



## **Predicting the Outcomes of NCAA Basketball Championship Game**

H.O. Stekler and Andrew Klein

Department of Economics  
George Washington University  
Washington DC 20052  
Phone: 202-994-1261  
Fax: 202-994-1947

RPF Working Paper No. 2011-003  
<http://www.gwu.edu/~forcpgm/2011-003.pdf>

July 7, 2011

RESEARCH PROGRAM ON FORECASTING  
Center of Economic Research  
Department of Economics  
The George Washington University  
Washington, DC 20052  
<http://www.gwu.edu/~forcpgm>

## Predicting the Outcomes of NCAA Basketball Championship Games

H.O. Stekler<sup>\*</sup>

[hstekler@gwu.edu](mailto:hstekler@gwu.edu)

Andrew Klein

[aklein88@gmail.com](mailto:aklein88@gmail.com)

Department of Economics  
George Washington University  
Washington DC 20052  
Phone: 202-994-1261  
Fax: 202-994-1947

\* corresponding author

# Predicting the Outcomes of NCAA Basketball Championship Games

## Abstract

This paper uses the difference in seeding ranks to predict the outcome of March Madness games. It updates the Boulier-Stekler method by predicting the outcomes by rounds. We also use the consensus rankings obtained from individuals, systems and poll. We conclude that the consensus rankings were slightly better predictors in the early rounds but had the same limitations as the seedings in the later rounds.

**Key words:** Sports forecasts, March Madness, Ranking methods, Expert forecasts, Consensus forecasts

## Predicting the Outcomes of NCAA Basketball Championship Games

Every year college basketball fans participate in March Madness. They enter office pools or bet on the outcome of the 63 games that constitute the National Collegiate Athletic Association's (NCAA) basketball championship tournament. One web site indicated that billions of dollars are bet on the outcomes of games during this tournament. Moreover, Google Scholar found over 18,000 references to the term "March Madness Betting". Not all of these references are relevant to economics, but at least some include optimal betting strategies, e.g. Metrick, 1995; Kaplan and Garstka, 2001; Clair and Letscher, (2007). So it is no surprise that there has been interest in methods for forecasting the outcomes of these games.

A variety of statistical methods have been used to forecast which team should win each game. Although there are many other approaches, one forecasting procedure is to use rankings based on the seeds of the teams as determined by the NCAA committee and to always select the team that has the better seeding. Boulier and Stekler (1999) showed that between 1985 and 1995 the higher-ranked men's basketball team beat lower-ranked opponents 73.5% of the time. There are three reasons to explore this approach further.

First, as we explain below, this method has limitations. Given the nature of the NCAA tournament (which will be explained below), this methodology can only be used to predict the outcomes of the first four (of six) rounds of the tournament. It can not be used in forecasting winners among the Final Four teams. Second, this method may be successful in some rounds of the tournament but not in others. Thus we explore the forecasting accuracy of this method by the rounds of the tournament. Finally, an entirely new data set that yields a different set of rankings is available to make forecasts about the outcomes of these games. These data are obtained from

an average of the teams' rankings of 29-45 (depending upon the year) individuals, systems, and polls. Predictions that could be obtained from these data have never previously been analyzed. Moreover, this approach can be applied to all six rounds of the tournament. In addition to being available for all six rounds, these consensus rankings may have a valuable forecasting characteristic. The macroeconomic forecasting literature has shown that the mean of a set of predictions is usually more accurate than any one given estimate. Thus, the difference in the consensus rankings may be more accurate than the difference in seeds, which is the ranking established by the NCAA committee.<sup>i</sup>

This paper updates the procedures of Boulier and Stekler to the period 2003-2010, including an analysis of the results by round, and determines whether the consensus rankings contain any information that is not already embodied in the seedings. The next section describes the structure of the NCAA tournament and is followed by a data and methodology section. This is followed by a discussion of the results and conclusions.

#### I. The Structure of the NCAA Tournament

Since 1985 the NCAA tournament has consisted of 64 teams.<sup>ii</sup> These teams are divided into four groups (regions) of 16 teams each. Within each region the teams are ranked 1 through 16, and in the first round the best team (ranked 1) plays the lowest-ranked team. The Number 2 plays Number 15, etc. Since this is a single elimination tournament, the winner plays in the second round; the loser goes home. If the difference in seedings yielded perfect forecasts, in the second round Number 1 would play Number 8 etc.<sup>iii</sup> It is thus possible to determine whether the difference in rankings has predictive value.

This process continues through two more rounds, yielding four regional winners (the Final Four) who then play two more rounds to determine the Nation Champion. The seedings in

the regional rounds cannot be used in making predictions for these final two rounds because they only refer to the team rankings in each region and do not measure the overall difference in the abilities of these teams. Thus the Boulier-Stekler method of using the difference in ranks as a forecast of the winner cannot be used in these rounds, but the consensus rankings do provide this information.

## II. Literature Review

A number of different methods for predicting the winner of a game in the NCAA tournament have been proposed. These predictors include a variety of information about the teams' performance during the regular season and various ratings systems. (See Carlin, 1996; Harville, 2001; Caudill, 2003; Kvam and Sokol, 2006). Currently, a frequently used methodology has been to use the difference in seeds either as a predictor by itself or in models which predict the probability that the better seeded team will win (Schwertmann et al., 1991; Schwertmann et al., 1996; Boulier and Stekler, 1999; Jacobson and King, 2009).

Boulier and Stekler presented two sets of results. The first was descriptive and provided information about the winning percentage by rank and the overall result that higher-ranked teams defeated lower-ranked opponents 73.5% of the time. These results were obtained by aggregating the results over the first four rounds of the tournament. The statistical method was a probit function, in which the outcome of a discrete event, (winning or losing), was related to the difference in the seeds of the two teams. This approach was applied in pseudo real time by recursively estimating the probits and evaluating the forecasts using the Brier Score.

### III. Data and Methodology

#### A. Data

The data cover the NCAA tournaments of 2003-2010<sup>iv</sup>. The seedings and the outcomes of the games were obtained from the NCAA Web site. The second set of data is the consensus (average) rankings of each Division I team obtained from 29-45 individuals, systems, and polls. These rankings are gathered by Kenneth Massey and are archived on his web site, *www.masseyratings.com*. The rankings that are used in this analysis were those that were available just before each NCAA tournament began.

#### B. Methodology

The methodology used in this analysis is similar to that of Boulier and Stekler. We first calculated the percentage of times when the better rated team won. This was done for both the seedings and the consensus rankings, yielding the descriptive statistics. A variation of the earlier procedure was to calculate these percentages for each round not just for the entire tournament as was done by Boulier and Stekler.

In order to establish the relationship between the difference in seedings (consensus rankings) and the probability that the better ranked team wins, we use probit models. These probits were calculated for the entire tournament and for each of the rounds.<sup>v</sup> A probit model relates the probability that a discrete event (winning or losing) is explained by either the difference in tournament seedings or the difference in consensus rankings.

Formally, the probit model is written as:  $\Pr(y=1|x) = \Phi(\mathbf{x}\mathbf{b})$ , where  $\Phi$  is the standard cumulative normal probability distribution;  $y = 1$  is a win; the  $x_i$ s are the explanatory variables, the difference in seedings and consensus rankings; and a one unit change in  $x$  increases the probability of a win by  $b$  units.

#### IV. Results

There were 504 games in the eight tournaments between 2003 and 2010, with 480 occurring in the first four rounds. The seedings and consensus rankings data were both available for only the 480 games in these rounds. In all but 21 of these 480 games, both rankings agreed on which team was considered better. These differences primarily occurred in the first round and involved the teams seeded eighth (seventh) playing the ninth (tenth) seed.

In the first four rounds, forecasts based on the difference in the consensus rankings were slightly more accurate but not significantly different than those based on the difference of the seeds, 73.8% versus 72.3% (Table 1). In Rounds 1-3, both measures provide predictions that are significantly different from chance but in Round 4, the Regional Championship, neither measure was significantly better than chance.

Only the consensus rankings were available for Rounds 5 and 6, the Final Four and Tournament Championship games. Those results are mixed. The results for Round 5, when 12 of the favored 16 favored teams won, are significantly different from zero; those for Round 6 are no better than chance, i.e. 5 of 8. If the results for those two rounds are combined, the better ranked team won 17 of the 24 games which is significantly better than chance. However, if the results based on the difference in consensus rankings for Rounds 4-6 are combined, the results, once again are not significant.<sup>vi</sup>

One explanation for the difference in the results for Rounds 1-3 and those for Rounds 4-6 is that the teams in the later rounds are more comparable in quality, thus making forecasting more difficult. The difference in rankings based on either set of rankings is much smaller in the later rounds. (Table 2).<sup>vii</sup>

The probits (presented in Table 3) related the outcomes of the game to the differences in the two types of rankings. The results are similar to the descriptive findings. Using the difference in either ranking, the relationship is significant and has the appropriate sign in each of the first three rounds. The relationship does not hold in the later rounds. Moreover, the availability of the consensus rankings does not contribute substantially to the information in the seedings and vice versa.<sup>viii</sup> However, it should be noted that the consensus rankings were slightly more accurate and were available for the later rounds.

## V. Conclusion

Using data for 2003-2010, this paper evaluated the Boulier-Stekler methodology, derived from another set of tournaments, for predicting the winners in the NCAA Tournaments. We showed that this approach was successful in predicting the overall winners and those of the first three rounds. Its record was no better than chance in the fourth Round.

Because these data were not available for the Final Four and Championship games, we also evaluated another set of data that was available for all six rounds. These data consisted of the average rankings of analysts, etc. We showed that the consensus rankings were slightly more accurate than the seedings in the early games, but had the same limitations as the seedings in the later rounds. We also showed that the quality of the teams, as measured by the consensus rankings, were more similar in those later games, thus making predictions more difficult.

<b>Table 1: Percentage of Time Better Ranked Team Won; By Seed and Consensus Rankings, By Round</b>			
		Seed	Consensus
<b>Overall</b>	Number of Games	480	504
	%	72.3	73.6 <sup>a</sup>
Round 1	Number of Games	256	256
	%	75.8	77.7
Round 2	Number of Games	128	128
	%	68.8	71.1
Round 3	Number of Games	64	64
	%	75.0	75.0
Round 4	Number of Games	32	32
	%	53.1	50.0
Round 5	Number of Games	N/A	16
	%	N/A	75.0
Round 6	Number of Games	N/A	8
	%	N/A	62.5

a. The winning percentage for the 480 Round 1-4 games was 73.8%.

**Table 2**  
**Magnitude of Difference in Seedings and**  
**Consensus Rankings by Round**

	Seedings	Consensus Rankings
Round 1	8	62
Round 2	5.6	25
Round 3	4.2	16
Round 4	2.7	11
Round 5	N/A	7.6
Round 6	N/A	4.8

<b>Table 3: Probit Estimates<sup>a</sup></b>			
	Overall		
Diff seed	-0.14 (8.40)		-0.05 (1.62)
Diff consensus		-0.02 (6.4)	-0.01 (3.09)

	Round 1			Round 2			Round 3		
Diff seed	-0.015 (7.64)		-0.06 (1.23)	-0.12 (2.63)		0.009 (0.166)	-0.21 (2.37)		-0.14 (0.96)
Diff consensus		-0.02 (6.40)	-0.01 (2.25)		-0.04 (4.53)	-0.04 (3.72)		-0.05 (2.80)	-0.02 (0.73)

	Round 4			Round 5			Round 6		
Diff seed	-0.1 (0.88)		0.062 (0.030)	NA		NA	NA		NA
Diff consensus		-0.03 (1.01)	-0.04 (0.86)		-0.03 (0.93)			0.04 (0.38)	

a. The number in parentheses is the t-ratio.

## Bibliography

- Boulier, B. and Stekler, H. O., (1999). Are Sports Seedings Good Predictors?: An Evaluation, *International Journal of Forecasting*, 15, 83-91.
- Carlin, B. P., (1996). Improved NCAA Basketball Tournament Modeling via Point Spread and Team Strength Information, *The American Statistician*, 50, 39-43.
- Caudill, S. B., (2003). Predicting Discrete Outcomes with the Maximum Score Estimator: The Case of the NCAA Men's Basketball Tournament, *International Journal of Forecasting*, 19, 313-317.
- Clair, B. and Letscher, D., (2007). Optimal Strategies for Sports Betting Pools, *Operations Research*, 55, 1163-1177.
- Harville, D. A., (2003). The Selection or Seeding of College Basketball or Football Teams for Postseason Competition, *Journal of the American Statistical Association*, 98, 17-27.
- Jacobson, S. H. and King, D.M., (2009). Seeding in the NCAA Men's Basketball Tournament: When is a Higher Seed Better?, *The Journal of Gambling Business and Economics*, 3, 63-87.
- Kaplan, E. H. and Garska, S. J., (2001). March Madness and the Office Pool, *Management Science*, 47, 369-382.
- Kvam, P. and Sokol, S. J., (2006). A Logistic Regression/Markov Chain Model for NCAA Basketball, *Naval Research Logistics*, 53, 788-803.
- Metrick, A., (2001). March Madness? Strategic Behavior in NCAA Basketball Tournament Betting Pools, *Journal of Economic Behavior and Organization*, 30, 159-172.
- Schwertman, N. C., McCready, T. A. and Howard, L., (1991). Probability Models for the NCAA Regional Basketball Tournaments, *The American Statistician*, 45, 35-38.

Schwertman, N. C., Schenk, K. L. and Holbrook, B. C., (1996). More Probability Models for the NCAA Regional Basketball Tournaments, *The American Statistician*, 50, 34-38.

---

End Notes

<sup>i</sup> Of course, the NCAA committee is aware of the rankings of the various analysts.

<sup>ii</sup> The actual tournament consists of only 64 teams, but a 65<sup>th</sup> team is selected and there is a play-in game that eliminates one of the lower-ranked teams.

<sup>iii</sup> The sum of the rankings of the two teams is 17 in the first round; with perfect results, the sum would be 9 in the next round.

<sup>iv</sup> The consensus data were not available prior to 2003. That factor determined the years that were selected for this study.

<sup>v</sup> These probits are estimated for the four rounds for which the seedings data are available and the six rounds for which the consensus data are available. There is an additional difference from the Boulier-Stekler procedure. They recursively estimated the probit models for each year while this paper calculates only a single ex post set of equations.

<sup>vi</sup> The coefficients of the probits that were estimated combining the data for rounds 4-6 were not significant.

<sup>vii</sup> As further evidence about the difficulty of forecasting the outcomes of games when the teams are similar in quality, we examined the first round games between teams ranked 7 and 10 and 8 and 9. The better ranked team won only 53% of the time.

<sup>viii</sup> This result is not surprising because the correlation between the two variables is .85.