

2001

# Identifying the NCAA Tournament "Dance Card"

B. Jay Coleman

*University of North Florida*, [jcoleman@unf.edu](mailto:jcoleman@unf.edu)

Allen K. Lynch

Follow this and additional works at: [http://digitalcommons.unf.edu/bmgt\\_facpub](http://digitalcommons.unf.edu/bmgt_facpub)



Part of the [Management Sciences and Quantitative Methods Commons](#)

---

## Recommended Citation

Coleman, B. Jay and Lynch, Allen K., "Identifying the NCAA Tournament "Dance Card"" (2001). *Management Faculty Publications*. 2.  
[http://digitalcommons.unf.edu/bmgt\\_facpub/2](http://digitalcommons.unf.edu/bmgt_facpub/2)

This Article is brought to you for free and open access by the Department of Management at UNF Digital Commons. It has been accepted for inclusion in Management Faculty Publications by an authorized administrator of UNF Digital Commons. For more information, please contact [Digital Projects](#).

© 2001 All Rights Reserved

# Identifying the NCAA Tournament “Dance Card”

---

B. JAY COLEMAN  
jcoleman@unf.edu

*Department of Management, Marketing, and  
Logistics  
University of North Florida  
4567 St. Johns Bluff Road, South  
Jacksonville, Florida 32224-2675*

ALLEN K. LYNCH  
lynch\_ak@mercer.edu

*Stetson School of Business and Economics  
Mercer University  
1400 Coleman Avenue  
Macon, Georgia 31207*

---

The NCAA Basketball Tournament selection committee annually selects the Division I men’s teams that should receive at-large bids to the national championship tournament. Although its deliberations are shrouded in secrecy, the committee is supposed to consider a litany of team-performance statistics, many of which outsiders can reasonably estimate. Using a probit analysis on objective team data from 1994 through 1999, we developed an equation that accurately classified nearly 90 percent of 249 “bubble” teams during that time frame and over 85 percent for the 2000 tournament. Given the NCAA Tournament’s nickname of the big dance, the equation is effectively the “dance card” that determined whether a team got an invitation from past committees and is also a tool that could aid decision making for future committees. The accuracy of the dance card, and the factors and weights included in it, suggest that the committee is fairly predictable in its decisions, despite barbs from fans, teams, and the media.

**A**n annual debate rages among followers of intercollegiate athletics regarding which of the approximately 310 National Collegiate Athletic Association

(NCAA) Division I men’s basketball teams should be among the 64 selected to participate in the postseason national championship tournament (the NCAA Tournament).

---

Copyright © 2001 INFORMS  
0092-2102/01/3103/0076/\$05.00  
1526-551X electronic ISSN  
This paper was refereed.

RECREATION AND SPORTS  
GAMES-GROUP DECISIONS—VOTING-COMMITTEES

## THE NCAA TOURNAMENT

Approximately 30 teams get automatic invitations (or bids) to the NCAA Tournament by winning their respective conference championships. (The number of Division I teams and the number of automatic bids occasionally changes from year to year.) However, the NCAA Tournament selection committee fills all remaining open (or at-large) slots in the field of 64 teams, making its final decisions during meetings that are not open to the public or the media. Although the committee chairpersons typically answer media questions after they announce the tournament field, they generally reveal little about their deliberations.

However, the general process by which the committee is supposed to come to its conclusions is public and is described on the NCAA's Web site [NCAA 1999]. As a part of this process, the committee is to refer to descriptive statistics and other information about the teams eligible for at-large selection. This information (Table 1), which for the most part outsiders can reasonably well estimate or compute, is com-

piled into a single report that summarizes the relevant information for all teams combined. This so-called nitty gritty report [NCAA 1999] represents the largely objective inputs into the otherwise subjective process of team selection.

The purpose of our research was to use the available objective information in the nitty gritty reports of the six college basketball seasons from 1994 through 1999, along with the ex post knowledge of which teams made the tournament in those years, to quantitatively model the selection criteria of the committee. In the jargon of college basketball, where the NCAA Tournament is often called the big dance, this model would effectively be the "dance card" that at least partially captures the factors that the committee has considered most important in past years. It also could be used as a decision aid in future selections or as a means of determining if future committees weigh factors differently.

The latter issue is important given that the committee changes composition each year. Although the committee's decisions

Overall winning percentage	Overall RPI rank (among all Division I teams)
Overall ratings percentage index (RPI)	Number of losses overall
Number of wins overall	Number of nonconference losses
Number of nonconference wins	Nonconference winning percentage
Conference winning percentage	Nonconference RPI rank
Nonconference RPI	Number of conference losses
Number of conference wins	Conference RPI rank
Conference RPI	Number of road losses
Number of road wins	Number of losses in the last 10 games
Number of wins in the last 10 games	Losses against teams ranked 1–25 in RPI
Wins against teams ranked 1–25 in RPI	Losses against teams ranked 26–50 in RPI
Wins against teams ranked 26–50 in RPI	Losses against teams ranked 51–100 in RPI
Wins against teams ranked 51–100 in RPI	Losses against teams ranked 101–150 in RPI
Wins against teams ranked 101–150 in RPI	Losses against teams ranked 151-up in RPI
Wins against teams ranked 151-up in RPI	

**Table 1: The NCAA's nitty gritty report for each team that is a candidate for an at-large selection includes information on 29 factors.**

therefore can be expected to vary somewhat from year to year, it is incumbent upon the committee to wield decisions that are reasonably consistent with those of past committees. The model we propose could contribute to that consistency by indicating how previous committees tended to make decisions. The model does not include variables to reflect differences among committees, since our objective was to create a model that could be universally applied in the future regardless of who might sit on the committee.

Several prior studies have been conducted on NCAA Tournament issues, such as predicting the margin of victory based on seedings [Smith and Schwertman 1999], predicting the probability of a team being among the tournament's final four teams [Carlin 1996; Schwertman, McCready, and Howard 1991; Schwertman, Schenk, and Holbrook 1996], and describing behavior exhibited in tournament betting pools [Metrick 1996]. However, these analyses were based on the assumption that the tournament field had already been selected. We have found no study addressing the question of how the committee weights various criteria when making its at-large selections.

#### Data

It is well known among those that follow college basketball that the most important statistic to the committee is the ratings percentage index (RPI) of each team, a metric that the NCAA devised to aid in the evaluation of teams. The RPI is roughly approximated as 25 percent of the team's winning percentage, plus 50 percent of its opponent's average winning percentage, plus 25 percent of its opponents' opponents' winning percentage.

Only those games played against fellow NCAA Division I member teams are considered in the calculation. The NCAA makes adjustments to the opponents' winning percentages to account for those games played against the team being evaluated. Similar adjustments are made to the winning percentages of the opponents' opponents, if indeed those teams have played the team being evaluated. Moreover, the NCAA gives bonuses based on such factors as wins against teams ranked in the top 50, beating good teams away from home, and playing a majority of non-conference games against top 50 opponents. It also gives penalties for losing to non-Division I teams or teams ranked below 150, for losing to bad teams at home, and for playing a majority of nonconference games against teams ranked below 150. Unfortunately, the NCAA keeps these adjustments confidential [SportsLine 1999b; Palm 1999].

Although the "true" RPI that the committee considers is unknown, because of this confidentiality, many sources try to approximate the statistic so that they can distribute RPI rankings through the media. One of the best known of these sources is Jerry P. Palm, who provides the college-basketball statistical information that is published on the *CBS Sportsline* Web site and is recognized as CBS's RPI guru [SportsLine 1999a]. (CBS has the broadcasting rights to the NCAA Tournament, as well as to the NCAA Tournament selection show.) Given Palm's position, we viewed his "CollegeRPI" Web site [Palm 1999] as a reliable source for the data used in our analysis.

In addition to RPI statistics for each team and conference, Palm generates data

## THE NCAA TOURNAMENT

for the 29 objective factors in the nitty gritty report available to the committee (Table 1). We collected data for each of these objective factors from Palm [1999] for 453 teams that had overall winning percentages of at least 50 percent for each of six regular seasons (those completed in March 1994 through March 1999) but that did not receive automatic bids to the tournament. The committee has never chosen a team with a winning percentage below 50 percent as an at-large selection [Palm 1999], and thus, this factor served as an initial filter for the selection model.

In addition to the objective information in Table 1, the nitty gritty report includes information for each team that is largely subjective or unavailable to those outside the committee. This includes advisory rankings by selected coaches in each region of the country, the combined number of wins and losses against "tournament" teams (teams that received automatic bids into the tournament or have already been selected to receive at-large invitations), the combined number of wins and losses against teams that are under serious consideration for at-large positions, and inju-

ries that may have affected a team's performance over the course of the season [NCAA 1999]. (The committee uses an iterative nomination-and-voting process to incrementally select at-large teams.) This input likely influences the selection process to some degree, but because it was unavailable outside of the committee, we did not include it in our selection model.

Finally, we collected information from various media sources regarding which of the 453 available at-large selections the committee actually picked to participate in the tournament in the six seasons studied.

### Methodology

We treated each of the 29 objective factors as a potential predictor variable for our model. In addition, we computed several additional variables that we hypothesized might have affected the committee's decisions (Table 2). The 20-win plateau has long been considered a total indicative of a very successful season. Television analysts for college basketball telecasts frequently opine that a team should have at least a break-even record within its conference to be selected for the tournament. A superficial measure of the difficulty of the

Binary variable reflecting whether a team had at least 20 wins

Binary variable reflecting whether a team had at least a break-even conference record (at least as many conference wins as conference losses)

Number of road games

Number of games played against teams ranked from 1–25 in RPI

Number of games played against teams ranked 26–51 in RPI

Number of games played against teams ranked 51–100 in RPI

Difference between numbers of wins and losses

Difference between numbers of conference wins and conference losses

Difference between numbers of wins and losses against teams ranked 1–25 in RPI

Difference between numbers of wins and losses against teams ranked 26–50 in RPI

Difference between numbers of wins and losses against teams ranked 51–100 in RPI

Difference between numbers of wins and losses against teams ranked 101–150 in RPI

Difference between numbers of wins and losses against teams ranked 151-up in RPI

**Table 2: We included 13 variables in our analysis in addition to the 29 in the nitty gritty report.**

schedule played by a given team might include the total number of road games and the total number of games played against teams ranked at various general positions based on the RPI (for example, teams ranked in the top 25). Finally, it could be argued that the total number of wins or losses (whether overall, within the conference, or against a certain level of competition) is not as important as the differential between wins and losses. The additional variables addressed each of these concerns.

We designated the variable to be predicted by the model as a binary variable representing whether the team was selected by the committee to participate in the tournament in the respective year. Given the limited values that this dependent variable could take on, we selected a probit analysis as an appropriate modeling approach. Before we estimated the probit model, however, we further filtered the data to make the modeling process and outcomes more representative of the actual process and outcomes. Specifically, we considered only teams with RPI rankings worse than 25 and better than 80 in developing and assessing the model. The highest ranked team not to receive a bid between 1994 and 1999 was Texas Tech, with a ranking of 29 in the RPI of 1997. The lowest ranked team to receive a bid in that period was New Mexico, with a ranking of 75 in the RPI of 1999. Our filtered data set thus bracketed these rankings, with a small buffer at each extreme. Limiting our analysis to such teams accomplished two things. First, since model performance hinges significantly on the number of teams that are correctly pre-

dicted to receive an at-large bid, eliminating the no brainers at both ends of the RPI spectrum avoided criticism that the accuracy of the model included ridiculously easy predictions. Second, this focused our model almost exclusively on bubble teams (a term commonly used by the media to describe teams for which a bid is questionable), since the true committee deliberation process is directed at these teams.

Using the probit analysis, we developed a model whose predicted values represented the z-scores associated with a bubble team's probability of making the tournament. We followed the process described in the appendix to identify a model that best predicted those teams that the committee selected and had at least a modicum of face validity (for example, a variable representing wins against opponents ranked in the top 25 shouldn't have a negative coefficient). The latter issue was important, given the potential uses of the model by the committee, the teams, the media, and the public.

**The Dance Card**

Using our analysis, we produced the following equation as the best estimate of the dance card the NCAA Tournament selection committee used between 1994 and 1999 (Appendix):

The z-score associated with the probability of a bubble team receiving a tournament bid

$$\begin{aligned}
 &= 3.0707459 - 0.074646(RPI\ Rank) \\
 &\quad - 0.012203(Conference\ RPI\ Rank) \\
 &\quad + 0.235189(Top\ 25\ Wins) \\
 &\quad + 0.1442626(Conference\ Wins—Losses) \\
 &\quad + 0.4093414(Top\ 50\ Wins—Losses) \\
 &\quad + 0.264996(Top\ 100\ Wins—Losses). \quad (1)
 \end{aligned}$$

## THE NCAA TOURNAMENT

We identified six variables from the dance card as statistically significant in the team-selection process. The first variable (*RPI Rank*) represents the overall ranking of a given team among all teams in Division I based on the RPI metric. The second (*Conference RPI Rank*) represents the ranking of a given team (among all Division I teams) based only on its performance against its conference competitors. The expected negative coefficients for each of these variables imply that as the rank value increases by one unit (implying that the team's rank has actually worsened by one position), the *z*-score drops, as does the associated chance of getting into the tournament.

The third variable (*Top 25 Wins*) represents the total number of victories against teams ranked in the top 25 (according to the RPI). The final three variables represent the difference between the number of wins and the number of losses against opponents within a team's conference (*Conference Wins—Losses*), against teams ranked 26 through 50 according to the RPI (*Top 50 Wins—Losses*), and against teams ranked 51 through 100 according to the RPI (*Top 100 Wins—Losses*). The positive coefficients of these variables indicate that the *z*-score, and the associated chance of receiving a tournament bid, increases as these variables increase.

To calculate or predict a given team's overall probability of getting a tournament bid, we first simply entered the relevant information for that team into the dance-card equation. We then converted the resulting *z*-score, or standard normal random variable, in each case into a probability of receiving a tournament bid using the

standard normal cumulative distribution function. In other words, a team's predicted probability of getting a bid was the area under the standard normal curve that was less than or equal to that team's predicted *z*-score. We then ranked bubble teams in each season according to these predicted probabilities to determine those teams that the committee should have selected for the tournament in that season, if indeed the committee was perfectly consistent. (Simply ordering the bubble teams according to their predicted *z*-scores would have yielded precisely the same rankings.) A similar process could be followed to determine the teams that should be selected in future seasons.

The dance card's coefficients represent the marginal effect on the *z*-score of a one-unit change in any of the six factors on the right side of the equation. However, users of the model would likely be interested in knowing the impact of a one-unit change in any factor on the associated probability of getting bid. The marginal impact of a one-unit change in an individual factor on the probability of tournament entry is not the same linear constant for all teams because it depends on the value of the specific factors for each team. For example, the dance card indicates that an additional top 25 win increases the *z*-score of a bubble team by 0.235 units. Suppose that the various factor values for some team are such that the additional top 25 win takes the team from a *z*-score of 0.00, which has an associated probability of 50 percent, to a *z*-score of 0.235, which has an associated probability of 59.3 percent. Thus, for that team, the additional win increases its chance of getting a bid by 9.3 percentage

points. Suppose that for another team, the additional top 25 win increases its z-score from 3.00, which has an associated probability of 99.86 percent, to a z-score of 3.235, which has an associated probability of 99.94 percent. In that case, the additional win yields only a 0.08 percentage point improvement in the chance of a bid. In other words, an additional top 25 win would be very beneficial for a somewhat marginal team but really wouldn't make much difference for a team that had already virtually guaranteed itself a bid to the tournament. A similar principle is true for the other factors in the model.

**Dance Card Accuracy**

We summarize the performance of the dance card across all six of the seasons we

studied in Table 3. We measured model accuracy as follows. After we placed all teams with automatic bids and all teams with RPI rankings of 25 or better in a given year's field of 64, we assigned the remaining bids to the teams with the highest predicted probabilities, as generated by the dance card. For example, in 1994 there were 30 automatic bids, and 18 teams ranked in the top 25 in RPI did not receive an automatic bid. We thus assigned all 48 of these teams to the 1994 tournament field before considering the dance card. We assigned the remaining 16 positions in the 1994 field to those 16 teams (out of the 41 bubble teams available in 1994) with the highest predicted probabilities according to the z-scores calculated using the dance

Year	Bubble teams	Correctly classified	Teams selected by the dance card, but not by the committee	Teams selected by the committee, but not by the dance card
1994	41	90.24%	Oklahoma (31)	Seton Hall (45)
1995	42	85.71%	Georgia Tech (37)	George Washington (56)
			St Joseph's (36)	Stanford (47)
1996	40	95.00%	Virginia Tech (38)	Manhattan (54)
			New Mexico State (46)	Minnesota (66)
1997	44	86.36%	Tulane (53)	Boston College (45)
1998	42	90.48%	Texas Tech (29)	Temple (35)
			West Virginia (49)	Oklahoma (48)
			Hawaii (51)	Georgetown (55)
			Hawaii (41)	Oklahoma (51)
1999	40	90.00%	Vanderbilt (44)	Western Michigan (59)
			Rutgers (43)	Oklahoma (49)
In-Sample Totals	249	89.55%	DePaul (45)	New Mexico (75)
Out-of-Sample			13 Teams	13 Teams
2000	41	85.37%	Kent (34)	Seton Hall (48)
			SW Missouri State (36)	Indiana State (49)
			Bowling Green (55)	Pepperdine (53)

**Table 3: We compared committee results with dance-card results (RPI rankings shown in parentheses) and found the dance card's accuracy of 85 to 95 percent to be consistent across seasons and to extend to the out-of-sample projections.**



card. We followed a similar process to select the field in each of the other five years in our sample. We then compared the teams that we assigned to the available at-large positions in each year using this procedure to the teams the committee actually selected to fill those same positions.

We found that by using the dance card we correctly classified 223 of the 249 bubble teams during the years 1994–1999 combined (an accuracy rate of 89.55 percent). This accuracy percentage became even

---

### The average person can use the model without knowing anything about statistics.

---

more impressive when we included our filter, which correctly classified all teams with RPI rankings better than 26 and worse than 79. By including all 1,650 teams that did not receive automatic bids to the tournament in our six-year sample and using the filter combined with the dance card, we correctly classified 98.42 percent of all available at-large teams in Division I men's basketball between 1994 and 1999.

The dance card model was also quite consistent across the six seasons, with between two and six teams misclassified in each year (Table 3). This provided some evidence that the model did not rely too heavily on the committee criteria of a single year and is general enough to be applied across years.

We also measured the dance card's accuracy by applying the model to the 2000 tournament as an out-of-sample evaluation. Using a similar procedure to that described above, we used the dance card to

correctly classify 35 of the 41 bubble teams (or 85.37 percent) for 2000. When we included the filter with the model, we correctly classified 283 (or 97.92 percent) of the 289 teams that participated in Division I but did not receive automatic bids to the tournament in the season ending in March 2000. In other words, the out-of-sample accuracy was equivalent to the in-sample accuracy for the seasons of 1995 and 1997 and similar to the overall accuracy of the model for those seasons included in the sample on which the model was built. The strength of the out-of-sample predictions lends additional credence to the dance card.

Finally, viewing the dance card from a different perspective makes its accuracy even more impressive. Any single error by the dance card by definition had to result in at least two teams being misclassified. That is, any team that the dance card indicated should get in but that the committee did not select had to be matched by a team that the committee did select and that the dance card indicated should not have received a bid. For example, the bid that Boston College received in 1996 should have gone to Tulane, according to the dance card. If we view this as effectively one misassignment and not two, the dance card missed on only 13 selections in the six-year combined sample and on only three selections in the out-of-sample season. Stated otherwise, the dance card, when combined with the filter, accurately predicted 192 of the 205 (or 93.7 percent) available at-large tournament slots for 1994–1999, and 32 of the 35 (or 91.4 percent) available at-large slots for the 2000 tournament.

### Discussion and Conclusion

The filtered probit model represented by the dance card appears to be a reasonable approximation of the criteria the NCAA Tournament selection committee weigh when making at-large selections for the tournament field. Not only was model performance quite good, particularly given that our analysis focused on teams for which a bid was most questionable, but the factors that we identified as most important had significant face validity. Clearly the committee considers the RPI,

---

### The dance card is an estimation of the committee's decision rule.

---

as the NCAA's own metric, heavily in its own right as a filter and an important predictor variable and as a means of identifying the importance of individual games. We offer no comment on whether using the RPI is reasonable or correct but only accept it as major factor in the process.

Although it is difficult to conclude from our analysis exactly how the committee weighed various factors (in that some of the dance card's variables may be simple proxies for something else that the committee did weigh), the other factors included in the model are consistent with those often highlighted by basketball observers. For example, the dance card rewards wins against good teams (teams ranked in the top 25) but does not penalize losses against such teams. However, for games against more marginal teams, it rewards wins and punishes losses. The fact that the win/loss differential in games against teams ranked 26 through 50 in the

RPI garnered a heavy weight is not surprising, because such teams represent some of the best of the other bubble teams. One would expect the committee to closely consider a given team's record in these games as a means of comparing potential at-large invitees. A similar point could be made regarding games against teams ranked 51 through 100 in the RPI, and as would be expected, the dance card gives less weight to those results. It seems likely that the differing weights placed on wins and losses against teams ranked at various positions in the RPI at least partially reflect some of the unreported subjective adjustments the NCAA makes to the RPI. Finally, our results imply that conference performance and the strength of a team's conference (as it affects a team's conference RPI rank) are significant factors. Again both are consistent with observations basketball followers commonly make.

An advantage of the dance card is that it is fairly easy to use for media, teams, fans, and committee members not familiar with statistical methods. All a user has to do is enter the appropriate factor values for a team or group of teams, and the model will provide a result (a z-score) for each team that can then be used to evaluate a team or rank a group of teams. The user doesn't have to convert the z-score into a probability to use the model. Instead, the user can regard the z-score as simply a strength or power index for the team. When evaluating a single team, a user can view a power index (z-score) value of 0.00 as a sort of cutoff figure. If the team has a positive power index, then the team has a good shot at getting a bid,

## THE NCAA TOURNAMENT

and the more positive the index, the better the chances are. If the power index is around zero, the team's chances are marginal. If the power index is negative, the team has very little chance of making the tournament. Finally, users can rank groups of teams by simply using the power indexes, since the order will be the same using z-scores or using the associated probabilities. Thus, the average person can use the model without knowing anything about statistics.

The dance card is not our model; it is an estimation of the committee's decision rule, and its accuracy was limited by the degree to which committees over six seasons were consistent in their decisions. In developing the model, we were also limited by our inability to consider subjective information the committee may have employed. However, its high degree of accuracy, its face validity, and its consistency across the six seasons examined indicate that committees were fairly consistent in making selections and tended to consider reasonable, objective criteria. This might surprise many observers, given the critiques of the tournament field that follow its announcement every year.

Whether the dance card will garner significant interest among parties to the selection process remains to be seen, but early indications suggest that it may. WJXT, the local (Jacksonville, Florida) CBS television affiliate, showed immediate interest when we contacted them on the eve of the 2000 NCAA Tournament selection announcements. We told them about the mathematical formula we had developed that could be used to help predict the teams that would be selected for the NCAA Tourna-

ment the next day. We were promptly featured as the lead news story on the 6:00 PM Sunday broadcast that immediately preceded the selection show at 6:30 PM, and we were graded on our predictions on the 11:00 PM broadcast. Given that media response and given that the dance card correctly identified over 90 percent of the at-large teams selected that night, we believe our model has a chance of gaining more widespread attention.

### APPENDIX

The probit model is one of several techniques available for predicting the occurrence of events (such as a tournament bid) that are captured by binary dependent variables. In the general probit technique, one assumes that a latent, unobserved variable exists. The value of the underlying latent variable remains unknown because only the occurrence of the associated event is recorded. The probit model is used to predict the value of the unobserved latent variable. Any predicted value for the latent variable in excess of a predefined threshold leads to prediction that the event will occur.

The predicted values for the latent variable that result from use of the finished model are effectively standard normal variables (or z-scores), not probability estimates, and are linear in each of the regressors. However, using the standard normal cumulative distribution function, one can convert these z-scores into predicted probabilities of the event occurring. By virtue of the nonlinear nature of the standard normal distribution, the predicted probabilities are not linear in any of the regressors used in the model.

To estimate the coefficients of the probit model that help to predict the underlying latent variable (the z-scores), one uses a maximum-likelihood technique. The resulting coefficient estimate for each regres-

Variable names	Coefficient estimate	t-statistic
Intercept	3.0707459	4.47
RPI rank	-0.074646	-6.11
Conference RPI rank	-0.012203	-1.76
Top 25 wins	0.235189	1.94
Conference wins-losses	0.1442626	3.80
Top 50 wins-losses	0.4093414	4.26
Top 100 wins-losses	0.264996	3.57
Log likelihood function =	-65.11	
Constrained log likelihood function =	-168.86	
Degrees of freedom =	242	
Pseudo R-squared =	0.6144	

**Table 4: All coefficients estimates for our best model for predicting the latent unobserved variable underlying a tournament bid were significant at the five-percent level in the direction of the expected sign.**

sor, however, provides the marginal impact of a one-unit increase in that regressor on the predicted value of the unobserved latent variable, and not on the probability of the occurrence of the event being modeled.

We selected the best model, via an iterative process, as the one with statistically significant coefficients (at the five-percent level) in the direction of the expected sign and the highest percentage of accurate designations (Table 4) using the bid-allocation process described in the text.

We found the pseudo  $R^2$  value of 0.6144 by comparing the value of the log likelihood function associated with the estimated model to the one that would exist if the right-hand side of the equation was limited to an intercept.

**References**

Carlin, B. P. 1996, "Improved NCAA basketball tournament modeling via point spread and team strength information," *The American Statistician*, Vol. 50, No. 1 (February), pp. 39-43.

Metrick, A. 1996, "March madness? Strategic behavior in NCAA basketball tournament betting pools," *Journal of Economic Behavior and Organization*, Vol. 30, No. 2 (August), pp. 159-162.

National Collegiate Athletic Association 1999, "Principles and procedures for establishing

the 64-team bracket," Retrieved on September 28, 1999 from (www.ncaachampionships.com/sports/bkm/i\_64teambracket.html).

Palm, J. P. 1999, "CollegeRPI.com—The college basketball RPI," Retrieved on March 9, 1999 from (www.collegerpi.com).

Schwertman, N. C.; McCready, T. S.; and Howard, L. 1991, "Probability models for the NCAA regional basketball tournaments," *The American Statistician*, Vol. 45, No. 1 (February), pp. 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*, Vol. 50, No. 1 (February), pp. 34-38.

Smith, T. and Schwertman, N. C. 1999, "Can the NCAA basketball tournament seeding be used to predict margin of victory?," *The American Statistician*, Vol. 53, No. 2 (May), pp. 94-98.

SportsLine USA, Inc. 1999a, "CBS SportsLine RPI," Retrieved on March 9, 1999 from (wysiwyg://35/http://www.sportsline.com/u/basketball/college/1998/polls/rpi/).

SportsLine USA, Inc. 1999b, "CBS SportsLine RPI: Frequently asked questions," Retrieved on March 9, 1999 from (www.sportsline.com/u/basketball/college/1998/polls/rpi/faq.htm).