An Article Submitted to

Journal of Quantitative Analysis in Sports

Manuscript 1039

A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament

Brady T. West*

*University of Michigan, Ann Arbor, bwest@umich.edu

A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament*

Brady T. West

Abstract

This paper first presents a brief review of potential rating tools and methods for predicting success in the NCAA basketball tournament, including those methods (such as the Ratings Percentage Index, or RPI) that receive a great deal of weight in selecting and seeding teams for the tournament. The paper then proposes a simple and flexible rating method based on ordinal logistic regression and expectation (the OLRE method) that is designed to predict success for those teams selected to participate in the NCAA tournament. A simulation based on the parametric Bradley-Terry model for paired comparisons is used to demonstrate the ability of the computationally simple OLRE method to predict success in the tournament, using actual NCAA tournament data. Given that the proposed method can incorporate several different predictors of success in the NCAA tournament when calculating a rating, and has better predictive power than a model-based approach, it should be strongly considered as an alternative to other rating methods currently used to assign seeds and regions to the teams selected to play in the tournament.

KEYWORDS: quantitative reasoning, ratings, sports statistics, NCAA basketball

*The author would like to thank Fred L. Bookstein for the original motivation behind this paper, his colleagues at the Center for Statistical Consultation and Research for their helpful thoughts and suggestions, and freelance writer Dana Mackenzie and four anonymous reviewers for thoughtful and constructive feedback on earlier drafts.

1. INTRODUCTION

Few times are as celebrated in American sports culture as the time known to fans of college basketball as "March Madness." March Madness, or the month-long extravaganza that is the National Collegiate Athletic Association (NCAA) Division I men's basketball tournament, features 65 teams of young men representing their respective universities and playing basketball games, strictly for one prize: the national championship of men's college basketball. For this unbelievably popular American social event to even take place, the 34 teams that are the most deserving to play for the national championship must first be selected from the pool of Division I men's basketball teams that did not win conference championships (the 31 conference champions have automatic bids), and the 65 teams must then be seeded in order to determine what teams will play each other and where the teams will play. Each year, a small NCAA-appointed committee considers and juggles several different numbers and factors (both subjectively and objectively) in deciding how to seed the 65 teams, and the resulting seedings and tournament pairings cause a great deal of controversy every year.

The 65 teams that earn a spot in the tournament each year are divided by the selection committee into three regions of 16 teams and one region of 17 teams, and the teams in each region are assigned a seed by the committee, ranging from 1 (highest, or strongest team) to 16 (lowest). In the first round of the tournament in each region, the 1 seed plays the 16 seed, the 2 seed plays the 15 seed, and so forth. In the region with 17 teams, a recently instituted opening round game pairs two teams with a 16 seed, and those two teams play to be the single 16 seed in that region. Smith and Schwertman (1999) used simple regression models fitted to historical tournament data (the first 12 years of the 64-team format) to show that a team's seed is a strong predictor of margin of victory in games played in the tournament. This work suggests that the seeds assigned by the committee are strong indicators of the relative strengths of the teams. The teams winning four straight games in each region win the regional championships and advance to the "Final Four," as it is referred to by the media, and compete in a single-elimination format for the national championship.

Because the selection committee's seedings are only partly objective and mostly subjective, it is quite important that the numerical ratings used objectively by the committee to make seeding decisions are strong indicators of actual success in the tournament. Alternative numerical rating methods at the selection committee's disposal can be compared by how well they predict success in the NCAA tournament, and a method specifically designed to produce a rating that represents an expectation of success in the tournament could be a quite useful tool for the committee. The American public expects that the NCAA tournament selection committee will seed the 65 teams selected for the tournament in a way that results in balanced regions and highly competitive games, and the use of established, objective statistical methods to develop rating methods based on predicted tournament success could be of great benefit to this committee (in addition to exciting practice for quantitative researchers).

The primary purposes of this paper are twofold: first, to discuss and compare different quantitative methods that have been designed to either assign a rating to an NCAA basketball team or predict actual success in the NCAA tournament; and second, to highlight the simplicity, flexibility, and effectiveness of a proposed rating method based on ordinal logistic regression and the expected value of a discrete random variable. The proposed rating method can be used to calculate the expected number of wins for teams that are selected for the tournament, and could therefore be used by the NCAA selection committee to establish both the seeds and regions of the selected teams according to predicted success.

2. POTENTIAL RATING METHODS

Stern (2004) provides a nice primer on computer ratings and rating methods with respect to college sports in general, and compares a handful of different rating methods proposed by statisticians in terms of their ability to correctly predict the outcomes of games in a variety of collegiate and professional sports. Stern correctly points out that human opinions and polls can tend to be biased. Despite this obvious bias, many of the NCAA men's basketball selection committee's decisions continue to be subjective in nature (NCAA 2005; Mihoces 2002). Objective rating methods based on well-established rules and relevant information can help to eliminate some of this bias, and some alternative rating methods considered in the context of college basketball are discussed in this section.

2.1 The RPI

A particular rating method that continues to receive a great deal of weight in selecting and seeding teams is the ratings percentage index, or RPI (NCAA 2005; www.collegerpi.com). In more NCAA sports than just men's basketball, this rating method is used to compare teams based on their winning percentages, and the winning percentages of their opponents. During the 2004-2005 NCAA basketball season, the NCAA basketball committee modified arbitrarily the RPI and made the decision to weigh road wins more heavily than home wins (NCAA 2005) in an effort to improve this tool. Despite this revision, the RPI continues to receive a great deal of criticism, due to its arbitrary calculation method and exclusion of other important factors that may be helpful in rating teams (Stern

2004). In spite of these criticisms, the RPI continues to play a large role every year in the selection and seeding of teams for the NCAA basketball tournament.

The ratings percentage index (RPI) for a given college basketball team *i* is calculated by the NCAA as follows:

$$\mathrm{RPI}_i = 0.25 \times \mathrm{WP}_i + 0.50 \times \mathrm{OAWP}_i + 0.25 \times \mathrm{OOAWP}_i \qquad (1)$$

In (1), WP = Winning Percentage, OAWP = Opponents' Average Winning Percentage, and OOAWP = Opponents' Opponents Average Winning Percentage. During the 2004-2005 season, the NCAA basketball committee approved a revision to this formula, weighting home wins by 0.6, home losses by 1.4, road wins by 1.4, and road losses by 0.6 when calculating the winning percentages. This revision was designed to take the site of a game into account when calculating the ratings. The use of these weights, as well as the weights in the original RPI formula in (1), however, appears to be quite arbitrary. The same formula in (1) was still used by the selection committee in calculating a team's RPI throughout the 2004-2005 season, but the winning percentages were calculated differently using the weights above. Despite its flaws (for example, winning a game against a poor opponent may actually cause a team's rating to decrease), the RPI is a simple example of a completely objective measure that does not take the bias of human polls into account when rating the basketball teams.

The RPI continues to receive a lot of weight from the NCAA selection committee when decisions are being made about selecting and seeding the teams, especially after the 2005 adjustment. According to Bob Bowlsby, the Director of Athletics at the University of Iowa and the chair of the selection committee in 2005, "The committee continues to view the RPI as one of many pieces of available information used in the process…we want the RPI to be the best tool possible. We believe this adjustment accomplishes that" (NCAA 2005).

2.2 Jeff Sagarin's Computer Ratings

The NCAA selection committee also uses another objective rating method to help distinguish between teams. Jeff Sagarin, who graduated with a degree in mathematics from the Massachusetts Institute of Technology and an MBA from Indiana University, has been supplying the selection committee with his unique computer ratings of the teams since 1984. In short, his overall ratings for NCAA basketball teams are based on a synthesis of two different ratings: Sagarin's personal modification of the chess rating system developed by Elo (1978), which only takes wins and losses into account, and a rating method developed by Sagarin known as the PURE POINTS method, which takes a team's scoring performance into account. It is worth mentioning here that Sagarin's computer ratings are respected enough by the NCAA to play a role in the calculation of the now infamous BCS computer ratings in college football. Sagarin computes his ratings[1] after every single Division I men's basketball game, meaning that the ratings are readily available for the selection committee at any point in the season, much like the RPI.

2.3 Previous Work by Other Quantitative Researchers

A survey of the statistics literature reveals several interesting ideas proposed by statisticians for either predicting success in the tournament or assigning ratings to the teams. In the context of an interesting tool for teaching probability to students with real-world data, Schwertman et al. (1991) introduced three ad hoc probability models that did very well at predicting the number of teams seeded 1, 2, 3, or 4 or higher that would win regional championships (based on the known seeds of 24 regional champions from six years of NCAA tournament data). Schwertman et al. (1996) then considered the use of eight additional logistic and ordinary least squares regression models to further develop the probability models, using linear and non-linear functions of seeding information and tournament outcome data from the first 10 years (1985-1994) that the tournament was played using a 64-team format. This paper determined the most effective probability models based on the empirical data, depending on the objective of the predictions (predicting winners of individual games, or predicting regional champions). Carlin (1996) built on the work of Schwertman and others to present a method that incorporates point spread information as well as the seeds of the teams in producing predicted probabilities of winning the regional championship (or advancing to the Final Four) for each team selected into the tournament. Carlin shows that his method provides the best fit to actual 1994 tournament results out of five possible methods, including the method of Schwertman et al. (1991).

Caudill and Godwin (2002) presented an interesting application of the "skewed logit" model (BarNiv & McDonald 1999), or a binary logit model based on an asymmetric error structure, in predicting outcomes in the NCAA tournament. Specifically, logit models (assuming a symmetric logistic distribution for the errors in the model), probit models (assuming a symmetric normal distribution for the errors), and skewed logit models with and without heterogeneous skewness (allowing the skewness parameter in the model to vary from observation to observation, or game to game) were fitted to NCAA tournament data from 1985 to 1998 (with 60 games per year, or 15 in each of the

[1] Exact details on Sagarin's method are not publicly available.

four regions leading to the four teams in the Final Four). Comparisons of the four fitted models using absolute prediction errors and likelihood ratio tests indicated that the model allowing for heterogeneous skewness as a function of the higher seed in each game resulted in better prediction probabilities, and provided an improved fit to the observed data.

Bassett (1997) provides a nice review of previous rating methods proposed in the statistics literature, focusing on methods that estimate ratings in linear models using least squares techniques, and proceeds to present a method based on least absolute error that can be used to estimate the ratings of teams. In this method, team ratings are formulated as parameters in a linear model predicting differences between home and away teams in the scores of games. Bassett proposes an estimator for the ratings based on least absolute error (as opposed to least squares, in order to reduce the influence of outliers), which he shows to be the median of a team's normalized scores (which are normalized to adjust for home-field advantage and opponent strength). Bassett applies his method to data from the 1993 National Football League (NFL) regular season, developing ratings for each team using all scores from the regular season. With some work, this method could be extended to the NCAA Division I regular season; the resulting ratings for the teams could be used by the selection committee for seeding purposes.

3. THE OLRE METHOD

This section proposes a simple and flexible rating method based on ordinal logistic regression (OLR) modeling and expectation (E) that can be used to calculate ratings for the teams selected to play in the NCAA tournament as a function of team-level predictor variables. The calculated ratings represent predictions of how many wins each team selected for the tournament will achieve. The method is *simple* in that it is based on the estimation of standard ordinal logistic regression models and calculation of expected values for discrete random variables, and it only requires the collection of team-level data for 64 teams in a given season. The method is *flexible* in that it easily allows one to incorporate additional auxiliary information for the teams in developing models that can be used to assign ratings to the teams.

The purpose of this proposed rating method is not to predict which teams will be selected for the NCAA tournament, but rather to predict *success* in the NCAA tournament (provided that the team is selected by the committee). Stern (2004) points out that the information used in a rating system can be effectively increased by utilizing information from previous seasons, and this method explores that notion further. The OLRE method considers a multivariate set of historical data collected on teams selected for the tournament at the ends of several regular seasons, and examines how patterns of success in the NCAA tournament can be derived using that information. The method then uses these patterns of success to develop ratings for the teams, where the ratings are essentially predictions of how successful a team would be in the tournament given their current team-level information. This method could effectively be utilized by the committee in helping to distinguish between the teams that are selected for the tournament when assigning seeds and regions to each team.

To begin, one can collect any amount of historical team-level data (e.g., winning percentage, wins against top-rated teams, point differential, etc.) at the end of the *regular* season (including conference tournaments) for the cohorts of 64 teams selected to participate in the NCAA tournament in prior years. It is important that the historical team-level data are collected prior to the onset of the NCAA tournament in any given year, and information from multiple seasons can be utilized. The loser of the now-instituted opening round game between the two 16 seeds in the region with 17 teams should be excluded from this data set, and this game is considered as an additional regular season game. In the application of the OLRE method discussed in this paper, the team-level variables considered were a team's winning percentage (WINPCT), a team's point differential for the season (points for minus points against, or DIFF), a strength of schedule metric computed by Jeff Sagarin (SAGSOS), and the number of wins against Top 30 teams according to Sagarin's ratings at the *end of the regular season* (TOP30WIN). Information on these team-level variables was collected for the cohorts of 64 teams selected for the NCAA tournament after each of the 2002-2003, 2003-2004, 2004-2005, and 2005-2006 regular seasons. The team-level information was based on games against Division I opponents only. One of the nice features of the OLRE method is that additional relevant team-level variables (e.g., road winning percentage) could be collected to strengthen the fits of the models and the resulting predictions based on the method.

The number of wins achieved by each team in the past tournaments (WINS) is then recorded in the data set described above. One can then consider the WINS variable as an ordinal dependent variable in an ordinal logistic regression model, with WINPCT, DIFF, SAGSOS, and TOP30WIN as the predictor variables. In this model, the predicted probability of team i winning j games (j = 0, 1, 2, 3, 4, 5) can be written as

$$\pi_{ij} = \frac{\exp(\alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta})}{1 + \exp(\alpha_j + \boldsymbol{x}_i'\boldsymbol{\beta})} - \sum_{k=0}^{j-1} \pi_{ik}, \tag{2}$$

where α_j is the intercept for the j-th outcome, $\boldsymbol{x}_i$ is a vector of values for team i on the team-level predictor variables, $\boldsymbol{\beta}$ is a vector of coefficients associated with the predictor variables, and the last term represents the cumulative sum of the predicted probabilities of winning k games ($k = 0,\ldots, j$-1; this term would be equal to 0 for $j = 0$). The predicted probability of achieving six wins would simply be calculated as 1 minus the sum of the six predicted probabilities based on the formula above. Any statistical software can be used to obtain maximum likelihood estimates of the coefficients in the model (PROC LOGISTIC in SAS/STAT Version 9.1.3 was used for this paper).

In the *current* year, at the end of the regular season and before the tournament begins, the same variables are collected once again for the 64 teams selected for the tournament (excluding the loser of the opening round game between the two 16 seeds). Predicted probabilities of obtaining zero, one, two, three, four, five, and six wins in the upcoming tournament are then calculated for each of the 64 teams, based on the ordinal logistic regression model estimated using the historical tournament data (2). The predicted probabilities of winning j games are thus based on the patterns of success for selected teams in the *previous* years. Because the model is fitted using historical data for teams selected for the tournament, this method does not apply to teams that are not selected for the tournament.

This results in a 64 × 7 matrix of predicted probabilities, where the sum of the predicted probabilities for each row (team) is equal to 1. However, the number of teams that will win 0, 1, 2, 3, 4, 5, and 6 games is known prior to the tournament, since 32 teams will win 0 games, 16 teams will only win one game, eight teams will only win two games, and so forth. Calculating the predicted probabilities based on the fitted ordinal logistic regression model does not guarantee that the column sums will be equal to these totals, which are known prior to the tournament. The predicted probabilities therefore need to satisfy known marginal constraints, where the sum of each row must be equal to 1, and the sums of each column must be equal to 32, 16, 8, 4, 2, 1, and 1, respectively.

One can therefore consider the 64 × 7 matrix of predicted probabilities as an observed contingency table, with cell probabilities that need to be estimated in a way that satisfies these known marginal constraints. A program written for the R software package (Lang 2004), capable of solving this problem by using maximum likelihood estimation to fit Multinomial-Poisson Homogeneous Models to contingency tables (Lang 2005), can be used to adjust the predicted probabilities for each team so that they satisfy the known marginal constraints of a tournament in any given year.[2]

Once the seven adjusted predicted probabilities have been calculated for each team, the definition of an expected value for a discrete random variable (e.g.,

---

[2] The R program used to implement this technique (`mph.fit`) is available from Joseph Lang (jblang@stat.iowa.edu), and the specific application of this program used to adjust the 64 × 7 = 448 predicted probabilities in 2006 based on the OLRE method is available from the author.

Rice 1995) can be applied to calculate the expected number of wins for a team in the tournament. The expected value of WINS can now be calculated for a given team i:

$$E_i[\text{WINS}] = \sum_{j=0}^{6} j \times \hat{\pi}_{ij} \tag{3}$$

These expected values, representing the expected number of wins for each team, are defined as each team's rating based on the OLRE method.

4. APPLICATION OF THE OLRE METHOD

Prior to the onset of the NCAA tournament in March 2006, data were collected on the following team-level predictor variables for the 192 teams selected into the tournament in 2003, 2004, and 2005: winning percentage at the end of the regular season (WINPCT), point differential at the end of the regular season (DIFF), Sagarin's strength of schedule metric (SAGSOS), and the number of wins against teams rated in the Top 30 based on Sagarin's ratings at the end of the regular season (TOP30WIN). Once again, it is important to note that information could be collected on additional team-level predictor variables (e.g., road winning percentage) in future applications of the OLRE method, but additional variables were not collected for this paper. The number of wins achieved in the 2003, 2004, and 2005 tournaments by each of the 192 teams was also recorded. An ordinal logistic regression model was fitted to the tournament outcome data, and equations similar to (2) that could be used to develop predicted probabilities of success in 2006 were obtained. The estimated coefficients in the model fitted to the 2002-2005 data are displayed in Table 1.

Information was then collected on the same four predictor variables for the 64 teams selected for the 2006 tournament (excluding the loser of the opening round game between the 16 seeds). Predicted probabilities of obtaining zero, one, two, three, four, five, and six wins were calculated for each of the 64 teams based on the estimated 2002-2005 regression model, and the predicted probabilities were adjusted to satisfy the known marginal constraints for the 64 teams (or the number of teams that would win 0, 1, 2, 3, 4, 5, and 6 games, as described earlier). The adjusted predicted probabilities were then used to calculate an expected number of wins in the 2006 tournament for each team, as described in (3). An expected number of wins for each team based on *unadjusted* predicted probabilities was also calculated, to assess whether the adjustment to satisfy the known constraints of the tournament would improve the predictive power of the OLRE method.

5. VALIDATION OF THE OLRE METHOD

To assess the predictive power of the OLRE method, the number of wins achieved by each team in the 2006 NCAA tournament was recorded. In 2006, ratings for the teams participating in the tournament were calculated *prior to the onset of the tournament* by applying the OLRE method to historical data from the 2002-2003, 2003-2004, and 2004-2005 seasons. The sum of squared errors between the actual number of wins and the expected number of wins based on the OLRE method (both unadjusted and adjusted to satisfy the known marginal constraints of the tournament) was then computed as a measure of the predictive power of the method.

To further validate the OLRE method, a program was written using the R software package to simulate the results of one million tournaments using the 64 teams selected (and their known seeds and pairings) in 2006.[3] The outcome of each game in the simulated tournaments was determined by calculating the probability that team i would defeat team j using the Bradley-Terry model (Bradley and Terry 1952), and then drawing a random value from the Uniform(0,1) distribution. If the random value was less than or equal to the probability based on the model, then team i would advance in the tournament. The strength parameters for the 64 teams used to calculate the probabilities of victory based on the Bradley-Terry model were the *ranks* of the final pre-tournament Sagarin ratings for the teams (where the top-rated team, Duke, was assigned a strength parameter of 64, and the bottom-rated team, Southern, was assigned a strength parameter of 1). The simulation resulted in an estimate of the joint probability distribution of wins for each team, and these distributions were used to calculate the expected number of wins in the 2006 tournament for each team using the formula in (3). The sum of squared errors between the actual number of wins and the expected number of wins based on the simulation was then computed for comparison with the OLRE method.

Table 2 presents the resulting sums of squared errors for each of the three different methods for calculating the expected number of wins (or ratings) in the 2006 tournament. The ratings based on the OLRE method (with an adjustment to the predicted probabilities to satisfy the marginal constraints defined by the known tournament outcomes) were found to have the smallest sum of squared errors, followed by the unadjusted OLRE ratings and then the expected number of wins based on the simulation. These results suggest that the OLRE method provides an improved fit to the observed tournament outcome data relative to a model-based method for calculating the expected number of wins for each team.

[3] The R program used for the simulation and all data sets referenced in this paper are available on the author's Web site (http://www.umich.edu/~bwest).

Table 1. Estimated Ordinal Logistic Regression Model in 2006

Predictor	Estimated Coefficient	Standard Error	*p*-value
WINPCT	-10.8485	3.0717	0.0004
DIFF	-0.0062	0.0023	0.0057
SAGSOS	-0.8758	0.1286	< 0.0001
TOP30WIN	0.2576	0.1102	0.0194
INTERCEPT 0	76.6847	11.0695	< 0.0001
INTERCEPT 1	78.7468	11.1610	< 0.0001
INTERCEPT 2	80.1007	11.2083	< 0.0001
INTERCEPT 3	81.2570	11.2450	< 0.0001
INTERCEPT 4	82.2572	11.2755	< 0.0001
INTERCEPT 5	83.1525	11.3015	< 0.0001

Nagelkerke $R^2 = 0.562$.
p-values are for Wald Chi-square statistics.

Table 2. Sums of Squared Errors for Assessing and Comparing Predictive Power

Method	Sum of Squared Errors
OLRE (unadjusted)	69.3565
OLRE (adjusted)	63.9152
Bradley-Terry	70.0237

OLRE (adjusted) refers to the estimation of predicted probabilities that satisfy the marginal constraints defined by the known tournament outcomes, per the work of Lang (2004).

6. CONCLUSIONS

This paper presents a brief review of some potential methods for predicting success in the NCAA tournament that have been proposed by both statisticians and non-statisticians, and proposes a simple and flexible rating method based entirely on ordinal logistic regression modeling and expectation (the OLRE method) that can be used to calculate team ratings representing the expected number of wins that a team selected for the tournament will achieve. An application of the computationally simple OLRE method using four team-level predictors is shown to result in ratings that had a closer fit to the observed tournament outcomes in 2006 than the expected number of wins based on the results of one million simulated tournaments according to the parametric Bradley-

Terry model for paired comparisons. This method thus has the potential to be a useful tool for the NCAA selection committee, in that it would allow the committee to develop seeds based on the number of wins that each selected team would be predicted to achieve in the tournament. Some of the methods reviewed in the paper rely on the seeds assigned by the committee for predicting success in the tournament, and these methods would therefore not be of any use to the committee in actually determining the seeds. In addition, the OLRE method also does not require any assumptions about a normal distribution of team strengths, as some previous methods have, and requires far less computation and data collection than the other methods considered. With some work, this method could easily be extended to more sports with season-end tournaments than just men's college basketball (e.g., women's college basketball or hockey); that possibility was not considered in this paper.

Additional applications of the proposed OLRE method could incorporate additional team-level predictors of success in the tournament, which would allow for the comparison of different models in terms of which does the best job of predicting success in future tournaments. The difficulty with investigating additional potential predictors, such as team free throw percentage, is that the OLRE method relies on team-level information available at the end of the regular season and *prior* to the onset of the NCAA tournament. All information collected for this paper was collected from free online resources after each regular season and prior to the onset of the tournament (see references), and historical records available online generally represent team-level data *after* the conclusion of the tournament. Additional assistance from the NCAA may be required to collect the necessary historical data and assess additional potential predictors.

Another important limitation of the OLRE method applied in this paper warrants discussion. Given that only three years of historical data were considered in the application of the OLRE method, there were only three outcomes of both five and six wins (national finalists and national champions) available for fitting the ordinal logistic regression model. This may have an adverse impact on the predicted probabilities of achieving five and six wins, and future applications of the OLRE method might consider the use of even more historical data to improve the predictions. Data were not collected prior to 2002 for the application of the OLRE method presented in this paper.

A simulation was considered in this paper to assess the predictive power of calculating an expected number of wins based on the parametric Bradley-Terry model for paired comparisons, and compare the predictive power of that method with that of the OLRE method. The OLRE method was found to result in ratings that had a better fit to the observed tournament outcomes in 2006 than the expected number of wins for each team based on the simulation of one million tournaments, using the known tournament seeds and pairings in 2006 and the ranks of Jeff Sagarin's pre-tournament ratings as strength parameters for the teams. However, the simulation used the same fixed strength parameters for the 64 teams for each simulated tournament, and the uncertainty of *estimating* team strengths with the Sagarin ratings rather than observing them directly was therefore not accounted for. A proper predictive simulation would require that the team strengths also be simulated from the sampling distributions for these strength parameters in each simulated tournament, and the information necessary to do this was not collected for this paper.

The proposed OLRE method explores the possibility of using standard statistical tools to determine possible multivariate patterns of success in the NCAA basketball tournament that can be used to predict success in future tournaments, and presents an interesting application of statistics to sports that is accessible to quantitative researchers of nearly all experience levels. The NCAA selection committee may find the proposed method to be a useful tool when assigning seeds and regions to teams that are selected for the tournament, and the simplicity and flexibility of the method will hopefully motivate additional research into models with improved predictive power.

REFERENCES

Barniv, R. and McDonald, J.B., "Review of categorical models for classification issues in accounting and finance," *Review of Quantitative Finance and Accounting*, 1999, 13, 39-62.

Bassett, G.W., "Robust Sports Ratings Based on Least Absolute Errors," *The American Statistician*, 1997, 51, 99-105.

Bradley, R.A., and Terry, M.E., "Rank Analysis of Incomplete Block Designs I. The Method of Paired Comparisons," *Biometrika*, 1952, 39, 324-345.

Berndt, E.R., Hall, B.H., Hall, R.E., and Hausman, J.A., "Estimation and inference in nonlinear structural models," *Annals of Economic and Social Measurement*, 1974, 3, 653-665.

Carlin, B.P., "Improved NCAA basketball tournament modeling via point spread and team strength information," *The American Statistician*, 1996, 50, 39-43.

Caudill, S.B., and Godwin, N.H., "Heterogeneous skewness in binary choice models: Predicting outcomes in the men's NCAA basketball tournament," *Journal of Applied Statistics*, 2002, 29 (7), 991-1001.

Elo, A.E., *The Rating of Chessplayers, Past and Present*, 1978, Arco, New York.

ESPN Bracketology with Joe Lunardi, 2006, Online HTML Document [http://sports.espn.go.com/ncb/bracketology].

Lang, J.B., “Homogeneous Linear Predictor Models for Contingency Tables,” *Journal of the American Statistical Association (Theory and Methods)*, 2005, 100, 121-134.

Lang, J.B., “Maximum Likelihood Fitting of Multinomial-Poisson Homogeneous (MPH) Models for Contingency Tables using MPH.FIT,” 2004, Online HTML document [www.stat.uiowa.edu/~jblang/mph.fitting/mph.hlp.description.htm].

Mihoces, G., “Committee, Not Computers, Makes Calls,” *The USA Today*, March 8, 2002.

Rice, J.A., *Mathematical Statistics and Data Analysis*, 1995, Duxbury, Belmont, California, 2nd Edition.

Schwertman, N.C., McCready, T.A., and Howard, L., “Probability models for the NCAA regional basketball tournaments,” *The American Statistician*, 1991, 45, 35-38.

Schwertman, N.C., Schenk, K.L., and Holbrook, B.C., “More probability models for the NCAA regional basketball tournaments,” *The American Statistician*, 1996, 50, 34-38.

Smith, T. and Schwertman, N.C., “Can the NCAA basketball tournament seeding be used to predict margin of victory?,” *The American Statistician*, 1999, 53, 94-98.

Stern, H.S., “On the Probability of Winning a Football Game,” *The American Statistician*, 1991, 45, 179-183.

Stern, H.S., “Statistics and the College Football Championship,” *The American Statistician*, 2004, 58, 179-185.

The NCAA News Online, “Men’s Basketball Committee Approves RPI Modifications,” January 3, 2005, Online HTML Document [www2.ncaa.org/media_and_events/association_news/ncaa_news_online/2005/01_03_05/division_i/4201n26.html].

Jeff Sagarin's Computer Ratings, 2002-2005 (Online HTML Documents).
[www.usatoday.com/sports/sagarin/bkt0203.htm]
[www.usatoday.com/sports/sagarin/bkt0304.htm]
[www.usatoday.com/sports/sagarin/bkt0405.htm]
[www.usatoday.com/sports/sagarin/bkt0506.htm]

Jerry Palm's College RPI Ratings, 2006 (Online HTML Document).
[www.collegerpi.com]