

Modeling multiple entry strategies in the NCAA tournament bracket pool

Tom Adams
Poollogic

Introduction

Single entry strategies for the NCAA tournament bracket pool have been explored in the statistical literature (Metrick 1996, Breiter and Carlin 1997, Kaplan and Garstka 2001, Clair and Letscher 2007, and Niemi et al. 2008), but in-depth analysis has never been extended to multiple entry strategies. Modeling suggests the favorability of certain single entry strategies, but the estimated profits have never been directly confirmed in the most common real-world bracket pools, those that lack upset incentives. It is difficult to design a single-entry hypothesis test with high statistical power. Multiple entries represent one way to increase the statistical power.

Materials and methods

Each year's pool (i.e., the tournament outcome and the opponent entries) was simulated 10,000 times to form a training set. For all years, an estimate of 200 opponent entries was used in the simulations.

The tournament outcome was simulated using a Markov model based on the win probabilities for each possible game which were calculated from Sagarin Predictor spreads assuming a normal distribution with a standard deviation of 11 (similar to the methods of Breiter and Carlin (1997)).

Opponent entries were simulated based on the advancement probabilities derived from the pick distribution (which is available before betting closes) for the ESPN or Yahoo bracket contest using the method of Clair and Letscher (2007).

The return on investment (ROI) of a potential entry was estimated based on its average ROI in the training set. The entry was optimized using a nearest neighbor hill climb. Only the last three rounds were optimized. The lower seed was the default winner in games that were not affected by the optimization.

Overall this method of optimizing a single entry is similar to Clair and Letscher (2007) with their ENORM function replaced by Monte Carlo estimation.

The next entry was optimized assuming all previous optimized entries had been played. 100 entries were thus iteratively optimized and added to the 200 opponent entries.

Acknowledgments

Special thanks to Brad Carlin, Ed Kaplan, and Bryan Clair for over a decade of email discussions of bracket pool strategy. And thanks to those pesky journalists who kept asking me about my results in my office pool (I am about even). This made me realize that this was a good question, but the answer proved little since I was betting only a few entries each year and the noise obscured the signal.

Results

The method was back-tested on ten years of data (2008-2017) from a real bracket pool with approximately 200 opponent entries. The number of opponent entries ranged from 178 to 241. 100 optimized entries were added to each pool for this "what if" analysis. The pool scoring was 1,2,4,8,16,32 for rounds 1 through 6. The fee was \$5 per entry. There were 4 prizes: 40%, 30%, 20%, and 10%.

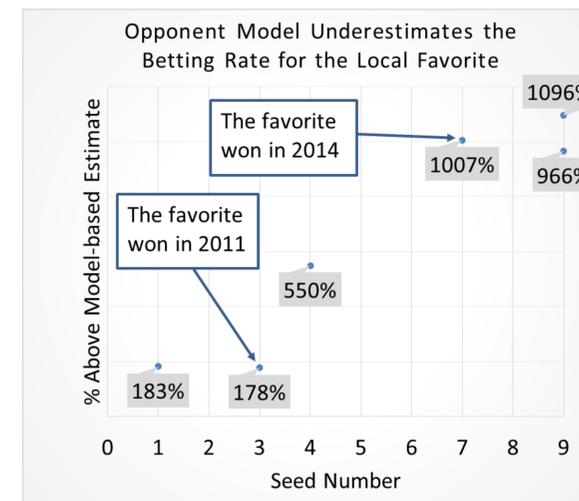
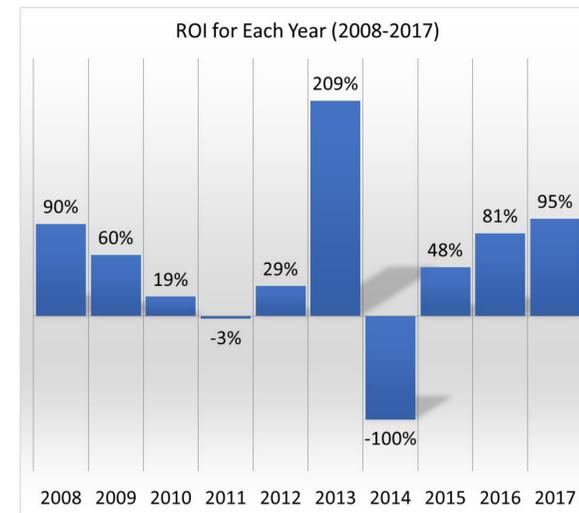
This analysis does not precisely model the real pool because of two issues: (1) In the real pool, ties were broken in favor of the closest a total championship score guess. In this back-test, it was assumed that the prize money was divided in the case of ties. This has the effect of somewhat reducing the standard deviation of returns in this back-test relative to the real pool, while leaving the mean unbiased, assuming average guessing ability. (2) Seven to eight percent of the pooled bets were used to fund a post-tournament donut party, and these funds are not removed from ROIs presented and analyzed here.

The ROI for the ten years is shown in the bar chart to the right. The yearly average ROI is \$263 or 53% of the \$500 of exposed capital. The standard deviation is \$396. The one-tailed t-test p-value is 0.032.

The two years with negative returns were years that the local favorite team (Connecticut) won the tournament. The worst year represented a total loss of all entry fees. The local team is over-backed relative to the ESPN and Yahoo pick distributions. Metrick (1996) observed this localized over-betting in his data and called it the "hometown" effect, but it sometimes extends to the entire state or region. The bias increases for higher seeds; this finding is consistent with Null (2016). See the plot to the right.

Merely correcting the bias would lead to less betting of the home team and a lower ROI when the local favorite wins. When the local favorite wins, the method does not compete well against the excessive number of opponent sheets that bet this favorite for champ and are awarded 32 points. Further analysis (not shown) indicated that the variance of returns could be reduced by making a hedging bet on the local favorite in the futures (sports betting) market, at least in the case of this 10 year back-test.

A new test set of 10,000 simulations for each year was generated to determine the method's internal estimate of the ROI of the 100 entries. The average was 72%, well above the 53% estimated from the back-test. This sort of bias is to be expected because the internal ROI estimates are based on the assumption that the tournament outcome model and opponent model used in the simulations are perfectly accurate. Comparison to the back-test provides an estimate of the size of this bias.



Conclusions

This research supports earlier findings that there are profit opportunities in bracket pools because opponent betting behavior differs from financial equilibrium behavior.

The simulations estimated an average ROI of 72% versus 53% for the back-test, a 26% bias.

The method used here is not exactly the same as methods used in the three papers that estimate the ROI in pools without upset incentives. Metrick (1996) used a tournament outcome model and an opponent model limited to the championship game. His tournament outcome model was derived from the betting market futures for the championship game. His opponent model was derived directly from the brackets bet in the pools he analyzed. Niemi et al. (2008) used some tournament outcome models similar to the one used here and an opponent model derived directly from the brackets in the pools analyzed. Clair and Letscher (2007) used very similar models but only estimated the ROI in the large ESPN and Yahoo bracket contests. Also, the metrics and/or procedures used for estimating an optimal bracket were different in all cases.

If any of these earlier methods were used to estimate the ROI for an optimized entry in an office pool before betting closed, then they would have a similar source of bias. Metrick (1996) and Niemi et al. (2008) used the entries from the pools analyzed as the opponent model, but these entries are typically not available before betting is closed. The pick distribution from the ESPN or Yahoo could be substituted. All the methods would assume perfect accuracy of the tournament outcome and opponent model. So there would be some bias toward overestimating the ROI of an optimized bracket. Lacking any other estimate, we can tentatively put this bias in the range of 26%. But, even with a 26% reduction, these methods would still result in an estimated positive return on investment.

The ROI in the current research might be improvable. The simulations used here could be improved. Breiter and Carlin (1997) used betting line spreads for the first round instead of Sagarin spreads. The betting line spreads are more accurate than the Sagarin spreads. The optimization used here was limited to the last three rounds of the tournament to reduce computer processing time, but it could be extended to all rounds. The last three rounds are the most important because of the top-heavy scoring rules, but extending the optimization to the first three rounds might lead to some additional improvement. The tournament outcome model could be improved by using information about absences of key players relative to the season. The pool size estimate might be improved by using the prior year's pool size. The opponent model could be corrected for the "hometown" effect using data like that available from Null (2016).

Further information

Email: tadamsmar@yahoo.com

Literature cited

Breiter, D. J., and B. P. Carlin. 1997. "How to Play Office Pools if You Must." *Chance* 10:5-11.
Clair, B., and D. Letscher. 2007. "Optimal Strategies for Sports Betting Pools." *Oper Res* 55:1163-1177.
Kaplan E. H., and S. J. Garstka. 2001. "March Madness and the Office Pool." *Manage Sci* 47:369-382.
Metrick, A. 1996. "March madness? Strategic behavior in NCAA Basketball Tournament Betting Pools." *J Econ Behav Organ* 30:159-172.

Niemi, J. B., B. P. Carlin, and J. M. Alexander. 2008. "Contrarian Strategies for NCAA Tournament Pools: A Cure for March Madness?" *Chance* 21:39-46.
Null, B. 2016. "Homer bias is real and it will derail your March Madness bracket." <https://www.cbssports.com/college-basketball/news/homer-bias-is-real-and-it-will-derail-your-march-madness-bracket/>. Accessed on August 4, 2017.