler, A. (n.d.). R's interface to NBA data. Retrieved July 29, 2020 from http://asbcllc.com/nbastatR/

[9] Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273286.

[10] Zermelo, E. (1929). Die berechnung der turnier-Ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Math Z* 29, 436460.

[11] Bradley, R. A., and; Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4), 324.

[12] Harville, D. (1976). Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects. *The Annals of Statistics*, 4(2), 384-395.

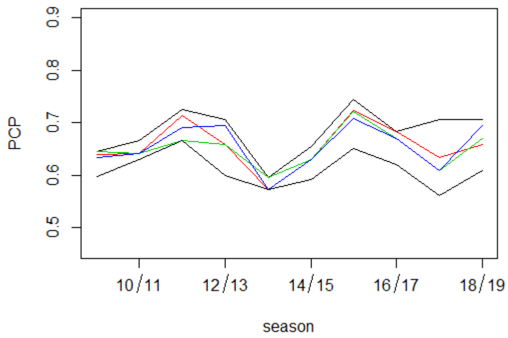
[13] Batchelder, W. H., Bershad, N. J., and Simpson, R. S. (1992). Dynamic paired-comparison scaling. *Journal of Mathematical Psychology*, 36(2), 185-212.

[14] Harville, D. A. (2003). The selection or seeding of college basketball or football teams for postseason competition. *Journal of the American Statistical Association*, 98(461), 17-27.

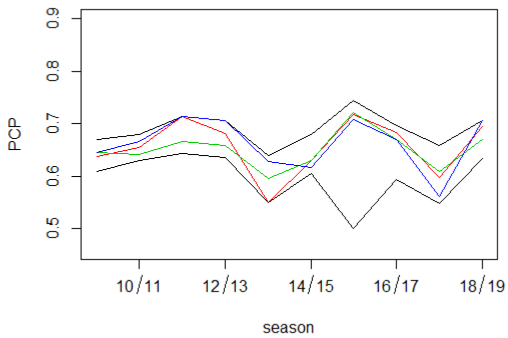
[15] Wang J. (2010). Consistent selection of the number of clusters via crossvalidation. *Biometrika*, 97(4), 893904.

[16] Lim, A., Chiang, C. T., and Teng, J. C. (2018). Estimating robot strengths with application to selection of alliance members in FIRST robotics competitions. *arXiv preprint arXiv:1810.05763*.

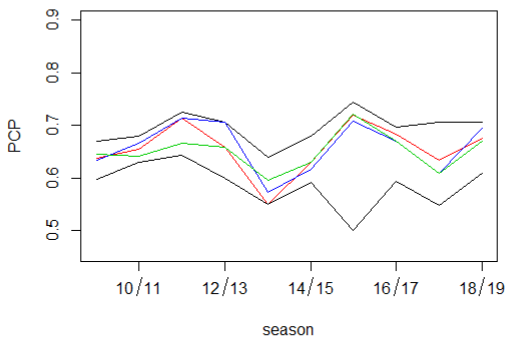
Figure 6.1: Proportion of correct predictions' of $\mathcal{M}_{\cdot\text{PCP}}$ (red line), $\mathcal{M}_{\cdot\text{PMSE}}$ (blue line), and $\mathcal{M}_{\cdot\text{BIC}}$ (green line) over ten seasons. Black lines are the highest and lowest proportions of correct predictions' among $\mathcal{M}_{\cdot}$ for $\cdot = 1, 2$, and T.



(a) $\mathcal{M}_1$



(b) $\mathcal{M}_2$



(c) $\mathcal{M}_T$

Table 6.1: The estimated proportion of correct predictions from the Dynamic Bradley-Terry models (cf. [6]) and the selected proposed models on playoffs data.

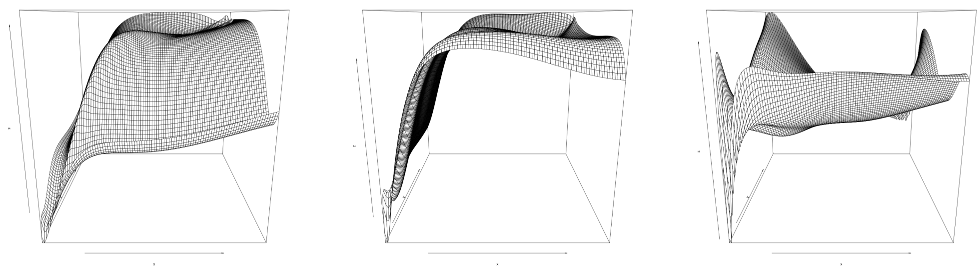
Model \ Season	09–10	10–11	11–12	12–13	13–14
DBT	0.6707	0.6667	0.6786	0.6353	0.5618
DBTS	0.5000	0.5926	0.5000	0.5647	0.5618
$\mathcal{M}_{1\text{PCP}}$	0.6402	0.6420	0.7143	0.6588	0.5730
$\mathcal{M}_{1\text{PMSE}}$	0.6341	0.6420	0.6905	0.6941	0.5730
$\mathcal{M}_{1\text{BIC}}$	0.6463	0.6420	0.6667	0.6588	0.5955
$\mathcal{M}_{2\text{PCP}}$	0.6402	0.6543	0.7143	0.6824	0.5506
$\mathcal{M}_{2\text{PMSE}}$	0.6463	0.6667	0.7143	0.7059	0.6292
$\mathcal{M}_{2\text{BIC}}$	0.6463	0.6420	0.6667	0.6588	0.5955
$\mathcal{M}_{T\text{PCP}}$	0.6382	0.6543	0.7143	0.6588	0.5506
$\mathcal{M}_{T\text{PMSE}}$	0.6341	0.6420	0.6667	0.6941	0.5843
$\mathcal{M}_{T\text{BIC}}$	0.6463	0.6420	0.6667	0.6588	0.5955
Model \ Season	**14–15**	**15–16**	**16–17**	**17–18**	**18–19**
DBT	0.5926	0.6744	0.5696	0.7073	0.5610
DBTS	0.5926	0.6744	0.5696	0.7073	0.5610
$\mathcal{M}_{1\text{PCP}}$	0.6296	0.7229	0.6835	0.6341	0.6585
$\mathcal{M}_{1\text{PMSE}}$	0.6296	0.7093	0.6709	0.6098	0.6951
$\mathcal{M}_{1\text{BIC}}$	0.6296	0.7209	0.6709	0.6098	0.6707
$\mathcal{M}_{2\text{PCP}}$	0.6296	0.7171	0.6835	0.5976	0.6951
$\mathcal{M}_{2\text{PMSE}}$	0.6173	0.7093	0.6709	0.5610	0.7073
$\mathcal{M}_{2\text{BIC}}$	0.6296	0.7209	0.6709	0.6098	0.6707
$\mathcal{M}_{T\text{PCP}}$	0.6296	0.7200	0.6835	0.6341	0.6768
$\mathcal{M}_{T\text{PMSE}}$	0.6049	0.6628	0.6582	0.5732	0.6341
$\mathcal{M}_{T\text{BIC}}$	0.6296	0.7209	0.6709	0.6098	0.6707

Table 6.2: Mean of proportion of correct predictions of Dynamic Bradley-Terry model (cf. [6]) and proposed models over the ten seasons.

DBT	DBTS	$\mathcal{M}_{1\text{PCP}}$	$\mathcal{M}_{1\text{PMSE}}$	$\mathcal{M}_{1\text{BIC}}$
0.6318	0.5824	0.6557	0.6548	0.6511

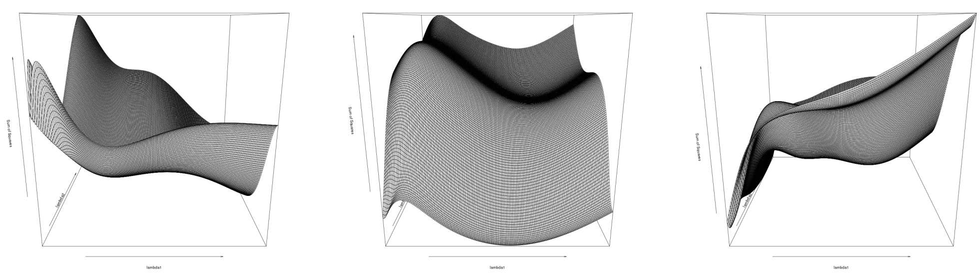
$\mathcal{M}_{2\text{PCP}}$	$\mathcal{M}_{2\text{PMSE}}$	$\mathcal{M}_{2\text{BIC}}$	$\mathcal{M}_{\text{PCP}}$	$\mathcal{M}_{\text{PMSE}}$	$\mathcal{M}_{\text{BIC}}$
0.6565	0.6628	0.6511	0.6560	0.6596	0.6511

Figure 6.2: Selected plots of minimized sum of squares value to λ



(a) 2013–2014 season (b) 2014–2015 season (c) 2016–2017 season

Figure 6.3: Selected plots of minimized sum of squares value to λ



(a) 2016–2017 season (b) 2017–2018 season (c) 2018–2019 season

Table 6.3: Ratio of $\hat{\sigma}_h^2$ to $\hat{\sigma}_0^2$ and $\hat{\sigma}_h^2$ to $\hat{\sigma}_0^2$ (rounded to 3 decimal places)

	Ratio of $\hat{\sigma}_h^2$ to $\hat{\sigma}_0^2$			
Model	M1R	M2R	M2R-i	M2R-ii
Case 1	0.013	0.000	0.000	0.000
Case 2	0.018	0.000	0.000	
Case 3	0.769	0.000	0.000	0.002
Case 4	0.011	0.002	0.002	

	Ratio of $\hat{\sigma}_v^2$ to $\hat{\sigma}_0^2$			
Model	M1R	M2R	M2R-i	M2R-ii
Case 1	0.000	0.000	0.000	0.000
Case 2	0.022	0.000	0.000	
Case 3	0.000	0.000	0.000	0.001
Case 4	0.020	0.001	0.001	

Table 6.4: Mean of PCP and BIC over ten seasons

	PCP			
Model	M1	M2	M2-i	M2-ii
Case 1	0.586	0.631	0.654	0.661
Case 2	0.655	0.661	0.657	
Case 3	0.632	0.639	0.664	0.646
Case 4	0.652	0.652	0.655	

	BIC			
Model	M1	M2	M2-i	M2-ii
Case 1	10105.43	10109.69	10125.14	9765.25
Case 2	9854.61	9838.85	9807.57	
Case 3	9838.49	9792.86	9793.35	9540.60
Case 4	9569.06	9559.94	9559.02	