

How Predictable is the Overall Voting Pattern in the NCAA Men's Basketball Post Tournament Poll?

After the completion of the National Collegiate Athletic Association's (NCAA) men's basketball tournament in 1993, a poll of coaches was taken so that those results – as well as those from the National Invitational Tournament (NIT) – could be considered when generating one final, subjective measure concerning that season's teams. Each and every year since then, this poll, as administered by *ESPN/USA Today*, has placed the team that has won the NCAA tournament as its number one team, and typically the runner-up in the championship game has finished in second place; the other two, Final Four participants are also usually voted as the third and fourth teams, respectively. However, the question remained: could an objective mechanism be discovered that matched this knowledgeable set of people's opinion reasonably well – across the entire list of teams appearing in that final poll – in an effort to uncover any irregularities in this final ranking? (While researching this question, it was discovered that on four previous occasions – 1953, 1954, 1974, and 1975 – a final poll of sportswriters, from the Associated Press (AP), was also taken after those particular NCAA championship games.) Since 2000, 31 coaches have voted in the *ESPN/USA Today* poll, where 25 points are awarded to the team appearing first, 24 for second, and so on, down to 1 point to any team appearing in the twenty-fifth spot, on any ballot, yielding a maximum point total of 775. (When we use the terms votes and vote totals in this article, these will be synonymous with the point total in this final poll.)

How does the NCAA tournament work? We will begin with a brief history of this ever-changing tournament. Eight teams were invited to compete in the inaugural NCAA men's basketball tournament in 1939. The number of invitations was doubled in 1951, and again in 1975, until 64 teams competed in the 1985 tournament after the last major expansion was implemented. One play-in game was added in 2001, and three more such preliminary contests were added in 2011, increasing the number of teams in this NCAA tournament to 68. Ignoring these play-in games, this single elimination tournament has four regions of 16 teams each. The NCAA tournament selection committee extends these prized invitations and seeds the teams in each region from the strongest (a 1-seed) to the weakest (a 16-seed). This committee considers many different measurements, concerning each team's performance that year, to determine these tournament seeds: won-loss record, conference affiliation, RPI ranking (a quantitative formula that incorporates strength of schedule, and ignores margin of victory), level of success in conference postseason tournaments, etc. These four regional champions then move on to the Final Four, where the NCAA champion is determined.

Previous Results

After examining the last 20 years of the final coaches' polls (1993-2012), there were only three times that the NCAA runner-up did not finish in second place: 1996, 2002 and 2004. In those three years, the runner-up appeared third in this poll, and there was a reasonably large difference between the runner-up's winning percentage and that of the Final Four participant who finished

second (and who the runner-up had not defeated in their previous tournament game): 1996, Syracuse (third place, 76.3%) and Massachusetts (second place, 94.6%); 2002, Indiana (third place, 67.6%) and Kansas (second place, 89.2%); and 2004, Georgia Tech (third place, 73.7%) and Duke (second place, 83.3%). Similarly, the two Final Four teams that were defeated by the NCAA championship game's participants typically were voted as the third and fourth place teams, except for the three times a team finished second in this poll, and the five times a team was ranked lower than that: 2000, where identical records of 22-14 (18-13 before the NCAA tournament began) landed those two, 8-seeds the eleventh (North Carolina) and sixteenth (Wisconsin) spots in the final poll; 2003, when a 3-seed (Marquette) with a final record of 27-6 was voted into sixth place; 2006, when an 11-seed (George Mason) was voted eighth (27-8); and finally in 2011, when Virginia Commonwealth (VCU), another 11-seed, finished in sixth place (29-12). (More about the latter two teams will be presented later in this article.)

Likewise, the other four teams who lost to the Final Four teams in the tournament's four regional finals (and complete what is referred to as the Elite Eight) are usually ranked between fifth and eighth place (or, they at least typically appear in the poll's list of top 10 teams), and the eight teams that were defeated by the Elite Eight complete the Sweet Sixteen, and typically appear within this final poll's top 20 teams. However, the alignment between the Elite Eight, and this poll's top eight teams, and the Sweet Sixteen, and the first sixteen teams in this poll, is not as strong as with the Final Four: 12 Elite Eight teams have finished as the ninth or tenth team, and another 15 were ranked between eleventh and eighteenth in the final poll over this 20 year period; 49 Sweet Sixteen teams were ranked between seventeenth and twenty-fourth, and six more between twenty-fifth and twenty-eighth. So, roughly five out of every six Elite Eight, and Sweet Sixteen, teams were ranked in the top eight, or top sixteen, teams in these 20 final polls. (Specific breakdowns for these outliers appear in Tables 1a and 1b.) From these observations, it appeared that perhaps a linear equation could be derived that would accurately forecast where the polled coaches would rank teams after the entire season was finished.

Table 1a – Counts for Elite Eight (rank > 8th) in final coaches' poll

Rank	9	10	11	12	13	14	15	16	17	18
Number	8	4	2	2	2	3	3	1	1	1

Table 1b – Counts for Sweet Sixteen (rank > 16th) in final coaches' poll

Rank	17	18	19	20	21	22	23	24	25	26	27	28
Number	8	6	7	8	5	5	4	6	3	0	2	1

Specifics for Predictive Models

To help separate teams that shared identical records and who also won the same number of post season games, like the aforementioned Final Four teams in 2000 (North Carolina and Wisconsin), or as in 2011 with Ohio State and San Diego State (34-3, two wins apiece), or, Notre Dame and George Mason (27-7, one win each), it was necessary to include another objective measure to go along with the two previously mentioned values: winning percentage (WP) and NCAA/NIT tournament success.

The power rating system, designed by Bob Carroll, Pete Palmer and Jim Thorn for professional football, has been recognized to be a fairly effective predictive system, as illustrated in a study by John Trono a few years ago. Each team's power rating (PR) is effectively the difference between the average number of points they have scored in each game and those that they have surrendered, and then the strength of schedule (SOS) component for each team is determined and added to this average point differential to produce the final rating. (It may take the software close to one hundred iterations to determine the final ratings because the SOS components must be repetitively recomputed until they converge, within a specified tolerance, to their final values.) Some teams have earned very high power ratings, and high positions in the final, regular season polls, but their records were not indicative of their team's ability, e.g. Maryland was eleventh in both polls (in 2001), with a 21-10 record (39th overall, percentage-wise, of the 322 teams in Division I that year); however, they had the 5th highest PR – and reached the Final Four as a 3-seed.

So, the first model (model-1) includes a team's WP and PR, where the raw WP value is multiplied by 100, so that it will be close in magnitude to the PR (whose typical values range between 90 and 120 for teams invited to either tournament), and seven tournament indicator variables that are associated with how far a team advanced in the NCAA Men's Championship. (Including six indicator variables for the NIT introduced a significant level of multicollinearity in this model, which is why those variables were excluded. This also seemed reasonable since only 42 teams, who were invited to the NIT, received votes in the final poll for the years contained in our training data set; only eight of those had a vote total larger than 15, and 21 teams received less than five.) Thus, the NCAA champion would have a one in the field designating they won the tournament, the tournament runner-up would have a zero in that field, but a one in the next field (since they were the runner-up); all other fields would be zero for these two teams. Two more teams would have a one indicating that they lost their Final Four game, and so on down to those 32 invited teams that lost their first round game (the NCAA play-in games are ignored) but would have a one in the field indicating an NCAA tournament appearance. Therefore, any team invited to the NCAA tournament would have only one non-zero indicator variable, and zeros for the other six indicator variables; seven zeros would represent a team that was not invited to the NCAA tournament.

Each team invited to the NCAA tournament would have two numeric values (WP and PR) and a single, non-zero indicator variable identifying that team's level of NCAA tournament success;

the other six zeroes would contribute nothing to the team's predicted vote total. The predictive equation would include ten different model coefficients, each of which will multiply one of the nine, specific variables associated with each team (excluding the intercept – which is simply added). The final predicted vote totals would be ordered, and the ranking produced could be compared against the actual ordering contained in the final coaches' poll.

To determine the set of ten model coefficients, a weighted least squares regression was performed, with the total number of actual votes cast for each team as the dependent variable, and the WP, PR, and seven indicator variables as the independent variables. Weights were initially determined by regressing the absolute value of the residuals from ordinary least squares regression on the predictor variables. Several iterations were performed using the residuals from the weighted least squares (WLS) to revise the weights until the estimated regression coefficients stabilized. (These values can be found in Table 2.) Since the number of voters has varied occasionally, voting totals have been normalized to the most recent maximum: 775. Because 35 to 45 teams normally receive at least one vote in this poll, it was deemed wise to only include those teams, instead of the more than 300 Division-I teams that are eligible to be invited to either tournament. The main reason to omit those teams not receiving any votes was to exclude a very large number of entries which would have zero as their dependent variable's value. (The following four teams have received at least one vote in the final polls, but were not invited to either tournament: Wisconsin-Milwaukee (23-4) in 1993, Akron (26-7) in 2007, College Basketball Invitational (CBI) Champ Tulsa (25-14) in 2008 as well as CBI champ Pittsburgh (22-17) in 2012.) Using this information, 658 teams have received at least one vote in the fifteen year training data set: 1993 to 2007. This training set will be used to determine model-1's coefficients, and subsequent years will be evaluated to assess the merits of the model.

After evaluating model-1 (where initial results looked reasonable), a possible anomaly was noticed; it was somewhat surprising to see that the coefficient for the NCAA runner-up indicator variable was larger than that associated with the NCAA champion. After further investigation, this is not that unreasonable since the champions' WP and PR values for those fifteen years in question were higher on average, so the champion's coefficient could afford to be smaller, and the predicted vote totals would still be 30 to 50 predicted votes higher than the runner-up.

Table 2 – Coefficients for model-1: WLS with indicator variables

Variables	Coefficient	Standard Error	p-value
Intercept	-1879.30700	111.79400	0
WP	7.23787	0.56217	0
PR	12.81086	0.86492	0

NCAA Champ	448.79253	32.09062	0
NCAA Runner-up	482.36326	24.06626	0
NCAA Final Four	376.41758	30.62105	0
NCAA Elite Eight	292.57112	21.66412	0
NCAA Sweet Sixteen	169.42528	12.29250	0
NCAA Round of 32	8.35231	10.58175	0.43020
NCAA Invitee	-31.28519	10.13553	0.00211

This led to the creation of model-2, which combined all seven indicator variables into two values: number of NCAA tournament wins + 1, and number of NIT tournament wins + 1. Some strong teams have lost their opening round game, most notably Duke and Missouri in 2012, both of which were 2-seeds, so the +1 is to reward invited teams over those not in either tournament. (Previously unpublished research efforts by John Trono on this topic strongly supported making a distinction between invited teams with zero wins, and those teams that were not invited.) Table 3 lists the values produced after applying the same aforementioned WLS technique to this model.

Table 3 – Coefficients for model-2: WLS with NCAA & NIT wins

Variable	Coefficient	Standard Error	p-value
Intercept	-2237.46505	94.30357	0
WP	7.34741	0.42781	0
PR	14.55405	0.83897	0
NCAA Win	99.38331	2.86699	0
NIT Win	30.01610	2.34287	0

Table 4 contains the Spearman correlation coefficients (SCC) for the models when evaluating the top 35, the top 25 and the top 15 teams in the final coaches' poll for the data in the training set (1993 to 2007). For comparison purposes, a baseline model was also created that essentially emphasizes tournament wins almost exclusively. In this model, the coefficients are one for WP and PR, and one hundred was the coefficient reflecting tournament success. Given the range of values for WP and PR, these two quantities only break ties for teams with identical tournament

win totals in the baseline model. (When more than one value appears in the win coefficient column in Table 4, the top value is for an NCAA win; the bottom value is for an NIT win.)

Table 4 – SCC values (highest are underlined) as well as model coefficients.

	Top 35	Top 25	Top 15	WP Coef	PR Coef	Win Coef	Intercept
Model-1	.83152	.84631	.78298	See Table 2			
Model-2	.85258	.85074	.83571	See Table 3			
Baseline	.72164	.64692	.44214	1.0	1.0	100.0	-350.0
NIT-3	.85444	.85146	.83852	7.58503	15.76744	96.05271	-2387.55079
NIT-4	.85635	.86077	.83738	7.34532	15.70023	95.80927	-2359.20723
NIT-5	.84996	.86090	.83536	7.22588	15.74452	96.05181	-2352.31166
NIT-10	.83079	.86003	.84381	7.04634	16.01035	92.28397	-2360.91580
NIT-0	.79160	.85395	.84726	6.94728	16.47848	88.12535	-2394.82389
NIT'	.86012	<u>.86423</u>	.84381	6.94728	16.47848	88.12535	-2394.82389
MC-2	.85989	.86005	.84143	6.34443	15.46155	97.51648 25.11566	-2500.0
MC-Best	<u>.86119</u>	.86110	<u>.85024</u>	6.68507	17.64763	88.24644	-2500.0

NCAA versus NIT

Since there are roughly 30 automatic NCAA tournament bids awarded to recognized conference champions, the remaining at large NCAA tournament teams are judged to be the most deserving teams not already invited. However, there are usually a small number of teams who are not invited, but may be just as deserving as the last few “bubble teams” who received an invitation, and those teams have an opportunity to redeem themselves in the NIT. Wins in the NIT will not be worth as much as NCAA wins, to those casting votes in the post tournament poll, so the question is: how much should they be worth? According to Table 4, the ratio of an NIT tournament win to an NCAA tournament win is roughly 0.3 (30.016098 / 99.383310). (A Delta method argument was used to construct a 95% confidence interval, 0.259 to 0.345, for this ratio.) To investigate how accurate that ratio is, we combined the two tournament win variables into one such that each NIT win value is multiplied by a constant (that is less than one).

Since the NIT champion must win just five (not six) games to be crowned NIT champion, six would appear as the value for said champ in model-2. If each NIT value was divided by two, this would produce a value of three for the NIT champion, implying that they reached the same level of achievement as a team that reached the Sweet Sixteen in the NCAA (i.e., two wins) but then lost their next game. Such NCAA tournament teams typically finish in the final poll's top 20, but no NIT champion, since 1993, has done so. Dividing the win total by three equates the NIT champ with an NCAA tournament team who only won one NCAA game; dividing by four places them halfway between those teams with one NCAA win and those who lost their first round game; dividing by five puts them slightly above those teams who lost their only NCAA game; dividing by ten makes them about halfway between an NCAA first round loser and a team who was not invited; and one could ignore NIT wins altogether. Those five possible NIT win multipliers (0.33, 0.25, 0.2, 0.1 and 0) were evaluated, and the results are listed as NIT-3, NIT-4, NIT-5, NIT-10, and NIT-0 in Table 4. Given where NIT champions have been ranked in the final coaches' polls since 1993, equating them to teams who won at most one NCAA tournament game seemed to be a fairly reasonable, rough equivalency – especially considering the perceived, lower overall quality of the NIT field.

Serendipity

After producing each model's coefficients, a separate program was used to complete the following tasks: read in those coefficients; order all teams in each year in the training set according to each team's predicted number of votes; compare those rankings with the final polls; and output the SCC values. However, during the side-study to examine the idea regarding the possibility for an appropriate NIT win multiplier, as was just described, the NIT-0 coefficients were used in the evaluation program and the NIT win multiplier constant was not changed from its previous value (0.25). Because the resulting SCC values were better than any other produced so far (see NIT' in Table 4), it came to our attention that perhaps we were being overly restrictive in our model by trying to reproduce the actual vote totals too closely when only the final ranking was truly what we were trying to match.

Given this revelation, we: examined all the model coefficients produced by all the combined win models; determined a range of values that would include all those produced; created a program that would randomly generate each of the model's three coefficients uniformly within those ranges; and evaluated the resulting rankings. During this Monte Carlo search, any set of coefficients which produced results as good as any previously generated, using the training data set, would be output, and a large number of random model's coefficients would be evaluated.

This strategy discovered many higher SCC values, which prompted several observations. First of all: which of the three SCC values is most important? Whenever any of the three SCC values was improved upon, those model's coefficients (and accompanying SCC values) were output. However, the model coefficients that produced the largest SCC value for the top 15 teams might be low with regard to the top 35, and vice versa. Therefore, we concluded that the maximum sum

of the three SCC values was the quantity to search for since the best coefficients should work consistently across the entire spectrum of teams receiving votes in the final poll. (We wish to remind everyone that we are not attempting to quantitatively determine which teams deserve specific rankings; we are simply trying to objectively generate rankings that closely match those of the polled coaches.)

After the initial Monte Carlo runs, the ranges were tightened several times and then rerun. When comparing the results produced by those runs, it was noticed that a relatively large SCC sum was produced by two different sets of model coefficients. This led to the conclusion that because we were now only concerned with the ranking, and not the predicted vote total used to produce the ranking, the only truly important relationships between these three model coefficients were the ratios between them. The ratio of win coefficient to PR coefficient was 5.0046 and 5.0099 for the two sets of model's coefficients that produced the large SCC sum of 2.57253. Likewise, the PR coefficient to WP coefficient ratio was 2.6327 and 2.6352. We named the set of model coefficients with the largest SCC sum discovered during the Monte Carlo searches MC-Best; it also had the highest top 15 and top 35 SCC values in the training set. (The Monte Carlo approach actually generated 26 sets of model coefficients that produced the same SCC values; all 26 coefficients were averaged, and the means became the model coefficients used in MC-Best – which also produced the aforementioned highest SCC sum value. MC-2 refers to the best Monte Carlo results when applied to model-2, where NCAA and NIT wins are separate values – not the one, combined value used in MC-Best.)

Results

The SCC value for the top 35 teams, in the five years following the training years (2008-12), was slightly higher (0.86706 versus 0.86119), and this SCC value, for the 2008 final poll (0.93480), was actually larger than any year in the training data set. (For this same five year period, the overall top 25 and top 15 SCC values were slightly lower than in the training data set: 0.83692 versus 0.86110, and 0.83321 versus 0.85024.) However, the top 15 SCC value was 0.95357 for 2008 and 2009, and roughly 0.90 for 2010. Given how much weight is given to NCAA tournament performance in the final poll, and therefore, in this predictive model as well, two 2-seeds losing their opening round game in the 2012 tournament heavily contributed to the smaller, top 15 SCC value (0.62321); likewise, the 11-seed in 2011 (VCU) was voted sixth, but they were predicted to be fifteenth, which was the primary reason for the low, top 15 SCC value (0.73036) that year. Otherwise, results from the last five years are quite similar to those from the training set.

George Mason reached the Final Four in 2006, had higher WP and PR values than VCU in 2011, and both teams were 11-seeds from the same mid-major conference (Colonial Athletic Association), and both teams lost in their conference's championship tournament game, so they both received "at large" NCAA invitations. Both teams defeated the 6-seed, the 3-seed and the 1-seed in their regions (VCU beat the 10-seed, and George Mason the 7-seed, in their second round

games) and both lost their Final Four game, so on paper, they had very identical resumes. However, the voters thought George Mason was eighth in 2006 (MC-Best predicted them to be ninth), and VCU finished sixth (predicted fifteenth by MC-Best). This is an example of what appears to be inconsistent behavior amongst those coaches who voted since there is really no quantitative reason to have VCU voted two positions higher than George Mason especially since VCU's final WP was lower (70% versus 77.14%), and their final PR was smaller (108.588 versus 112.465) – and both teams lost their only Final Four game. The other three Final Four teams (since 1993) that did not receive enough votes to make the top four in the final polls were seeded higher than VCU, yet were ranked lower (or the same) in those final coaches' polls. In 2000, 8-seed North Carolina was voted eleventh, and 8-seed Wisconsin was voted sixteenth, though MC-Best predicted they would be ninth and eleventh, respectively; both teams shared identical final records, and Wisconsin's PR was slightly smaller than North Carolina's. Both of their PR values were also larger than VCU's by roughly 8.5 points, which helps to explain why MC-Best predicted them to finish in a higher position in this poll than VCU. (In 2000, Wisconsin and North Carolina were also the first teams to reach the Final Four with seed numbers larger than five since the final coaches' poll began in 1993; perhaps the coaches who voted back then were somewhat unsure how high to place such teams.) Number three seed Marquette's PR in 2003 was about the same as those two 8-seeds from 2000, but Marquette's WP was almost 20 points higher; they were voted as the sixth team in 2003 – and MC-Best agreed exactly.

As mentioned in this article's first paragraph, four final polls were taken before 1993. With a few caveats, MC-Best displayed a reasonable level of success predicting those final rankings. In all four of these years, only the top 20 teams were listed in the final polls, and one important distinction that was made is that in the early 1950s, the NIT and NCAA tournaments were equally prestigious, so the teams reaching the same level of achievement in either were rewarded similarly in this updated MC-Best evaluation. The SCC value for the top 10 teams in 1953 was 0.91515 as four team's positions were predicted exactly, five more were off by one, and the other team was off by three. The other caveat is that since only one team from each conference could be invited to the NCAA tournament before 1975, many teams who could not be invited were still ranked highly in the final poll even though they did not participate in either post season championship. There were four such teams in the top 20 in 1953, and MC-Best predicted them to be seven or more positions away from the actual final poll ranking. The other six tournament-bound teams, in the second ten of the final top 20 ranked teams, differed only by two, zero, one, one, three, and four spots, respectively.

There were similar results for the other years as well. In 1954, the top 10 SCC value was 0.83333, and in 1974, five of the top 10 teams were ranked exactly, one was off by one, one by two, one by three, and two more that who did not compete in either tournament, were off by eight and ten. As the prestige associated with the NIT tournament slowly dwindled, its win multiplier was reduced from being the same as the NCAA win coefficient in the early 1950s, to one half of that value in the early 1970s, and finally to one quarter – as used in MC-Best – since

1993. (One half seemed reasonable given that only 25 teams could be invited to the 1974 NCAA tournament, and therefore, many strong teams were still available to compete in the NIT. The number of NCAA invitations increased to 32 in 1975.) In 1975's final poll, five more of the top 10 teams matched exactly, three teams were off by one, one by two and North Carolina State, who was on probation – and could not compete in either tournament – was predicted to be a dozen spots lower than their final poll ranking (of seventh). Using an unmodified MC-Best model to predict what the final polls might have been could be interesting, but doing so for the years prior to 1985, when the NCAA began to invite 64 teams to its tournament – or even more problematically, before 1975, when NCAA invitation restrictions seriously limited who could be invited – would probably always generate rankings which would suffer from many teams not being rewarded as much as the voters back then might have done (for the two reasons just described).

Model Limitations

Table 4 illustrated that the baseline model, which essentially uses a team's WP and PR to break the ties when using a large, multiplying coefficient for tournament wins, isn't exemplary with regards to many of the teams in the top 35, but it does match the top five quite well. Table 5 illustrates how closely the baseline model matched the top five teams in the past 20 final coaches' polls (since 1993), comparing those results against MC-Best, as well as the next ten, top teams (in two groups of five). Because the baseline model can guarantee correctly selecting the top team by multiplying that team's six wins by a large value, this simplified baseline model can always match the final poll as long as the coaches voting in it continue to choose the NCAA tournament champion as that season's best team. However, after the top three or four teams, it is obvious that each team's ranking in the final poll is based on more than the number of NCAA tournament wins they achieved (after you examine the lower ranks in Table 5) since the baseline model has fewer close matches for the teams ranked from sixth to fifteenth place than the MC-Best model does. (The average difference between each top 35 team's actual and predicted position was roughly 3.5 for MC-Best and 5.0 for the baseline model).

Table 5 – Difference counts between actual ranks for two models: baseline/MC-Best.

Rank/Diff.	0	1	2	3	4	5-9	>9
1	15/13	0/2	---	---	---	---	---
2	12/10	3/3	0/2	---	---	---	---
3	10/8	4/3	0/3	1/0	0/1	---	---
4	12/3	2/8	0/3	---	0/1	1/0	---
5	9/4	3/6	1/4	0/1	2/0	---	---

6-10	20/29	23/21	7/12	7/4	3/5	8/3	7/1
11-15	7/12	19/25	11/15	4/8	4/7	26/7	4/1

Vote Matching

As mentioned previously, 775 is the (normalized) total number of votes for the typically unanimous number one team in the final coaches' poll that occurs after the NCAA tournament. Since the Monte Carlo method was not trying to necessarily match the actual number of votes cast, but focused instead on maximizing the SCC sum, the prediction equation's intercept value was essentially ignored when generating random model coefficients. Table 3 shows that the models, combining the two tournament win values into one, have intercept values close to -2400, and since the predicted number of votes for the top team seemed to be about one hundred points higher than expected, we tried -2500 as the MC-Best intercept value. This produced a reasonable vote total, with roughly 3.5 teams, on average, that did receive votes in the final poll receiving none according to the model (negative predicted vote totals were changed to zero), and six teams, on average, who did not receive any votes had positive, predicted vote totals. On average, the difference between the actual number of votes, and the predicted number in the training set was 64 votes. This increased to 72.5 when applied to the next five years of final polls.

In the fifteen years in the training data set, 44 teams received votes on average, and with this intercept value MC-Best predicted 46.5 teams would get at least one vote. For the subsequent five final polls, only three ranked teams were excluded, but about ten teams each year were predicted to be present in said poll. Decreasing the intercept by 100 would exclude many of those latter teams, but would also remove teams who did actually appear on at least one voter's ballot, from the predicted final pool as well. For the training set, increasing the intercept by 25 would allow 16 of the 52 excluded teams to be present, but would again increase the number of imposters receiving predicted final poll votes as well. (The highest ranked team excluded was ranked twenty-ninth and the highest predicted rank for a team not in the final poll was twenty-fourth.)

Concluding Thoughts

Several models have been presented here that attempt to predict the group behavior of those coaches who've been asked to vote in the final ESPN/USA Today Men's basketball poll, i.e. after the NCAA championship tournament has concluded. Model-2 performed the best of the purely statistical models; it was a definite improvement over model-1, with a Spearman correlation coefficients (SCC) sum of 2.53903 versus 2.46081, and it only requires four variables instead of nine. After this model was modified to combine the NCAA and NIT win totals into one variable, this updated model's three coefficients were computed (to best match the actual vote totals), and

then several Monte Carlo evaluations were performed in hopes of increasing the correlation between the rankings produced in this fashion and the ranking represented by the final poll. (Once the actual point totals in the polls were seen to be less important than the relative team ordering, the computed model's coefficients could be randomly varied, within close proximity of that established set of parameter values.) Using the SCC sum – with regards to the top 35, 25 and 15 teams in the actual polls – as the value to be maximized, coefficients were found (model MC-Best) that yielded SCC values that were all quite close to 0.86, with the highest SCC sum of 2.57253.

An interesting thought experiment was designed **after** the coefficients in model MC-Best were found. This experiment would examine the consistency between the last coaches' poll taken at the end of the regular season, after all conference tournaments have been completed, and the final poll. If these two polls near the end of the season are fairly consistent, then perhaps by setting the tournament win coefficient to zero in model MC-Best, the resulting ranking would match the penultimate poll fairly closely. A more detailed presentation of these results, as well as year by year, detailed breakdowns for all final polls since 1993, can be found at the corresponding web page (<http://academics.smcvt.edu/jtrono/finalpoll.html>), but suffice it to say that the SCC values were 0.64867, 0.60343, and 0.61205 (for the top 35, 25, and 15 teams, respectively), and so Model MC-Best doesn't predict the coaches' penultimate poll as well. The coaches seem to place much more emphasis on NCAA tournament performance when casting their votes in the final poll as indicated by the power rating (PR) weight being roughly 2.6 times larger than the winning percentage (WP) weight, and the weight for NCAA tournament wins being five times larger than the PR weight in MC-Best. (At first glance, it could be conjectured that when the coaches vote in the penultimate poll, they place less emphasis on a team's perceived strength, as captured by the PR, and more emphasis on their won-loss record; perhaps this specific relationship can be investigated further in a subsequent study.)

However, the prediction equation does perform fairly accurately in both polls if you consider that the SCC for the top 35 increases to 0.73228 when you halve the largest difference for any team in each year, since it seems that there is typically at least one top 35 team whose ranking is quite different than what is predicted. (Said outlier was ranked somewhere in the last ten spots in the top 35 teams in 14 of the last 20 years. The six outlier differences that were above 30 were as follows: 32, 33, 35, 38, 48, and 48.5.) Table 6 summarizes the number of teams that differed by certain amounts for all 20 years in both the final – and penultimate – polls when utilizing the MC-Best coefficients.

Table 6 – Accuracy of predicted poll positions.

Rank Difference	1993-2007 Final Poll (Training Set)	2008-2012 Final Poll (Predicted)	1993-2012 Final Poll (Overall)	1993-2012 Pen. Poll (Overall)
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	1
11	1	1	1	1
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	1	1	1	1
16	1	1	1	1
17	1	1	1	1
18	1	1	1	1
19	1	1	1	1
20	1	1	1	1
21	1	1	1	1
22	1	1	1	1
23	1	1	1	1
24	1	1	1	1
25	1	1	1	1
26	1	1	1	1
27	1	1	1	1
28	1	1	1	1
29	1	1	1	1
30	1	1	1	1
31	1	1	1	1
32	1	1	1	1
33	1	1	1	1
34	1	1	1	1
35	1	1	1	1
36	1	1	1	1
37	1	1	1	1
38	1	1	1	1
39	1	1	1	1
40	1	1	1	1
41	1	1	1	1
42	1	1	1	1
43	1	1	1	1
44	1	1	1	1
45	1	1	1	1
46	1	1	1	1
47	1	1	1	1
48	1	1	1	1
49	1	1	1	1
50	1	1	1	1

0	107	32	139	71
1	104	41	145	125
2	79	32	111	83
3	44	18	62	64
4	48	13	61	58
5	31	10	41	46
6	32	4	36	38
7	7	5	12	38
8	19	4	23	38
9	9	4	13	26
10	12	3	15	14
11-15	30	7	37	64
16-20	6	2	8	26
21-25	2	2	4	15
26-30	0	0	0	6
>30	0	0	0	6

Overall, model-2 captured the coaches voting practices quite accurately, and the Monte Carlo technique increased its performance by the following amount: from an SCC sum of 2.53903 (0.85258+0.85074+0.83571) to 2.57253 (0.86119+0.86110+0.85024). Utilizing the MC-Best coefficients, over 61% of the teams, in the five years that follow the 15 years in the training data set, are at most two places away in the predicted final poll (versus the actual final coaches' poll), and only 17.5% are more than five places away. Only considering the top 15 teams in the same five year period, each team is only 1.53 places from their actual ranking, and 92% were less than four away: VCU in 2011 had the second largest difference (nine) and the 2012 Florida State team, a 3-seed who had a strong season – but who lost their second round, NCAA tournament game – had the largest difference (eleven): they finished as the fifteenth team, but were predicted to only be the twenty-sixth team that year.

In a way, our work parallels that by Jay Coleman, Michael Dumond and Allen Lynch, where they attempted to quantitatively determine which teams will be invited to the NCAA championship (by the tournament selection committee). They did not focus on which teams **should** be invited, as we have not focused on who **should** be awarded the designation as the top team, the second place team, and so on, at the end of any given season.

Given the strong correspondence between the MC-Best predictions and the final coaches' poll, perhaps there aren't any obvious cases of irregular voting impacting this final poll. Even the aforementioned case of the two 11-seeds (George Mason in 2006, and VCU in 2011) can be somewhat explained away by examining the actual vote totals those two teams accrued in those final coaches' polls. VCU's sixth place finish in the polls was actually quite precarious, as their total of 555 was only one point larger than the team who ranked seventh, and only 23 and 24 points more than the teams that finished just below them in eighth and ninth place, respectively. George Mason received 564 points, but three losing teams in the Elite Eight were 1-seeds, and the other 1-seed (Duke) lost after reaching the Sweet Sixteen. Those team's superior seasons, and reasonably strong performances in the NCAA tournament, still influenced the coaches enough so that those 1-seeds were ranked from fourth to seventh in the final poll: Duke was seventh, and 22 points ahead of George Mason – who had 24 more points in the final poll than the ninth team. However, in VCU's case, only one 1-seed made the Elite Eight, and two of the other three losing Elite Eight teams were 2-seeds, and the fourth was a 5-seed; all three of the latter ended up ranked below VCU because the voters were not as impressed with those teams' overall seasons as they were with the four strong 1-seeds who just missed reaching the Final Four in 2006 (and all of which finished ahead of George Mason). VCU winnings its play-in game may also have contributed to the voters ranking them as sixth, but if VCU's point total had been 25 less, they would have finished as the tenth team in 2011, much closer to where MC-Best predicted they should be ranked. This is just one illustration of how misleading comparing positions in the poll can be, rather than comparing the actual vote totals, since George Mason actually received 8 more votes in 2006 than VCU did in 2011, yet George Mason finished in eighth place in the final poll, whereas VCU finished sixth.

Further Reading

Carroll, B., Palmer, P., Thorn, J. 1988. *The Hidden Game of Football*. New York, New York: Warner Books.

Coleman, B. J., Lynch, A. K. 2001. Identifying the NCAA Tournament 'Dance Card'. *Interfaces* 31(3):76-86.

Coleman, B. J., Dumond, J. M., and Lynch, A. K. 2010. Evidence of Bias in NCAA Tournament Selection and Seeding. *Management and Decision Economics* 31:431-452.

Trono, J. A. 2010. Rating/Ranking Systems, Post-Season Bowl Games, and 'The Spread'. *Journal of Quantitative Analysis in Sports* 6(3), Article 6. DOI:10.2202/1559-0410.1220.