COMPARING SEVERAL MODELING METHODS ON NCAA MARCH MADNESS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Su Hua

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Statistics

July 2015

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Comparing Several Modeling Methods on NCAA March Madness

By

Su Hua

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Rhonda Magel
Co-Chair

Dr. Gang Shen
Co-Chair

Dr. Seung Won Hyun

Dr. Kenneth Magel

Approved:

7/10/2015
Date

Rhonda Magel
Department Chair

ABSTRACT

This year (2015), according to the AGA's (American Gaming Association) research, nearly about 40 million people filled out about 70 million March Madness brackets (Moyer, 2015). Their objective is to correctly predict the winners of each game. This paper used the probability self-consistent (PSC) model (Shen, Hua, Zhang, Mu, Magel, 2015) to make the prediction of all 63 games in the NCAA Men's Division I Basketball Tournament. PSC model was first introduced by Zhang (2012). The Logit link was used in Zhang's (2012) paper to connect only five covariates with the conditional probability of a team winning a game given its rival team. In this work, we incorporated fourteen covariates into the model. In addition to this, we used another link function, Cauchit link, in the model to make the predictions. Empirical results show that the PSC model with Cauchit link has better average performance in both simple and doubling scoring than Logit link during the last three years of tournament play.

In the generalized linear model, maximum likelihood estimation is a popular method for estimating the parameters; however, convergence failures may happen when using large dimension covariates in the model (Griffiths, Hill, Pope, 1987). Therefore, in the second phase in this study, Bayesian inference is used for estimating in the parameters in the prediction model. Bayesian estimation incorporates prior information such as experts' opinions and historical results in the model. Predictions from three years of March Madness using the model obtained from Bayesian estimation with Logit link will be compared to predictions using the model obtained from maximum likelihood estimation.

ACKNOWLEDGMENTS

It is with immense gratitude that I acknowledge the support and help of my advisors, Dr. Rhonda Magel and Dr. Gang Shen. Thank you for giving me continuous guidance and support throughout this whole research process. This dissertation could not have been completed without the effort from you. I would like to thank all my committee members, Dr. Seung Won Hyun, and Dr. Kenneth Magel, as well as Dr. Changhui Yan and Mr. Curt Doetkott. At last, I would like to thank all professors and staff from the Department of Statistics.

I would also like to thank my parents. They were always supporting me and encouraging me with their best wishes.

Finally, I would like to thank my wife, Qian Wen. She was always there cheering me up and stood by me through the good times and bad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

1. INTRODUCTION

1.1. Research Objective

NCAA March Madness is a phenomenon that catches sports fans' eyes from the second week of March through the first week of April. The NCAA tournaments are an American tradition that sends millions of fans into a synchronized frenzy each year. It's this chaos that gives the tournament its March Madness nickname. This work will focuses on bracketing the NCAA Men's Division I Basketball Tournament based on probability self-consistent (PSC) model. This model was first introduced by Zhang (2012). The Logit link was used in his paper to connect the five covariates with the conditional probability of a team winning a game given its rival team. Maximum likelihood estimation was applied to estimate the unknown coefficients.

In this work, we will first employ the Cauchit link to the PSC model for use in bracketing for March Madness (Shen et al., 2015), we will also consider using more covariates in the model. Bracket development using the Cauchit link and Logit link will be completed and compared with the actual results from March Madness over a 3 year period of time.

In the second phase, we will develop Bayesian estimation in place of maximum likelihood estimation for use in the PSC model with Logit link. Bracketing with Logit link will be done using Bayesian inference and compared to the actual results obtained from March Madness.

1.2. The Playing Rule and Structure

The National Collegiate Athletic Association (NCAA) Men's Division I Basketball Tournament, more commonly known as March Madness, is a single-elimination tournament that starts each March. Before describing the model, we provide a short introduction to the NCAA March Madness for those who might not be familiar.

Currently 68 college basketball teams are playing in each year. They are divided into four regions (East, South, Midwest, and West) and each team is ranked from 1 to 16 in its region. Eight of 68 teams first play four games, which are called the First4. Finally, 64 teams are determined and brackets are filled out. After six rounds, namely, Round64 (Rd64), Round32 (Rd32), Sweet16, Elite8, Final4 and the Championship, the national title is awarded to the team with six wins. In Rd64, 64 teams play 32 games, 32 teams play 16 games in Rd32, 16 teams play 8 games in Sweet16, 8 teams play 4 games in Elite8, 4 teams play 2 games in Final4, and 2 teams fight for the Championship. Starting with Rd64, there are 63 games played each year (NCAA Basketball Championship, 2015). Figure 1 shows the NCAA Men's Division I Basketball Tournament bracket and complete tournament results in 2014-2015 season.

1.3. Qualifying Procedure

There are more than three hundred eligible Division I teams only, and 68 teams make it into the March Madness. The 68 qualifying teams are from two bids: Automatic bids; and at-large bids (2015 NCAA Basketball Tournament, 2015). There are 32 teams who qualify from automatic bids, with 31 of these bids granted to the winner of the conference tournament championship. The only exception is the Ivy League which does not hold a conference tournament. For this league, the bid goes to the team with the best regular-season record. However, if two or more teams are tied for the best regular-season record, the league will hold a one-game playoff between the top two (or a series of such playoffs if more than two teams are tied) (NCAA basketball selection process, 2015). Table 1 shows the Automatic qualifiers in 2015 March Madness.

The remaining 36 at-large bids are granted by the NCAA Selection Committee to the teams it feels are the best 36 teams that did not receive automatic bids. Even though each conference receives only one automatic bid, the selection committee may select any number of at-large teams

from each conference (NCAA basketball selection process, 2015). The at-large teams generally come from college basketball's top conferences, including the ACC, The American, Atlantic-10, Big 12, Big East, Big Ten, Conference USA, Mountain West, Pac-12, and SEC. Table 2 is at-large qualifiers in 2015 March Madness (2015 NCAA Basketball Tournament, 2015).

Table 1. Automatic qualifiers for the 2015 NCAA March Madness

Conference	**Team**	**Conference**	**Team**
ACC	Notre Dame	MAC	Buffalo
America East	Albany	MEAC	Hampton
A–10	VCU	Missouri Valley	Northern Iowa
American	SMU	Mountain West	Wyoming
Atlantic Sun	North Florida	Northeast	Robert Morris
Big 12	Iowa State	Ohio Valley	Belmont
Big East	Villanova	Pac-12	Arizona
Big Sky	Eastern Washington	Patriot	Lafayette
Big South	Coastal Carolina	SEC	Kentucky
Big Ten	Wisconsin	Southern	Wofford
Big West	UC Irvine	Southland	Stephen F. Austin
Colonial	Northeastern	SWAC	Texas Southern
C-USA	UAB	Summit	North Dakota State
Horizon	Valparaiso	Sun Belt	Georgia State
Ivy League	Harvard	West Coast	Gonzaga
MAAC	Manhattan	WAC	New Mexico State

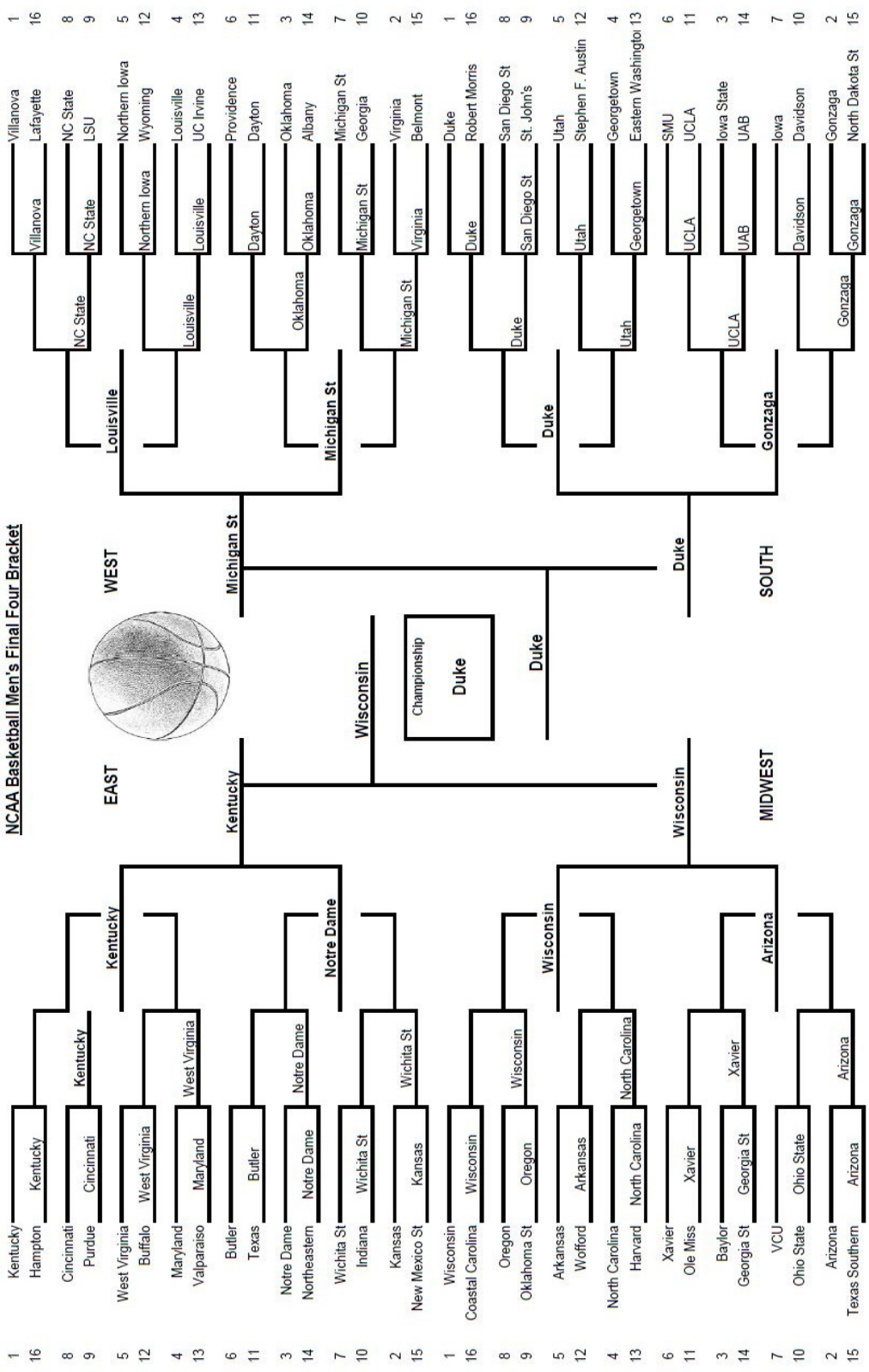

Figure 1. NCAA 2015 March Madness bracket with complete tournament results (This template is downloaded from: www.samplewords.com/ncaa-blank-printable-tournament-bracket/)

Table 2. At-large qualifiers for the 2015 NCAA March Madness

Conference	Team	Conference	Team
ACC	Virginia	Big 12	Baylor
ACC	Louisville	Big 12	Oklahoma State
ACC	North Carolina State	Big East	Butler
ACC	Duke	Big East	Providence
ACC	North Carolina	Big East	Georgetown
American	Cincinnati	Big East	St. John's
Atlantic 10	Dayton	Big East	Xavier
Atlantic 10	Davidson	Missouri Valley	Wichita State
Big 10	Maryland	Mountain West	Boise State
Big 10	Purdue	Mountain West	San Diego State
Big 10	Indiana	Pac 12	Utah
Big 10	Michigan State	Pac 12	UCLA
Big 10	Iowa	Pac 12	Oregon
Big 10	Ohio State	SEC	Louisiana State
Big 12	Kansas	SEC	Georgia
Big 12	West Virginia	SEC	Arkansas
Big 12	Texas	SEC	Ole Miss
Big 12	Oklahoma	West Coast	BYU

Before the bracket of 64 teams is put together each year, eight teams - the four lowest-seeded automatic qualifiers and the four lowest-seeded at-large teams will play in the First Four. The winners of these games advance to the Rd64. The two winning teams from automatic bids will be seeded 16 and the winners from at-large bids will be seeded 11 (NCAA Basketball Championship, 2015). The First Four games played in 2015 March Madness are shown in Table 3 (2015 NCAA Basketball Tournament, 2015).

Table 3. First Four games in 2015 NCAA March Madness

At-large	Automatic	At-large	Automatic
West Region (11)	Midwest Region (16)	East Region (11)	South Region (16)
BYU	Hampton	Boise State	North Florida
Ole Miss	Manhattan	Dayton	Robert Morris

In terms of bracketing the result of NCAA Men's Division I Basketball Tournament, this work will only focus on the last six rounds of the tournament beginning with Rd64. The First4 games will be excluded.

1.4. Bracket Scoring System

Two types of scoring systems will be considered: one is the doubling scoring system, and the other is the simple scoring system (Shen et al., 2015). There are 6 rounds in the tournament. Under the doubling points system, for each correct pick, one point will be awarded in the first round, two points will be awarded in the second round, four points will be awarded in the third round, and so on. In this system, one might not care about the individual rounds since predicting the correct champion is worth as much as the entire Rd64 combined. The system puts more weight on later rounds rather than the first several rounds.

Under the simple system: each correct pick will be awarded one point regardless of which round. In this system, one really cares about every single game in the whole tournament with no prediction discrimination existing among rounds. Table 4 is the summary of the NCAA Men's Division I Basketball Tournament scoring system.

Table 4. Scoring systems

	Rd64	Rd32	Sweet16	Elite8	Final4	Championship	total
Number of games	32	16	8	4	2	1	63
Simple	1	1	1	1	1	1	63
Doubling	1	2	4	8	16	32	192

2. LITERATURE REVIEW

For predicting outcomes of games in the NCAA tournament, Schwertman, Schenk and Holbrook (1996) used seed position to estimate the probability of each of the 16 seeds winning the regional tournament. It seems reasonable to use some function of seed positions because seeds were determined by a consensus of experts.

Magel and Unruh (2013) used both least squares regression and logistic regression to determine the key factors that influence the NCAA Men's Division I college basketball games. Least squares regression was used to develop a model to predict point spread, and logistic regression was used to develop a model to estimate the probability of winning a game. Magel and Unruh found four in-game statistics to be significant in both the least squares regression model and the logistic regression model. The in-game statistics found to be significant were free throw attempts, defensive rebounds, assists and turnovers.

Nelson (2012) used a logistic regression model to predict the probability that the higher seeded team beat the lower seeded team in each of the 63 games in March Madness. Bayesian inference was used in identifying the model that best fits the data as well as finding the coefficients of regression. The prior density for each coefficient β_i $(i = 0,1, \dots, n)$ was assumed follow normal distribution with mean μ_i and variance σ_i^2, where $\mu_0 = \mu_1 = \cdots = \mu_n$ and $\sigma_0^2 = \sigma_1^2 = \cdots = \sigma_n^2$ in Bayesian estimation.

Rating Percentage Index, commonly known as the RPI, is a rating system based on a team's wins and losses and its strength of schedule (Rating Percentage Index, 2015). It is the method that NCAA Basketball Selection Committee used to pick at-large bids and determine the seed in the tournament. The current formula is given as follows:

$$\text{RPI} = (\text{WP} \times 0.25) + (\text{OWP} \times 0.50) + (\text{OOWP} \times 0.25) \quad (1)$$

where WP is Winning Percentage, OWP is Opponents' Winning Percentage and OOWP is Opponents' Opponents' Winning Percentage. One can fill out a bracket using the value of RPI. In a single game, the team with higher RPI will go to the next round. Therefore, the team with the highest RPI value will win the entire tournament.

Jeff Sagarin has been providing ratings for USA TODAY since 1985 (Sagarin ratings, 2015). He uses each team's regular season statistics to create a single rating for each team. Exact details for this method are not publicly available, so one cannot know the exact method behind this rating system. The rating is available on USA TODAY and one can complete the bracket with it.

Pomeroy's College Basketball Ratings were first published in 2003 by Ken Pomeroy (Pomeroy ratings, 2015). This rating was built upon Pythagorean winning percentage (Pyth) which has the formula:

$$\text{Pyth} = \frac{AdjO^x}{AdjO^x + AdjD^x} \tag{2}$$

where AdjO is the adjusted offensive efficiency, an estimate of the offensive efficiency (points scored per 100 possessions) that a team would have against the average Division I defense; AdjD is the adjusted defensive efficiency, an estimate of the defensive efficiency (points allowed per 100 possessions) that a team would have against the average Division I offense; and x is an exponent that is empirically determined. This x was assigned 10.25 since 2012. However it was updated recently. Equivalent to the RPI and Sagarin's rating, before the March Madness, Pyth for each NCAA Division I team is available on kenpom.com and one can complete the bracket with it.

West (2006, 2008) proposes an ordinal logistic regression model and expectation (restricted OLRE model) on π_{ik}, the probability of team i has k winnings in the tournament as follows:

$$\pi_{ik} = \frac{\exp(\alpha_k + \boldsymbol{x_i}'\boldsymbol{\beta})}{1 + \exp(\alpha_k + \boldsymbol{x_i}'\boldsymbol{\beta})} - \sum_{j=0}^{k-1} \pi_{ik} \tag{3}$$

where α_k is the intercept for k winnings with $k = 0, 1, \ldots, k$. $\boldsymbol{x_i}$ is a vector of values for team i on the predictor variables, $\boldsymbol{\beta}$ is a vector of coefficients associated with the predictor variables, and rthe last term presents the cumulative sum of the probabilities of winning j games ($j = 0, 1, \ldots, k - 1$). This term would be equal to 0 for k = 0. Putting all $\pi_{ik} (i = 1, 2, \ldots, 64; k = 0, 1, \ldots, 6)$ in a 64 × 7 matrix, West (2006) requires the sums of each column must be equal to 32, 16, 8, 4, 2, 1, and 1, respectively. Zhang (2012) rewrote the restricted OLRE model in the form of restricted proportional odds model. The second term on the right hand side of equation (3) is moved to the left side, then the model (3) can be written as follows

$$\begin{cases} P(Z_i \leq k) = \dfrac{\exp(\alpha_k + \boldsymbol{x_i'}\boldsymbol{\beta})}{1 + \exp(\alpha_k + \boldsymbol{x_i'}\boldsymbol{\beta})}, k = 0, \ldots, 5 \\ P(Z_i \leq 6) = 1 \end{cases} \tag{4}$$

Subject to

$$\sum_{i=1}^{64} P(Z_i \geq k) = 2^{6-k}, k = 1, 2, \ldots, 6 \tag{5}$$

where $P(Z_i \geq k)$ is the probability that team i wins at least k games. All of the probabilities $P(Z_i \geq k)$, can be put into a matrix having 64 rows and 6 columns. Each row would represent one of the 64 teams. Column 1 would be the estimated probabilities for each team winning at least 1 game. Column 2 would be the estimated probabilities for each team winning at least 2 games. Columns 3 to 6 would be the estimated probabilities for each team winning at least 3 to 6 games. The sum of column 1 equals 32. The sum of columns 2 through 6 equals 16, 8, 4, 2, and 1, respectively. The restriction on the sum of the column does not guarantee the legitimacy of the

model. It is still possible that given two teams playing each other in the first round that the probabilities of each winning the game will not add up to 1 (Shen et al., 2015).

Zhang (2012) filled out his bracket by using the probability self-consistent (PSC) model with Logit link function. $I_{ij}^{(k)}$ is an indicator variable denoting the result of the game between team i and team j in the k^{th} round of the tournament and $p_{ij}^{(k)}$ be the conditional probability of team i defeating team j in the k^{th} round , i.e.,

$$I_{ij}^{(k)} = \begin{cases} 1, & if\ i^{th}\ \text{team wins;} \\ 0, & if\ j^{th}\ \text{team wins.} \end{cases} \tag{6}$$

$$p_{ij}^{(k)} = P\left(I_{ij}^{(k)} = 1 | Z_j \geq k-1, Z_i \geq k-1\right) \tag{7}$$

Then a logistic conditional probability model has been structured as follows,

$$\log \frac{p_{ij}^{(k)}}{1-p_{ij}^{(k)}} = (\boldsymbol{x_i} - \boldsymbol{x_j})' \boldsymbol{\beta^{(k)}} \quad \text{k} = 1,2,\ldots,6 \tag{8}$$

where $(\boldsymbol{x_i} - \boldsymbol{x_j})$ is the vector of the predictor variables spread between team i and team j, and $\boldsymbol{\beta^{(k)}}$ are the associated coefficients in the k^{th} round. These logistic conditional probability models imply

$$\begin{aligned} P(Z_i \geq k) &= P(Z_i \geq k-1)(Z_i \geq k | Z_i \geq k-1) \\ &= P(Z_i \geq k-1) \sum_{j \in O_i^{(k)}} P\left(Z_j \geq k-1\right) p_{ij}^{(k)} \end{aligned} \tag{9}$$

where $O_i^{(k)}$ is the set of all the rival teams that team i may encounter in the k^{th} round. Z_i is self-consistent which means

$$\sum_{j \in U_i^{(k)}} P\left(Z_j \geq k\right) = 1 \tag{10}$$

where $U_i^{(k)} = \bigcup_{j=0}^{(k)} O_i^{(j)}$, and $O_i^{(0)} = i$. Note $P(Z_i \geq 1) = p_{ij}^{(1)}$, then $P(Z_i \geq k)(k > 1)$ can be computed iteratively based on $p_{ij}^{(k)}$ in the logistic regression model. Once all the $P(Z_i \geq k), i = 1, \ldots, 64$ and $k = 1, \ldots, 6$ are obtained, then the team that has the largest $P(Z_i \geq k)$ will be picked as the winning team in $U_i^{(k)}$ of the k[th] round. Like the restricted OLRE model, the PSC model can be written into a matrix form as well. The probabilities $P(Z_j \geq k)$ can be placed in a 64 × 6 matrix. In the probability matrix of the PSC model, not only the sums of each column are required to be 32, 16, 8, 4, 2 and 1, respectively, but also the $P(Z_i \geq k)$ for all possible teams in $U_i^{(k)}$ have to add up to 1.

3. PROBABILITY SELF-CONSISTENT MODEL WITH CAUCHIT LINK

3.1. Introduction of Cauchit Link

In the PSC model, $p_{ij}^{(k)}$ denotes the conditional probability of team i winning against team j in the k^{th} round. Instead of using the Logit link function, this study will use another link function, Cauchit link, to connect the linear predictor with conditional probability $p_{ij}^{(k)}$.

Cauchit link function is another symmetric link function for binary response (Koenker and Yoon, 2009). When using it in the PSC model, the conditional probability model (8) can be structured as follows:

$$\tan(\pi\left(p_{ij}^{(k)} - \frac{1}{2}\right) = \left(\boldsymbol{x_i} - \boldsymbol{x_j}\right)'\boldsymbol{\beta}^{(k)} \text{ k} = 1,2,\dots,6 \tag{11}$$

Comparing with the Logit distribution, the Cauchit distribution has heavier tails, hence the Cauchit link is useful when the value for linear prediction is extreme in either direction. Figure 2 shows the plots of both Cauchit link function and Logit link function. The y axis represents the conditional probability $p_{ij}^{(k)}$ and x axis denotes the linear predictor $\left(\boldsymbol{x_i} - \boldsymbol{x_j}\right)'\boldsymbol{\beta}^{(k)}$, where the vector $(\boldsymbol{x_i} - \boldsymbol{x_j})$ represents difference of the covariates (such as the average assists per game in regular season, and adjusted offensive efficiency from Pomeroy's Ratings) between team i and team j, and $\boldsymbol{\beta}^{(k)}$ are the associated coefficients. In a single game, if two teams have approximately the same strength, they will have very similar values in the covariates, then the vector $(\boldsymbol{x_i} - \boldsymbol{x_j})$ is around $\mathbf{0}$. Therefore, the value of linear predictor of the probability that team i winning the game should be close to zero. In this case, the two link functions, Logit and Cauchit link, have similar performance. However, the difference between the two link functions appears when the linear predictor of team i winning the game has extreme (small negative or large positive) values. In the PSC model, a very large absolute value of linear predictor implies the strengths' of two teams are

not equal since one team must have larger value in some of the covariates than the other team. In extreme cases, when using Logit link function, the conditional probability of the weak team beats the strong team ($p_{ij}^{(k)}$) will be close to 0, while the Cauchit link has a larger value on $p_{ij}^{(k)}$. The Cauchit link gives the weak team a chance of winning the game.

To illustrate it, let us consider a simple example with only two covariates: average assists per game in regular season and adjusted offensive efficiency (AdjO) in Pomeroy's Ratings. Assuming the coefficients for these two covariates are 0.1 and 0.2, respectively, if the weak team (i) has value 14 and 97 for these two covariates, while strong team (j) has larger values of 22 and 118 respectively, then the linear predictor for computing $p_{ij}^{(k)}$ is derived as $(14-22)\times 0.1 + (97-118)\times 0.2 = -5$. When using the Cauchit link functions, $p_{ij}^{(k)}$, the probability of the weak team beating the strong team is 0.0628. Using the Logit link function, the probability is 0.0067.

Overall, we believe it is more appropriate to use the Cauchit link function in sports events rather than Logit link function because this does happen in sports, especially in a one game elimination tournament such as March Madness.

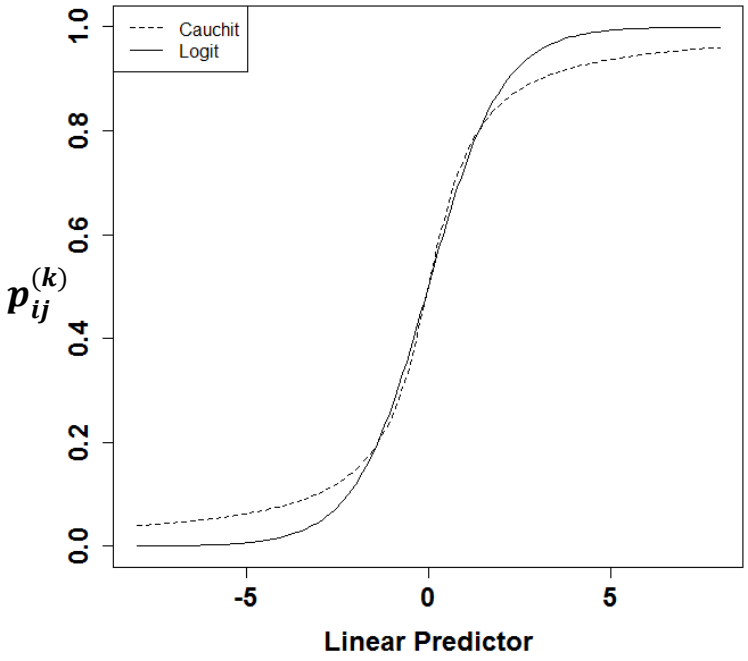

Figure 2. Plots of Cauchit link function and Logit link function

3.2. Application

Magel and Unruh (2013) determined that four in-game statistics such as defensive rebounds and free throw attempts in the regular season are significant in predicting the game results, while in Zhang's research (2012), five candidate covariates, including only one regular seasonal in-game statistics (Assist to Turnover Ratio in Regular Season) were used. Therefore in this study, the total of fourteen covariates, including eight regular seasonal average statistics (ESPN, 2015), were considered for possible usage in the PSC model with Cauchit link. Other than these eight covariates, meanwhile, seed number (ESPN, 2015), ASM (Team rankings, 2015), SAGSOS (Sagarin ratings, 2015), Pyth, AdjO and AdjD (Pomeroy ratings, 2015) have been considered as covariates. Out of these six variables, Seed numbers are decided by the NCAA Basketball Selection Committee based upon the Rating Percentage Index (RPI). ASM is short for

average scoring margin. It highly correlated to the winning percentage. SAGSOS measures the team's opponents' strength. Pyth, AdjO and AdjD come from the Pomeroy Rating system. All the data in our work are collected from 2002-2003 season through 2013-2014 season (12 seasons). The covariates are listed in Table 5.

Table 5. Covariates used in the model

FGM	Field Goals Made Per Game in Regular Season
3PM	3-Point Field Goals Made Per Game in Regular Season
FTA	Free Throws Made Per Game in Regular Season
ORPG	Offensive Rebounds Per Game in Regular Season
DRPG	Defensive Rebounds Per Game in Regular Season
APG	Assists Per Game in Regular Season
PFPG	Personal Fouls Per Game in Regular Season
SEED	Seed Number
ASM	Average Scoring Margin
SAGSOS	Sagarin Proxy for Strength of schedule (Sagarin ratings)
ATRATIO	Assist to Turnover Ratio in Regular Season
Pyth	Pythagorean Winning Percentage (Pomeroy ratings)
AdjO	Adjusted Offensive Efficiency (Pomeroy ratings)
AdjD	Adjusted Defensive Efficiency (Pomeroy ratings)

3.3. Model Selection

Three models are constructed using PSC method with Cauchit link to predict the results in March Madness. The first model was developed for predicting all 32 Rd64 games. The second model was developed for predicting all 16 Rd32 games. Round3 through 6 (Sweet16, Elite8, Final4 and Championship) are combined into one model to overcome the convergence problem in the MLE.

To select the best model that can explain the data in each round (Rd64, Rd32, Sweet16 - Championship), the corrected Akaike Information Criterion (AICc) is applied for model selection with all possible combinations of predictive variables being considered. The computation form is

$$AICc = -2\log -likelihood + \frac{2kN}{k-N-1} \tag{12}$$

k is the number of parameters and N is the number of games involved in fitting the model. Comparing with form of AIC, AICc can be written as

$$AICc = AIC + \frac{2k(k+1)}{N-k-1} \tag{13}$$

Burnham and Anderson (2012) suggested using AICc when the number of covariates is large, especially, when the ratio $\frac{\mathrm{N}}{\mathrm{k}} \leq 40$. The total number of models is $\sum_{k=1}^{14}\binom{14}{k}$=16384 in each round (Rd64, Rd32, Sweet16 - Championship). We will use the model with the smallest AICc as our prediction model in each round.

3.4. Prediction Result

To predict the 2015 NCAA March Madness, 384 Rd64 games (from 2002-2003 season through 2013-2014 season) were used to fit the conditional probability model (8) in order to predict the $p_{ij}^{(k)}$ for 32 Rd64 games. There were 192 Rd32 games (from 2002-2003 season through 2013-2014 season) used to fit the conditional probability model (8) in order to predict the $p_{ij}^{(k)}$ for 16 Rd32 games in 2015 March Madness. There were 96 Sweet16 games, 48 Elite8 games, 24 Final4 games and 12 Championship games (from 2002-2003 season through 2013-2014 season) combined, the total of 180 games, to fit the conditional probability model (8) in order to predict the $p_{ij}^{(k)}$ for the rest of games in 2015 March Madness. Once the predicted $p_{ij}^{(k)}$ was computed, we can put it into model (9) to derive the $P(Z_i \geq k)$ for $k > 1$ (Note $P(Z_i \geq 1) = p_{ij}^{(1)}$).

Table 6. Summary of the three models for PSCM with Cauchit link (2015 March Madness)

Rd64	coefficients	Std. Error	p-value
FGM	0.2565	0.0960	0.0076
DRPG	-0.1617	0.0928	0.0813
APG	-0.2073	0.1027	0.0435
Seed	0.0827	0.0508	0.1035
AdjO	0.3579	0.0693	<0.0001
AdjD	-0.4175	0.0802	<0.0001
SAGSOS	-0.1268	0.0631	0.0446

Rd32	coefficients	Std. Error	p-value
FTA	-0.1031	0.0613	0.0929
Seed	0.2634	0.0964	0.0063
AdjO	0.4165	0.1080	<0.0001
AdjD	-0.5301	0.1337	<0.0001

Sweet16 - Championship	coefficients	Std. Error	p-value
seed	0.1498	0.0794	0.0591
AdjO	0.3489	0.0889	<0.0001
AdjD	-0.3594	0.0983	0.0003
ATRATIO	-2.4113	1.0658	0.0237

Table 6 has the estimated coefficients for the selected model in each round (Rd64, Rd32, Sweet16 - Championship). It is not hard to imagine that some of the in-game statistics are correlated. For instance, usually, with more assists, the team will make more field goals, so these two variables are positively correlated. Hence, the slight collinearity issue cannot be avoided in this study. Even though the slight collinearity does not reduce the predictive power or reliability of the model, it affects the interpretation of the coefficients. It is no surprise that the variable seed, AdjO and AdjD have been selected in all three models. Seed number is decided by the NCAA Basketball Selection Committee and AdjO and AdjD are from Pomeroy's College Basketball Ratings. Both of them can be treated as the experts' opinions. The in-game statistics in each model can be thought of as the adjustment of the experts' judgments.

Table 10 shows the probability matrix using PSC model with Cauchit link. One can fill out the bracket based upon this matrix. The team predicted to advance to the k+1 round is the team with the highest $P(Z_i \geq k)$ in set $U_i^{(k)}$. For instance, in Rd32, $U_i^{(k)} = \{1, 2, 3, 4\}$,$P(Z_1 \geq 2) = 0.9253$, $P(Z_2 \geq 2) = 0.0014$, $P(Z_3 \geq 2) = 0.0604$, $P(Z_4 \geq 2) = 0.0129$. The team predicted to advance to the Sweet16 from these 4 teams is team 1 (Kentucky) since it has the highest $P(Z_i \geq k) = 0.9253$.

Figure 3 gives the predicted bracket based on the probability matrix. Matching up with the true bracket, the wrong teams predicted are highlighted. The accuracy for each round with single and doubling scoring systems are given in Table 7. To compare the Cauchit link PSC model with Logit link PSC model and restricted OLRE model, the other two models were developed using the same covariates. Three different models were constructed for the Logit link PSC model as well. To make the comparison of three methods on equal terms, we used all fourteen variables as possible covariates and the same model selection criteria in the Logit link PSC model (restricted OLRE model used all fourteen covariates directly since the model selection is not required in this method). Table 8 and 9 gives the estimated coefficients for the Logit link PSC model and restricted OLRE model.

Table 7. Prediction accuracy for PSCM with Cauchit link (2015 March Madness)

	Rd64	Rd32	Sweet16	Elite8	Final4	Championship	total	PCT
Correct pick	26	11	5	1	0	0	43	68.25%
Simple	26	11	5	1	0	0	43	68.25%
Doubling	26	22	20	8	0	0	76	39.58%

Table 8. Summary of the three models for PSC model with Logit link (2015 March Madness)

Rd64	coefficients	Std. Error	p-value
ORPG	0.0954	0.0601	0.1128
DRPG	-0.0952	0.0637	0.1347
Seed	0.0626	0.0426	0.1420
AdjO	0.2662	0.0385	<0.0001
AdjD	-0.3135	0.0448	<0.0001
SAGSOS	-0.1126	0.0490	0.0214

Rd32	coefficients	Std. Error	p-value
FTA	-0.0943	0.0519	0.0691
Seed	0.2279	0.0710	0.0013
AdjO	0.3693	0.0647	<0.0001
AdjD	-0.4623	0.0795	<0.0001

Sweet16 - Championship	coefficients	Std. Error	p-value
seed	0.1428	0.0675	0.0345
AdjO	0.3526	0.0620	<0.0001
AdjD	-0.3354	0.0653	<0.0001
ATRATIO	-2.3455	0.8918	0.0085

Table 9 Summary of the restricted OLRE model for 2015 March Madness

Intercepts	Value	Standard error
α_1	-15.2813	1.176
α_2	-13.3504	1.1796
α_3	-11.8942	1.1932
α_4	-10.6899	1.2105
α_5	-9.6174	1.2319
α_6	-8.5883	1.2681

Covariate	FGM	3PM	FTA	AdjO	AdjD	ORPG	DRPG
Coefficient	0.2024	0.0962	0.0473	0.5386	-0.6605	0.0455	-0.0510
Standard error	0.0659	0.0690	0.0367	0.0383	0.0443	0.0555	0.0533
Covariate	PFPG	Seed	ASM	SAGSOS	ATRATIO	APG	Pyth
Coefficient	-0.0104	-0.0128	-0.2026	-0.0616	0.7375	-0.1337	-11.8962
Standard error	0.0532	0.0360	0.0429	0.0282	0.6478	0.0709	1.4901

The selected models for both PSC methods returned similar covariates and coefficients except the model for Rd64 in this year. However, when predicting the $p_{ij}^{(k)}$, it will have greater predicted value when using Cauchit link function than Logit link function if the covariates between team i and team j have large absolute values. The restricted OLRE model developed 6 different models (one for each actual round). The six models shared the same coefficients of the covariates, but each model has its own intercept. All fourteen covariates were used in the restricted OLRE model. However, the convergence problem occurs in this year's prediction with this method. The coefficient of the variable Pyth has a large negative value, which means with higher value of Pyth, the higher the probability that a team will lose. It contradicts with the definition of Pyth in the Pomeroy's ratings . Therefore, the prediction based upon the restricted ORLE model is not reliable in this year.

From the probability matrix (Table 10, 11), it is clear that the probability self-consistency holds for the PSC model with both link functions. For instance, in Rd64, for all 32 games, when two teams are playing each other, the sum of the probabilities of each of the two teams winning the game is one. Furthermore, let us take Midwest Region as an example, all sixteen teams in this region have a chance to get the regional championship. The summation of $P(Z_i \geq 4)$ for all sixteen teams (highlighted in red color in Table 10) is equal to one. The restricted OLRE model does not have this property. From the matrix in Table 12, it is apparent that, the restricted OLRE model does not necessarily satisfy the probability self-consistency. For example, every paired team in the Rd64 has $(Z_i \geq 1) + P(Z_j \geq 1) \neq 1$. In the first game of the Midwest region, the winning probabilities for Kentucky and Hampton are 0.9988 and 0.0346 respectively. The summation of these two probabilities is not equal to one. The comparison of accuracy for these three methods is found in Section 5.

Table 10. PSC model (Cauchit link) probability matrix in 2015 March Madness using MLE

	Seed	TEAM	k=1	k=2	k=3	k=4	k=5	k=6
MIDWEST	1	Kentucky	0.9677	0.9253	0.8722	0.8055	0.6451	0.5658
	16	Hampton	0.0323	0.0014	0.0001	0.0000	0.0000	0.0000
	8	Cincinnati	0.8074	0.0604	0.0373	0.0099	0.0010	0.0001
	9	Purdue	0.1926	0.0129	0.0042	0.0006	0.0000	0.0000
	5	West Virginia	0.7603	0.4419	0.0426	0.0108	0.0011	0.0001
	12	Buffalo	0.2397	0.1156	0.0089	0.0015	0.0001	0.0000
	4	Maryland	0.4940	0.1568	0.0136	0.0028	0.0003	0.0000
	13	Valparaiso	0.5060	0.2858	0.0211	0.0034	0.0003	0.0000
	6	Butler	0.5684	0.1425	0.0497	0.0072	0.0008	0.0001
	11	Texas	0.4316	0.2938	0.1899	0.0343	0.0061	0.0016
	3	Notre Dame	0.9385	0.5596	0.2467	0.0389	0.0053	0.0009
	14	Northeastern	0.0615	0.0041	0.0003	0.0000	0.0000	0.0000
	7	Wichita St	0.8652	0.7056	0.3529	0.0574	0.0082	0.0015
	10	Indiana	0.1348	0.0271	0.0043	0.0004	0.0000	0.0000
	2	Kansas	0.8567	0.2528	0.1542	0.0271	0.0045	0.0010
	15	New Mexico St	0.1433	0.0145	0.0020	0.0002	0.0000	0.0000
WEST	1	Wisconsin	0.9562	0.9038	0.8004	0.2998	0.0853	0.0533
	16	Coastal Carolina	0.0438	0.0036	0.0004	0.0000	0.0000	0.0000
	8	Oregon	0.4915	0.0439	0.0115	0.0010	0.0001	0.0000
	9	Oklahoma St	0.5085	0.0487	0.0209	0.0028	0.0003	0.0000
	5	Arkansas	0.8645	0.1660	0.0178	0.0016	0.0001	0.0000
	12	Wofford	0.1355	0.0126	0.0008	0.0000	0.0000	0.0000
	4	North Carolina	0.8809	0.7724	0.1442	0.0249	0.0036	0.0006
	13	Harvard	0.1191	0.0490	0.0041	0.0003	0.0000	0.0000
	6	Xavier	0.5430	0.1821	0.0177	0.0023	0.0002	0.0000
	11	Ole Miss	0.4570	0.1272	0.0134	0.0019	0.0002	0.0000
	3	Baylor	0.7828	0.6283	0.1088	0.0277	0.0048	0.0010
	14	Georgia St	0.2172	0.0624	0.0044	0.0004	0.0000	0.0000
	7	Virginia Commonwealth	0.3451	0.0317	0.0102	0.0014	0.0001	0.0000
	10	Ohio State	0.6549	0.0794	0.0386	0.0077	0.0010	0.0002
	2	Arizona	0.9620	0.8874	0.8067	0.6280	0.2313	0.1848
	15	Texas Southern	0.0380	0.0015	0.0001	0.0000	0.0000	0.0000
EAST	1	Villanova	0.9590	0.8973	0.7361	0.4609	0.3444	0.0760
	16	Lafayette	0.0410	0.0022	0.0001	0.0000	0.0000	0.0000
	8	NC State	0.6233	0.0619	0.0112	0.0014	0.0002	0.0000
	9	LSU	0.3767	0.0386	0.0065	0.0008	0.0001	0.0000
	5	Northern Iowa	0.9369	0.8020	0.2125	0.0678	0.0266	0.0034
	12	Wyoming	0.0631	0.0076	0.0004	0.0000	0.0000	0.0000
	4	Louisville	0.8858	0.1780	0.0324	0.0058	0.0012	0.0001
	13	UC Irvine	0.1142	0.0124	0.0008	0.0001	0.0000	0.0000
	6	Providence	0.6556	0.1156	0.0156	0.0019	0.0003	0.0000
	11	Dayton	0.3444	0.0633	0.0076	0.0009	0.0001	0.0000
	3	Oklahoma	0.9283	0.8150	0.2294	0.0651	0.0246	0.0031
	14	Albany	0.0717	0.0061	0.0004	0.0000	0.0000	0.0000
	7	Michigan St	0.7437	0.0964	0.0273	0.0040	0.0007	0.0001
	10	Georgia	0.2563	0.0265	0.0073	0.0010	0.0002	0.0000
	2	Virginia	0.9489	0.8741	0.7122	0.3904	0.2821	0.0583
	15	Belmont	0.0511	0.0030	0.0002	0.0000	0.0000	0.0000
SOUTH	1	Duke	0.9472	0.8122	0.5260	0.3433	0.1254	0.0200
	16	Robert Morris	0.0528	0.0031	0.0002	0.0000	0.0000	0.0000
	8	San Diego St	0.7236	0.1466	0.0397	0.0105	0.0014	0.0001
	9	St. John's	0.2764	0.0382	0.0058	0.0009	0.0001	0.0000
	5	Utah	0.8272	0.7531	0.3770	0.2231	0.0706	0.0103
	12	Stephen F. Austin	0.1728	0.1104	0.0213	0.0046	0.0005	0.0000
	4	Georgetown	0.8942	0.1320	0.0297	0.0076	0.0010	0.0001
	13	Eastern Washington	0.1058	0.0045	0.0003	0.0000	0.0000	0.0000
	6	SMU	0.6520	0.2498	0.0658	0.0137	0.0019	0.0002
	11	UCLA	0.3480	0.0917	0.0180	0.0026	0.0003	0.0000
	3	Iowa State	0.9192	0.6519	0.1890	0.0443	0.0070	0.0007
	14	UAB	0.0808	0.0066	0.0005	0.0000	0.0000	0.0000
	7	Iowa	0.4533	0.0679	0.0249	0.0043	0.0005	0.0000
	10	Davidson	0.5467	0.0817	0.0147	0.0016	0.0001	0.0000
	2	Gonzaga	0.9430	0.8465	0.6868	0.3435	0.1106	0.0163
	15	North Dakota St	0.0570	0.0040	0.0003	0.0000	0.0000	0.0000
		summation	32	16	8	4	2	1

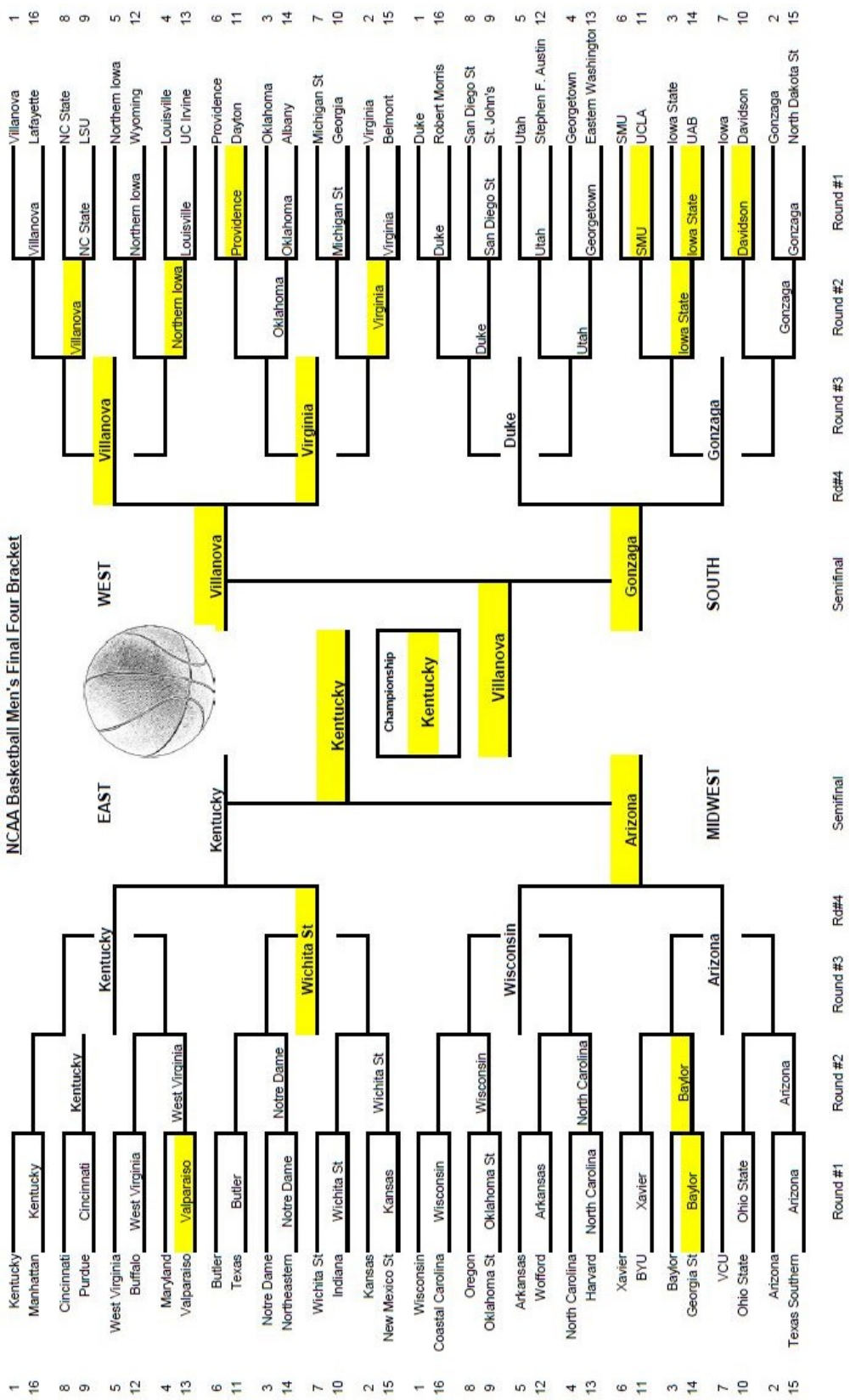

Figure 3. PSC model (Cauchit link) bracket in 2015 March Madness using MLE

Table 11. PSC model (Logit link) probability matrix in 2015 March Madness using MLE

	Seed	TEAM	k=1	k=2	k=3	k=4	k=5	k=6
MIDWEST	1	Kentucky	0.9989	0.9971	0.9923	0.9728	0.7667	0.6976
	16	Hampton	0.0011	0.0000	0.0000	0.0000	0.0000	0.0000
	8	Cincinnati	0.7143	0.0024	0.0014	0.0004	0.0000	0.0000
	9	Purdue	0.2857	0.0005	0.0002	0.0000	0.0000	0.0000
	5	West Virginia	0.7562	0.4172	0.0035	0.0009	0.0000	0.0000
	12	Buffalo	0.2438	0.1147	0.0005	0.0001	0.0000	0.0000
	4	Maryland	0.4330	0.1577	0.0010	0.0002	0.0000	0.0000
	13	Valparaiso	0.5670	0.3105	0.0011	0.0001	0.0000	0.0000
	6	Butler	0.4191	0.1150	0.0408	0.0006	0.0000	0.0000
	11	Texas	0.5809	0.3689	0.2268	0.0076	0.0008	0.0002
	3	Notre Dame	0.9476	0.5152	0.2636	0.0062	0.0005	0.0001
	14	Northeastern	0.0524	0.0008	0.0000	0.0000	0.0000	0.0000
	7	Wichita St	0.8894	0.7171	0.3289	0.0072	0.0005	0.0001
	10	Indiana	0.1106	0.0226	0.0033	0.0000	0.0000	0.0000
	2	Kansas	0.8098	0.2500	0.1357	0.0039	0.0004	0.0001
	15	New Mexico St	0.1902	0.0104	0.0009	0.0000	0.0000	0.0000
WEST	1	Wisconsin	0.9901	0.9833	0.9216	0.3429	0.0560	0.0357
	16	Coastal Carolina	0.0099	0.0003	0.0000	0.0000	0.0000	0.0000
	8	Oregon	0.4269	0.0061	0.0017	0.0000	0.0000	0.0000
	9	Oklahoma St	0.5731	0.0103	0.0046	0.0002	0.0000	0.0000
	5	Arkansas	0.8274	0.1694	0.0039	0.0001	0.0000	0.0000
	12	Wofford	0.1726	0.0116	0.0000	0.0000	0.0000	0.0000
	4	North Carolina	0.8226	0.7388	0.0672	0.0049	0.0001	0.0000
	13	Harvard	0.1774	0.0802	0.0008	0.0000	0.0000	0.0000
	6	Xavier	0.5232	0.1686	0.0038	0.0002	0.0000	0.0000
	11	Ole Miss	0.4768	0.1187	0.0033	0.0002	0.0000	0.0000
	3	Baylor	0.8561	0.6653	0.0525	0.0081	0.0003	0.0001
	14	Georgia St	0.1439	0.0475	0.0004	0.0000	0.0000	0.0000
	7	Virginia Commonwealth	0.4621	0.0055	0.0016	0.0001	0.0000	0.0000
	10	Ohio State	0.5379	0.0256	0.0121	0.0012	0.0000	0.0000
	2	Arizona	0.9967	0.9688	0.9263	0.6422	0.1747	0.1358
	15	Texas Southern	0.0033	0.0000	0.0000	0.0000	0.0000	0.0000
EAST	1	Villanova	0.9972	0.9831	0.8204	0.5294	0.3811	0.0670
	16	Lafayette	0.0028	0.0000	0.0000	0.0000	0.0000	0.0000
	8	NC State	0.5080	0.0079	0.0013	0.0001	0.0000	0.0000
	9	LSU	0.4920	0.0090	0.0012	0.0001	0.0000	0.0000
	5	Northern Iowa	0.9565	0.8250	0.1655	0.0530	0.0205	0.0010
	12	Wyoming	0.0435	0.0042	0.0000	0.0000	0.0000	0.0000
	4	Louisville	0.8627	0.1644	0.0115	0.0015	0.0002	0.0000
	13	UC Irvine	0.1373	0.0064	0.0000	0.0000	0.0000	0.0000
	6	Providence	0.6177	0.0886	0.0057	0.0003	0.0000	0.0000
	11	Dayton	0.3823	0.0594	0.0028	0.0001	0.0000	0.0000
	3	Oklahoma	0.9456	0.8497	0.2045	0.0452	0.0157	0.0007
	14	Albany	0.0544	0.0022	0.0000	0.0000	0.0000	0.0000
	7	Michigan St	0.6622	0.0341	0.0100	0.0009	0.0001	0.0000
	10	Georgia	0.3378	0.0069	0.0018	0.0001	0.0000	0.0000
	2	Virginia	0.9882	0.9588	0.7752	0.3693	0.2386	0.0318
	15	Belmont	0.0118	0.0001	0.0000	0.0000	0.0000	0.0000
SOUTH	1	Duke	0.9809	0.8437	0.5977	0.3853	0.1563	0.0159
	16	Robert Morris	0.0191	0.0002	0.0000	0.0000	0.0000	0.0000
	8	San Diego St	0.7400	0.1377	0.0359	0.0070	0.0005	0.0000
	9	St. John's	0.2600	0.0184	0.0025	0.0002	0.0000	0.0000
	5	Utah	0.7026	0.6653	0.3083	0.1678	0.0516	0.0036
	12	Stephen F. Austin	0.2974	0.1764	0.0271	0.0050	0.0003	0.0000
	4	Georgetown	0.9445	0.1581	0.0285	0.0061	0.0005	0.0000
	13	Eastern Washington	0.0555	0.0002	0.0000	0.0000	0.0000	0.0000
	6	SMU	0.6645	0.2986	0.0619	0.0116	0.0011	0.0000
	11	UCLA	0.3355	0.1092	0.0132	0.0014	0.0001	0.0000
	3	Iowa State	0.8763	0.5879	0.1566	0.0381	0.0052	0.0001
	14	UAB	0.1237	0.0044	0.0001	0.0000	0.0000	0.0000
	7	Iowa	0.6276	0.0546	0.0204	0.0028	0.0002	0.0000
	10	Davidson	0.3724	0.0326	0.0063	0.0004	0.0000	0.0000
	2	Gonzaga	0.9783	0.9124	0.7415	0.3742	0.1277	0.0101
	15	North Dakota St	0.0217	0.0004	0.0000	0.0000	0.0000	0.0000
		summation	32	16	8	4	2	1

Table 12. Restricted OLRE model probability matrix in 2015 March Madness using MLE

	Seed	TEAM	k=1	k=2	k=3	k=4	k=5	k=6
MIDWEST	1	Kentucky	0.9988	0.9890	0.9458	0.8235	0.6112	0.3788
	16	Hampton	0.0346	0.0090	0.0029	0.0008	0.0002	0.0001
	8	Cincinnati	0.3773	0.0987	0.0283	0.0080	0.0026	0.0011
	9	Purdue	0.2734	0.0689	0.0201	0.0056	0.0018	0.0008
	5	West Virginia	0.7354	0.2813	0.0856	0.0255	0.0085	0.0035
	12	Buffalo	0.3473	0.0896	0.0258	0.0073	0.0024	0.0010
	4	Maryland	0.4669	0.1290	0.0369	0.0106	0.0035	0.0015
	13	Valparaiso	0.1833	0.0458	0.0136	0.0038	0.0012	0.0006
	6	Butler	0.6543	0.2196	0.0645	0.0189	0.0063	0.0026
	11	Texas	0.6363	0.2084	0.0609	0.0178	0.0059	0.0024
	3	Notre Dame	0.8999	0.5279	0.1995	0.0647	0.0223	0.0089
	14	Northeastern	0.0497	0.0128	0.0041	0.0011	0.0003	0.0002
	7	Wichita St	0.8098	0.3626	0.1171	0.0357	0.0120	0.0049
	10	Indiana	0.3366	0.0865	0.0249	0.0070	0.0023	0.0010
	2	Kansas	0.9296	0.6148	0.2581	0.0877	0.0307	0.0123
	15	New Mexico St	0.0809	0.0206	0.0064	0.0017	0.0006	0.0003
WEST	1	Wisconsin	0.9960	0.9647	0.8431	0.5898	0.3238	0.1571
	16	Coastal Carolina	0.0460	0.0118	0.0038	0.0010	0.0003	0.0002
	8	Oregon	0.5096	0.1456	0.0417	0.0120	0.0039	0.0017
	9	Oklahoma St	0.4893	0.1375	0.0394	0.0113	0.0037	0.0016
	5	Arkansas	0.6385	0.2097	0.0614	0.0180	0.0059	0.0025
	12	Wofford	0.0959	0.0243	0.0075	0.0020	0.0006	0.0003
	4	North Carolina	0.8797	0.4806	0.1727	0.0549	0.0187	0.0075
	13	Harvard	0.1498	0.0375	0.0113	0.0031	0.0010	0.0005
	6	Xavier	0.6512	0.2176	0.0639	0.0187	0.0062	0.0026
	11	Ole Miss	0.4179	0.1118	0.0320	0.0091	0.0030	0.0013
	3	Baylor	0.8679	0.4565	0.1601	0.0504	0.0172	0.0069
	14	Georgia St	0.1608	0.0402	0.0121	0.0033	0.0011	0.0005
	7	Virginia Commonwealth	0.6870	0.2420	0.0720	0.0212	0.0071	0.0029
	10	Ohio State	0.5287	0.1537	0.0441	0.0127	0.0042	0.0018
	2	Arizona	0.9956	0.9614	0.8303	0.5670	0.3034	0.1450
	15	Texas Southern	0.0365	0.0094	0.0030	0.0008	0.0003	0.0001
EAST	1	Villanova	0.9917	0.9301	0.7260	0.4157	0.1905	0.0839
	16	Lafayette	0.0076	0.0020	0.0007	0.0002	0.0001	0.0000
	8	NC State	0.5123	0.1467	0.0421	0.0121	0.0040	0.0017
	9	LSU	0.4038	0.1071	0.0307	0.0087	0.0028	0.0012
	5	Northern Iowa	0.8106	0.3636	0.1175	0.0359	0.0121	0.0049
	12	Wyoming	0.0549	0.0141	0.0045	0.0012	0.0004	0.0002
	4	Louisville	0.6824	0.2387	0.0708	0.0209	0.0069	0.0029
	13	UC Irvine	0.1285	0.0323	0.0098	0.0027	0.0009	0.0004
	6	Providence	0.6607	0.2238	0.0659	0.0194	0.0064	0.0027
	11	Dayton	0.3004	0.0763	0.0221	0.0062	0.0020	0.0009
	3	Oklahoma	0.9212	0.5877	0.2383	0.0797	0.0277	0.0111
	14	Albany	0.0374	0.0097	0.0031	0.0008	0.0003	0.0001
	7	Michigan St	0.7056	0.2561	0.0768	0.0227	0.0076	0.0031
	10	Georgia	0.4844	0.1356	0.0388	0.0111	0.0037	0.0016
	2	Virginia	0.9829	0.8667	0.5695	0.2637	0.1056	0.0440
	15	Belmont	0.0432	0.0112	0.0036	0.0009	0.0003	0.0002
SOUTH	1	Duke	0.9753	0.8189	0.4829	0.2026	0.0772	0.0316
	16	Robert Morris	0.0535	0.0137	0.0043	0.0012	0.0004	0.0002
	8	San Diego St	0.5197	0.1498	0.0430	0.0124	0.0041	0.0017
	9	St. John's	0.5729	0.1740	0.0502	0.0146	0.0048	0.0020
	5	Utah	0.8762	0.4733	0.1688	0.0535	0.0182	0.0073
	12	Stephen F. Austin	0.2225	0.0556	0.0164	0.0045	0.0015	0.0007
	4	Georgetown	0.7485	0.2934	0.0900	0.0269	0.0090	0.0037
	13	Eastern Washington	0.0710	0.0181	0.0057	0.0015	0.0005	0.0002
	6	SMU	0.5849	0.1800	0.0520	0.0151	0.0050	0.0021
	11	UCLA	0.5360	0.1569	0.0451	0.0130	0.0043	0.0018
	3	Iowa State	0.9152	0.5696	0.2259	0.0748	0.0259	0.0104
	14	UAB	0.1194	0.0300	0.0092	0.0025	0.0008	0.0004
	7	Iowa	0.5995	0.1876	0.0544	0.0158	0.0052	0.0022
	10	Davidson	0.4809	0.1342	0.0384	0.0110	0.0036	0.0015
	2	Gonzaga	0.9641	0.7571	0.3957	0.1520	0.0557	0.0226
	15	North Dakota St	0.0684	0.0175	0.0055	0.0015	0.0005	0.0002
		summation	32	16	8	4	2	1

4. BAYESIAN INFERENCE

In the previous section, there is only one model being used to estimate the results of the entire last four rounds using the PSC method and either the Cauchit or Logit link. However, we can imagine how different between the Sweet16 and Championship on the atmosphere around the arena, the pressures in each player and coach. These can reflect in the spread statistics. So using only one model to explain all games in these four rounds is inappropriate. Therefore if we can develop a model for each round separately, especially on the last few rounds, the estimation would be more accurate.

In the PSC model, insufficiently small sample size is still a large obstacle for developing a model for each round separately. When using maximum likelihood estimation, Griffiths et al. (1987) found that there is significant bias for small samples due to the convergence issue. Even though twelve season's data were collected as training data, there is only one Championship game, two Final4 games, four Elite8 games and eight Sweet16 games in each season, respectively. Hence, the sample sizes for each of the last 4 rounds are not large enough. Developing a separate model for last four rounds could not be accomplished by using maximum likelihood estimation. In this case, another estimation method, Bayesian Inference has been considered to estimate the conditional probability $p_{ij}^{(k)}$.

The one advantage of Bayesian inference is that it does not have the convergence issue when using small size samples. By using Bayesian estimation, we can construct a model for each round separately. In addition, the other advantage of using the Bayesian inference is that it can augment the precision of the prediction for the current year's March Madness by the incorporation of prior information such as experts' opinions and historical results. When there is good prior information, the Bayesian approach will enable one to form the prior distribution with the

consideration of this information. To illustrate these two advantages, the PSC model with Logit link under Bayesian inference will be introduced in this section (In SAS 9.3, Cauchit link is not a build-in link function in the "PROC MCMC" statement, so we only use Logit link function in our study).

Based on the Bayes theorem, we can write (Rashwan and El dereny, 2012)

$$\mathrm{P}(\text{parameters}|\text{data}) \propto \mathrm{P}(\text{parameters})\mathrm{P}(\text{data}|\text{parameters})$$

The first term on the right-hand side is called prior density, often described as "what is known" about the parameters before estimation. The second term is the joint distribution of the observed random variable y given estimated parameters, known as likelihood function. The left-hand side is the posterior density of the parameters given the current data. The posterior density can be seen as a mixture of the prior information and current information (Green, 2003). The Bayesian inference for logistic regression analysis usually follows three steps: The likelihood function of the data is written down; appropriate prior distribution is found over estimated parameters to form the posterior distribution over all parameters; and samples are drawn from posterior distribution and used to estimate the mean value as the coefficients.

4.1. The Likelihood Function

Suppose y is a random variable that follows Bernoulli distribution with probability p. If n independent random variables $(y_1, y_2, \dots, y_n)$ are observed, the generalized linear model is given by

$$g(\mathrm{p}_i) = x_{i1}\beta_1 + \cdots + x_{ik}\beta_k = \eta_i, \quad i = 1, \dots, n \tag{14}$$

where $x_{i1}, \dots, x_{ik}$ are the k covariates, $\beta_0, \beta_1, \dots, \beta_k$ are the estimated intercept and coefficients corresponding to each covariates and n is the total number of games. Here $g(.)$ is a link function,

which connect the linear predictor η_i and the response variable. Using the Logit link, the model can be expressed as

$$g(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \eta_i \tag{15}$$

Solving the equation w.r.t p_i, the result can be written as:

$$p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \tag{16}$$

and the distribution of $f(y_i|\eta_i)$ is given by

$$f(y_i|\eta_i) = (p_i)^{y_i}(1-p_i)^{1-y_i} = \left(\frac{\exp(\eta_i)}{1+\exp(\eta_i)}\right)^{y_i}\left(\frac{1}{1+\exp(\eta_i)}\right)^{1-y_i} \tag{17}$$

Therefore the likelihood function for given covariates and corresponding coefficients is

$$\mathrm{L}(\mathbf{Y}|\boldsymbol{X},\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i|\eta_i) = \exp\left(\sum_{i=1}^{N} y_i\eta_i\right)\prod_{i=1}^{n}\left(\frac{1}{1+\exp(\eta_i)}\right) \tag{18}$$

4.2. The Prior Distribution of Logistic Coefficients

A prior distribution for a parameter is often assessed by experts' judgments and historical results. In the basketball games, even though there exists plenty of prior information, there are only few related with the model coefficients $\boldsymbol{\beta}$. Usually, it is very difficult to introduce a prior distribution on the estimated coefficients $\boldsymbol{\beta}$ immediately. However, we can form a prior distribution on the winning probability, and then relate this prior distribution to the coefficients $\boldsymbol{\beta}$ to get the prior density of $\boldsymbol{\beta}$ indirectly.

Breiter and Carlin (1997) computed the winning probabilities based on seed number. The probability of seed i team beats seed j team is defined as $h_{ij} = \frac{j}{i+j}$. For instance, probability of team with seed number 1 beats the team with seed number 16 is $\frac{16}{16+1} = 0.941$. The seed number is decided by the NCAA selection committee according to the performance in the regular season.

Therefore, this information can be used to form the prior density in Bayesian inference. Table 13 gives the probability matrix based upon the seed number. Notice this probability matrix is identical for each region and tournament year.

In general, let us introduce a variable $h_{ij}^{(l)}$ to represent the probability of team i beats team j for game l in the samples. One possible prior density on winning probability $h_{ij}^{(l)}$ can be formed by assuming each $h_{ij}^{(l)}$ has mean $\hat{h}_{ij}^{(l)}$ and variance $\hat{h}_{ij}^{(l)}(1-\hat{h}_{ij}^{(l)})$, where $\hat{h}_{ij}^{(l)}$ is computed by

$$\hat{h}_{ij}^{(l)} = \frac{seed_j^{(l)}}{seed_i^{(l)} + seed_j^{(l)}} \tag{19}$$

In (19) $seed_i^{(l)}$and $seed_j^{(l)}$ denote the seed number of team i and team j in game l respectively, moreover, $l = 1, \dots, \text{n}$, where n is the total sample size, for example, twelve seasons data were collected in our study, then n=384 for Rd64, n=192 for Rd32, n=96 for Sweet16, n=48 for Elite8, n=24 for Final4, and n=12 for Championship. $\hat{h}_{ij}^{(l)}$ is observed from last twelve seasons in our study. Table 14 gives the matchup information for Championship in recent twelve years. For example, in the last year's March Madness, Connecticut (i) with seed number 7 faced number 8 seeded team Kentucky (j) in the Championship, the winning probability $h_{ij}^{(12)}$ can be assumed follow some distribution with mean $\hat{h}_{ij}^{(12)} = \frac{8}{7+8} = 0.4$ and variance $\hat{h}_{ij}^{(12)}\left(1-\hat{h}_{ij}^{(12)}\right) = 0.24$.

Table 13. Probability matrix based upon the seed number

i \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	.500	.667	.750	.800	.833	.857	.875	.889	.900	.909	.917	.923	.929	.933	.938	.941
2	.333	.500	.600	.667	.714	.750	.778	.800	.818	.833	.846	.857	.867	.875	.882	.889
3	.250	.400	.500	.571	.625	.667	.700	.727	.750	.769	.786	.800	.813	.824	.833	.842
4	.200	.333	.429	.500	.556	.600	.636	.667	.692	.714	.733	.750	.765	.778	.789	.800
5	.167	.286	.375	.444	.500	.545	.583	.615	.643	.667	.688	.706	.722	.737	.750	.762
6	.143	.250	.333	.400	.455	.500	.538	.571	.600	.625	.647	.667	.684	.700	.714	.727
7	.125	.222	.300	.364	.417	.462	.500	.533	.563	.588	.611	.632	.650	.667	.682	.696
8	.111	.200	.273	.333	.385	.429	.467	.500	.529	.556	.579	.600	.619	.636	.652	.667
9	.100	.182	.250	.308	.357	.400	.438	.471	.500	.526	.550	.571	.591	.609	.625	.640
10	.091	.167	.231	.286	.333	.375	.412	.444	.474	.500	.524	.545	.565	.583	.600	.615
11	.083	.154	.214	.267	.313	.353	.389	.421	.450	.476	.500	.522	.542	.560	.577	.593
12	.077	.143	.200	.250	.294	.333	.368	.400	.429	.455	.478	.500	.520	.538	.556	.571
13	.071	.133	.188	.235	.278	.316	.350	.381	.409	.435	.458	.480	.500	.519	.536	.552
14	.067	.125	.176	.222	.263	.300	.333	.364	.391	.417	.440	.462	.481	.500	.517	.533
15	.063	.118	.167	.211	.250	.286	.318	.348	.375	.400	.423	.444	.464	.483	.500	.516
16	.059	.111	.158	.200	.238	.273	.304	.333	.360	.385	.407	.429	.448	.467	.484	.500

Table 14. Matchup information in NCAA March Madness Championship games from season 2002-2003 through 2013-2014 (n=12).

Season	l	Team i (seed number)	Team j (seed number)	Mean of $h_{ij}^{(l)}$	Variance of $h_{ij}^{(l)}$
02 - 03	1	Syracuse (3)	Kansas (2)	0.400	0.24
03 - 04	2	Georgia Tech (3)	Connecticut (2)	0.400	0.24
04 - 05	3	Illinois (1)	North Carolina (1)	0.500	0.25
05 - 06	4	UCLA (2)	Florida (3)	0.600	0.24
06 - 07	5	Florida (1)	Ohio State (1)	0.500	0.25
07 - 08	6	Kansas (1)	Memphis (1)	0.500	0.25
08 - 09	7	Michigan State (2)	North Carolina (1)	0.333	0.222
09 - 10	8	Butler (5)	Duke (1)	0.167	0.139
10 - 11	9	Connecticut (3)	Butler (8)	0.727	0.198
11 - 12	10	Kentucky (1)	Kansas (2)	0.667	0.222
12 - 13	11	Louisville (1)	Michigan (4)	0.800	0.16
13 - 14	12	Connecticut (7)	Kentucky (8)	0.533	0.249

$\log\left(\frac{h_{ij}^{(l)}}{1-h_{ij}^{(l)}}\right)$ was assumed to be normal distribution, the mean and variance derived as

$$log\left(\frac{h_{ij}^{(l)}}{1-h_{ij}^{(l)}}\right)\sim N(\mu_l,\sigma_l^2)=N\left(log\frac{\hat{h}_{ij}^{(l)}}{1-\hat{h}_{ij}^{(l)}},\frac{1}{\hat{h}_{ij}^{(l)}\left(1-\hat{h}_{ij}^{(l)}\right)}\right) \tag{20}$$

by Delta method (Oehlert, 1992).

Now define a column vector $\boldsymbol{h_{ij}}=(h_{ij}^{(1)},h_{ij}^{(2)},\ldots,h_{ij}^{(n)})'$, then $log\left(\frac{\boldsymbol{h_{ij}}}{1-\boldsymbol{h_{ij}}}\right)$ follows a multivariate normal distribution with dimension n.

$$log\left(\frac{\boldsymbol{h_{ij}}}{1-\boldsymbol{h_{ij}}}\right)\sim MVN_n\left(\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{bmatrix},\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \ldots & 0 & \sigma_n^2\end{bmatrix}\right) \tag{21}$$

(21) is the prior density for $log\left(\frac{\boldsymbol{h_{ij}}}{1-\boldsymbol{h_{ij}}}\right)$. Now we need to connect the probability $\boldsymbol{h_{ij}}$ with the coefficient $\boldsymbol{\beta}$ to find the prior density on $\boldsymbol{\beta}$.

Let $\pi(\boldsymbol{\beta})$ denotes the prior distribution of $\boldsymbol{\beta}$. To introduce a $\pi(\boldsymbol{\beta})$ with respect to $\boldsymbol{h_{ij}}$, a logistic function is used to connect $\mathbf{h_{ij}}$ and $\boldsymbol{\beta}$,

$$\log\left(\frac{\mathbf{h_{ij}}}{1-\mathbf{h_{ij}}}\right)=\boldsymbol{X'\beta} \tag{22}$$

where $\mathbf{X}$ is the covariate matrix. The least square approximation for $\boldsymbol{\beta}$ can be written as:

$$\boldsymbol{\beta}=(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\log\left(\frac{\mathbf{h_{ij}}}{1-\mathbf{h_{ij}}}\right) \tag{23}$$

The density of $log\left(\frac{\boldsymbol{h_{ij}}}{1-\boldsymbol{h_{ij}}}\right)$ has already been derived in (21). According to the basic theory of linear algebra on expected value and covariance matrices, the prior distribution of $\boldsymbol{\beta}$ is developed as

$$\boldsymbol{\beta}\sim MVN_n\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{bmatrix},(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \dots & 0 & \sigma_n^2\end{bmatrix}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')'\right)$$

$$= MVN_n\left(\mu_{\boldsymbol{\beta}},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right) \tag{24}$$

Then the probability density function for $\boldsymbol{\beta}$ can be written as

$$\pi(\boldsymbol{\beta}) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\mu_{\boldsymbol{\beta}}\right)'\boldsymbol{\Sigma}_{\boldsymbol{\beta}}{}^{-1}\left(\boldsymbol{\beta}-\mu_{\boldsymbol{\beta}}\right)\right] \tag{25}$$

Notice that this is not the only way that we can form the $\pi(\boldsymbol{\beta})$ based on $h_{ij}^{(k)}$, for instance, we can assume $h_{ij}^{(l)}$ follows some distribution with mean $\hat{h}_{ij}^{(l)}$ and variance $\left[\hat{h}_{ij}^{(l)}\left(1-\hat{h}_{ij}^{(l)}\right)\right]^2$. In this case, the prior distribution of $\boldsymbol{\beta}$ has distribution:

$$\boldsymbol{\beta}\sim MVN_n\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{bmatrix},(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \tag{26}$$

The covariance matrix of prior distribution $\pi(\boldsymbol{\beta})$ is only related with the covariate matrix $\mathbf{X}$.

4.3. The Posterior Distribution of Logistic Coefficient

The posterior distribution is derived by multiplying the density function of prior distribution (25) by the full likelihood function (18). Then the posterior distribution is defined by

$$\pi(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{X}) \propto \pi(\boldsymbol{\beta})\mathrm{L}(\mathbf{Y}|\boldsymbol{X},\boldsymbol{\beta})$$

$$= \prod_{\mathrm{j}=1}^{\mathrm{k}}\left[\frac{1}{\sigma_{\mathrm{j}}\sqrt{2\pi}}\exp\left(-\frac{(\beta_{\mathrm{j}}-\mu_{\beta_{\mathrm{j}}})^2}{2\sigma_j^2}\right)\right]\exp\left(\sum_{i=1}^{N}y_i\eta_i\right)\prod_{i=1}^{N}\left(\frac{1}{1+\exp(\eta_i)}\right) \tag{27}$$

Equation (27) represents the posterior probability distribution of $\boldsymbol{\beta}$, and under these distribution statistical inference can be carried out by using Bayesian method.

4.4. Estimation

The Bayesian point estimate is the parameter vector that minimizes the expected loss function. If the loss function is quadratic form $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2$, then the mean of the posterior probability distribution in (27) is the "minimum expected loss" (MELO) estimator (Green, 2003).

4.5. Sampling Algorithms

Random sampling from posterior distribution is a key step in Bayesian analysis. Nonetheless, only a few well-known probability distributions are ready for use. So some sampling methods have to be used in Bayesian estimation.

The MCMC procedure is a general purpose Markov chain Monte Carlo (MCMC) simulation procedure that is designed to fit Bayesian models. In SAS 9.3 (SAS Institute, Cary NC), the statement "PROC MCMC" uses Metropolis algorithm with a normal proposal distribution to obtain the posterior samples as default (How PROC MCMC Works, 2015). The Metropolis algorithm is named after the American physicist and computer scientist Nicholas C. Metropolis. It is used to obtain random samples from any arbitrary distribution of any dimension (Chib and Greenberg, 1995). Suppose we want to draw samples from a multivariate distribution with full joint posterior probability density function $\pi\,(\beta_1, \beta_2, \dots, \beta_k)$. The first step of Metropolis algorithm is to initialize the sample value for each random variable. The next step is to generate a candidate sample $\beta_1^{(c)}, \beta_2^{(c)}, \dots, \beta_k^{(c)}$ from the proposal distribution $q(.\,|\beta_1^{(i-1)}, \beta_2^{(i-1)}, \dots, \beta_k^{(i-1)})$. And then calculate an acceptance probability

$$\alpha = \min\left(1, \frac{\pi\,(\beta_1^{(c)}, \beta_2^{(c)}, \dots, \beta_k^{(c)} | \mathbf{Y}, \mathbf{X})}{\pi\,(\beta_1^{(i-1)}, \beta_2^{(i-1)}, \dots, \beta_k^{(i-1)} | \mathbf{Y}, \mathbf{X})}\right) \tag{28}$$

Now the acceptance probability for this candidate sample is α (Introduction to Bayesian Analysis Procedures, 2015). The algorithm is self-repeating, so theoretically, it can be carried out as long as required. The algorithm is given in Figure 4. Suppose the m data vectors are accepted

$\left(\beta_1^{(l)}, \beta_2^{(l)}, \dots, \beta_k^{(l)}.\ l = 1,2, \dots, m\right)$ from posterior distribution. By MCMC method, then we can find accurate Bayesian estimate for vector **β** is:

$$\widehat{\boldsymbol{\beta}} \approx \frac{\sum_{l=1}^{m} \beta^{(l)}}{m} \tag{29}$$

Initialize $\beta_0^{(0)}, \beta_1^{(0)}, \dots, \beta_k^{(0)}$
for iteration i = 1, 2, . . . **do**
  Propose$\beta_0^{(c)}, \beta_1^{(c)}, \dots, \beta_k^{(c)} \sim$ q(. $|\beta_1^{(i-1)}, \beta_2^{(i-1)}, \dots, \beta_k^{(i-1)}$)
  Acceptance Probability:

$$\text{á} = \min\left(1, \frac{\pi\,(\beta_1^{(c)}, \beta_2^{(c)}, \dots, \beta_k^{(c)} | \mathbf{Y}, \mathbf{X})}{\pi\,(\beta_1^{(i-1)}, \beta_2^{(i-1)}, \dots, \beta_k^{(i-1)} | \mathbf{Y}, \mathbf{X})}\right)$$

  u $\sim$ Uniform (u; 0, 1)
  if u $<$ α **then**
    Accept the candidate sample: $\beta_1^{(i)}, \beta_2^{(i)}, \dots, \beta_k^{(i)} \leftarrow \beta_1^{(c)}, \beta_2^{(c)}, \dots, \beta_k^{(c)}$
  else
    Reject the proposal: $\beta_1^{(i)}, \beta_2^{(i)}, \dots, \beta_k^{(i)} \leftarrow \beta_1^{(i-1)}, \beta_2^{(i-1)}, \dots, \beta_k^{(i-1)}$
  end if
end for

Figure 4. Metropolis algorithm

4.6. Application

The Bayesian estimation first applies on the conditional probability model (8) in order to predict the $p_{ij}^{(k)}$ in 2015 NCAA March Madness. We will use all covariates mentioned on Section 3. However, in Pomeroy's Ratings, Pyth is derived from both AdjO and AdjD (Pomeroy ratings, 2015). To avoid the severe collinearity problem caused by high correlation between Pyth and AdjO, and Pyth and AdjD, the variable Pyth was excluded. Meanwhile, the variable seed number was used to generate the prior density of coefficients $\boldsymbol{\beta}$ (19). Therefore, it was removed also. Now, the total of twelve covariates were considered in the Bayesian inference. Six models were

developed using PSC method with Logit link. The first model was developed for predicting all 32 Rd64 games. The second model was developed for predicting all 16 Rd32 games. The third, fourth, fifth and sixth round models ware developed for predicting 8 Sweet16 games, 4 Elite8 games, 2 Final4 games and 1 championship game, respectively. All of the data used in this work were collected from 2002-2003 season through 2013-2014 season (12 seasons).

The number of samples from the Metropolis algorithm was assigned 10,000,000 in this work. However, we will not use all these samples because the procedure of Metropolis algorithm cannot guarantee all 10,000,000 samples are independent and identically distributed. Although the Markov chain eventually converges to the target distribution, the initial samples may not converge to the desired distribution (Metropolis–Hastings algorithm, 2015). For this reason, the first half of the iterations were used as a burn-in period, resulting in an initial 5,000,000 iterations thrown away. By doing this, one can guarantee the remaining 5,000,000 samples are from an identical distributions. Since the samples are drawn iteratively, the correlation between two samples from successive iterations is very high. To acquire independent samples, a common strategy is to thin the MCMC in order to reduce sample autocorrelations (Link and Eaton, 2011). In our study, the thinning rate was assigned as 2,000, which means for the remaining 5,000,000 samples, we only keep one sample in every 2,000 samples simulation. Consequently, $\frac{(10{,}000{,}000-5{,}000{,}000)}{2{,}000} = 2{,}500$ samples were collected for each model and now these samples can be considered as from an independent samples from identical distributions. Table 15 gives the summaries of the posterior distribution for the Rd64 model. It reports the posterior mean, standard deviation, 2.5% and 97.5% percentile for the distribution. The mean value is used as estimated coefficients in Bayesian estimation. These 2.5% and 97.5% percentile are known as credible interval in Bayesian statistics (Jaynes, 1976), which is analogous to confidence interval in Non-Bayesian statistics used for

interval estimation. As an example, Figure 5, shows the trace, autocorrelation and the univariate density plot for the first two covariates (FGM, 3PM). The trace plot appears stationary, the autocorrelations are close to zero, and the density plots are relatively smooth. Therefore the 2,500 samples can be assumed from an independent identical distribution. Table 16 gives the estimated coefficient (samples means) for all six models. Once the predicted $p_{ij}^{(k)}$ are computed, using the procedures in Section 3, we need to put the conditional probability $p_{ij}^{(k)}$ into model (9) to derive the $P(Z_i \geq k)$ for $k > 1$ (Note $P(Z_i \geq 1) = p_{ij}^{(1)}$), and then put all $P(Z_i \geq k)$ (k = 1,2, ...,6) into the probability matrix (Table 18). The accuracy using simple and doubling scoring systems for year 2015's bracket (Figure 6) using Bayesian estimation is given in Table 17.

Table 15. Summary table for posterior distribution in 2015 March Madness Rd64 model

Posterior Summaries					
Parameter	**N**	**Mean**	**Standard Deviation**	**2.5% percentile**	**97.5% percentile**
FGM	2500	0.0527	0.0603	-0.0563	0.1774
3PM	2500	0.00648	0.0637	-0.1102	0.138
FTA	2500	0.00512	0.0346	-0.0611	0.0756
ORPG	2500	0.025	0.0482	-0.0666	0.1196
DRPG	2500	-0.0071	0.0512	-0.103	0.0989
APG	2500	-0.0294	0.0667	-0.1587	0.1091
PFPG	2500	0.0404	0.0509	-0.0553	0.144
ATRATIO	2500	0.4239	0.6404	-0.795	1.6478
AdjO	2500	0.2066	0.0338	0.1379	0.2714
AdjD	2500	-0.2361	0.0387	-0.3159	-0.1652
ASM	2500	-0.1216	0.045	-0.2107	-0.0351
SAGSOS	2500	-0.1096	0.0396	-0.1885	-0.0356

Table 16. Estimated coefficients (mean of posterior distribution) for 2015 March Madness using Bayesian inference (standard deviation of posterior distribution given in parenthesis)

Parameter	Rd64	Rd32	Sweet16	Elite8	Final4	Champs
FGM	0.0527 (0.0603)	0.1134 (0.0845)	0.1109 (0.1405)	-0.0225 (0.1946)	0.2021 (0.3965)	0.1956 (1.949)
3PM	0.00648 (0.0637)	0.0793 (0.0861)	0.1223 (0.1272)	-0.1671 (0.2332)	0.6263 (0.5706)	0.7705 (1.0141)
FTA	0.00512 (0.0346)	-0.0265 (0.0531)	0.0257 (0.0739)	0.0535 (0.1223)	-0.1492 (0.2029)	-0.3023 (0.356)
ORPG	0.025 (0.0482)	0.0198 (0.0688)	0.0468 (0.1027)	-0.00949 (0.1212)	0.245 (0.4387)	0.279 (0.4724)
DRPG	-0.00712 (0.0512)	-0.0444 (0.0667)	0.0613 (0.0973)	-0.1072 (0.1947)	0.2132 (0.2277)	0.1641 (3.9657)
APG	-0.0294 (0.0667)	-0.0498 (0.0905)	0.00773 (0.1272)	-0.049 (0.217)	-0.00232 (0.4206)	0.4318 (4.9482)
PFPG	0.0404 (0.0509)	-0.0388 (0.0704)	0.0649 (0.1011)	0.019 (0.1443)	-0.0623 (0.2886)	-0.067 (1.3861)
ATRATIO	0.4239 (0.6404)	-0.2554 (0.9118)	-0.5118 (1.1752)	-0.00566 (2.2451)	-3.4968 (3.3122)	-9.0191 (29.9698)
AdjO	0.2066 (0.0338)	0.2231 (0.0464)	0.3192 (0.0719)	0.2856 (0.0999)	0.4547 (0.1833)	0.3545 (0.5545)
AdjD	-0.2361 (0.0387)	-0.2736 0.0568)	-0.365 (0.0896)	-0.2905 (0.123)	-0.5438 (0.1926)	-0.2109 (1.286)
ASM	-0.1216 (0.045)	-0.0985 (0.0604)	-0.2262 (0.0988)	-0.1008 (0.1342)	-0.4241 (0.264)	-0.3006 (0.6069)
SAGSOS	-0.1096 (0.0396)	-0.0899 (0.0567)	-0.0614 (0.0881)	-0.0948 (0.1101)	-0.2334 (0.2256)	-0.1654 (0.4296)

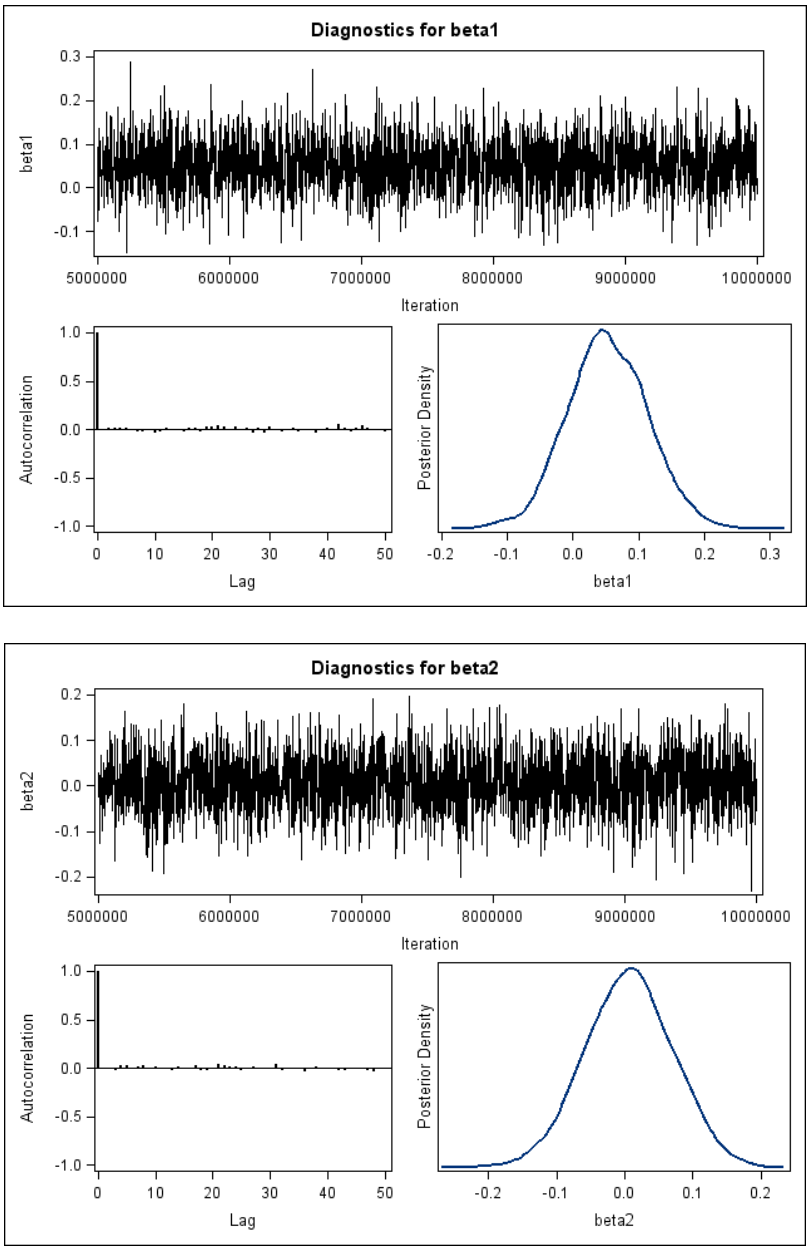

Figure 5. Diagnostics plots for posterior distribution on covariate FGM (beta1) and 3PM (beta2)

Table 17. Prediction accuracy for PSC model (Logit link) using Bayesian estimation in 2015 March Madness

	Rd64	Rd32	Sweet16	Elite8	Final4	Championship	total	PCT
Correct pick	26	9	5	2	0	0	42	66.67%
Simple	26	9	5	2	0	0	42	66.67%
Doubling	26	18	20	16	0	0	80	41.67%

Table 18. PSCM (Logit link) probability matrix in 2015 March Madness using Bayesian Est.

	Seed	TEAM	k=1	k=2	k=3	k=4	k=5	k=6
MIDWEST	1	Kentucky	0.9345	0.8381	0.7667	0.6674	0.2136	0.0324
	16	Hampton	0.0655	0.0049	0.0001	0.0000	0.0000	0.0000
	8	Cincinnati	0.5607	0.1065	0.0355	0.0130	0.0065	0.0016
	9	Purdue	0.4393	0.0505	0.0182	0.0040	0.0009	0.0002
	5	West Virginia	0.6414	0.3985	0.1023	0.0498	0.0251	0.0109
	12	Buffalo	0.3586	0.1622	0.0202	0.0048	0.0016	0.0002
	4	Maryland	0.6070	0.2951	0.0473	0.0131	0.0083	0.0030
	13	Valparaiso	0.3930	0.1442	0.0097	0.0010	0.0004	0.0001
	6	Butler	0.4574	0.1887	0.0591	0.0106	0.0053	0.0010
	11	Texas	0.5426	0.2404	0.0977	0.0160	0.0136	0.0092
	3	Notre Dame	0.8445	0.5518	0.2168	0.0459	0.0157	0.0043
	14	Northeastern	0.1555	0.0191	0.0007	0.0000	0.0000	0.0000
	7	Wichita St	0.6633	0.3136	0.1347	0.0294	0.0093	0.0007
	10	Indiana	0.3367	0.0999	0.0296	0.0019	0.0015	0.0013
	2	Kansas	0.8806	0.5666	0.4605	0.1430	0.1253	0.0700
	15	New Mexico St	0.1194	0.0198	0.0009	0.0001	0.0000	0.0000
WEST	1	Wisconsin	0.9373	0.8180	0.5685	0.3489	0.1976	0.0246
	16	Coastal Carolina	0.0627	0.0062	0.0003	0.0000	0.0000	0.0000
	8	Oregon	0.5385	0.0995	0.0415	0.0051	0.0044	0.0037
	9	Oklahoma St	0.4615	0.0764	0.0294	0.0063	0.0047	0.0024
	5	Arkansas	0.7718	0.3411	0.0804	0.0153	0.0067	0.0017
	12	Wofford	0.2282	0.0416	0.0010	0.0001	0.0000	0.0000
	4	North Carolina	0.7951	0.5466	0.2762	0.0759	0.0515	0.0280
	13	Harvard	0.2049	0.0707	0.0028	0.0003	0.0001	0.0000
	6	Xavier	0.5500	0.2490	0.0628	0.0164	0.0098	0.0031
	11	Ole Miss	0.4500	0.1629	0.0299	0.0065	0.0045	0.0019
	3	Baylor	0.7593	0.5224	0.1937	0.0737	0.0635	0.0504
	14	Georgia St	0.2407	0.0656	0.0027	0.0004	0.0000	0.0000
	7	Virginia Commonwealth	0.6704	0.1606	0.0656	0.0156	0.0114	0.0029
	10	Ohio State	0.3296	0.0654	0.0146	0.0023	0.0005	0.0001
	2	Arizona	0.9445	0.7707	0.6306	0.4333	0.2180	0.0207
	15	Texas Southern	0.0555	0.0033	0.0001	0.0000	0.0000	0.0000
EAST	1	Villanova	0.9663	0.8206	0.7300	0.5028	0.3355	0.2444
	16	Lafayette	0.0337	0.0017	0.0000	0.0000	0.0000	0.0000
	8	NC State	0.5452	0.0971	0.0483	0.0122	0.0083	0.0069
	9	LSU	0.4548	0.0805	0.0365	0.0059	0.0043	0.0040
	5	Northern Iowa	0.8511	0.6502	0.1357	0.0733	0.0409	0.0292
	12	Wyoming	0.1489	0.0320	0.0004	0.0000	0.0000	0.0000
	4	Louisville	0.6417	0.2480	0.0464	0.0148	0.0064	0.0040
	13	UC Irvine	0.3583	0.0699	0.0027	0.0002	0.0001	0.0000
	6	Providence	0.6870	0.2098	0.0756	0.0255	0.0101	0.0066
	11	Dayton	0.3130	0.0662	0.0076	0.0014	0.0001	0.0001
	3	Oklahoma	0.9032	0.7106	0.4309	0.1352	0.1262	0.1134
	14	Albany	0.0968	0.0135	0.0002	0.0000	0.0000	0.0000
	7	Michigan St	0.5560	0.1698	0.0709	0.0130	0.0086	0.0075
	10	Georgia	0.4440	0.0948	0.0304	0.0072	0.0041	0.0029
	2	Virginia	0.8944	0.7192	0.3838	0.2085	0.0916	0.0344
	15	Belmont	0.1056	0.0162	0.0005	0.0000	0.0000	0.0000
SOUTH	1	Duke	0.8671	0.6464	0.4670	0.2991	0.1177	0.0928
	16	Robert Morris	0.1329	0.0145	0.0004	0.0000	0.0000	0.0000
	8	San Diego St	0.4048	0.1490	0.0341	0.0150	0.0049	0.0030
	9	St. John's	0.5952	0.1901	0.0757	0.0275	0.0122	0.0087
	5	Utah	0.6834	0.4418	0.1937	0.1053	0.0307	0.0200
	12	Stephen F. Austin	0.3166	0.0974	0.0152	0.0036	0.0001	0.0001
	4	Georgetown	0.8737	0.4422	0.2132	0.1290	0.0529	0.0385
	13	Eastern Washington	0.1263	0.0186	0.0008	0.0000	0.0000	0.0000
	6	SMU	0.4630	0.1963	0.0636	0.0302	0.0035	0.0021
	11	UCLA	0.5370	0.2254	0.0901	0.0258	0.0156	0.0135
	3	Iowa State	0.7604	0.5285	0.3381	0.1450	0.0806	0.0667
	14	UAB	0.2396	0.0498	0.0046	0.0004	0.0001	0.0001
	7	Iowa	0.5376	0.1935	0.0682	0.0263	0.0071	0.0047
	10	Davidson	0.4624	0.1532	0.0466	0.0049	0.0017	0.0015
	2	Gonzaga	0.8630	0.6314	0.3882	0.1879	0.0367	0.0215
	15	North Dakota St	0.1370	0.0219	0.0006	0.0000	0.0000	0.0000
		Column summation	32	16	8	4	2	1

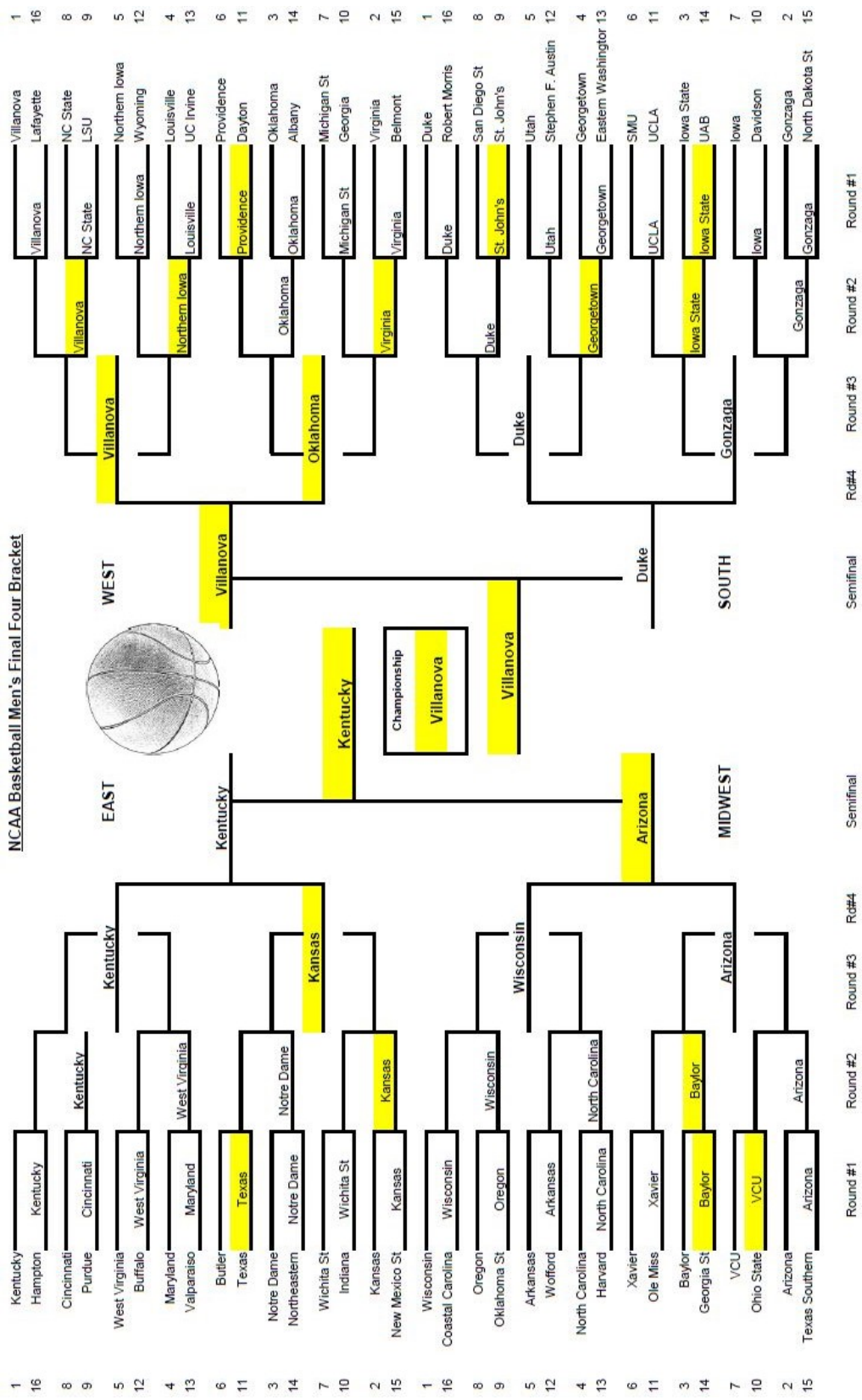

Figure 6. PSC model (Logit link) bracket in 2015 March Madness using Bayesian estimation

5. COMPARISON OF PREDICTION ACCURACY IN THE PAST THREE YEARS

The precision in using the restricted OLRE model (2006, 2008), the PSC model with Cauchit link (MLE), the PSC model with Logit link (MLE, Bayesian estimation) are computed for the last three years. Moreover, other popular prediction methods or rating system such as Pomeroy and RPI ratings are also incorporated into the comparison. In each year, under MLE, the Logit link and Cauchit link PSC methods used all fourteen variables as candidate covariates with same model selection criteria. Both of these methods developed three models: one for predicting all 32 Rd64 games; one for predicting all 16 Rd32 games; and one for predicting all 15 remaining games.

Under Bayesian estimation, twelve covariates were used to construct six models under the PSC method with Logit link. The first model was developed for predicting all 32 Rd64 games. The second model was developed for predicting all 16 Rd32 games. The third, fourth, fifth and sixth models were developed for predicting 8 Sweet16 games, 4 Elite8 games, 2 Final4 games and 1 championship game, respectively. Two prior densities have been used considered in Bayesian inference. Prior 1 was introduced in Section 4, It assumes $h_{ij}^{(l)}$ has mean $\hat{h}_{ij}^{(l)}$ and variance $\hat{h}_{ij}^{(l)}\left(1-\hat{h}_{ij}^{(l)}\right)$. The other prior assumes $h_{ij}^{(l)}$ has mean $\hat{h}_{ij}^{(l)}$ and variance $\left[\hat{h}_{ij}^{(l)}\left(1-\hat{h}_{ij}^{(l)}\right)\right]^2$. The variance of this prior distribution is not related with observed value $\hat{h}_{ij}^{(l)}$. It is only related to the covariate matrix $\boldsymbol{X}$. The restricted OLRE model developed one model for each actual round as well. For Pomeroy and RPI ratings, we just complete the bracket with the value of Pyth and RPI (In a single game, the team with higher Pyth or RPI value will win the game).

Table 19 shows the prediction accuracy in the past three years. In our work, to predict 2015 March Madness and compute the accuracy, March Madness data from 2002-2003 season through 2013-2014 season (12 seasons) are used as the training data to fit the PSC model and restricted OLRE model (Since Pomeroy and RPI rating systems do not need historical data, Pyth and RPI

value from the current year were used to make the predictions). Fourteen covariates were considered in the MLE (Logit, Cauchit link and restricted OLRE model), and twelve covariates were used in the Bayesian estimation model (Logit link). Similarly, 11 seasons' data (from 2002-2003 season through 2012-2013) were used as training data to make predictions and compute the accuracy for 2014 March Madness. And 10 seasons' data (from 2002-2003 season through 2011-2012) were used as training data to make predictions and compute the accuracy for 2013 March Madness. Notice when using Bayesian estimation to predict 2013 and 2014 March Madness, the sample sizes for the last round are only 10 and 11 for these two years. These sample sizes are less than the number of covariates used in the model, and therefore $\boldsymbol{X}'\boldsymbol{X}$ is a singular matrix in (24) when forming the prior density. Hence, the last two rounds have to be combined to acquire a non-singular matrix under Bayesian inference.

Overall, the PSC models has better performances than other popular prediction methods. For some years, the restricted OLRE model has similar accuracy when comparing with the PSC models. However, the restricted OLRE model uses all covariates to fit the model, and this may lead to a risk when using maximum likelihood estimation. The high dimension of covariates results in a greater chance to have the convergence problem. For this reason, the restricted OLRE model has a bad performance for 2015 March Madness. Different from maximum likelihood estimation, Bayesian estimation can avoid the convergence problem even with a high dimension of covariates.

When using the Logit link function, Bayesian estimation with prior 1 and prior 2 correctly predicted almost three and one more games, respectively, than MLE on average because it incorporates the experts' opinions as a prior information. Also in Bayesian estimation, high dimensional parameter was used to explain the data while not having the large risk of convergence issues.

Comparing the PSC model with two different link functions, Cauchit link had better performance in 2014 March Madness because that year's tournament had more lower-seeded teams winning against their higher-seeded opponents. Indeed the two teams with seed number 7 (Connecticut) and 8 (Kentucky) went to the final. Each of these teams beat four higher seeded opponents after Rd64 and then met in the final. This was the first time that the championship game did not have a team with seed number 1,2 or 3. As we mentioned in Section 3, when using Cauchit link, there is always a chance that a weak team can beat a strong team. Therefore when such a case happens several times in a year of the tournament, Cauchit link would have done better.

Table 19. The summary of the prediction accuracy from the 2013 through 2015 March Madness

2013 March Madness	PSCM Logit link (MLE)	PSCM Logit link (Bayesian-prior1[a])	PSCM Logit link (Bayesian-prior2[b])	PSCM Cauchit Link (MLE)	restricted OLRE model	Pomeroy	RPI
Correctly Predicted	47	47	42	47	48	41	34
Simple System	74.60%	74.60%	66.67%	74.60%	76.19%	65.08%	53.97%
Doubling System	75.00%	77.60%	61.98%	75.00%	78.13%	61.46%	28.13%

2014 March Madness	PSCM Logit link (MLE)	PSCM Logit link (Bayesian-prior1)	PSCM Logit link (Bayesian-prior2)	PSCM Cauchit Link (MLE)	restricted OLRE model	Pomeroy	RPI
Correctly Predicted	34	43	40	39	38	38	36
Simple System	53.97%	68.25%	63.50%	61.90%	60.32%	60.32%	57.14%
Doubling System	53.65%	40.10%	35.93%	58.33%	32.29%	31.25%	31.77%

Table 19. The summary of the prediction accuracy from the 2013 through 2015 March Madness (continued)

2015 March Madness	PSCM Logit link (MLE)	PSCM Logit link (Bayesian-prior1)	PSCM Logit link (Bayesian-prior2)	PSCM Cauchit Link (MLE)	restricted OLRE model	Pomeroy	RPI
Correctly Predicted	43	42	45	43	24	43	43
Simple System	68.25%	66.67%	71.43%	68.25%	38.10%	68.25%	68.25%
Doubling System	43.23%	41.67%	48.96%	39.58%	19.80%	40.63%	45.83%

Average	PSCM Logit link (MLE)	PSCM Logit link (Bayesian-prior1)	PSCM Logit link (Bayesian - prior2)	PSCM Cauchit Link (MLE)	restricted OLRE model	Pomeroy	RPI
Correctly Predicted	41.3	44	42.3	43	36.7	40.7	37.7
Simple System	65.61%	69.84%	67.15%	68.25 %	58.20%	64.55%	59.79 %
Doubling System	57.29%	53.12%	48.96%	57.64 %	43.41%	44.45%	35.24 %

a: prior 1 assumes $h_{ij}^{(l)}$ follows some distribution with mean $\hat{h}_{ij}^{(l)}$ and variance $\hat{h}_{ij}^{(l)}\left(1-\hat{h}_{ij}^{(l)}\right)$

b: prior 2 assumes $h_{ij}^{(l)}$ follows some distribution with mean $\hat{h}_{ij}^{(l)}$ and variance $\left[\hat{h}_{ij}^{(l)}\left(1-\hat{h}_{ij}^{(l)}\right)\right]^2$

6. DISCUSSION

Based upon the bracket accuracy of past three seasons' tournament, PSC model has better performance than other methods (restricted OLRE model, Pomeroy, RPI). Comparing two link functions in PSC model, Logit link successfully predicts 124 games out of 189 games and the Cauchit link has 129 games. They have very similar performance when there is no surprises during the whole tournament. However, like 2014 March Madness, with several weaker teams beating stronger teams, the performance of Cauchit link exceeded the performance of the Logit link in both simple and doubling scoring systems.

Bayesian estimation using two different priors for the PSC model with Logit link successfully predicts 132 and 127 games out of 189 games, respectively. Compared with 124 games of maximum likelihood estimation in Logit link. The models using Bayesian estimation have better accuracy in the simple scoring system. However, for the doubling points system, they are no better than the MLE. The reason for this may be that only the seed number is accounted for in the prior density. Seed number may not be significantly important for teams that have already won several tournament games. Additionally, when the prior information is based on empirical data, such as winning probability from the last decade rather than information based only on experts' opinions, then the Bayesian methods are usually less controversial (Bland and Altman, 1998). In future research, one could use a rating system based on both historical results and experts' opinions. The performance of Bayesian inference might be improved in this model. In addition, in SAS 9.3, Cauchit link is not a build-in link function in the "PROC MCMC" statement. The Cauchit link function could not be used in the PSC model using Bayesian estimation. Future research could include building this link function and thereby possibly increasing prediction accuracy.

When using Bayesian estimation, the models with two different prior densities yielded different accuracies. The model with prior 2 has better performance in the 2015 March Madness, while the model with prior 1 has better performance in the two previous years. Hence, future research could also be conducted using different priors.

REFERENCES

[1] Bland, J. M., Altman, D. G. (1998). *Bayesians and frequentists*. BMJ, 317, Oct, 1998, p.1151–1160.

[2] Breiter, D. J., Carlin, B. P. (1997), *How to Play Office Pools If You Must*, CHANCE, 10:1, p.5-11.

[3] Burnham, K. P., Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* 2nd edition, Springer-Verlag, London, United Kingdom.

[4] Chib, S., Greenberg, E. (1995), *Understanding the Metropolis-Hastings Algorithm*, The American Statistician, Vol. 49, No. 4. Nov., 1995, p. 327-335.

[5] ESPN. Retrieved March 16, 2015, from http://espn.go.com/mens-college-basketball/statistics

[6] Green, W.H. (2003), *Econometrics Analysis*, 5th edition, Prentice Hall, NJ.

[7] Griffiths, W. E., Hill, R. C., Pope, P. J. (1987), *Small Sample Properties of Probit Model Estimators*, Journal of the American Statistical Association, 82, p. 929-937.

[8] How PROC MCMC Works. Retrieved May 15, 2015, from http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mcmc_sect019.htm

[9] Introduction to Bayesian Analysis Procedures. Retrieved May 15, 2015, from http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introbayes_sect007.htm

[10] Jaynes, E. T. (1976). *Confidence Intervals vs Bayesian Intervals*, Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, pp. 175 et seq

[11] Koenker, R., Yoon, J. (2009). *Parametric links for binary choice models: A Fisherian – Bayesian colloquy.* Journal of Econometrics, Vol. 152, Issue 2, Oct. 2009, p. 120–130

[12] Link, A.W., Eaton, J. M.(2011), *On thinning of chains in MCMC*, Methods in Ecology and Evolution, Vol. 3, Issue 1, p.112–115

[13] Magel, R., Unruhm S. (2013). *Determining Factors Influencing the Outcome of College Basketball Fames.* Open Journal of Statistics, Vol. 3 No. 4, 2013, p. 225-230

[14] Metropolis–Hastings algorithm. Retrieved May 12, 2015, from http://en.wikipedia.org/wiki/Metropolis%E2%80%93Hastings_algorithm

[15] Moyer, C. (2015, March 12). *Americans To Bet $2 Billion on 70 Million March Madness Brackets This Year, Says New Research. American Gaming Association*. Retrieved from http://www.americangaming.org/newsroom/press-releases/americans-to-bet-2-billion-on-70-million-march-madness-brackets-this-year

[16] Nelson, T.B. (2012). *Modeling the NCAA Tournament through Bayesian logistic regression*. Unpublished Thesis Paper, Duquesne University, Pittsburgh, PA.

[17] NCAA basketball tournament selection process. Retrieved March 20, 2015, from http://en.wikipedia.org/wiki/NCAA_basketball_tournament_selection_process

[18] NCAA Men's Division I Basketball Championship. Retrieved April 10, 2015, from http://en.wikipedia.org/wiki/NCAA_Men%27s_Division_I_Basketball_Championship

[19] Oehlert, G. W. (1992), *A Note on the Delta Method*, The American Statistician, Vol. 46, No. 1, p. 27-29.

[20] Pomeroy ratings. Retrieved March 16, 2015, from http://kenpom.com

[21] Rashwan, N. I., El dereny, M. (2012), *The Comparison between Results of Application Bayesian and Maximum Likelihood Approaches on Logistic Regression Model for prostate cancer Data*, Applied Mathematical Sciences, Vol. 6, 2012, no. 23, p.1143 – 1158

[22] Rating Percentage Index. Retrieved April 27, 2015, from http://en.wikipedia.org/wiki/Rating_Percentage_Index

[23] Sagarin ratings. Retrieved March 16, 2015, from http://www.usatoday.com/sports/ncaab/sagarin/

[24] Schwertman, C.N., Schenk, L.K., Holbrook, C.B. (1996). *More Probability Models for the NCAA Regional Basketball Tournaments.* The American Statistician, Vol. 50, No. 1. , p. 34-38

[25] Shen, G., Hua, S., Zhang, X., Mu, Y., Magel, R. (2015). *Predicting Results of March Madness Using the Probability Self-Consistent Method.* International Journal of Sports Science, 5(4), p.139-144

[26] Team rankings. Retrieved March 16, 2015, from https://www.teamrankings.com/ncaa-basketball/stat/average-scoring-margin

[27] West, B.T. (2006). *A simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament*. Journal of Quantitative Analysis in Sports, 2, 3, p. 3-8.

[28] West, B.T. (2008). *A simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament: Updated Results from 2007*. Journal of Quantitative Analysis in Sports, 4, 2, p. 6-8.

[29] Zhang, X. (2012). *Bracketing NCAA Men's Division I Basketball Tournament*. Unpublished Thesis Paper, North Dakota State University, Fargo, ND.

[30] 2015 NCAA Men's Division I Basketball Tournament. Retrieved April 20, 2015, from http://en.wikipedia.org/wiki/2015_NCAA_Men%27s_Division_I_Basketball_Tournament

APPENDIX A. R CODE FOR PROBABILITY SELF-CONSISTENCY MODEL WITH CAUCHIT LINK

```
options(max.print=100000000)

g<- function(x) !all(x==0)

setwd("E:/NDSU/Su's research/regular season")

BB<-read.csv("NCAA_U5D.csv",header=T)

m<- 12*64

X<- BB[1:m,2:15]

Zeros <- rep(0,14)

Ones <- rep(1,14)

Cs<-expand.grid(Map(c, Zeros, Ones))

Cs<-cbind(Cs,NA, NA)

colnames(Cs) <-

c("V1","V2","V3","V4","V5","V6","V7","V8","V9","V10","V11","V12","V13","V14","AICc","

converge")

tn<- nrow(Cs)


#1st round#

a<- seq(from=1, to=m-1, by=2)

b<- seq(from=2, to=m, by=2)

Z1<- as.matrix(X[a,]-X[b,]);

Y1<- as.matrix(cbind(BB$R1[a],BB$R1[b]))

rowSums(Y1)
```

```
n<- m/2
C1<- Cs
mod.0<- glm(Y1~ 0,family=binomial(link=cauchit))
C1[1,15]<- -2*sum(Y1*log(0.5)); C1[1,16]<- mod.0$converged
for (j in 2:tn) {
WX1<- Z1*matrix(as.numeric(C1[j,1:14]),nrow=n,ncol=14,byrow=T)
WX1<- WX1[,apply(WX1,2,g)]
mod<-glm(Y1~0+WX1,family=binomial(link=cauchit))
fit <- mod$fitted.values
p.hat<- c(fit, 1-fit)
p<- sum(as.numeric(C1[j,1:14]))
C1[j,15]<- -2*sum(Y1*log(p.hat))+2*p*n/(n-p-1)
C1[j,16]<- mod$converged
        }

C1.new <- C1[order(C1$AICc),]
C1.new[1,]
WX.1<- Z1*matrix(as.numeric(C1.new[1,1:14]),nrow=n,ncol=14,byrow=T)
WX.1<- WX.1[,apply(WX.1,2,g)]
cc1.ind<- colnames(WX.1)
mod1<- glm(Y1~0+WX.1,family=binomial(link=cauchit))
```

```
#2nd round#
id2<- which(BB$R1==1)
a<- seq(from=1, to=m/2-1, by=2)
b<- seq(from=2, to=m/2, by=2)
Z2<- as.matrix(X[id2[a],]-X[id2[b],]);
Y2<- as.matrix(cbind(BB$R2[id2[a]],BB$R2[id2[b]]))
rowSums(Y2)
C2<- Cs
n<- m/4
mod.0<- glm(Y2~ 0,family=binomial(link=cauchit))
C2[1,15]<- -2*sum(Y2*log(0.5)); C2[1,16]<- mod.0$converged
for (j in 2:tn) {
WX2<- Z2*matrix(as.numeric(C2[j,1:14]),nrow=n,ncol=14,byrow=T)
WX2<- WX2[,apply(WX2,2,g)]
mod<-glm(Y2~ 0+WX2,family=binomial(link=cauchit))
p<- sum(as.numeric(C2[j,1:14]))
fit <- mod$fitted.values
p.hat<- c(fit, 1-fit)
C2[j,15]<- -2*sum(Y2*log(p.hat))+2*p*n/(n-p-1)
C2[j,16]<- mod$converged
}
C2.new <-C2[order(C2$AICc),]
C2.new[1,]
```

```
WX.2<- Z2*matrix(as.numeric(C2.new[1,1:14]),nrow=n,ncol=14,byrow=T)

WX.2<- WX.2[,apply(WX.2,2,g)]

cc2.ind<- colnames(WX.2)

mod2<- glm(Y2~ 0+WX.2,family=binomial(link=cauchit))



#3rd, 4th, 5th, 6th round#

id2<- which(BB$R2==1)

a<- seq(from=1, to=m/4-1, by=2)

b<- seq(from=2, to=m/4, by=2)

Z3<- as.matrix(X[id2[a],]-X[id2[b],]);

Y3<- as.matrix(cbind(BB$R3[id2[a]],BB$R3[id2[b]]))


id2<- which(BB$R3==1)

a<- seq(from=1, to=m/8-1, by=2)

b<- seq(from=2, to=m/8, by=2)

Z4<- as.matrix(X[id2[a],]-X[id2[b],]);

Y4<- as.matrix(cbind(BB$R4[id2[a]],BB$R4[id2[b]]))


#5th round#

id2<- which(BB$R4==1)

a<- seq(from=1, to=m/16-1, by=2)

b<- seq(from=2, to=m/16, by=2)
```

```
Z5<- X[id2[a],]-X[id2[b],];

Y5<- cbind(BB$R5[id2[a]],BB$R5[id2[b]])


#6th round#

id2<- which(BB$R5==1)

a<- seq(from=1, to=m/32-1, by=2)

b<- seq(from=2, to=m/32, by=2)

Z6<- X[id2[a],]-X[id2[b],];

Y6<- cbind(BB$R6[id2[a]],BB$R6[id2[b]])


Z=as.matrix(rbind(Z3,Z4,Z5,Z6))

Y=as.matrix(rbind(Y3,Y4,Y5,Y6))


rowSums(Y)

C6<- Cs

n<- m/8+m/16+m/32+m/64

mod.0<- glm(Y~ 0,family=binomial(link=cauchit))

C6[1,15]<- -2*sum(Y*log(0.5)); C6[1,16]<- mod.0$converged

for (j in 2:tn) {

WX<- Z*matrix(as.numeric(C6[j,1:14]),nrow=n,ncol=14,byrow=T)

WX<- WX[,apply(WX,2,g)]

mod<-glm(Y~ 0+WX,family=binomial(link=cauchit))

p<- sum(as.numeric(C6[j,1:14]))
```

```
fit <- mod$fitted.values

p.hat<- c(fit, 1-fit)

C6[j,15]<- -2*sum(Y*log(p.hat))+2*p*n/(n-p-1)

C6[j,16]<- mod$converged

}

C6.new <- C6[order(C6$AICc),]

C6.new[1,]


WX<- Z*matrix(as.numeric(C6.new[1,1:14]),nrow=n,ncol=14,byrow=T)

WX<- WX[,apply(WX,2,g)]

cc.ind<- colnames(WX)

mod3<- mod4<- mod5<- mod6 <- glm(Y~ 0+WX, family=binomial(link=cauchit))

cc1<- ncol(WX.1); cc2<- ncol(WX.2);

cc3<- cc4<- cc5<- cc6<- ncol(WX)


########## Bracketing ############

p<- matrix(NA,nrow=64,ncol=6)   #Probability matrix#

w<- BB[(m+1):(m+64),2:15]



#1st round prediction#

a1<- seq(from=1, to=63, by=2)

b1<- seq(from=2, to=64, by=2)
```

```
z<- w[a1,colnames(w) %in% cc1.ind]-w[b1,colnames(w) %in% cc1.ind]; n<- nrow(z)
p[a1,1] <- temp<-
atan(rowSums(z*matrix(as.numeric(mod1$coef),nrow=n,ncol=cc1,byrow=T),na.rm=T))/pi+0.5
p[b1,1] <- 1-temp

#2nd round prediction#
nu<- 2^(6-2)
for (t in 1:nu) {
ri<-2^(2-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;

for (s in (ini+1):mi) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc2.ind]-w[mi+i,colnames(w) %in% cc2.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod2$coef),nrow=n,ncol=cc2,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,1]*p[mi+i,1]*pc
            }
p[s,2]<- psum
             }

for (s in (mi+1): ui) {
psum<- 0
```

```
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc2.ind]-w[ini+i,colnames(w) %in% cc2.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod2$coef),nrow=n,ncol=cc2,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,1]*p[ini+i,1]*pc
            }
p[s,2]<- psum
                }
     }

for (t in 1:nu) {
ri<-2^(2-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;
cat(sum(p[(ini+1):ui,2]), "\n")
            }

#3rd round prediction#
nu<- 2^(6-3)
for (t in 1:nu) {
ri<-2^(3-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;

for (s in (ini+1):mi) {
psum<- 0
for (i in 1:ri) {
```

```
z<- w[s,colnames(w) %in% cc.ind]-w[mi+i,colnames(w) %in% cc.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod3$coef),nrow=n,ncol=cc3,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,2]*p[mi+i,2]*pc
            }
p[s,3]<- psum
             }

for (s in (mi+1): ui) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc.ind]-w[ini+i,colnames(w) %in% cc.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod3$coef),nrow=n,ncol=cc3,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,2]*p[ini+i,2]*pc
            }
p[s,3]<- psum
                }
    }

for (t in 1:nu) {
ri<-2^(3-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;
cat(sum(p[(ini+1):ui,3]), "\n")
```

```
        }

#4th round prediction#
nu<- 2^(6-4)
for (t in 1:nu) {
ri<-2^(4-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;

for (s in (ini+1):mi) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc.ind]-w[mi+i,colnames(w) %in% cc.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod4$coef),nrow=n,ncol=cc4,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,3]*p[mi+i,3]*pc
        }
p[s,4]<- psum
         }

for (s in (mi+1): ui) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc.ind]-w[ini+i,colnames(w) %in% cc.ind]; n<- nrow(z)
```

```
pc<-
atan(rowSums(z*matrix(as.numeric(mod4$coef),nrow=n,ncol=cc4,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,3]*p[ini+i,3]*pc
            }
p[s,4]<- psum
                }
    }

for (t in 1:nu) {
ri<-2^(4-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;
cat(sum(p[(ini+1):ui,4]), "\n")
            }

#5th round prediction#
nu<- 2^(6-5)
for (t in 1:nu) {
ri<-2^(5-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;

for (s in (ini+1):mi) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc.ind]-w[mi+i,colnames(w) %in% cc.ind]; n<- nrow(z)
```

```
pc<-
atan(rowSums(z*matrix(as.numeric(mod5$coef),nrow=n,ncol=cc5,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,4]*p[mi+i,4]*pc
            }
p[s,5]<- psum
             }

for (s in (mi+1): ui) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc.ind]-w[ini+i,colnames(w) %in% cc.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod5$coef),nrow=n,ncol=cc5,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,4]*p[ini+i,4]*pc
            }
p[s,5]<- psum
                }
    }

for (t in 1:nu) {
ri<-2^(5-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;
cat(sum(p[(ini+1):ui,5]), "\n")
            }
```

```
#6th round prediction#
nu<- 2^(6-6)
for (t in 1:nu) {
ri<-2^(6-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;

for (s in (ini+1):mi) {
psum<- 0
for (i in 1:ri) {
z<- w[s,1:14]-w[mi+i,1:14]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod6$coef),nrow=n,ncol=cc6,byrow=T),na.rm=T))/pi+0.5
psum<- psum+p[s,5]*p[mi+i,5]*pc
            }
p[s,6]<- psum
             }

for (s in (mi+1): ui) {
psum<- 0
for (i in 1:ri) {
z<- w[s,colnames(w) %in% cc.ind]-w[ini+i,colnames(w) %in% cc.ind]; n<- nrow(z)
pc<-
atan(rowSums(z*matrix(as.numeric(mod6$coef),nrow=n,ncol=cc6,byrow=T),na.rm=T))/pi+0.5
```

```
psum<- psum+p[s,5]*p[ini+i,5]*pc
            }
p[s,6]<- psum
                }
    }

for (t in 1:nu) {
ri<-2^(6-1); ini<-(t-1)*2*ri; mi<-ini+ri; ui<-ini+2*ri;
cat(sum(p[(ini+1):ui,6]), "\n")
            }

Ex<- rowSums(p)
AC<- rowSums(BB[(m+1):(m+64),16:21])

Q<- matrix(NA,nrow=64,ncol=6)   #bracteting matrix#

for (r in 1:6) {
nu<- 2^(6-r)
for (t in 1:nu) {
ri<-2^r; ini<-(t-1)*ri; mi<-ini+1; ui<-ini+ri;
idmax<- which.max(p[mi:ui,r])+ini
for (s in mi:ui) {
Q[s,r]<- (s==idmax)*1
```

```
        }
       }
      }
colSums(Q)

da<- data.frame(BB[(m+1):(m+64),1],round(p,4),round(Ex,4),Q,AC)
write.csv(da,file="NCAA2014_pred_Cauchy_nn.csv")

###Bracketing and Computing Accuracy###
R<- as.matrix(BB[(m+1):(m+64),16:21])

1-(sum(abs(Q-R))/2)/63  #accuracy of bracketing using single scoring system#
ds<- matrix(c(1,2,4,8,16,32),6,1)
  1-  sum((abs(Q-R)%*%ds)/2)/192  #accuracy of bracketing using double scoring system#
```

APPENDIX B. R AND SAS CODE FOR BAYESIAN INFERENCE WITH LOGIT LINK

B.1. R Code Part

```
options(scipen=999)
setwd("F:/NDSU/Su's research/Bayesian/seed_wtpyth/2015/round5")
BB<-read.csv("round1_matrix.csv",header=T)
attach(BB)
n<-12*64/2
pp<- 1/( p*(1- p))
v<- diag(pp[1:n])
H1<- as.matrix(BB[1:n,1:12])
Ex<-solve(t(H1)%*%H1)%*%t(H1)%*%log(p/(1-p))
Var<-(solve(t(H1)%*%H1)%*%t(H1))%*%v%*%t((solve(t(H1)%*%H1)%*%t(H1)))
round(Ex, 5)
round(Var, 5)
```

B.2. SAS Code Part

```
data U5D;
infile "F:\NDSU\Su's research\Bayesian\NCAA_U5D.csv" dsd missover dlm=',' firstobs=2;
informat Team $30.;
input TEAM  FGM  _3PM  FTA  ORPG DRPG APG  PFPG  seed  AdjO  AdjD  ASM
      SAGSOS  ATRATIO  Pyth  R1  R2  R3  R4  R5  R6  season
$;
a=_n_;
```

```
run;

data
temp1 (rename=(Team=Team1 FGM=FGM1 _3PM=_3PM1 FTA=FTA1 ORPG=ORPG1
DRPG=DRPG1 APG=APG1 PFPG=PFPG1 AdjO=AdjO1 AdjD=AdjD1 ASM=ASM1
SAGSOS=SAGSOS1 ATRATIO=ATRATIO1 Pyth=Pyth1 R1=y seed=seed1))
temp2 (rename=(Team=Team2 FGM=FGM2 _3PM=_3PM2 FTA=FTA2 ORPG=ORPG2
DRPG=DRPG2 APG=APG2 PFPG=PFPG2 AdjO=AdjO2 AdjD=AdjD2 ASM=ASM2
SAGSOS=SAGSOS2 ATRATIO=ATRATIO2 Pyth=Pyth2 seed=seed2));
set U5D;
if mod(a,2)=1 then output temp1;
if mod(a,2)=0 then output temp2;
run;

data R1;
merge temp1 temp2;
FGM=FGM1-FGM2; _3PM=_3PM1-_3PM2; FTA=FTA1-FTA2; ORPG=ORPG1-ORPG2;
DRPG=DRPG1-DRPG2; APG=APG1-APG2;PFPG=PFPG1-PFPG2;ATRATIO=ATRATIO1-
ATRATIO2;
AdjO=AdjO1-AdjO2;AdjD=AdjD1-AdjD2;ASM=ASM1-ASM2;SAGSOS=SAGSOS1-
SAGSOS2;Pyth=Pyth1-pyth2;
p=1-seed1/(seed1+seed2);
where season^='14_15';
```

```
keep TEAM1 TEAM2          FGM    _3PM  FTA    ORPG DRPG APG    PFPG ASM SAGSOS
ATRATIO                   season y p AdjO       AdjD;
run;


proc mcmc data=R1 nbi=5000000 nmc= 5000000 nthin=2000   diag=(mcse ess)
outpost=R1_out; seed=0;
ARRAY beta[12];
parms beta 0;
ODS output PostSummaries=Post;
ARRAY mu0[12]
(0.01473        0.0002 0.00774        0        0.0248 -0.01029        -0.00382        0.11102
        0.06422         -0.07324        0.00225         0.05093);
ARRAY sigma0[12,12]
(0.00749        0.00109         0.00086         -0.00228        -0.00219        -0.00351        -
0.00123         0.00643         -0.00062        -0.00099        -0.00102        0.00001
0.00109         0.00861         0.00133         0.00034         -0.00091        -0.00092        -
0.00205         -0.00463        -0.00047        -0.001          -0.00094        0.00001
0.00086         0.00133         0.00251         -0.00115        -0.00128        0.00001         -
0.0013          0.00705         -0.0006         -0.00026        -0.00028        0.00038
-0.00228        0.00034         -0.00115        0.00486         0.00087         -0.00034
0.0003          0.01111         0.00008         0.00038         0.00021         -0.00028
-0.00219        -0.00091        -0.00128        0.00087         0.00554         -0.00104
0.00199         0.01632         0.00067         0.00003         -0.0005         -0.00058
```

-0.00351 -0.00092 0.00001 -0.00034 -0.00104 0.00918 -
0.00052 -0.04945 0.00088 -0.00031 -0.00035 -0.00054
-0.00123 -0.00205 -0.0013 0.0003 0.00199 -0.00052
0.00516 0.01562 0.00078 -0.00008 -0.00018 -0.00046
0.00643 -0.00463 0.00705 0.01111 0.01632 -0.04945
0.01562 0.86795 -0.0059 -0.0055 -0.01317 -0.00237
-0.00062 -0.00047 -0.0006 0.00008 0.00067 0.00088
0.00078 -0.0059 0.0025 -0.002 -0.00233 -0.00258
-0.00099 -0.001 -0.00026 0.00038 0.00003 -0.00031 -
0.00008 -0.0055 -0.002 0.00314 0.00316 0.00278
-0.00102 -0.00094 -0.00028 0.00021 -0.0005 -0.00035 -
0.00018 -0.01317 -0.00233 0.00316 0.00449 0.00276
0.00001 0.00001 0.00038 -0.00028 -0.00058 -0.00054 -
0.00046 -0.00237 -0.00258 0.00278 0.00276 0.00367);

prior beta ~ MVN(mu0, sigma0);

p =

logistic(beta[1]*FGM+beta[2]*_3PM+beta[3]*FTA+beta[4]*ORPG+beta[5]*DRPG+beta[6]*A
PG+beta[7]*PFPG+beta[8]*ATRATIO+beta[9]*AdjO+beta[10]*AdjD+beta[11]*ASM+beta[12
]*SAGSOS);

model y ~ binary(p);

run;

run;